UNIVERSITAT DE VALÈNCIA

FACULTAT DE FILOSOFIA I CIÈNCIES DE L'EDUCACIÓ

Programa de Doctorado en Ética y Democracia

Inteligencia artificial responsable. Humanismo tecnológico y ciencia cívica

Tesis Doctoral

Presentada por:
Antonio Luis Terrones Rodríguez

Dirigida por:
Dr. Francisco Arenas Dolz

Valencia, septiembre de 2020

Sentimos que aun cuando todas las posibles cuestiones científicas hayan recibido respuesta, nuestros problemas vitales todavía no se han rozado en lo más mínimo.
(Wittgenstein, 2003: 181)

Un espectro anda al acecho entre nosotros y sólo unos pocos lo han visto con claridad. No se trata del viejo fantasma del comunismo o del fascismo, sino de un nuevo espectro: una sociedad completamente mecanizada, dedicada a la máxima producción y al máximo consumo materiales y dirigida por máquinas computadoras. En el consiguiente proceso social, el hombre mismo, bien alimentado y divertido, aunque pasivo, apagado y poco sentimental, está siendo transformado en una parte de la maquinaria total.
(Fromm, 1970: 65)

ÍNDICE

TERCERA PARTE

ÁMBITOS DE APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL RESPONSABLE EN EL MUNDO CONTEMPORÁNEO

CAPÍTULO 6

CAPÍTULO 7

CAPÍTULO 8

AGRADECIMIENTOS

Esta investigación es fruto de un largo y arduo camino iniciado en 2016. No ha sido fácil desarrollar este trabajo, pues lo he elaborado muy lejos de mis seres queridos, a los que siempre he necesitado para que me transmitan su ilusión y confianza. Agradezco a mi familia la confianza depositada en mí desde el principio para seguir caminando hacia adelante en el mundo de Sofía, que siempre me ha apasionado. Este trabajo supone un gran logro personal y familiar por toda la confianza que mis seres queridos han depositado en mí. Es un regalo en forma de esfuerzo que les ofrezco en señal de agradecimiento por todo lo que me han dado a lo largo de mi vida.

A mis queridos colegas Julián García, Stéphane Vinolo y Dennis Schutijser, que durante mi estancia en Ecuador siempre me han empujado para que no pierda de vista el principal objetivo de este trabajo. Siempre han sido un excelente apoyo en las dificultades académicas y laborales a las que me he tenido que enfrentar y que en ocasiones suponían un obstáculo para seguir con el cumplimiento de este objetivo formativo y profesional.

A mi amigo Fernando Villacampa Salinas, porque gracias a su escucha atenta he aprendido a vivir lejos de los míos y tener la suficiente entereza psicológica para afrontar el reto que supone una tesis doctoral. Los largos ratos de conversación y paciencia mutua han hecho más fácil el proceso de investigación de este trabajo. Y a otros tantos amigos a lo largo y ancho del mundo que me han brindado enriquecedoras enseñanzas.

A mi director de tesis, Francisco Arenas Dolz, por su disponibilidad desde el comienzo para dirigir esta investigación, orientarme en un ámbito tan novedoso del conocimiento y transmitirme la ilusión por el trabajo bien hecho.

INTRODUCCIÓN

> Se trata del mandato de la cautela, en vista del carácter revolucionario que adopta la mecánica de la elección de alternativas bajo el signo de la tecnología, con su inherente «ir a por todas», tan ajeno a la evolución.
>
> (Jonas, 1995: 72)

El sueño de las máquinas pensantes que Alan Turing presentó en la revista *Mind* en 1950 se está haciendo realidad. La inteligencia artificial (IA) está irrumpiendo con fuerza en diversidad de esferas de la vida humana para someterlas a una configuración tecnológica nunca antes conocida. La IA tiene por objeto de estudio que las computadoras hagan las mismas cosas que puede hacer el cerebro humano, como razonar, reconocer visualmente, asociar, predecir, planificar, controlar, etc. (Boden, 2017: 11-12). La introducción de intelectos sintéticos en el mundo humano ha despertado un gran interés en investigadores procedentes de numerosos ámbitos. Esta introducción ha suscitado profundas incertidumbres en torno a las problemáticas que la IA puede presentar, dependiendo del terreno de despliegue. Las reacciones no se han hecho esperar y ya pueden identificarse grupos que sostienen diversas perspectivas: tecnooptimistas y tecnopesimistas, bioconservadores y bioprogresistas, entre otros.

Debido a las incertidumbres y reacciones, el objetivo principal de este trabajo se centra en la justificación de la necesidad de llevar a cabo un ejercicio reflexivo sobre el fenómeno de la IA desde una perspectiva crítica, a la vez que optimista, para impulsar un proyecto de responsabilidad. El campo de la IA está traspasando todas las esferas de la vida humana y, por lo tanto, debería someterse a una profunda consideración desde la ética. El ser humano actual es producto de la cuarta revolución tecnológica (Floridi, 2011) y comparte el mundo con artefactos tecnológicos que incorporan IA, una situación que obliga a pensar mirando al futuro desde la responsabilidad. La aproximación al mundo de los intelectos sintéticos impone, en este tiempo de enormes avances tecnológicos, comprometedoras exigencias

éticas y políticas caracterizadas por la innovación para el establecimiento de criterios de responsabilidad.

El capítulo poshumanista en el que las fuerzas tecnológicas condicionarán la vida se presenta como un nuevo relato en la historia de la humanidad. Este espíritu poshumanista está motivando importantes investigaciones en diversas áreas que serán abordadas en este trabajo. El principal impulso del relato poshumanista se ha desarrollado gracias a las investigaciones en el campo de la IA. Los sistemas artificiales han contribuido al enriquecimiento de muchos procesos para lograr eficacia y velocidad, aunque todavía queda un largo camino por recorrer. Los investigadores involucrados en las áreas de despliegue de la actividad de la IA son conocedores del gran potencial de los intelectos sintéticos y reconocen que la inteligencia es un arma muy poderosa que permite abrir nuevos horizontes de posibilidades nunca antes conocidos.

Cada vez es mayor la bibliografía, los congresos, coloquios, conferencias, etc., sobre IA, lo que muestra el gran interés que está despertando este campo de estudio en las últimas décadas. El título de esta tesis doctoral, *Inteligencia artificial responsable. Humanismo tecnológico y ciencia cívica*, plantea la incorporación de criterios de responsabilidad ante los desafíos que la transición transhumanista presenta en el desarrollo de la IA. No obstante, esta incorporación de criterios implica un requisito filosófico ineludible, a saber, el desarrollo de un humanismo tecnológico y el cultivo de habilidades cívicas y democráticas en el entorno científico, dentro de un marco regulatorio, como premisas esenciales para promover un modo de pensar e innovar que permita dar respuesta a las exigencias de este tiempo. Por ello, las preguntas centrales que motivan este trabajo son las siguientes: ¿Es realizable un diseño tecnológico que permita restablecer la centralidad de lo humano como medida del desarrollo de la IA en el contexto actual? ¿Es posible el cultivo de habilidades cívicas y democráticas en el proceso de generación de conocimiento científico en el ámbito de la IA?

Los resultados presentados no se centrarán en detalles estrictamente técnicos del ámbito de la IA, sino más bien en ofrecer una revisión sinóptica de aquellas áreas consideradas más importantes y que han abierto profundos debates jurídicos, éticos, políticos, filosóficos o teológicos. En el trasfondo de los temas esbozados existen ciertas polémicas y un numeroso listado de investigadores e investigadoras que han tratado de dar respuestas a través de su participación en un enérgico debate que ha servido de hilo conductor para esta tesis. No se hará una revisión exhaustiva de la bibliografía especializada sobre IA publicada en los últimos tiempos, sino que se asumirán los planteamientos de algunas investigaciones como punto de partida para enriquecer la reflexión y favorecer así la construcción de una propuesta que permita al conocimiento científico adoptar una perspectiva ética y dote de sentido cívico y democrático a la tecnología. El ejercicio reflexivo debe ser una tarea de toda la ciudadanía y no exclusivamente de los especialistas en la materia, pues corresponde a todos los agentes involucrados por esta actividad pensar las bases de un futuro de prosperidad y libertad.

Para el cultivo de la responsabilidad es fundamental reconocer la importancia del conocimiento científico en el ámbito tecnológico como un recurso y bien de interés público y cuestionar la brecha que existe entre quienes crean, regulan y usan las tecnologías. Se planteará la necesidad de una adaptabilidad práctica que debe estar presente en los procesos deliberativos y participativos, donde sean considerados los testimonios de los grupos de interés. Se trata de promover una investigación e innovación responsables, entendiendo que la actividad tecnológica tiene un ineludible impacto social y ético. Este tipo de innovación e investigación teje un hilo conductor entre la aceptabilidad y la deseabilidad de los procesos de generación de conocimiento, promoviendo una ética aplicada fundamentada en el diálogo. La actividad científica puede caracterizarse así como una comunicación constructiva para favorecer la comprensión de la ciudadanía en aquellos asuntos que son de interés público, y por lo tanto por el impulso de una ciencia cívica. Por ello es importante el cultivo de habilidades cívicas y comunicativas y la construcción de puentes de diálogo entre los grupos de interés en la actividad de la IA, situando a la ciudadanía en el centro de las preocupaciones. En ese sentido, resulta fundamental la incorporación de aspectos

valorativos y consideraciones morales en la práctica tecnológica para impulsar evaluaciones con carácter ético. La actividad dedicada a la gestión del ámbito tecnológico debe ser repartida entre las esferas que configuran un modelo de innovación abierto y responsable (MIAR), sostenido en el paradigma de la quíntuple hélice (mercado, Estado, academia, comunidad y medio ambiente). El cultivo de una inteligencia artificial responsable (IAR), que es la propuesta central de este trabajo, tiene como pilares fundamentales el restablecimiento de un humanismo tecnológico, donde la ética tiene un peso regulativo para defender que la persona es un fin en sí mismo, y la importancia de una ciencia cívica capaz de promover habilidades cívicas y democráticas que contribuyan a sacar lo mejor de nosotros mismos.

La IAR se construye desde la defensa de espacios de encuentro deliberativo en los que los asuntos científicos relativos a la IA se abordan de forma cooperativa por medio de la confluencia entre diversas perspectivas y saberes, gracias a metodologías de investigación-acción de carácter participativo que nos proporcionan una visión más amplia sobre la IA y el conocimiento científico en general, donde la ciudadanía ve reflejados sus deseos, creencias y perspectivas. Esto permitiría democratizar el conocimiento científico y por lo tanto no privilegiar los enfoques de arriba abajo de carácter monopolista y oligopolista.

Este trabajo tiene tres partes. La primera parte, de carácter más filosófico, está dedicada al marco teórico, donde se presenta tanto la propuesta de un humanismo tecnológico, premisa fundamental de una IAR, como el principio de responsabilidad, propuesto por Hans Jonas, referente filosófico esencial para la incorporación de criterios éticos de responsabilidad en el contexto tecnológico. El objetivo de la segunda parte consistirá en definir y presentar algunas de las perspectivas éticas sobre el concepto de IA y en desarrollar el concepto de IAR, inspirado en una ciencia cívica y en un modelo de innovación abierto y responsable, pilares fundamentales para interpretar los desafíos que impone este tiempo. En la tercera y última parte se abordarán algunos de los desafíos futuros a los que conducen los intelectos sintéticos, profundizando en cuatro de ellos: el desafío transhumanista, el desafío en el ámbito de las profesiones, el desafío de los coches autónomos y finalmente el desafío militar y de la ciberseguridad. Existen también otros

desafíos que se derivan del despliegue de la IA, aunque por motivos de concreción y brevedad únicamente se estudiarán los señalados.

En el capítulo 1 se recogen algunos de los testimonios filosóficos que han observado la tecnología desde perspectivas plurales con el objetivo de poner de relieve cuáles son aquellas notas características que configuran la relación entre el ser humano y el universo tecnológico. Uno de los propósitos de esta parte gira en torno a la posibilidad de complementar los postulados de responsabilidad de Jonas a través de la incorporación de otras miradas desde la esfera filosófica para enriquecer una perspectiva ética sobre la IA. La tarea se centra en la búsqueda de la complementariedad entre distintos enfoques y por ello el planteamiento de Jonas es fundamental en la formulación de una propuesta de humanismo tecnológico que contribuya al fortalecimiento de un fundamento para la IAR. Esa tarea de complementariedad y enriquecimiento de la propuesta de responsabilidad para el contexto de la IA exige la consideración de otros pensadores que han reflexionado sobre la tecnología. Entre las figuras que se destacan en el capítulo 1 se encuentran principalmente: José Ortega y Gasset, Martin Heidegger, Peter Sloterdijk y Gilbert Simondon. Estos y otros pensadores han contribuido a impulsar un humanismo tecnológico, optimista y a la vez crítico, que observa la tecnología como una oportunidad para adquirir un compromiso de responsabilidad con la existencia que condiciona las exigencias de su despliegue.

A partir de un humanismo tecnológico es posible promover una IAR en el contexto tecnológico actual. Los desafíos de la IA exigen el planteamiento de un humanismo de este tipo, que asuma un compromiso con miras al futuro. En ese sentido, el humanismo tecnológico representa el imperativo de un tiempo de desafíos tecnológicos y a la vez una premisa ineludible en el planteamiento de la IAR. Es una premisa para la responsabilidad ante un tiempo de exigencias que no pueden ser esquivadas. Los importantes y profundos avances que ha experimentado el campo de la tecnología conducen a un escenario novedoso para la humanidad que demanda una nueva contextualización del humanismo. A la vez, este humanismo tecnológico es crítico porque es conocedor del límite y de su condición condicionada, es decir, de su posición de carestía que le invita a imaginar lo que

puede ser y hasta dónde puede llegar. Esta invitación permite entender que los límites no se encuentran situados en un plano negativo, sino más bien como un avistamiento del lado razonable de la tecnología, como un espacio desde el que hacer posible el florecimiento humano, en el marco de unos criterios cívicos y democráticos.

Para seguir en la senda de la fundamentación filosófica, el capítulo 2 está dedicado a profundizar en el principio de responsabilidad formulado por el filósofo Hans Jonas. El postulado del filósofo alemán sirve como hilo conductor para fundamentar la propuesta de una IAR, por supuesto salvando las distancias con la perspectiva de responsabilidad que Jonas propone en la década de los años 70 del siglo pasado. El ideal motivador de la obra de Jonas nace de la exigencia de responsabilidad ante el poder técnico que está desarrollando la humanidad. En 1979 se publicó su obra principal, *El principio de responsabilidad*, que se convertiría en un importante punto de referencia para la filosofía de la tecnología y la bioética. El principio de responsabilidad se caracteriza por la confluencia de varias investigaciones que van desde el gnosticismo hasta la crítica al utopismo marxista y la ideología del progreso, la preocupación por el enorme poder de la técnica y el imperativo de una nueva ética que supere el fundamento antropocéntrico, pasando por la filosofía de la vida y la naturaleza. El planteamiento jonasiano es interesante porque invita a pensar sobre un contexto tecnológico nuevo, lo que ineludiblemente exige a la humanidad el planteamiento de una nueva ética a la altura de las problemáticas de este tiempo. La novedad del contexto del siglo XX del que parte este pensador se caracteriza por la fuerte influencia y condicionamiento que el poder de la técnica ha ejercido sobre la acción humana, derivando en su transformación. Esta situación se encuentra fuertemente marcada por la convicción de progreso indefinido de carácter ideológico y utópico que la propia técnica representa. Tal es la magnitud del poder técnico que es capaz de afectar y amenazar todas las esferas en las que la vida humana se desenvuelve y, lo que es aún más preocupante para Jonas, la propia biosfera.

Es importante aclarar que este trabajo no está orientado de forma exclusiva a dar cuenta de las implicaciones que el poder tecnológico tiene sobre la biosfera en términos ecológicos, sino más bien a reflexionar sobre las consecuencias del progreso tecnológico, y

concretamente de la IA, en la vida de los seres humanos. Jonas no centra su trabajo en un análisis de la IA, sino en el fenómeno tecnológico en general y en la gran capacidad transformadora que éste tiene. No obstante, el fundamento filosófico que Jonas proporciona al concepto de IAR es esencial, ya que es capaz de orientar la actividad tecnológica en el campo de la IA asumiendo exigencias que son necesarias en la actualidad. Entre las contribuciones más importantes del principio de responsabilidad jonasiano para la IAR se encuentran las siguientes: el reconocimiento del gran poder transformador de la tecnología, el cuidado por aquello que es susceptible de vulnerabilidad, la necesidad de una mirada crítica sobre la acción tecnológica y la reorientación de la ética mediante la incorporación de nuevos criterios de mayor amplitud de reconocimiento. Existen otras aportaciones muy interesantes que también brinda el principio de responsabilidad, aunque son más cercanas a una filosofía de la vida y no tanto a una propuesta de ética aplicada al campo de la IA. Así pues, en este trabajo se reconoce la importante aportación que Jonas hace a la filosofía de la tecnología para fundamentar una nueva ética, y por ello se recoge el testigo del espíritu que este filósofo quiere transmitir y se contextualiza en el ámbito de la IA salvando aquellas diferencias que la especificidad del tema exige.

En la segunda parte de este trabajo se expone el concepto de IA. Para realizar un ejercicio de reflexividad filosófica, es importante en primer lugar un acercamiento al objeto de estudio, en este caso, la IA. El capítulo 3 aclara el concepto de IA, los antecedentes históricos, el actual desarrollo y los pronósticos para el futuro que presenta. Los orígenes de los sistemas artificiales distan mucho de los asombrosos avances experimentados en los últimos años a raíz de importantes descubrimientos, tanto en el campo de la computación como en el de la neurociencia. Estos asombrosos avances representan para la humanidad un nuevo escenario para el estudio de iniciativas, muchas de las cuales no son ni siquiera imaginables en este momento, pues en ocasiones han sido objeto de los relatos de ciencia ficción más destacados.

Al igual que es necesario realizar una tarea de clarificación conceptual del objeto de estudio, también lo es de aquellas perspectivas éticas que han arrojado luz sobre el fenómeno de la IA. En ese sentido, el capítulo 4 está dedicado al análisis ético. La reflexión

ética en este ámbito ha cobrado fuerza durante la última década y cada vez es mayor la bibliografía existente. El interés ético se ha incrementado debido a los avances en las ciencias computacionales y el impacto que han tenido en algunos ámbitos de la actividad humana, como la política o las biotecnologías, entre otros. La IA está dejando una profunda huella en diversas esferas de la vida humana que deben ser objeto de reflexión para vislumbrar en el horizonte un trasfondo filosófico en el que sean puestas de relieve las problemáticas y asuntos que giran en torno a la vida.

En la actualidad puede observarse el surgimiento de grupos de expertos y organizaciones comprometidos con el estudio de los intelectos sintéticos y que ofrecen un enriquecimiento de la discusión desde una perspectiva ética. Además, con el tiempo se han ido perfilando distintas posiciones arraigadas en las diversas tradiciones éticas que reflexionan sobre el fenómeno moral. Entre las voces contemporáneas que han contribuido a este debate se encuentran figuras como Raymond Kurzweil, Nick Bostrom, Shannon Vallor, Bill Hibbard y Rajakishore Nath y Vineet Sahu, entre otros. Kurzweil se sitúa en la estela del utilitarismo ético, ya que considera a la IA como un medio tecnológico desde el que alcanzar la felicidad para el mayor número. En cambio, Bostrom entiende la necesidad de introducir criterios de normatividad ética en el universo artificial, aunque reconoce la dificultad de llevar a cabo esta empresa, debido a que este campo no ha avanzado lo suficiente para poder emular el comportamiento humano tal y como esperarían sus planteamientos. Vallor, partiendo de la ética aristotélica de las virtudes, defiende la introducción de valores éticos en los intelectos sintéticos, aunque también reconoce una dificultad, a saber, la relatividad moral que se encuentra en las tradiciones culturales. En cuanto a Hibbard, propone el aprendizaje de valores por medio de la experiencia, también conocido como «aprendizaje profundo» (*deep learning*), un camino posible desde el que pueden introducirse patrones de comportamiento ético en los sistemas artificiales. Por último, Nath y Sahu se sitúan en el lado contrario de los anteriores planteamientos, ya que niegan con rotundidad la posibilidad de considerar los intelectos sintéticos desde una perspectiva ética, pues afirman que no son propiamente agentes morales. Como puede observarse, son diversas las perspectivas éticas que han reflexionado sobre fenómeno de la

IA. La propuesta de una IAR toma distancia de estos cinco planteamientos y pone el acento en el diseño de un marco deliberativo adecuado para que participen los diversos grupos de interés, subrayando el contexto ético y político en el que se generan los conocimientos científicos en el ámbito de la IA.

En el capítulo 5 puede encontrarse el núcleo de nuestra propuesta, la inteligencia artificial responsable. El reconocimiento de la actividad tecnológica como una actividad humana que recae dentro del ámbito de la moralidad es un aspecto fundamental para la formulación de una propuesta de responsabilidad. La actividad tecnológica tiene implicaciones morales porque se despliega en contextos sociales y políticos susceptibles de reflexión en torno al principio ético de la responsabilidad. No obstante, la tarea de reconocimiento de responsabilidad no es fácil, ya que por diversos motivos, que pueden ser económicos o estrictamente técnicos, en ocasiones se torna complejo realizar esta empresa. En este capítulo se identifican algunos de los aspectos de la IA que son susceptibles de discusión ética dentro del campo tecnológico, así como los impactos que determinados mecanismos tecnológicos avanzados tienen sobre la democracia. Se clasifican estos impactos desde cuatro perspectivas políticas: *laissez-faire*, optimismo, escepticismo y esencialismo. Dado que ninguna de esas cuatro perspectivas representa una alternativa política desde la que fundamentar una IAR, en este trabajo se propone una nueva vía. Esta alternativa novedosa surge de un modelo de generación de conocimiento que proporciona un terreno fértil desde el que formular prácticas deliberativas que favorecen el cultivo de las habilidades cívicas y el fortalecimiento de la democracia: el modelo de innovación abierta y responsable.

No es la primera vez que se menciona en el ámbito académico el concepto de «inteligencia artificial responsable», aunque sí lo es la fundamentación filosófica que se hace del mismo y, por lo tanto, su conceptualización. La tarea de fundamentación es clave, aunque otros planteamientos emplean este concepto con cierta debilidad. Por un lado, la tecnóloga Virginia Dignum formula su concepto de IAR a partir de tres premisas que giran en torno a la formación de los expertos, la posibilidad de que los algoritmos representen valores humanos y la participación de los grupos de interés a través de mecanismo de

transparencia y rendición de cuentas. Y, por otro lado, la Declaración de Montreal dota al concepto de IAR de un carácter exclusivamente deontológico. Esta declaración es muy interesante en el sentido prescriptivo y también por el espíritu deliberativo que promueve en el seno de la sociedad canadiense. No obstante, a pesar de que ambos planteamientos son muy enriquecedores para el campo de la IA, no representan una alternativa política que encuentre su fundamento en un marco deliberativo concreto ni tampoco en una teoría ética en particular, por lo que podrían incurrir superficialmente el primero en positivismo y el segundo en deontologismo.

El MIAR se plantea en esta tesis desde un amplio escenario que permite reconocer las complejidades que caracterizan la realidad. Desde este modelo se pueden generar conocimientos innovadores que cuenten con un importante grado de legitimidad. El paradigma de la quíntuple hélice sirve como soporte fundamental para el MIAR, identificando cinco esferas esenciales para generar conocimiento: mercado, Estado, academia, comunidad y medio ambiente. La importancia de este paradigma radica en que las cinco esferas o subsistemas mencionados intercambian conocimientos innovadores que permiten hacer frente a los desafíos del presente.

Frente a otros modelos de generación de conocimiento, la novedad del modelo de la quíntuple hélice consiste en el reconocimiento de la necesidad de integrar una quinta esfera, el medio ambiente, como elemento fundamental a tener en cuenta para cualquier innovación. Además, recogiendo el espíritu de la cuádruple hélice, donde se incorpora la esfera de la comunidad, se pone en valor que el conocimiento innovador que exige el momento actual, frente a los desafíos de la IA, no puede surgir de un modelo de carácter monopolista u oligopolista que entienda el diseño de las tecnologías como una empresa dogmática y cerrada al mundo. Las sinergias entre las esferas que configuran nuestras complejas realidades deben ser consideradas en el contexto de prácticas de innovación social y búsqueda de alternativas que exigen los desafíos del presente y el futuro. Por ello, el papel de la comunidad es muy importante en este ejercicio deliberativo del MIAR, que debe proporcionar conocimientos útiles e innovadores para la resolución de problemas. De esa forma, al incorporar a la comunidad y, en general, a los grupos de interés en el ejercicio

generador, se estarán cultivando habilidades cívicas y también fortaleciendo los mecanismos democráticos más fundamentales.

La tercera parte está dedicada a estudiar algunos de los desafíos más importantes del impacto de la IA en las diversas esferas de la vida humana. Aunque existen otros desafíos que también son relevantes, debido al interés por concretar y destacar algunos de los más vinculantes, en esta tesis se han seleccionado los siguientes: el desafío transhumanista, el desafío de las profesiones, el desafío de los coches autónomos y finalmente los desafíos militares y la ciberseguridad. En ese sentido, la propuesta de la IAR será contextualizada en estos ámbitos.

Un nuevo campo de estudio para la filosofía, que ha despertado un progresivo interés en los últimos años, es el de las tecnologías destinadas al mejoramiento de la especie. En *Meditación de la técnica*, José Ortega y Gasset hizo hincapié en la doble imposición que existe entre la naturaleza y el ser humano, pues mientras la primera le condiciona su vida al segundo, el segundo se ve empujado a crear una sobrenaturaleza. El ser humano vive inmerso en circunstancias, rodeado de una naturaleza que le impone unas necesidades que debe satisfacer, como protegerse del frío o comer. La vida se encuentra ligada a la necesidad y el humano se empeña por superar sus carencias porque quiere vivir bien. Debido a estas necesidades el humano pone en marcha una serie de actividades para satisfacerlas. No se trata únicamente de estar en el mundo, sino de estar bien, por ello el humano se las ingenia para construir esta sobrenaturaleza.

Además de ser *homo sapiens*, el ser humano también es *homo faber*, como sostiene Hannah Arendt en *La condición humana* (2012). En la diferenciación que esta autora establece entre labor, trabajo y acción, como aquellas actividades en las que el ser humano despliega su vida, se pone de relieve esa nota de creación que caracteriza al ser humano respecto a los demás animales. El *homo faber* crea un mundo que los seres humanos comparten entre sí, desarrolla su producción mediante la evaluación, elección y empleo de los medios adecuados para alcanzar determinados fines. La relación del *homo faber* con el medio es la de adueñamiento y uso de la naturaleza, por ello, también posee la capacidad de

crear y destruir sus obras de consumo. Además, el individuo se encuentra dentro de una esfera política en la que se descubre a sí mismo a través del discurso y la acción. En ese sentido, para la filósofa alemana la acción es la actividad que se da entre los individuos sin la mediación entre cosas, es una condición humana arraigada en la política. Por lo tanto, el ser humano es creador de realidades y también potencialmente político en el sentido en que comparte un mundo de pluralidad en la esfera política en la que debe desarrollar su vida en común.

Las NBIC (nanotecnologías, biotecnologías, tecnologías de la información y ciencias cognitivas) contribuyen a hacer realidad las principales tesis del transhumanismo, un movimiento cultural e intelectual que tiene como finalidad última mejorar la condición humana mediante el empleo de las tecnologías más avanzadas entre las que se encuentra la IA. Los defensores del transhumanismo reconocen en las tecnologías un medio legítimo para la búsqueda del beneficio de la especie humana, asumiendo la tarea del mejoramiento como un imperativo moral que no puede menospreciarse. El proyecto transhumanista representa una vía de transición hacia un escenario de condiciones que van más allá de lo humano en términos biológicos, y hasta incluso políticos, caminando hacia la escritura de un nuevo relato, lo poshumano. El transhumanismo promueve una transformación radical del ideal terapéutico tradicional de la medicina para proponer el de perfeccionamiento, lo que supondría un cambio de paradigma.

Las reacciones frente a los mecanismos de mejoramiento promovidos desde las tesis transhumanistas no se han hecho esperar y existen diversas perspectivas enfrentadas, aunque aquí se realiza una clasificación de los postulados de algunos pensadores dentro de dos grupos claramente diferenciados: los bioconservadores, entre los que se encuentran Jürgen Habermas, Michael Sandel, Steven J. Jensen y José Luis Widow; y los bioprogresitas, como Peter Sloterdijk, Raymond Kurzweil y Nick Bostrom.

La IA juega un papel muy importante para el logro de las tesis transhumanistas, ya que es un medio tecnológico avanzado para alcanzar muchos de los ideales de esta corriente de pensamiento y para impulsar otros mecanismos tecnológicos que planteen nuevos horizontes de trascendencia desde lo biológico hacia lo tecnológico. Toda tecnología tiene un carácter político dada su incorporación en el entramado social, por lo que las tecnologías orientadas al mejoramiento de la especie plantean importantes interrogantes en el terreno de lo político. Debido a los impactos, es importante la incorporación de criterios de responsabilidad en el despliegue de estas tecnologías. El concepto de IAR es fundamental para la puesta en marcha del ejercicio práctico de las tesis transhumanistas, pues permite orientar los postulados del transhumanismo hacia el cultivo de habilidades cívicas y democráticas, y hacia una mejora de las condiciones de vida de los seres humanos y la biosfera. En ese sentido, los postulados transhumanistas pueden asumir un compromiso de cuidado deliberativo desde una perspectiva cívica y democrática.

El capítulo 7 trata sobre las profesiones y el impacto que ocasionará la automatización impulsada desde los intelectos sintéticos. En primer lugar, la incorporación de la IA a las profesiones puede llevarse a cabo a través de dos caminos: innovación y automatización. Este capítulo se centra en la segunda, la automatización, es decir, en el proceso tecnológico que produce tecnologías avanzadas que integran IA para realizar aquellas tareas que los humanos desarrollan en las profesiones. No obstante, es importante mencionar que la sustitución de las tareas laborales no es nada nuevo en la historia del trabajo, pues desde la Revolución Industrial las máquinas vienen sustituyendo a los humanos en muchas actividades. Sin embargo, en la actualidad la diferencia radica en el factor cognitivo de estas máquinas, pues gracias a la IA es posible el desarrollo de nuevas actividades que no necesariamente tengan que ver con lo rutinario.

Los datos arrojados por un estudio de la Federación Internacional de Robótica (2018) destacan el importante crecimiento de ventas anual que ha experimentado la industria robótica entre los años 2013 y 2017, representando en términos porcentuales un 114 %. Datos como los de este informe ofrecen un claro indicador para realizar un temprano ejercicio de reflexión, pues permiten avistar en el horizonte profundas transformaciones que

deben invitar a emprender acciones a todas las instituciones y organizaciones en las que la sociedad civil desarrolla sus actividades. Las profesiones juegan un papel fundamental en la vida, principal motivo por el que la automatización debe someterse a un importante debate público fomentando la participación democrática y ciudadana en las discusiones con expertos, lo cual potenciaría modelos de innovación social caracterizados por la inclusión y la responsabilidad pública. El ejercicio deliberativo entre los grupos de interés de las esferas de la quíntuple hélice permitirá promover formas alternativas en el manejo de las problemáticas derivadas de la automatización.

Conforme se van sucediendo nuevos avances en el campo de la IA, las profesiones van experimentando diferentes grados de transformación. En los casos más extremos muchas profesiones desaparecerán, mientras que en otros cambiarán por completo como consecuencia de la integración de la IA en sus procedimientos. Esto pone de relieve que, aunque todas las profesiones presenten diferencias entre sí, todas tienen en común un componente, a saber, que sirven para hacer frente a los retos cotidianos.

Hay quienes ven en la automatización una amenaza para los trabajadores con capacidades laborales limitadas y escasos conocimientos. No obstante, no cabe duda que debido a los intelectos sintéticos las profesiones que requieren una mayor formación y especialización también se encuentran en el punto de mira. Ante este panorama de incertidumbre han surgido dos perspectivas claramente diferenciadas, los tecnooptimistas y los tecnopesimistas. Los primeros se sitúan en la línea de un optimismo exacerbado de corte positivista, pues entienden que sin lugar a dudas la tecnología generará nuevas habilidades y fuentes de empleabilidad. En cambio, los tecnopesimistas dudan de aquello que sostienen los primeros y observan el fenómeno de la automatización desde la desconfianza y en ocasiones promueven tesis catastrofistas.

En ese sentido, es importante plantear una alternativa fundamentada en el humanismo tecnológico y en el principio ético de la responsabilidad que observe el fenómeno de la automatización desde una óptica de optimismo crítico que reconoce sus posibilidades y limitaciones. Esa alternativa pasa necesariamente por una IAR que permita su

aprovechamiento para las profesiones y sea capaz de plantear nuevos horizontes de realización y crecimiento humano. A partir de aquí pueden plantearse innovadoras políticas y estrategias económicas alternativas en el centro de laboratorios abiertos, laboratorios ciudadanos, para definir el valor del tiempo humano como un momento de liberación profesional y el cultivo de nuevos habilidades. La redefinición de los modelos educativos y de la propia tarea de educar también se sitúa en la estela de una IAR promovida desde una preocupación cívica y democrática.

El capítulo 8 está dedicado a otro de los desafíos considerados prioritarios en este trabajo, los automóviles autónomos. Dentro del espectro de los vehículos autónomos pueden encontrarse aviones, barcos, coches, etc., aunque la reflexión girará únicamente en torno al caso de los coches. Este tipo de vehículos se caracterizan principalmente por el control remoto desde un operador, o por su circulación autónoma, sin la necesidad de intervención humana. La conducción autónoma es posible gracias a unos niveles óptimos de seguridad que permiten intercambiar en tiempo real información con el entorno físico en el que se movilizan.

Los coches autónomos integran IA para analizar el entorno donde realizan sus desplazamientos. A través del *deep learning*, y mediante una serie de sensores combinados, estos vehículos son capaces de reconocer la información necesaria. Posteriormente esta información es procesada mediante una acumulación de datos para poder circular en entornos cada vez más complejos. Uno de los retos de las investigaciones en este campo consiste en proporcionar unos sistemas lo suficientemente perfeccionados en materia de autonomía que permitan una movilidad completamente autónoma en entornos de gran complejidad. Por lo tanto, no se trata de automóviles con componentes de control humano, sino de automóviles con un alto nivel de automatización y conocimiento del entorno.

El desarrollo de los sistemas artificiales y el uso de las tecnologías de la información y la comunicación contribuyen en el diseño de los coches sin conductor mediante conocimientos adquiridos y procesados a partir de bases de datos. Estos datos facilitan la eficiencia de los sistemas de transporte y contribuyen a mejorar las medidas de seguridad

vial de estas tecnologías autónomas. Los expertos en este campo entienden que la puesta en circulación de los coches sin conductor mejorará la calidad de vida de la ciudadanía en diversos ámbitos y además favorecerá una reducción de las emisiones de CO_2, pues este tipo de vehículos se mueven gracias a energías renovables o a la electricidad. Son por lo tanto más limpios, en términos medioambientales, y seguros, en término viales, y a la vez fomentan una nueva cultura de la movilidad y contribuyen a la eliminación de determinadas conductas viales que dificultan el tráfico. No obstante, a pesar de la opinión de los expertos, es necesaria la generación de espacios de transparencia comunicativa para que la ciudadanía comprenda estas tecnologías y lleve a cabo una acción propositiva para mejorar sus condiciones de vida, pues "con el fin de que la sabiduría no sea engreída y altanera, debemos pensarla conjuntamente con la responsabilidad" (Domingo Moratalla, 1991: 6). Esta participación activa de la ciudadanía es posible desplegarla en laboratorios abiertos, que favorezcan el cultivo de una dimensión cívica en el contexto de esta tecnología emergente y promuevan la legitimidad democrática de las innovaciones.

Si la tecnología más avanzada impulsada por la IA está transformando el mundo del trabajo, el de las tecnologías destinadas a la mejora humana y la cultura de la movilidad, lo mismo sucede en el terreno militar y de la seguridad. Las dinámicas del campo de batalla comienzan a modificarse a partir de los sistemas artificiales. El avance de la robótica militar es muy notable, aunque el nivel de investigación no se encuentra tan avanzado como en otros ámbitos. La robótica militar está impulsando un cambio en la concepción de la guerra.

Expertos como Peter Warren Singer (2009a; 2009b) califican a este fenómeno como la «deshumanización del campo de batalla». La creciente autonomía de los robots militares avanza en sintonía con la IA, aunque no con la misma intensidad que en otros ámbitos. Uno de los intereses más importante de la autonomía de los sistemas artificiales en materia militar se encuentra en los vehículos aéreos no tripulados (UAV, siglas en inglés de *Unmanned Aerial Vehicle*), lo que comúnmente se conoce como «drones». Esta tecnología comenzó a utilizarse en las guerras desde finales de la primera mitad del siglo XX y ha sido empleada por países como EE. UU. o Alemania. La Alianza Tecnológica Colaborativa

Robótica (RCTA, siglas en inglés de *Robotics Collaborative Technology Alliance*) basa la autonomía de estos vehículos en cinco categorías humanas que emulan los sistemas artificiales: pensar (adaptación del razonamiento para la táctica), mirar (centrar la atención en la percepción de la situación), moverse (en base a parámetros de seguridad y adaptación al territorio), hablar (comunicación eficiente) y trabajar (interacción con el entorno físico). El ámbito militar está experimentado un proceso de robotización que impactará considerablemente sobre la cultura y las instituciones de defensa. La IA es ya un componente indispensable de los ejércitos y de las políticas de seguridad.

La introducción de criterios de responsabilidad cívica y democrática en el diseño y uso de sistemas artificiales en el ámbito de las Fuerzas Armadas resulta fundamental, especialmente ante los retos que presenta la Agenda 2030. En ese sentido, la función de las Fuerzas Armadas en el contexto de la integración de sistemas artificiales puede resignificarse y orientarse hacia el cultivo de las habilidades cívicas y democráticas, profundizando de ese modo el compromiso con los derechos humanos y los ODS, así como por un respeto con los límites planetarios. Entre las medidas que podrían adoptarse para promover la IAR en el ámbito de la defensa y la seguridad, se encuentran las siguientes: lucha contra la destrucción del medio ambiente, mejora de la comunicación y la calidad de los datos, mecanismos de transparencia para fortalecer el acceso de la ciudadanía a la información, mejor conocimiento para el desempeño del ejercicio profesional, fortalecimiento de las misiones humanitarias, etc.

Por último, el capítulo 9 está dedicado a la ciberseguridad, pues el desarrollo de las tecnologías ha ocasionado la aparición de un nuevo escenario, el ciberespacio, donde los conflictos se desarrollan en escenarios que no son estrictamente físicos. Este cambio de escenario va acompañado de una variación en la forma de concebir la violencia. El entorno de la guerra está cambiando y por lo tanto sus efectos también, ya que no se limitan a los daños estrictamente físicos. Las nuevas tecnologías permiten ampliar el potencial y espectro de actuación, pues el espacio virtual permite una deslocalización de los ataques. En ese sentido, la demanda de seguridad es creciente ante el desarrollo de tecnologías con enorme potencial para generar conflictos. Debido a la complejidad del ciberespacio y a la

fragilidad de la seguridad frente al enorme potencial tecnológico, es necesario impulsar mecanismos de seguridad que cuenten con legitimidad política. En ese sentido, las innovaciones en el terreno de la seguridad deben ser fruto de un trabajo inclusivo y participativo de las esferas que conforman la quíntuple hélice. La confluencia de diversas perspectivas permite el establecimiento de redes para dar una respuesta más aproximada a los desafíos y al mismo tiempo promueve una interrelación cívica entre los grupos e instituciones de interés en el marco de los derechos humanos y los ODS.

Este trabajo representa una invitación dirigida a la ciudadanía en general, y a los tecnólogos y personas del mundo de la filosofía en particular, para que orienten sus reflexiones hacia un asombroso medio tecnológico que está transformando la vida. La intención de este trabajo es la de incidir en los aspectos éticos y políticos de estas tecnologías avanzadas y ponerlas al servicio del cultivo de las habilidades cívicas y el fortalecimiento de la democracia a través de la generación de espacios abiertos a la deliberación y el debate sobre cuestiones científicas. Frente a la irrupción de la IA y la dictadura de los algoritmos (Lasalle, 2019), se erige una alternativa ética para hacer frente a los desafíos, presentes y futuros, desde la responsabilidad.

PRIMERA PARTE

HUMANISMO TECNOLÓGICO Y RESPONSABILIDAD ÉTICA

CAPÍTULO 1

HUMANISMO TECNOLÓGICO

> Antes que una confrontación y antagonismo entre la técnica y la libertad, lo que se impone en nuestra época es una superación de semejante antítesis para lograr que ellas se fecunden mutuamente y de su conjunción nazca un nuevo destino para el hombre. Ello significaría sentar las bases y explicar el sentido de un nuevo humanismo –el auténtico humanismo de nuestros días: el humanismo técnico– donde esa técnica, como producto de la libertad humana, queda reconciliada con la propia libertad que la origina y, en lugar de destruirla, la potencia y multiplica como exponente del don más humano que distingue y caracteriza al existir del hombre. Efectivamente: así como la tecnocracia, en tanto que es producto de la *ratio technica* que la sustenta es orientada por una voluntad de amor, aquella tecnocracia puede ser utilizada para ayudar al hombre y a los pueblos en la difícil aunque irrenunciable tarea de ser dueños y gestores de su propio destino mediante el ejercicio de la libertad.
>
> (Mayz Vallenilla, 1984: 258-259)

Este primer capítulo representa un espacio de reflexión y fundamentación filosófica para el concepto de inteligencia artificial responsable (IAR). La atención se centrará en algunos postulados filosóficos que han dirigido la mirada sobre el fenómeno tecnológico en la condición humana. Tal ejercicio permite promover un basamento para elaborar la propuesta de un humanismo tecnológico que afronte el enorme potencial trasformador que la inteligencia artificial (IA) tiene para la vida desde una perspectiva crítica y optimista. Es el momento de perseguir un equilibrio entre la dimensión estrictamente técnica de la tecnología y las exigencias morales que existen en el universo humano. En ese sentido, este equilibrio permitirá cultivar desde la tecnología aquellas habilidades que pueden suponer una beneficiosa contribución cívica y democrática para la ciudadanía.

La filosofía de la tecnología, más allá de sus antecedentes en el siglo XIX con la filosofía de la técnica, ha entrado en escena en el seno de las ciencias humanas sobre todo en el siglo XX. Así, la sorprendente expansión tecnológica por todos conocida, y más concretamente en los últimos tiempos el avance de la IA, demanda una nueva interpretación

y definición del ser humano, y por lo tanto un nuevo concepto de humanismo para este nuevo tiempo.

El propósito de este capítulo consiste en el enriquecimiento de los postulados de Hans Jonas, que serán expuestos posteriormente y que suponen un elemento central para la fundamentación teórica del concepto de IAR. Se partirá del reconocimiento y asunción de que todo pensamiento debería ser revisado y enriquecido con el paso del tiempo, si así lo exige la realidad. La complementariedad y el enriquecimiento de la propuesta jonasiana responden a un claro interés por formular un humanismo tecnológico que fortalezca sus rasgos más esenciales desde las contribuciones que diversos filósofos han ofrecido. Es importante aclarar que no se promueve la invitación a una clara oposición a la tecnología, sino que más bien se propone una mirada a esta desde dentro, con pensadores como Gilbert Simondon, y desde fuera, con maestros como Heidegger. La falsa oposición entre la técnica y la cultura no conduce a ninguna parte. Así pues, el humanismo tecnológico no representa una oposición a la técnica, sino una preocupación para pensarla en sus amplias dimensiones. A propósito de esto es importante rescatar unas palabras de Simondon:

> Este estudio está animado por la intención de suscitar una toma de conciencia del sentido de los objetos técnicos. La cultura se ha constituido en sistema de defensa contra las técnicas; ahora bien, esta defensa se presenta como una defensa del hombre, suponiendo que los objetos técnicos no contienen realidad humana […] La toma de conciencia de los modos de existencia de los objetos técnicos debe ser efectuada por el pensamiento filosófico, que se encuentra en la posición de tener que cumplir en esta obra un deber análogo al que cumplió en la abolición de la esclavitud y la afirmación del valor de la persona humana. La oposición que se ha erigido entre la cultura y la técnica, entre el hombre y la máquina, es falsa y sin fundamentos; sólo recubre ignorancia o resentimiento. Enmascara detrás de un humanismo fácil una realidad rica en esfuerzos humanos y en fuerzas naturales, y que constituye el mundo de los objetos técnicos, mediados entre la naturaleza y el hombre […] La mayor causa de alienación en el mundo contemporáneo reside en este desconocimiento de la máquina, que no es una alienación causada por la máquina, sino por el no-conocimiento de su naturaleza y de su esencia, por su ausencia de mundo de significaciones, y por su omisión en la tabla de valores y de conceptos que forman parte de la cultura (Simondon, 2008: 31-32).

No se sugerirá bajo ningún concepto lo que Simondon define como un «humanismo fácil«, a saber, esa oposición inútil entre la cultura y la técnica en la que el humanismo ve a la máquina como a un enemigo; ni tampoco en la que el ámbito técnico considera el ser humano como ausente en la incidencia del desarrollo maquinal, al estilo que plantea el determinismo de Jacques Ellul (2003). Por lo tanto, ni romanticismo humanista, ni tecnocracia determinista, sino que se promoverá la formulación de una propuesta que supere las absurdas enemistades entre el humano y la máquina, un humanismo tecnológico. En esta propuesta es clave la fundamentación filosófica para abordar el fenómeno de la tecnología y elaborar estrategias encaminadas hacia la propuesta de una IAR. Es esencial entender que la propuesta de un humanismo tecnológico favorece la comprensión de la IA y el cultivo habilidades cívicas y democráticas desde el principio ético de la responsabilidad.

Así pues, ¿cómo no discutir la posibilidad de un humanismo tecnológico si cada vez es más difícil afirmar que existe una diferencia ontológica radical entre la naturaleza y la cultura? Unas palabras de Andrés Vaccari resultan útiles para entender la necesidad de promover un humanismo de este tipo, superando así aquellos postulados que encuentran un soporte fundamental en la diferenciación ontológica radical de las esferas de lo natural y lo cultural.

> Recientemente las ciencias de la evolución han comenzado a considerar la tecnología como un factor cuasi-biológico en el desarrollo de ciertos rasgos morfológicos y cognoscitivos característicos de la especie. Por ejemplo, se especula que la fabricación y uso de herramientas ha tenido un rol central en la diferenciación de los hemisferios cerebrales (Ambrose, 2001), en el desarrollo del pensamiento causal (Wolpert, 2003) y en la evolución del lenguaje (Corballis, 1999). Todo esto ha problematizado profundamente la división metafísica entre naturaleza y cultura (Vaccari, 2010: 3).

Es importante reivindicar un humanismo tecnológico que parta del ser humano corriente, de carne y hueso, como diría Miguel de Unamuno (2017), sin pretensiones de rechazo ni humanización de la tecnología, sino con la esperanza de ponerla al servicio del progreso humano, cultivando sus ricas potencialidades para la felicidad y el bienestar de la ciudadanía, pero siempre teniendo presente la noción de límite.

> Entre los extremos del entusiasmo y de la crítica descalificadora, consideramos que la actitud más razonable es abogar por un humanismo tecnológico que parta no de lo que debe ser el hombre, sino del hombre concreto, espacial y temporal. No es la amenaza de la tecnología ni el deseo de humanizarla lo que impulsa el humanismo tecnológico, sino la forma de encauzarla en provecho de los seres humanos […] El hombre es un ser tecnológico y las tecnologías son ampliaciones del ser humano. El humanismo tecnológico es una corriente intelectual, cultural y social que apuesta por la formación profesional o continua y las tecnologías para la evolución humana […] La tecnología debe ser un medio más a disposición de las mujeres y hombres del siglo XXI, con el objetivo de potenciar los valores humanos, a través de la igualdad de acceso al conocimiento y la interconexión de las diferentes culturas del mundo. En definitiva, el humanismo tecnológico se concentra en desarrollar talento y valores para que las personas y la comunicación evolucionen en el entorno tecnológico digital (Parra y Arenas-Dolz, 2015: 94-95).

El ser humano debe arrojar una luz responsable sobre la IA, pero esa luz implica una superación y complementación con los postulados de Jonas, porque, aunque son considerados importantes y necesarios, es esencial entender que son insuficientes y demandan interpretarse a la luz del tiempo actual. Como sostiene José María Lasalle, es fundamental promover un pacto entre la técnica y el ser humano, con el objetivo de subordinar la tecnología a un nuevo humanismo que sea capaz de controlar la voluntad de poder que subyace tras la técnica (Lasalle, 2019).

1.1. La sociedad del conocimiento como punto de partida

No resulta fácil definir un concepto tan amplio como sociedad del conocimiento, pero es importante contribuir a su esclarecimiento para poder identificar y contextualizar el punto de partida del humanismo tecnológico. El concepto de sociedad del conocimiento propone generalmente que el saber y el conocimiento son los aspectos principales que predominan en el gobierno de las estructuras y organización de las sociedades actuales. Además, estos dos aspectos suponen una mercancía que se traduce en la moneda de cambio para garantizar el bienestar y el progreso de la ciudadanía. La utilización del término «mercancía« puede resultar incómoda para el lector. Sin embargo, es importante mencionar

que hoy en día el conocimiento es objeto de intercambio económico en el terreno de las patentes, las licencias, los datos, etc. Estos últimos despiertan el interés en numerosas empresas que estudian el mercado y los patrones de comportamiento de los sujetos de una sociedad.

A pesar de que el conocimiento, propiamente dicho, no representa algo totalmente novedoso, es significativo señalar que la velocidad, la transmisión y los impactos, que en la actualidad lo caracterizan, exigen establecer una clara diferencia con tiempos pasados. Así pues, aunque el conocimiento siempre ha sido determinante, la diferencia en este momento estriba en los factores que decretan su generación y aplicación. En ese sentido, con acierto, se denomina a la sociedad actual sociedad del conocimiento, donde el aprendizaje supone uno de los mecanismos fundamentales para su garantía. La clave del crecimiento individual y colectivo reside en la constante adaptación a un medio que exige de continuos aprendizajes en materia de conocimiento.

La tecnología supone el principal motor de esta sociedad, pues en su despliegue es celebrada la confluencia de una cantidad de conocimientos que permiten incidir en la firme aparición de ideas, servicios, instituciones, prácticas y culturas que condicionan la vida de los seres humanos. Sin embargo, el impacto de los conocimientos científicos y tecnológicos es relativo y depende de las circunstancias y condiciones que lo motivan y extienden. En ese sentido, los productos tecnológicos destacan por su complejidad y sugieren la necesidad de un humanismo tecnológico que permita cultivar un sentido beneficioso de los conocimientos para el ser humano y la biosfera. Como afirma José Luis Mateo, el conocimiento se sostiene sobre dos pilares fundamentales: la investigación, desarrollo e innovación (I+D+i) y la enseñanza, como una pieza esencial para la transmisión y garantía de los conocimientos existentes (2006: 148). Por lo tanto, resulta primordial promover la responsabilidad ética en estos pilares del conocimiento como garantía de bienestar y beneficio para la ciudadanía. En primer lugar, es importante establecer referentes morales para la actividad tecnológica en el campo científico y tecnológico y construir un vínculo entre la aceptabilidad y deseabilidad de los procesos de investigación e innovación y sus resultados. Esta prioridad será abordada en uno de los apartados del capítulo 5 dedicado a la

IAR y la *Responsible Research and Innovation* (RRI). En segundo lugar, es inevitable señalar que el proyecto del humanismo tecnológico debe consistir en el poder de la inteligencia colectiva, pues hay que considerar la relevancia de la educación en la actualidad. El *Massachusetts Institute of Technology* (MIT) fundó en 2006 el *Center for Collective Intelligence*, un organismo que tiene como finalidad reunir a diferentes investigadores con el propósito de poder esclarecer cómo la tecnología está cambiando la forma de trabajar colectivamente, pues entendió que la inteligencia colectiva abre un abanico de posibilidades a organizaciones e instituciones para pensar de manera diferente (Parra y Arenas-Dolz, 2015: 115). Los seres humanos han cooperado desde sus inicios y la articulación de mecanismos y procesos que permitan fortalecer una comunicación intelectual que favorezca la generación de conocimiento en los espacios educativos supone un reto en la actualidad.

> La meta de la inteligencia colectiva es el reconocimiento mutuo y el enriquecimiento de las personas. Para ello es necesario que esas personas puedan conversar e interactuar, lo que resulta muy sencillo hoy en día gracias a la tecnología. Internet favorece el intercambio de ideas y conocimientos. Con la Web 2.0 aparecen nuevas formas de relacionarse, en las que los consumidores pasan a ser también creadores, como consecuencia de la facilidad para portar información […] El conocimiento absoluto no es posible. Es por esta razón que resulta casi vital la colaboración para «el reconocimiento y el enriquecimiento mutuo de las personas, y no el culto de comunidades fetichizadas o hipostasiadas». El intercambio de conocimiento y experiencias, donde la diferencia es una manera de enriquecer el saber, nos aleja de una uniformada de pensamiento […] De esta manera, si juntamos todos esos microsaberes, crearemos una inteligencia colectiva, que parte del principio de que cada persona sabe sobre algo. Por tanto nadie tiene el conocimiento absoluto. De lo que resulta fundamental la inclusión y participación de los conocimientos de todos (Parra y Arenas-Dolz, 2015: 123).

En definitiva, la sociedad del conocimiento representa un terreno fértil desde el que impulsar un humanismo tecnológico que contribuya en el beneficio de la ciudadanía y sus condiciones de vida desde criterios de responsabilidad ética. No obstante, se derivan importantes cuestionamientos de carácter ético y político que deben ser abordados desde un

humanismo reconocedor de los límites y potencialidades de la tecnología como un medio de garantía cívica y democrática en el contexto de la generación y transmisión de los conocimientos.

1.2. Necesidad de un sustrato humano

Desde hace ya unos cuantos siglos el ser humano se ha acostumbrado a convivir en este mundo con máquinas que le superan en muchas de sus características, ha desnaturalizado su entorno y vive en un medio tecnológico (Ellul, 2003; 2004). Por ejemplo, la catapulta impulsaba con una mayor fuerza que el ser humano aquellos pesos, la excavadora mueve cantidades de tierra superiores, el coche se desplaza a una mayor velocidad y la IA es capaz de gestionar una cantidad de datos muy superior y en una menor cuantía de tiempo. A pesar de esta costumbre que se ha ido forjando con el paso del tiempo, aparecen excepciones, sobre todo cuando algunas máquinas nos superar en el terreno de la inteligencia, a saber la IA. Mientras que el primer vuelo intercontinental de Pan American en 1958 entre Nueva York y París causaba asombro, las computadoras y su complejo poder están despertando profundas incertidumbres en el seno de las sociedades actuales. El origen de esta incertidumbre estriba en la superioridad de la capacidad intelectual, pues por primera vez en el curso de la historia la humanidad convive con máquinas que le superan intelectualmente. Así pues, en medio de un escenario totalmente novedoso cabría formular la siguiente pregunta: ¿Qué lugar queda para lo humano?

La incertidumbre está justificada tras un innegable progreso de la humanidad que se ha visto acompañado de aspectos negativos como la guerra, las violaciones de los derechos humanos, la pobreza, etc. Y una vez más existe la necesidad de poner en tela de juicio una invención, en este caso la IA, que destaca por su gran potencial transformador para la vida. Además, en la senda del progreso los avances tecnológicos se han incorporado y establecido con el paso del tiempo a través de la integración en las actividades cotidianas. En ese sentido, se ha normalizado la presencia de máquinas en la vida, aunque esta vez es diferente, pues existe una añadidura que radica en lo indicado en el párrafo anterior, a saber, máquinas con una inteligencia superior. El humano se acomoda en este tipo de

convivencia, que en ocasiones se caracteriza por la nebulosa de un sonambulismo tecnológico (Winner, 2008) que demanda someterse a un ejercicio hermenéutico de carácter crítico.

En medio de todo el abanico de especies que existen en la faz de la tierra, el ser humano ha destacado desde sus orígenes por su inteligencia, pero en ese terreno ya no se encuentra solo, comparte espacio intelectual con las máquinas. En ese sentido, se torna necesario pensar el lugar del ser humano en el mundo, destacando aquellos aspectos que merecen un reconocimiento, poniendo en valor lo propiamente humano frente al resto de las especies y las máquinas inteligentes. Pues sin una tarea de cultivo y valor de lo propiamente característico, qué sentido tendría la humanidad. Esta tarea hace referencia al legado que dejará la humanidad en los capítulos posteriores de la evolución.

> Podemos ser más audaces en nuestro razonamiento y especular que sí, por mucho que nos cueste aceptarlo, la especie humana se extinguirá y dejará paso a seres artificiales que nos habrá superado. Es atrevido pensar así. Hablamos de un *relevo* en nuestra preeminencia, e incluso existencia. La nueva pregunta relevante toma un cariz romántico: ¿qué *legado* dejaremos los humanos a los siguientes eslabones evolutivos? (Latorre, 2019: 14).

El recorrido de la evolución humana está plagado de errores y aciertos, con una experiencia de miles de años que ha permitido perfilar una inteligencia superior a la del resto de las especies. Y en medio de toda esa evolución también se ha dado lugar un progreso moral como consecuencia de un enriquecimiento basado en ideales éticos cada vez más exigentes. En ese sentido, puede afirmarse que a pesar de todas las guerras, genocidios, hambrunas, etc., estamos mejor que antes, pues la humanidad ha estado constantemente sometida a un proceso de perfeccionamiento ético con unos niveles cada vez más altos de exigencia. Y es precisamente este recorrido experimental el que permite al ser humano estar dotado de una inteligencia de la que pueden extraerse innumerables ideales éticos que sin duda interesan para ennoblecer el mundo de la inteligencia artificial.

Así pues, no sería afortunado rechazar el valor de lo humano frente a las máquinas, pues tenemos mucho que aportar. El progreso moral que experimenten los agentes implicados en la actividad de la IA tendrá un impacto beneficioso sobre sus diseños y programaciones, sin duda. Hablar de la necesidad del legado humano en este contexto supone hablar de un ineludible debate ético que tiene que celebrarse como fruto de la responsabilidad que nuestra especie ha venido forjando a lo largo del curso histórico frente a un sinnúmero de desafíos. Esta defensa del legado humano en el posterior desarrollo de los intelectos sintéticos se sitúa en la estela de un optimismo crítico que reconoce la posibilidad de un futuro que demanda un profundo ejercicio hermenéutico al que habrá que dedicar tiempo y profundidad, y que se caracterizará por la responsabilidad. Como afirma Latorre: «los humanos han desarrollado la capacidad de apreciar la sutileza, el compromiso con la búsqueda de la verdad, la oposición amistosa. Demos una oportunidad a los valores de la Ilustración» (2019: 19).

A continuación se esbozarán con brevedad las reflexiones filosóficas de unos pensadores que aportan distintas perspectivas teóricas sobre la técnica que nos ayudarán para plantearnos las posibilidades de fundamentación de un humanismo tecnológico. La confluencia de estas miradas desde la antropología de Ortega y Gasset, la crítica ontológica de Heidegger, la antropotécnica de Heidegger y la filosofía de la tecnología de Simondon, permite originar un ejercicio de hermenéutica crítica para aguzar los sentidos frente al despliegue de los intelectos sintéticos en un mundo con desafíos apremiantes.

1.3. Condición técnica, sobrenaturaleza y autoproyección

Según Carl Mitcham, José Ortega y Gasset es el primer filósofo en ocuparse con profundidad de la cuestión tecnológica (1989: 58). La obra más importante del filósofo español donde se aborda el tema de la técnica es *Meditación de la técnica*. En el prólogo, Ortega señala que esta obra nace a partir de un curso desarrollado en el año 1933 en la Universidad de Verano de Santander. Sin embargo, tras ese curso, las lecciones dictadas fueron fragmentadas y publicadas en artículos dominicales en el periódico *La Nación* de Buenos Aires en el año 1935. Finalmente, *Meditación de la técnica* se publicó en 1939

junto con otro ensayo, *Ensimismamiento y alteración*. También existe otro importante documento en relación con este tema, la conferencia *El mito del hombre allende la técnica*, pronunciada por Ortega en 1951 en la ciudad alemana de Darmstadt. Como señala Josep M. Esquirol, la reflexión de Ortega sobre la técnica se «inserta en el núcleo de su filosofía de la vida y de su comprensión de la condición humana» (2011: 15). Por ello sus postulados son fundamentales para este capítulo, donde se esboza la reflexión sobre un humanismo tecnológico. Pero hay una obra muy importante en la historia del pensamiento de Ortega, anterior a su *Meditación de la técnica*, y que marcará un antes y un después, *La rebelión de las masas* de 1929. En esta obra identifica a la técnica como la generadora del hombre-masa. Aunque los tratamientos que recibe la técnica en las dos últimas obras mencionadas son diferentes, coinciden al proporcionar unas orientaciones fundamentales para poder comprender la influencia que tiene dicha técnica sobre el humano y su vida. La técnica se encuentra en una posición relacional con la existencia, articulándose a partir de tres conceptos clave: necesidad, extrañamiento y proyecto.

El humano vive inmerso en unas circunstancias, rodeado de la naturaleza que le impone unas necesidades que debe satisfacer, como protegerse del frío o comer. La vida se encuentra ligada a la necesidad y el humano se empeña por cubrir esas necesidades porque quiere vivir. En ese sentido, pone en marcha una serie de actividades para satisfacer tales necesidades, estableciendo así las condiciones requeridas para que puedan ser satisfechas, y crea, por ejemplo, sistemas de cultivo, habitáculos con los que poder resguardarse del frío, etc. Este tipo de actividades significan la puesta en suspenso de las necesidades más primarias, como señala el propio filósofo, «calefacción, agricultura y fabricación de carros o automóviles no son, pues, actos en que satisfacemos las necesidades, sino que, por el pronto, implican lo contrario: una suspensión de aquel repertorio primitivo de haceres en que directamente procuramos satisfacerlas» (1965: 18).

A partir de esta consideración sobre la necesidad, surge la primera definición que Ortega hace en su obra:

> Es la técnica, que podemos, desde luego, definir como la reforma que el hombre impone a la naturaleza en vista de la satisfacción de sus necesidades […] Es, pues, la técnica, la reacción enérgica contra la naturaleza o circunstancia que lleva a crear entre ésta y el hombre una nueva naturaleza puesta sobre aquélla, una sobrenaturaleza. Conste, pues: la técnica no es lo que el hombre hace para satisfacer sus necesidades […] La técnica es la reforma de la naturaleza, de esa naturaleza que nos hace necesitados y menesterosos, reforma en sentido tal que las necesidades quedan, a ser posible, anuladas por dejar de ser problema su satisfacción (1965: 21-22).

De esta definición puede deducirse que el ser humano no se adapta a las circunstancias que le vienen dadas y por eso reacciona ante las mismas para estar bien, de modo que no se resigna. Pues no se trata únicamente de estar en el mundo, sino de estar bien, por lo que el ser humano se las ingenia para construir esa sobrenaturaleza que menciona el filósofo español. Pero el verdadero empeño consiste en estar bien en el mundo, de modo que reúne todas sus fuerzas para garantizar ese bienestar. Lo objetivamente superfluo se convierte entonces en lo que es visto como únicamente necesario (1965: 27). La afirmación de Ortega sirve para matizar lo expuesto anteriormente, pues las necesidades biológicas, objetivamente hablando, no son necesidades como tales para el humano, sino que más bien se convierten en verdaderas necesidades cuando condicionan su estar en el mundo, que en ese sentido es una necesidad considerada subjetiva, pues ese estar en el mundo es lo que garantiza el posterior bienestar y lo que es aceptado como superfluo.

Si recordamos el concepto de extrañamiento, es importante señalar que el ser humano crea un mundo diferente al anterior mundo que le es dado, porque no se siente perteneciente a ese mundo, sino un extraño y no percibe comodidad. Esta situación de extrañeza y falta de comodidad provoca el surgimiento de la voluntad, el empeño de construir un nuevo mundo, una nueva naturaleza, una sobrenaturaleza en la que se vea reflejado y que le aleje del extrañamiento. Esquirol interpreta que la idea de extrañamiento es similar a la de ser «arrojado» de Heidegger (Esquirol, 2011: 26).

El concepto de necesidad se encuentra estrechamente ligado al de bienestar, que es relativo al tiempo, al espacio, a la cultura, etc. Varía con el tiempo y con las gentes, motivos que dificultan su delimitación conceptual. Así pues, el concepto de bienestar es variable, ya que se encuentra ligado a una idea filosófica de cómo es entendida la vida. Pero también hay que mencionar que el carácter cambiante del bienestar va acompañado del cambio de técnica y aquí es donde pueden rescatarse estas palabras de Ortega:

> […] basta con que cambie un poco sustancialmente el perfil del bienestar que se cierne sobre el hombre, que sufra una mutación de algún calibre la idea de la vida, de la cual, desde la cual y para la cual hace el hombre todo lo que hace, para que la técnica tradicional cruja, se descoyunte y tome otros rumbos (Ortega, 1965: 32).

La técnica genera un vacío, ya que trata de ahorrar esfuerzo. Piénsese en el mundo actual, donde no paran de inventarse artefactos tecnológicos que sirven para la liberación de ciertas tareas, como por ejemplo la colocación de un tornillo en la cadena de montaje de una fábrica de coches. Los seres humanos desean satisfacer sus necesidades y estar bien con el mínimo esfuerzo. En resumen, los actos técnicos no consisten en realizar un esfuerzo para satisfacer de forma directa las necesidades, ya sean objetivas o subjetivas (superfluas), sino que son aquellos en los que se emprende una reacción frente a las circunstancias que requieren de esfuerzo, en primer lugar, para inventar y, después, para ejecutar un plan o proyecto que ha sido definido previamente. Ese plan o proyecto al que se hace referencia debe permitir lo siguiente:

1. Asegurar la satisfacción de las necesidades, por lo pronto, elementales.
2. Lograr esa satisfacción con el mínimo esfuerzo.
3. Crear posibilidades completamente nuevas produciendo objetos que no hay en la naturaleza del hombre. Así, el navegar, el volar, el hablar con el antípoda mediante el telégrafo o la radiocomunicación (Ortega, 1965: 34).

El humano consigue enfrentarse a las circunstancias y desafiarlas por medio de la capacidad reformadora que ofrece la técnica, consiguiendo de esa manera reducir el esfuerzo impuesto por tal circunstancia, que en definitiva es dominada para la creación de una sobrenaturaleza. Además, en esa acción técnica, caracterizada por el ahorro de esfuerzo, también se encuentra presente la búsqueda de seguridad, pues las circunstancias conducen a un espacio de incertidumbre y de inseguridad que dificulta el pleno desenvolvimiento y produce extrañamiento.

Ortega alerta de la decadencia cultural a la que está conduciendo el progresismo basado en la fe ciega e irreflexiva en un progreso técnico. La falta de reflexividad provoca una confusión en la humanidad, pues la sobrenaturaleza es identificada como naturaleza, perdiendo enteramente la conciencia de la técnica que está siendo utilizada. Este constante predominio de la técnica en la vida ha llevado a no poder vivir materialmente sin la técnica. Esta alerta que hace Ortega es perfectamente acorde con el objetivo perseguido en este capítulo, a saber, esbozar la necesidad de un humanismo tecnológico en el contexto de la IA que permita formular un concepto de IAR en la búsqueda del beneficio para la humanidad, contribuyendo de ese modo a una reorientación del sentido de la tecnología en beneficio del progreso y el florecimiento humano.

Ya se han abordado los conceptos de necesidad y extrañamiento para poder realizar una aproximación al pensamiento de Ortega y su esbozo de la técnica, pero aún queda reflexionar sobre el concepto de proyecto, que se encuentra vinculado con la propuesta de una razón vital e histórica.

El ahorro de esfuerzo que promueve la técnica crea posibilidades para emplear el tiempo y cultivar una imaginación proyectiva. Puesto que la técnica produce una liberación de esfuerzo, ¿en qué ocupamos ese tiempo libre? Según el filósofo español, ahí es donde el humano debe inventar su vida, hacerla él mismo, como si fuera un «artesano de su propia vida», expresión que emplea Adela Cortina (2013), aludiendo a Séneca, y crear así el propio relato de su vida, proyectándose a sí mismo. La técnica se encuentra ligada al concepto de lo que significa ser humano, pues adquiere un carácter antropológico y

ontológico. La filosofía de la tecnología de Ortega se construye desde su idea de la vida humana, entendida ésta como un fenómeno que da forma al significado de la relación activa con las circunstancias, es decir, como un activo creador de esas circunstancias; es un proyecto de vida que forja su razón de ser en la interacción con las circunstancias.

La vida humana no está completamente determinada por la naturaleza; al contrario, la persona tiene que crearse a sí misma elaborando un proyecto de vida. El humano desliza su existencia desde una actividad autointerpretativa y autocreativa, poniendo el acento en él mismo, pero también en las circunstancias que lo motivan a reaccionar proyectando. En esta existencia activa se halla presente una imaginación creativa que pone su poder a disposición del proyecto personal que procura realizarse. Una vez que se ha decidido qué proyecto asumir y emprender, son necesarios recursos de diversa índole. En ese sentido, para Ortega la técnica supone una apertura de nuevas posibilidades orientadas a hacer la vida, a través de la escritura de un relato entendido como proyecto. La vida no está definida por naturaleza, sino que está abocada a ser un puro proyecto que está por realizarse y es producto de su imaginación creativa.

A partir de la idea de vida entendida como proyecto podría identificarse al humano como un *homo faber* que no se limita únicamente a la producción material, sino también como un *faber* que se encarga de autoproyectarse y de escribir el relato de su propia existencia. El filósofo español está aplicando el esquema técnico al hacerse del humano, entendido como proyecto, poniendo el acento sobre la idea de construcción, y por eso utiliza la expresión «autofabricarse»: «de ahí que nuestra vida sea pura tarea e inexorable quehacer» (Ortega, 1965: 51). La forma según la cual Ortega entiende la vida se fundamenta en su modo de entender la razón, un modo que está en profunda conexión con la experiencia de la vida, pues a partir de ella se nutre. Su idea de vida fue expuesta en el *Discurso para los Juegos Florales de Valladolid* en 1906 y en ella entiende el vivir como «más vivir», como aumento del propio ser o «henchimiento» (Conill, 2016: 75). La postura de Ortega acerca de la vida se encuentra inspirada en el pensamiento de Nietzsche. En este sentido, como la vida es adaptación a las circunstancias, es también creación, es atrevimiento y voluntad vital.

El modo de entender la vida que tiene Ortega se elabora desde su nueva filosofía de la razón vital, que en cierta medida recoge el testigo de varios aspectos del pensamiento nietzscheano. Además, la aportación de Ortega se centra principalmente en su reflexión sobre la crisis de los deseos y la necesidad que plantea a la hora cultivar y forjar los proyectos de vida, pues los deseos tecnológicos marginan el verdadero deseo, el deseo de ser sí mismo, y desplazan la preocupación por un proyecto personal. Así pues, en la crisis de los deseos orteguiana se hace hincapié en que los deseos superfluos alimentan un vacío interior. El ser humano se encuentra desconcertado, saturado ante tanta tecnología, y en cierto sentido alimenta un deseo artefactual. Sin embargo, existe una inquietud por la conciencia de la principal ilimitación, una ilimitación ante el superávit de posibilidades inherentes en la tecnología. Precisamente desde ese enfoque orteguiano de necesidad de cultivar un proyecto vital es posible tejer un hilo conductor con el humanismo tecnológico.

Es importante destacar la opinión de Antonio Diéguez (2017), para quien el pensamiento de Ortega, y concretamente su obra *Meditación de la técnica*, fuente de su filosofía de la tecnología, ha sido escasamente tratada pese a su carácter vanguardista. Algunos intentos de la *Revista de Occidente* en el año 2000 y de la Fundación Ortega-Marañón, que organizó un congreso internacional sobre la técnica en Ortega, muestran un claro interés por rescatar el pensamiento orteguiano para interpretarlo a la luz del presente. Sin embargo, la filosofía de la técnica del filósofo español también puede ser sometida a crítica. Según Diéguez, la parte de Ortega más susceptible de crítica es la que tiene que ver con las tres fases históricas de la técnica.

Ortega distingue tres fases en su despliegue, tomando como punto de apoyo la idea que el ser humano ha tenido de su propia técnica. Las denomina «técnica del azar», «técnica del artesano» y «técnica del técnico». Si bien a grandes rasgos las dos últimas fases pueden interpretarse como una descripción simplificada de la separación que la tecnología basada en la ciencia supuso frente a las técnicas tradicionales, resulta dudoso que alguna vez existiera algo así como la técnica del azar –en el ser humano, al menos–, en la que el «inventar no es un previo y deliberado buscar soluciones» (Diéguez, 2017: 168-169).

Esta crítica de Diéguez se encuentra completamente justificada, pues los avances de la investigación antropológica han demostrado que las herramientas propias del tiempo del primate ya requerían de cierta planificación, suponiendo que de azarosas tendrían más bien poco. Las herramientas que el humano ha utilizado desde sus inicios son una clara evidencia de que la capacidad de planificación ha estado presente en su vida desde hace mucho tiempo. Otra de las críticas que recibe la filosofía de la tecnología de Ortega cobra fuerza en la comparativa con el pensamiento de Heidegger. Se le acusa de ser ciertamente optimista y, en términos generales, superficial. En cambio, como sostiene Diéguez, la postulación que Ortega lleva a cabo entre su reflexión de la técnica y otras reflexiones, como las del raciovitalismo, demuestran claramente que su filosofía está muy bien articulada y que cuenta con un hilo conductor indiscutible. Por lo tanto, más allá de las críticas que haya recibido el pensamiento sobre la técnica de Ortega, es importante destacar que su aportación a la filosofía de la tecnología es innegable y que fue muy novedosa para su tiempo, pues no hay que olvidar que data de 1933, año en que dictó su curso sobre la técnica en la Universidad de Verano de Santander. En definitiva, el pensamiento de Ortega se caracteriza por una gran riqueza y claridad, y por lo tanto sirve acertadamente para articular la propuesta de un humanismo tecnológico.

1.4. Carácter hermenéutico ontológico y crítico de la tecnología

El tema central de la filosofía de Heidegger es el Ser, es decir, la pregunta acerca de su sentido. Influido principalmente por Parménides y Aristóteles, Heidegger dedica su empeño a fundar una nueva ontología en el contexto de la modernidad, provocando el resurgimiento del problema central de la metafísica. La filosofía de la técnica de Heidegger hay que situarla en la estela de la búsqueda de la comprensión del Ser.

Desde la década de los años treinta del siglo pasado, el tema de la técnica comienza a formar parte de las reflexiones de Heidegger. En varios de sus textos –*Introducción a la metafísica*, *La época de la imagen del mundo*, *Acerca del evento*, etc.– el filósofo alemán comienza a considerar a la técnica de forma explícita. Sin embargo, no será hasta 1953, año en que pronunció la conferencia *La pregunta por la técnica*, cuando este tema filosófico se

constituirá como un marco referencial del pensamiento heideggeriano. Con el paso del tiempo el filósofo de Friburgo se ocupará de conjugar con mayor interés la técnica con la metafísica.

> El modo heideggeriano de aproximación a la técnica es, como no podía ser menos, la óptica de la historia del ser. Es lo que diferencia la visión heideggeriana de la técnica de tantas y tantas reflexiones sobre ella alentadas por su preponderancia en la vida moderna. El pensamiento de la metafísica como historia del ser es, sin duda, lo que induce a Heidegger a percibir en la técnica un fenómeno que supera con mucho la visión banal que de ella solemos tener (Rodríguez, 1991: 176).

Heidegger justifica la pertinencia para establecer ese vínculo entre la técnica y el Ser pues, para él, la técnica moderna se ha convertido en la metafísica del presente. La pregunta por el Ser le conduce necesariamente a la pregunta por la técnica moderna. El sentido de la técnica en Heidegger puede encontrarse en los escritos sobre la metafísica, porque es ahí donde estriba la problemática moderna de la técnica. En ese sentido, la pregunta por la técnica en el contexto del pensamiento de Heidegger contribuye de manera positiva a la formulación de un humanismo tecnológico, no desde el pesimismo sobre ésta, sino más bien para enriquecer el ejercicio de una hermenéutica ontológica aplicable en el contexto de la IA actual y estudiar sus implicaciones sobre el ser humano.

La pregunta por la técnica de Heidegger aparece en el contexto histórico de la Europa de la posguerra, donde surgió la necesidad de someter a discusión la relación existente entre política y tecnología. La reflexión que el filósofo de Friburgo formula sobre la técnica no se hace desde una representación instrumental, sino ontológica: «la técnica no es la misma cosa que la esencia de la técnica» (1994: 78). De esa forma la reflexión sobre la técnica es principalmente filosófica, pues la esencia de la técnica trasciende el Ser en sí mismo, y no radica en lo estrictamente técnico. En ese sentido, la relación de libertad con la técnica solo será posible si se la cuestiona. Así pues, Heidegger pone en tela de juicio la simpleza de la visión moderna sobre la técnica, pues, según él, radica exclusivamente en una concepción instrumental que no revela su esencia. Restringir la visión sobre la técnica a un «medio para fines» supone determinarla exclusivamente en el plano instrumental, reconociendo en ella

una causalidad eficiente, en la estela del pensamiento aristotélico. Para el de Friburgo la técnica no es solo un medio instrumental, sino un modo de desvelar porque se despliega en el centro del producir, un desocultamiento.

Sin embargo, para Heidegger el modo de desvelar de la técnica moderna se diferencia de la técnica de los antiguos porque «reposa en la ciencia exacta de la naturaleza» (1994: 17). Este modo de desvelar estriba en un provocar, pues se le exige a la naturaleza aquello que puede ofrecer al hombre, no solo mostrando lo que ella es capaz de proporcionar, sino sacándole el máximo partido de modo provocador. La diferencia entre la noción de la naturaleza de los antiguos y de los modernos es que los primeros permitían que la naturaleza se mostrara a la luz, mientras que actualmente ella es sometida a la lógica de lo cuantificable y vista como un potencial fondo disponible de recursos (*Bestand*). No se trata de esperar que la naturaleza se manifieste, sino de extraer de ella el máximo provecho en un tono provocador. La técnica moderna es al mismo tiempo «pro-ducción» y «pro-vocación».

Esta relación del ser humano sobre la naturaleza, a través de su intervención técnica, altera su significado sobre las cosas existentes a priori. Por ejemplo, el río que suministra energía a una represa hidroeléctrica ya deja de ser él mismo y pasa a adquirir otro significado. El valor del río no radica en él mismo, sino en la energía que suministra. La visión heideggeriana sugiere que la existencia del ser humano se caracteriza por el cultivo de una técnica que desoculta, pero no como un simple quehacer o un artefacto. Como señala Heidegger, hay «una interpelación que provoca, que coliga al hombre a solicitar lo que sale de lo oculto como existencias» (1994: 19). A esto el filósofo alemán los denomina *Ge-stell*, un término que puede ser entendido como estructuración, invención o creación. En ese sentido, *Ge-stell* designa la forma de desocultar que domina la esencia de la técnica moderna y que verdaderamente no es nada técnico.

Para Heidegger la modernidad se caracteriza por el encantamiento del mundo procedente de la técnica, y hace el siguiente diagnóstico:

> Lo que es ahora, se encuentra marcado por el dominio de la esencia de la técnica moderna, dominio que se manifiesta ya en todos los campos de la vida por medio de características que pueden recibir distintos nombres como funcionalización, perfección, automatización, burocratización e información (1988: 116-117).

La actualidad del pensamiento de Heidegger para pensar un humanismo tecnológico radica en su concepto de «serenidad». Del pensamiento heideggeriano puede extraerse una visión pesimista y determinista de la tecnología, que al parecer no es del todo cierta. En torno a la idea de pesimismo tecnológico es importante señalar unas palabras de Fernando Broncano:

> El pesimismo tecnológico, que no deberíamos confundir con el pesimismo general filosófico, es un modo de pensar de la historia del cambio tecnológico bajo categorías de un viaje que he llamado de regreso, de búsqueda de un estado esencial y puro a través de la historia. La innovación es algo que le sucede al sujeto de la historia, no un medio que dispone para el control de su destino. Para el pesimismo tecnológico no hay liberación en la historia que no sea la vuelta a un estado fundamental perdido. En unos casos la propia tecnología será el pecado original que explica la pérdida, en otros un hecho accidental que simplemente aleja o retrasa el cumplimiento del destino. No aceptaría que se me acuse de emplear categorías excesivas para la interpretación del pensamiento tecnológico. No soy yo, sino los pensadores contemporáneos de la tecnología quienes han unido la tecnología al plano antropológico, metafísico o de destino de la historia (2001: 55).

Es importante aclarar que Heidegger nunca ha reivindicado una vuelta utópica a un estado originario, una posición que durante mucho tiempo estuvo asociada al movimiento ludita del siglo XIX. En lo que respecta a la noción de retorno, la opinión de Heidegger es la siguiente:

> ¿Retornar? ¿Un renacimiento moderno de la Antigüedad? Sería absurdo y, por otra parte, imposible. El pensamiento griego no puede ser más que un punto de partida. La relación del pensamiento griego con nuestro mundo moderno no ha sido jamás tan presente […] Yo he escrito que la técnica moderna no ha sido completamente extraña a la Antigüedad, en donde encuentra su origen esencial (Towarnicki y Palnier, 1969).

Además, existe otra consideración errónea, tal como señala Broncano, sobre la concepción de la heideggeriana de la técnica, que nos lleva a pensarla como «esencialmente antidemocrática y antihumana» (2001: 87). Para Broncano la posición pesimista de Heidegger da la sensación de que alimenta una posición de pasividad y desinterés de la técnica como destino, implicando de ese modo el peligro de una racionalidad autoritaria. No obstante, es importante aclarar que el filósofo alemán no concibe la historia como un destino fatal del Ser. Recuérdese, por ejemplo, que en *Ser y tiempo* Heidegger considera al ser humano como un ser proyectado, arrojado al mundo, un ser-ahí que está inmerso en un constante quehacerse. Para seguir con la refutación del argumento de Broncano sobre el carácter pasivo y antidemocrático del pensamiento de Heidegger, es pertinente destacar unas líneas de Félix Duque sobre la serenidad:

> Serenidad no es «entrega» a las cosas (una especie de «bajar la guardia», cansado de «vigilar y castigar»: eso no sería sino una inversión que daría igual; un derrotismo victimista –y consumista: comer y gozar en vista de la catástrofe inminente– en vez de la «victoria» sobre la «naturaleza» que anunciara Bacon). Tampoco es pasiva resignación «historicista» (en plan: si algo ha ocurrido es porque tenía que ocurrir: más vale «acomodarse» a lo que hay). Pues ese mostrenco acomodo bien podría llamarse, nada menos, democracia (cf. la entrevista con *Der Spiegel*), entendida –interpreto yo– como un «neoliberalismo» que da valor a cosas y hombres (troquelados ambos según las necesidades la industria y del capital) sólo si «funcionan» como productos mecánicos (o electrónicos); que «usa» las palabras como instrumentos por «de fuera» bien empaquetados y científicamente unívocos y por «de dentro» retóricamente insinuantes y llenos de «maquinaciones» para persuadir de que «esto» es el «progreso» (y cualquier cambio de ese –verdadero– inmovilismo, una argucia «reaccionaria») (1996: 218).

Heidegger plantea el concepto de serenidad como una salida para la totalidad totalizante del Ser en el tiempo de la tecnificación planetaria. Para el de Friburgo pensar consiste en escuchar la voz del Ser, pero no siempre el ejercicio filosófico implica verdaderamente pensar, pues la filosofía no escapa a la interpretación técnica del pensar:

> La creciente falta de pensamiento reside así en un proceso que consume la médula misma del hombre contemporáneo: su huida ante el pensar. Esta huida ante el pensar es la razón de la falta de pensamiento. Esta huida ante el pensar va a la par del hecho de que el hombre no la quiere ver ni admitir. El hombre de hoy negará incluso rotundamente esta huida ante el pensar. Afirmará lo contrario. Dirá –y esto con todo derecho– que nunca en ningún momento se han realizado planes tan vastos, estudios tan variados, investigaciones tan apasionadas como hoy en día. Ciertamente. Este esfuerzo de sagacidad y deliberación tiene su utilidad, y grande. Un pensar de este tipo es imprescindible. Pero también sigue siendo cierto que éste es un pensar de tipo peculiar (2002: 18).

Pero ¿dónde radica la peculiaridad del tipo de pensar al que se refiere Heidegger?

> Su peculiaridad consiste en que cuando planificamos, investigamos, organizamos una empresa, contamos ya siempre con circunstancias dadas. Las tomamos en cuenta con la calculada intención de unas finalidades determinadas. Contamos de antemano con determinados resultados. Este cálculo caracteriza a todo pensar planificador e investigador (2002: 18).

Para Heidegger resultaría una necedad la oposición ciega al mundo técnico (2002: 27), pues los aparatos técnicos se han vuelto indispensables para nuestra vida cotidiana. Frente a esta situación de indispensabilidad, el filósofo de Friburgo sugiere un juego como salida, entre permitir que los objetos técnicos formen parte de nuestra vida, y al mismo tiempo mantenerlos fuera para dejarlos descansar, es decir, «la serenidad para con las cosas» (2002: 28). No obstante, es importante aclarar que el término «serenidad» puede traducirse al castellano de otra forma, pues como señala Duque, el término «desasimiento» sugiere otra conceptualización de lo que pretende expresar Heidegger. El desasimiento sugerido por Duque nada tiene que ver con entregarse a las cosas con una actitud pasiva frente al mundo técnico, ya que se situaría en la estela del pesimismo y determinismo tecnológico que anteriormente señalaba Broncano. Con este concepto en ningún momento el filósofo alemán invita a una actitud de pasividad frente a la técnica; al contrario, critica con vehemencia a aquellos que toman distancia de la técnica, negándola o ignorándola.

Para Heidegger la salida del peligro de la técnica, al que alude en *La pregunta por la técnica*, es posible dentro de la propia técnica, pues hará posible pensar la salida para el engranaje (*Gestell*). Ha sido la técnica la que ha provocado la salida temporal del Ser del escenario del mundo bajo un pensar calculador imperante. Sin embargo, será también por medio de la técnica, a través de un pensar meditativo, que el Ser volverá y se vinculará a un ser humano que lo reclama. Esta vuelta y acontecimiento apropiador es denominado por Heidegger «evento» (*Ereignis*).

En el próximo capítulo se esbozará la propuesta de Hans Jonas en torno a una fundamentación axiológica de la técnica desde el principio ético de la responsabilidad. Esta propuesta de Jonas encuentra su origen en el pensamiento heideggeriano, por lo que se torna fundamental analizar la influencia de Heidegger en el debate contemporáneo acerca de una fundamentación ética de la tecnología en el contexto de un humanismo tecnológico.

La génesis de concepto de responsabilidad de Jonas estriba en el concepto de «cuidado» (*Sorge*) propuesto por Heidegger. Existe un estrecho vínculo entre cuidado y responsabilidad, presuponiendo, de antemano, una determinada concepción ontológica de la técnica. La influencia de Heidegger sobre Jonas se encuentra en que para el segundo la ética para una civilización tecnológica debe fundamentarse en la ontología. En ese sentido, Jonas no se refiere a la responsabilidad en un sentido subjetivo antropocéntrico, sino que más bien su propuesta ética se origina en el Ser y no en el hacer –en sentido meramente instrumental–. Así pues, la ontología heideggeriana se encuentra muy presente en los postulados de Jonas en la relación entre ética y tecnología. Los contenidos y criterios a la hora de elegir los valores que forman parte de la dimensión axiológica de la tecnología en Jonas evidencian un antecedente fundamental, a saber, la influencia de Heidegger en lo relativo al lugar que ocupa el cuidado en su filosofía del Ser.

La reflexión de Heidegger sobre la técnica moderna permite formular un humanismo tecnológico con bases en una hermenéutica ontológica de carácter crítico. Para esclarecer los impactos en el universo humano y la biosfera es esencial realizar un ejercicio que reconozca la presencia de un desocultamiento que revela la impronta que se proyecta sobre los objetos tecnológicos en la actualidad. El pensamiento heideggeriano contribuye así

desde el cuidado a un pensar meditativo que favorece el planteamiento de un humanismo tecnológico.

1.5. La antropotécnica como proyección humana

Peter Sloterdijk es de esos filósofos atrevidos en sus planteamientos y así lo demuestra en *Normas para el parque humano*. Esta obra es fruto de un texto que expuso en un seminario al poco tiempo de la muerte de Emmanuel Lévinas y que despertó un intenso debate filosófico, sobre todo con Jürgen Habermas. La conferencia tuvo lugar el 17 de julio de 1999 en el castillo de Elmau, en Baviera, con motivo del Simposio Internacional «*Jenseits des Seins / Exodus from Being / Philosophie nach Heidegger*». Sloterdijk formula esta reflexión como una respuesta a la *Carta sobre el humanismo* de Heidegger y realiza un diagnóstico sobre la capacidad crítica del humanismo que entiende como tradicional y que ha desembocado, según él, en un naufragio como escuela y proyecto de domesticación del género humano. Sloterdijk invita a pensar la posibilidad de emprender nuevos caminos vinculados a un aprovechamiento de la tecnología para la elaboración de un nuevo relato humano.

A partir de las lecturas del *Político* de Platón y de *Carta sobre el humanismo* de Heidegger, Sloterdijk considera que hay que superar el proyecto humanista y plantear un nuevo relato contextualizado en la era tecnológica, concretamente en el tiempo de la ingeniería genética. En su reflexión pone el acento en el concepto de domesticación, entendido como la técnica por la cual el ser humano ha conseguido a través de la educación y la cultura, establecer una clara diferenciación entre los del mismo género, algo que él denomina «antropotécnica».

En la base del humanismo se encuentra la tradición literaria, que se remonta al tiempo de la cultura grecorromana, la cual ha sido artífice de la construcción de los cimientos de la *humanitas* a partir de técnicas de domesticación por medio de las letras. La alfabetización conseguida gracias a la literatura sirvió como vínculo de unión entre los seres humanos para la conformación de comunidades, «a partir de entonces, los pueblos se organizaron al modo

de asociaciones forzosas de amistad completamente alfabetizadas, vinculadas bajo juramento a un canon de lecturas establecido en cada espacio nacional» (Sloterdijk, 2011: 199). Sin embargo, el tiempo literario en el que se tejían lazos de unión entre las gentes en base a las letras, se ha quedado atrás, pues la irrupción del poder mediático y de la cultura de masas transporta a un nuevo escenario, donde las bases para la conformación de sociedades ya no se sostienen sobre los pilares tradicionales del humanismo literario, sino que van más allá, caminan hacia un tiempo posliterario, hacia un poshumanismo que presenta revelaciones muy diferentes a las de un tiempo pasado. No obstante, esto no quiere decir que la literatura tenga que ser subestimada o que pierda valor, sino que su capacidad rectora ha pasado a un segundo plano. Escribe Sloterdijk:

> De ningún modo la literatura ha tocado por ello a su fin, pero se ha diferenciado completamente en forma de una subcultura *sui generis*, y los días en que era sobreestimada como portadora de los espíritus nacionales han acabado. La síntesis social no es ya –tampoco en apariencia– principalmente un asunto de libros y cartas. Los nuevos medios de la telecomunicación político-cultural han pasado entre tanto a ocupar una posición rectora y con ellos han reducido a modestas dimensiones el esquema de las amistades nacidas de la escritura (Sloterdijk, 2011: 199).

La organización educativa y domesticadora planteada por la sociedad literaria, que sirvió como soporte para las estructuras económicas y políticas, es cosa del pasado y se abre un nuevo escenario de posibilidades ante la irrupción de la tecnología y su influencia sobre los diversos ámbitos de la vida humana. Es tiempo de un nuevo humanismo. Es oportuno mencionar que las pretensiones del humanismo eran las de domesticar a un ser humano que se encontraba bestializado por motivos de diversa índole, y orientarlo por el buen camino, refiriéndose principalmente a una función educativa: «el tema latente del humanismo es, por tanto, la domesticación del hombre, y su tesis latente dice así: las lecturas adecuadas amansan» (Sloterdijk, 2001: 202). Para el humanismo, el individuo alfabetizado era el que podía recibir el privilegio de ser considerado como propiamente humano, diferenciándose así de las bestias. En ese sentido, la escritura es vista como una herramienta para establecer una clara frontera entre los seres humanos, realizándose esta separación por medio de la domesticación que es emprendida por un Amo, siendo éste el

que destina el *instrumentum vocale* para la constitución del mundo de los humanos (Duque, 2002: 122).

En este sentido, la *humanitas* del humanismo tiene una función educativa, según Sloterdijk, convirtiendo los actos en disposiciones habituales para hacer el bien, esto es, en virtudes, alejándose así de las prácticas salvajes que son propias de las masas. Esta diferenciación entre actividades humanizadas y deshumanizadoras procede del tiempo de los romanos:

> Lo que los romanos cultivados llamaron *humanitas* sería impensable sin esta exigencia de abstenerse de la cultura de masas en los teatros de la crueldad. Si alguna vez el humanista se extraviaba en medio de la multitud vociferante, sólo era para comprobar que también él es un hombre y que por ello puede verse infectado por la bestialización […] con ello queda dicho que la humanidad consiste en seleccionar para el desarrollo de la propia naturaleza los medios amansadores y prescindir de los medios desinhibitorios […] lo que se ventila con dicha cuestión es nada menos que una antropodicea, esto es, una determinación del hombre a la vida de su franqueza biológica y su ambivalencia moral (Sloterdijk, 2011: 202-203).

La crítica de Sloterdijk está dirigida a unas relaciones de dominación entre los humanos que se encuentran presentes en la obra *Político* de Platón, caracterizada por la delegación de un poder superior. En cambio, Sloterdijk recoge el testigo del Zaratustra de Nietzsche (1985) para destacar que las antropotécnicas pueden ser autoreferenciales y en ellas se despliegan procedimientos en los que unos «crían» a otros y también se crían a sí mismos:

> Cuando Zaratustra camina por la ciudad en la que todo se ha vuelto más pequeño, se da cuenta del resultado de una política de crianza exitosa e inadvertida hasta entonces: los hombres han conseguido –según le parece–, gracias a una habilidosa combinación de ética y genética, criarse a sí mismos en pequeño. Se han sometido ellos mismo a la domesticación y han puesto en marcha consigo mismo una cría selectiva orientada a una sociabilidad típica de animales domésticos (Sloterdijk, 2011: 213).

Entre líneas puede leerse lo que Sloterdijk pretende destacar, a saber, una denuncia del monopolio de la cría que es principalmente impulsada por Nietzsche y su crítica cultural, y donde este monopolio de la cría está gestionado por sacerdotes y profesores. Sloterdijk hace esto con la intención de rescatar el espíritu de una crítica a la cultura humanista para plantear la posibilidad de ver más allá de la misma. Es cierto que el papel de la lectura fue muy importante para los pueblos, pero tras ese poder literario se escondía un poder de domesticación del ser humano que el humanismo no ha sabido enfrentar para que no derive en una selección excluyente. Ser objeto de selección implica que el ser humano no adquiera un papel protagónico y soberano en el terreno de su florecimiento.

A partir del *Político* y la *República* de Platón se forjó un discurso y una forma de pensar en la que la comunidad humana era vista como un parque zoológico, donde el espacio y la vida estaban gestionados por la zoopolítica con un pastor que controla el rebaño. La política se trata como un conjunto de reglas que proporcionan una organización a ese parque de humanos. Para Sloterdijk los seres humanos poseen una dignidad caracterizada por un poder soberano para cuidar de sí mismos, generando así un sentimiento de comunidad al margen del lugar en el que se encuentren ubicados, tratándose de un automantenimiento:

> Si hay una dignidad humana que merezca ser expresada por la meditación filosófica, ello es debido a que los hombres no son sólo mantenidos en los parques temáticos políticos, sino que ellos mismos se mantienen en aquéllos. Los hombres son seres que se protegen y se cuidan a sí mismos, que –independientemente de dónde vivan– generan en torno a ellos un efecto de parque. Ya sea en parques urbanos, nacionales, cantonales o ecológicos, en todas partes los hombres tienen que formarse una opinión sobre el modo en el que quepa regular su automantenimiento (Sloterdijk, 2011: 217).

Pero la fórmula leer-educar-domesticar ha perdido su vigencia con la aparición de nuevas formas de comunicación y tecnología, o como señala Duque, «han perdido su función epistolar y, por ende, humanista» (2002: 127). Se abre paso la era poshumanista. Siguiendo el hilo de este nuevo tiempo y de la justificación de la necesidad de asumirlo, Sloterdijk rescata la reflexión de Heidegger en la *Carta sobre el humanismo* y realza la

importancia de cuestionar el paradigma tradicional de humanismo, cegado por la idea de que el ser humano y su autorrepresentación filosófica en el humanismo significan una solución para los grandes retos que enfrenta el mundo. Es necesario un cuestionamiento del concepto humanista, lo que posiblemente conduzca a la renuncia del concepto tradicional que ha supuesto serias catástrofes como las dos guerras mundiales.

El humanismo europeo y, en concreto, la modernidad ilustrada parten del principio del ser humano como centralidad. Es precisamente este principio de centralidad el objeto de reflexión de la *Carta sobre el humanismo*. En lo referente al legado de Heidegger, Sloterdijk está de acuerdo con el antiguo rector de la Universidad de Friburgo en que el proyecto humanista ha fracasado, pues las guerras mundiales, el Holocausto y las bombas atómicas parecen no haber sido acontecimientos históricos lo suficientemente convenientes para proporcionar enseñanzas:

> […] ¿qué domestica o educa todavía al hombre cuando fracasa el humanismo como escuela de modelar al ser humano? ¿Qué domestica o educa al hombre cuando sus anteriores esfuerzos por domesticarse a sí mismo le han conducido principalmente a tomar el poder sobre todo lo ente? ¿Qué domestica o educa al hombre cuando, tras todos los experimentos anteriores de educar al género humano, sigue sin estar claro quién o qué educa a los educadores y para qué? ¿O es que la pregunta por el cuidado y el modelado del hombre ya no puede plantearse, de forma competente, en el marco de las meras teorías de la domesticación y la educación? (Sloterdijk, 2011: 209).

En ese sentido, la aportación de Heidegger supone para Sloterdijk un punto de inflexión desde el que repensar el proyecto humanista y desde el que arrojar una mirada vigilante para la búsqueda de claridad cognitiva y de esfuerzo para cuestionar los presupuestos sobre los que se erige. Para Heidegger la auténtica esencia del ser humano reside en la verdad, en su relación con el Ser, pensando al humano como claro y pastor del Ser, con una propiedad lingüística que le sirve para esclarecer el Ser y decirlo. Como Esquirol señala: «Sloterdijk ve ahí un intento de apaciguamiento más ambicioso todavía que el humanista. El habitar heideggeriano evocaría «un atento acercamiento del oído para el cual el hombre tiene que ser más silencioso y dócil de lo que es el humanista leyendo a sus clásicos» (2011: 183). No

obstante, Sloterdijk ve demasiada limitación en el planteamiento de Heidegger para los nuevos tiempos, y lo que recoge de él es la capacidad para plantear la necesidad de ir más allá del humanismo tradicional. Se sitúa en la senda de su lenguaje cuando utiliza »el claro del bosque», observando al ser humano despejado para una interpretación biologicista, donde habla de la posibilidad de un nacimiento prematuro y fracasado. El claro sería esa casa del lenguaje donde el ser humano cobija al ser y se cobija, además de generar una agrupación de casas donde los seres humanos se cobijan y amansan a sí mismos. Esta visión de Sloterdijk supone una diferenciación con la interpretación ontológica heideggeriana, pues en su caso se acerca más a lo óntico y biológico, y hay que contextualizarla en el actual escenario caracterizado por el impresionante poder tecnológico. El universo de posibilidades que brinda la tecnología, y concretamente la biotecnología, representa una pragmática vía desde la que desarrollar nuevas antropotécnicas a través de los postulados transhumanistas. Este universo de posibilidades exige un código de las antropotécnicas para que el ser humano participe consciente y activamente en el cultivo de la responsabilidad.

No podría entenderse el pensamiento de Sloterdijk sin su disputa filosófica con Habermas. Es abundante la bibliografía sobre su enfrentamiento intelectual a raíz del coloquio celebrado en el castillo de Elmau, Baviera, en julio de 1999. Para poder entender esta discusión es importante conocer la amplitud del contexto en el que se desarrolló, teniendo en cuenta las críticas que Sloterdijk había vertido anteriormente contra la teoría de la acción comunicativa de Habermas, hasta el punto de referirse a la misma como una religión civil. En parte, Sloterdijk llevaba razón, la teoría de la acción comunicativa habermasiana había llegado en un momento en el que la sociedad alemana de los años 60 demandaba cierta paz religiosa en el campo de las ciencias sociales y fue precisamente lo que consiguió con un discurso ético-religioso que servía para dulcificar la teoría crítica tradicional. Para Sloterdijk la filosofía de Habermas se encuentra cargada de contradicciones, pues el segundo dice renunciar a toda forma metafísica y religiosa, aunque esas formas estarían implícitas en sus postulados.

Otro aspecto a destacar en la discusión tiene que ver con el uso del lenguaje empleado por parte de Sloterdijk, quien, como se ha podido comprobar a lo largo de las últimas páginas, emplea términos como «domesticación y cría de los seres humanos», «selección», «rebaño», «pastor» etc., un lenguaje que en el contexto alemán implica la existencia de un cierto lastre debido a la cercanía con el nacionalsocialismo. La utilización de ciertos términos y esta línea argumentativa es de lo que se sirvió Habermas, que en aquel tiempo era un ícono de la conciencia antinazi de Alemania, para lanzar su crítica sobre Sloterdijk, acusándolo de estar cercano al pensamiento nazi.

Ante las graves acusaciones, Sloterdijk se defendió en una entrevista para el diario *La Nación* de Buenos Aires. En esta entrevista Sloterdijk sostiene que el eugenismo era una tendencia cultural que formaba parte del pensamiento moderno, identificando así sus bases en el progresismo. Además, Sloterdijk considera que el eugenismo es una idea de la izquierda clásica practicada cotidianamente, y que lo que la derecha fascista realizó fue un exterminio racista, de modo que eugenesia y exterminio no guardarían ninguna similitud. Por lo tanto, el eugenismo sería un procedimiento de reflexión orientado a mejorar las condiciones de la próxima generación, según Sloterdijk.

Más allá de las polémicas provocadas por el discurso de Sloterdijk, no debería menospreciarse la reflexión que presenta como respuesta a la *Carta sobre el humanismo* de Heidegger, pues constituye una importante reflexión sobre las complejas relaciones que el ser humano tiene con la técnica y las posibilidades que ésta le proporciona para intervenir en el desarrollo futuro de la especie. La experiencia histórica reciente ha demostrado que el proyecto humanista inaugurado por Platón ha resultado ser un fracaso y que es necesario establecer un nuevo relato humanista que no pierda de vista el universo tecnológico. La tecnología representa una oportunidad ineludible de enriquecimiento humano, donde el florecimiento de las potencialidades del ser humano como especie puedan verse impulsadas en el presente y futuro.

1.6. Impulso de un tiempo

Retomando la aclaración que se hacía al comienzo de este capítulo, el humanismo tecnológico no asume las tesis de aquellos discursos que versan sobre la deshumanización de la técnica, reivindicando la supuesta existencia de una esencia humana inmutable. El discurso de la deshumanización de la técnica suele implicar una demonización de ésta y la exclusión de ciertas responsabilidades. El impulso para formular un humanismo tecnológico nace, no de una mirada sobre la técnica vista como amenaza, ni tampoco de la pretensión de su humanización, sino de un reconocimiento de la misma como un universo de posibilidades que hay que enfrentar desde los límites humanos para cultivar la responsabilidad en un contexto cívico y democrático.

El humanismo tecnológico parte de una reflexión sobre la teoría de las dos culturas. Charles Percy Snow (2013) sugiere con optimismo la necesidad de construir un vínculo entre científicos y literatos en beneficio de la humanidad. La propuesta de encuentro cultural de Snow radica en la urgencia de transformar la educación de su tiempo para adecuarla a las nuevas posibilidades y retos que plantea la revolución científica. Frente a Snow, es importante señalar que John Brockman (1996) tomó distancia de la propuesta del primero, pues entendió que la conciliación de las dos culturas no puede circunscribirse exclusivamente al ejercicio comunicativo y divulgativo de los científicos para el gran público. Pues en realidad hay aspectos propios de las humanidades sobre los cuales la tradición científica debería reflexionar para enriquecer su actividad, otorgándoles un espacio que evite la colonización cognoscitiva y cultivando un carácter más plural en términos epistemológicos y metodológicos. Frente a una tradición de pensamiento que afirma la existencia de dos culturas, separando de forma errónea y tajante las ciencias de las humanidades, como si de dos dimensiones del conocimiento humano se trataran, hay que señalar que no existen dos culturas diferenciadas, como si una perteneciera al conocimiento científico y otra al conocimiento humanista vinculado con aquello que se denomina «tradición» (Huxley, 2017). Según esta visión las humanidades mirarían al pasado y serían las guardianas de la tradición, mientras que las ciencias se encargarían de mirar al futuro y

asumir la tarea del progreso. Concebir las ciencias y las humanidades como compartimentos separados, ajenos y cerrados, representa un serio obstáculo para el humanismo tecnológico. Este humanismo reconoce al ser humano en todas sus dimensiones, pues el conocimiento es un elemento esencial del ser humano y no puede considerarse de forma compartimentada, sino integral.

El ser humano no debe prescindir de ninguna de sus dimensiones para perseguir el enriquecimiento de su condición, pues todas las partes se encuentran imbricadas; es un ser integral, donde las partes están interrelacionadas y sus condiciones son variadas. Es importante reconocer este tema de las dimensiones imbricadas para no caer en la simplicidad de desestimar ciertos saberes que también forman parte del conocimiento y la existencia. Tampoco debe esquivarse la idea de que el ser humano desarrolla un ejercicio hermenéutico para poder entender la realidad a la que se enfrenta, y es precisamente en ese ejercicio donde surge un importante vínculo entre los saberes que el humano ha formulado para dotarse de una condición de integridad.

Los importantes y profundos avances que ha experimentado el campo de la tecnología se sitúan frente a un escenario novedoso para la humanidad que requiere de una contextualización del humanismo. Este proyecto humanista no se entiende como una recreación de los humanismos del pasado, sino que asume la responsabilidad de un nuevo tiempo, representando un compromiso ante el fenómeno tecnológico. Se trata de un nuevo espíritu que es impulsado en medio de la oleada tecnológica. Por ello es fundamental la consideración de este humanismo como un compromiso de responsabilidad ante los desafíos tecnológicos, y concretamente ante los retos de la IA, que exigen la puesta en valor de lo humano. A propósito de la contextualización del humanismo en el tiempo actual, José Mª Aguirre Romero señala lo siguiente:

> No quiero entender el Humanismo y la tradición humanista como una invención o recreación cultural del pasado, sino como una apertura a nuevas situaciones, a nuevos espacios que se abrían ante los hombres. No quiero pensarlo como un compromiso con los tiempos antiguos, sino como un compromiso con su propio tiempo. Prefiero considerar el Humanismo más

> como un impulso que como un depósito, más como una energía que como un cúmulo de conocimientos eruditos (2002: 8).

El humanismo siempre ha nacido de impulsos motivados por los contextos concretos, fuertemente marcados por el espacio y el tiempo, por ejemplo, el humanismo renacentista se caracteriza por una apertura grandiosa del conocimiento con la explosión de las artes y las ciencias. Este impulso humanista nace del deseo de buscar respuestas y enfrentar ese sonambulismo tecnológico (Winner, 2008) que cada vez está más presente en la vida. Es un impulso hacia el esclarecimiento para la búsqueda de posibilidades que permitan un florecimiento con miras al futuro, pero sin olvidar el legado histórico como reconocimiento. El pasado no puede representar un obstáculo desde el que reformular el presente y construir el futuro a partir de una mirada posibilitadora que arroje luz a este tiempo tecnológico en beneficio del ser humano.

Este humanismo debe ir acompañado de una superación del discurso de las dos culturas, que dificulta una formación integral del ser humano y promueve seres fragmentarios y epistemes con carencias comunicativas y de encuentro (Aguirre Romero, 2002). El humanismo se aleja de su razón de ser cuando renuncia a sus pretensiones de integridad y a la totalidad como horizonte, que es precisamente donde la tecnología tiene mucho que aportar en favor del proyecto de un nuevo humanismo.

1.7. El infranqueable reconocimiento de los límites

Se han mostrado diversos relatos acerca de la relación del ser humano con la tecnología que pueden servir como condición de posibilidad para construir un humanismo tecnológico responsable con el futuro. Las reflexiones expuestas sirven como punto de referencia y nunca como una estricta determinación, pues contribuyen a esclarecer un importante aspecto de la condición humana.

Retomando lo expuesto por Ortega en *Meditación de la técnica*, podría afirmarse que el ser humano posee un gran poder imaginario que permite ir más allá de la realidad presente para plantear un abanico de posibilidades. José Luis Molinuevo acierta cuando, recuperando la alegoría de la caverna de Platón, afirma que «somos productores y consumidores de imágenes» (2004: 170). Esta afirmación significa que la vida se encuentra en la frontera entre dos mundos que se funden entre sí gracias a la posibilidad, el mundo real y el mundo de la imaginación.

El humanismo tecnológico es un humanismo que reconoce los límites porque es consciente del imperativo de la responsabilidad. Es limitado en el sentido en que orienta su actividad tecnológica de forma responsable, incorporando la ética en sus procedimientos, evitando así la superación de unos límites que ponen en riesgo aquellos valores y objetos que son esenciales para la vida. Es un humanismo limitado porque es más intensivo que extensivo, ya que el interés se encuentra orientado hacia el henchimiento y no hacia el capricho ilimitado e impulsivo (Molinuevo, 2004: 170). Asume su finitud y reconoce los límites, comprometiéndose con una responsabilidad consciente. Aquello que se designa como «humano» puede ser reescrito dentro de unos límites tecnológicos que asuman los desafíos con conciencia y compromiso, para que aquel relato que intente promover la imaginación incorpore ineludiblemente premisas éticas. En ese sentido, algo de razón llevaba Sartre cuando sostenía que el ser humano es libre para hacerse y que eso implica una responsabilidad de la que no puede despojarse (Bello Reguera, 2007: 43). También debería recordarse a Ortega, pues no hay que olvidar que el filósofo español insiste a lo largo de su obra en que el humano se hace a partir de unas circunstancias, y con él podría afirmarse que las circunstancias empujan a la toma de conciencia de los límites y a la orientación de la acción en función de los mismos. El humanismo tecnológico representa una visión realista, pues reconoce los límites de la condición y ciertas distopías en la acción tecnológica, lo que en el contexto de este trabajo significaría promover un pensamiento sobre la necesidad de asumir responsabilidad frente al poder subyacente de la técnica.

Este humanismo no es idealista, pues no parte de la totalidad de la humanidad como premisa fundante, sino del reconocimiento de un individuo, situándolo en una posición relacional, alejándose de la visión de sustancia o subsistencia, propia del cartesianismo (Molinuevo, 2004: 178). El ser humano es un ser necesitado, por eso el humanismo es del individuo, pues reconoce una necesidad relacional e identifica en la tecnología una oportunidad de enriquecimiento existencial. Frente a la postura de un humanismo relacional y reconocedor de sus necesidades, se encuentra un humanismo anticuado, de otro tiempo, que es el causante de los grandes males que la humanidad ha sufrido, precisamente por no reconocer al ser humano como un ser necesitado, sino más bien como un ser auténtico y esencialista.

Antes se han mencionado los límites como una de las fronteras características del ser humano. Pues bien, es precisamente en la observación de esos límites donde se encuentra el reconocimiento de la necesidad y la búsqueda de relaciones. Molinuevo también habla de un humanismo del límite, siguiendo a Eugenio Trías (2000). Esta condición limítrofe allana el camino para hablar de un humanismo tecnológico. El distanciamiento del humanismo esencialista e idealista permite abrir la puerta a una perspectiva que reconoce las limitaciones. El conocimiento de los límites no radica en un soporte para proyectos que asuman la necesidad de llegar a los límites de las posibilidades, sino más bien para aprovechar las posibilidades que la condición de seres limitados nos ofrece. Es en ese punto donde reside el momento en el que el ser humano se convierte en ser técnico y comienza a ver la posibilidad de una sobrenaturaleza, como diría Ortega (1965). Así pues, la tecnología invita a repensar la realidad y a observar nuevas posibilidades de realización de la existencia en la búsqueda de un beneficio humano.

Se ha destacado que el humanismo tecnológico es un humanismo reconocedor del límite y a la vez de su condición condicionada, es decir, de su carácter de necesidad que le invita a imaginar lo que puede ser. Esta invitación permite entender que los límites no se encuentran situados en un plano negativo, sino más bien son un avistamiento de los beneficios de la técnica como espacio desde el que hacer posible el florecimiento humano a partir de la asunción de responsabilidad.

Pese a los enormes beneficios que ofrece la tecnología para el ser humano, también es importante destacar la noción de límite que debe estar presente en toda creación. Los beneficios tecnológicos suelen ir acompañados de unos costes, por lo que es fundamental considerar aquellos límites que pueden contribuir a valorar los impactos y a reflexionar sobre las implicaciones de la acción tecnológica. En ese sentido, y en relación a los beneficios y costes, Friedrich Rapp sostiene lo siguiente:

> Sin embargo, tales beneficios tienen sus costos inevitables. Es este costo el que en las naciones industrializadas a menudo se experimenta como una carga. Las quejas son muy conocidas: problemas ecológicos, agotamiento de recursos, carrera armamentista y la sensación de alienación en un mundo conformado por una división del trabajo altamente racionalizada.
>
> Se introdujo la tecnología moderna para hacer la vida más fácil. Pero, debido a su carácter concreto y objetivo, y porque en última instancia refleja la estructura del mundo físico, la tecnología moderna inevitablemente conduce a nuevas limitaciones en la libertad y la acción humana. Debido a que el dominio del entorno físico requiere ajustarse a los procesos tecnológicos puestos en acción, ahora la lógica interna del mundo tecnológico reemplaza las limitaciones anteriores. La adaptación a los principios de la acción tecnológica, junto con los efectos alienantes que inevitablemente producen, es un precio a pagar por los beneficios de la tecnología moderna (1985: 431).

La conciencia del límite empuja a poder ser para mantenerse a flote frente a las circunstancias de la vida. La tecnología pone de relieve las limitaciones del ser humano, y permite un reconocimiento, brindando la posibilidad de imaginar, de poder ser. Ahí es donde radica la necesidad de un humanismo tecnológico, en esa visión optimista de la tecnología, pero no acrítica, donde los inventos tecnológicos presentan un universo de posibilidades apremiantes para la existencia. Frente a este optimismo crítico tecnológico, también es necesario dar cuenta de que existen productos que no benefician al ser humano, y a los que hay que anteponer un humanismo tecnológico, no con las pretensiones de humanizar la técnica, o de que las humanidades fagociten el mundo tecnológico, sino más bien como reorientación de las intencionalidades que se encuentran detrás de determinados impulsos técnicos. En definitiva, se trata de reorientar la acción tecnológica en función de

un nuevo humanismo que no intenta representar un ser humano de corte idealista que domine el mundo, sino un ser respetuoso con la realidad, responsable y consciente de sus límites y posibilidades en función de las circunstancias, que reconozca en la tecnología su ser marginal para proyectarse en ella y obtener el mayor beneficio con responsabilidad y sentido del límite. Como señala Molinuevo: «no se trata de ir con las tecnologías al límite de las posibilidades humanas, sino de extraer las posibilidades del límite humano» (2004: 190).

1.8. La importancia del contexto y su comprensión

El humanismo tecnológico se construye en base a diversas apreciaciones filosóficas que han sido consideradas relevantes en la exposición de este primer capítulo. En esa línea, el pensamiento de Gilbert Simondon representa un importante soporte del humanismo tecnológico, ya que concede prioridad al reconocimiento de los procesos que constituyen la individuación. Además, este planteamiento nos permite esclarecer en muchos casos las implicaciones que pueden llegar a tener determinadas tecnologías para la vida, y más aún, tener en cuenta el potencial de la IA.

La filosofía simondiana sostiene que el esquema subjetivista, que separa radicalmente sujeto de objeto, subjetividad de objetividad, podría ser cuestionado a la luz de los objetos técnicos, pues la historia ha demostrado que el sujeto pensante no representa siempre un punto de partida sólido, ya que su intencionalidad se encuentra también determinada por aquellos objetos técnicos a los que supuestamente pretende determinar. Se trataría de un proceso relacional donde sujeto y objeto se determinan mutuamente. Escribe Simondon:

> El método consiste en no intentar componer la esencia de una realidad mediante una relación conceptual entre dos términos extremos, y en considerar toda verdadera relación como teniendo rango de ser. La relación es una modalidad del ser; es simultánea respecto a los términos cuya existencia asegura. Una relación debe ser captada como relación en el ser, relación del ser, manera del ser y no simple relación entre dos términos a los que podríamos conocer adecuadamente mediante conceptos ya que tendrían una efectiva existencia separada. Es porque los términos son concebidos como sustancias que la relación es relación entre

términos, y el ser es separado en términos porque es primitivamente, anteriormente a todo examen de individuación, concebido como sustancia (Simondon, 2009: 37).

El conocimiento de esta constitución relacional se torna necesario en un mundo en el que la IA está transformando diversas esferas de la vida, suponiendo que el humanismo tecnológico tenga que hacer frente a diversos desafíos. La identificación de este proceso relacional y la determinación dialéctica de la individuación permite la toma de conciencia de las profundas transformaciones que la tecnología representa para la existencia humana y de los usos que pueden hacerse de la misma, tanto desde una perspectiva beneficiosa para la ciudadanía, como perjudicial.

Los contextos son determinantes y constituyen tanto lo humano como lo no humano. Debe asumirse la idea que afirma que el sujeto no determina enteramente al objeto, pues ahí también se encuentra presente, y ejerce influencia, un medio que es condicionante. La constitución ontológica de la tecnología no responde únicamente a la intencionalidad del creador, ya que se inserta en un medio condicionante. Para vislumbrar las fuerzas e intencionalidades que se encuentran tras la IA, es necesario reconocer que son diversos los factores que influyen en la constitución. Más allá de la simple relación sujeto y objeto, subjetivo y objetivo, hay más aspectos que demandan ser considerados, pues la individuación es fruto de un proceso relacional. La individuación representa el acto de reconocimiento de la existencia de un medio asociado, a la vez natural y técnico, que incide, condiciona, y/o determina la constitución de los objetos técnicos, que provocan una interacción con los sujetos. El ser técnico se autocondiciona y a la vez condiciona a los sujetos, generando un conjunto de posibilidades de interacción entre el mundo y el objeto, entre el objeto y el mundo, el humano y el objeto, el objeto y el humano, el humano y el mundo, etc., que en definitiva se refieren a la pluralidad de factores que emergen en la constitución del ser técnico. Lo dicho hasta ahora significa que el ser de la técnica no es impuesto exclusivamente por el humano, como si de una esencia estática se tratara, sino que el devenir va marcando su concreción en base a diversidad de factores. Este reconocimiento permite poner de relieve que la tecnología puede ser un mecanismo beneficioso para el ser humano que es consciente de los contextos y sus limitaciones.

La comprensión de la técnica no puede desligarse del mundo humano, pues es una expresión de su relación con la realidad. Cuando crea objetos técnicos ejerce una función inventiva que supone una anticipación y un acercamiento a la vida. La creación de objetos técnicos es una invención de vida, una obra, una suerte de incidencia sobre la realidad en la creación de nuevas formas que cuentan con finalidades, casi siempre, meditadas e intencionales, pero que también se van constituyendo en el devenir plural.

1.9. Democracia *versus* tecnocracia

Desde la segunda mitad del siglo XX la tecnocracia se ha venido forjando como un paradigma que ha colonizado diversas instituciones humanas. Lo técnico ha prevalecido sobre lo verdaderamente humano y razonable. Jordi Pigem define qué es la tecnocracia:

> En los últimos treinta años la democracia ha ido siendo desplazada por la tecnocracia. La tecnocracia, como su nombre indica, es el control de la economía y de la sociedad a partir de criterios no humanos, sino exclusivamente técnicos. Tal como un ordenador funciona aplicando algoritmos (secuencias de reglas y cálculos), la tecnocracia solo atiende a modelos abstractos y secuencias de fórmulas y estadísticas. […] La tecnocracia presume de eficiencia. A corto plazo y en ámbitos estrictamente cuantificables, parece obtenerla. A largo plazo y desde una perspectiva más amplia, vemos que no. Al reducirlo todo a abstracciones, pierde de vista el mundo real y en vez de eficiencia genera negligencia (Pigem, 2013: 16 y 25-26).

El paradigma tecnocrático, dominante en la actualidad, de fundamento positivista y cientificista, afirma que la solución a los grandes males de la humanidad ocasionados por el poder exacerbado de la técnica radica en la propia técnica. Esta forma de pensar tecnocrática sugiere que la técnica es una suerte de antídoto para todos los problemas, pues entiende que no existe otra alternativa. Sin embargo, la superación de los retos humanos vinculados con las diversas esferas de la vida no vendrá precisamente de la tecnología, como señala el Papa Francisco en la encíclica *Laudato si'*:

> Buscar solo un remedio técnico a cada problema ambiental que surja es aislar cosas que en la realidad están entrelazadas y esconder los verdaderos y más profundos problemas del sistema mundial (2015: 35).
>
> Cualquier solución técnica que pretendan aportar las ciencias será impotente para resolver los graves problemas del mundo si la humanidad pierde su rumbo (2015: 62).
>
> Las soluciones meramente técnicas corren el riesgo de atender a síntomas que no responden a las problemáticas más profundas (2015: 46).

La complejidad de los desafíos actuales debe ser afrontada desde una superación del paradigma estrictamente tecnocrático, pues un posicionamiento exacerbado en esa línea puede suponer una suerte de fundamentalismo arraigado en la deshumanización y artificialidad para la búsqueda de soluciones pragmáticas. Por ello es fundamental reconocer que los problemas del mundo no pueden explicarse de forma estrictamente técnica, sino que requieren de conocimientos entrelazados, entre los que necesariamente deben encontrarse las humanidades, que contribuyen a fomentar el pensamiento crítico (Nussbaum, 2010). En ese sentido los desafíos exigen la formulación de preguntas de largo alcance que vayan más allá de la especialización exacerbada en lo técnico, en definitiva, precisan de una hermenéutica crítica aplicada a los entornos de generación de conocimiento científico en el ámbito tecnológico.

Es razonable que la tecnología contribuya a hacer la vida más fácil, pero la imposición del paradigma tecnocrático insinúa una desconfianza a ese respecto:

> El paradigma tecnocrático, al poner datos y cosas por delante de las personas, nos presenta una realidad unidimensional, empobrecida, ajena y enajenada, mientras eclipsa la realidad del mundo vivo y de la interioridad. Darnos cuenta de ella no es una invitación a regresar al pasado, sino a abrirnos a una nueva interpretación del ser que sea coherente con nuestra experiencia (Pigem, 2018: 107).

Así pues, el enriquecimiento de la experiencia humana en el contexto tecnológico debe originarse en un humanismo tecnológico que observe en la cultura democrática una oportunidad para impulsar una superación del paradigma tecnocrático. Se trata pues de

enriquecer la experiencia a partir del cultivo de habilidades cívicas y democráticas en los entornos tecnológicos para plantear una alternativa a la reificación tecnocrática:

> Es la tendencia a reducir todo lo que se manifiesta en nuestra experiencia a simples objetos cuantificables (y, ahora, digitalizables), controlables y manipulables. La reificación, que empieza en la menta y en la mirada, nos exilia de un mundo de cualidades, relaciones y matices, intrínsecamente dinámico y vibrante, para llevarnos a un mundo de cosas y cifras, un mundo cuadriculado y estático. Nuestra experiencia de la realidad, y la realidad misma, se vuelve más pobre y más controlable. La lógica tecnocrática aplana la profundidad de la presencia de las personas y de todos los seres (Pigem, 2018: 93).

Frente a esta mirada reduccionista y mecanicista de la tecnología que conduce a una violencia epistemológica en el acto de conocer el mundo, es necesario anteponer una visión democrática que permita restaurar y enriquecer el pensamiento científico, una democracia tecnológica que favorezca los procesos de generación de conocimiento en el entorno de la IA para fortalecer la búsqueda de respuestas a las problemáticas y desafíos del ser humano.

1.10. Un proceder pragmático y fronético

El humanismo tecnológico posee una raíz pragmática y fronética, donde el conocimiento teje un hilo conductor con la experiencia en la búsqueda de la resolución de problemas. Este humanismo se fundamenta, en cierto modo, en la experiencia humana donde la técnica representa un elemento constitutivo, según John Dewey (1948). Debido a su condición técnica, el ser humano puede orientar las creaciones tecnológicas hacia la resolución de problemas. Es importante proponer una fundamentación pragmática de la técnica, que Dewey describió en 1930 de la siguiente manera:

> «Tecnología« significa todas las técnicas inteligentes por las que las energías de la naturaleza y del hombre son dirigidas y utilizadas en la satisfacción de las necesidades humanas, no se puede limitar a unas pocas, exteriores y relativamente mecánicas formas. A la vista de sus posibilidades, la concepción tradicional de experiencias es obsoleta (Dewey, 1984: 277).

Una tecnología al servicio del beneficio humano destaca la dimensión pragmática del humanismo tecnológico en términos vitales. En ese sentido, la condición técnica dispone la acción tecnológica generando mecanismos que beneficien el camino para el provecho humano. Además, como hay una invitación evidente a la utilidad, las acciones tecnológicas tendrían que incorporar criterios de responsabilidad para no desviarse del pensamiento útil. No obstante, cuando se menciona la utilidad no se está pensando en ningún momento en una utilidad propia de la esfera de la racionalidad instrumental, que tanto criticaban Theodor W. Adorno y Max Horkeimer (2010; 2016). Las herramientas tecnológicas pueden servir de impulso hacia la realización de aquellos proyectos considerados beneficiosos para la existencia humana y la protección de toda forma de vida.

La formación de ideas colectivas contribuye a una cultura de encuentro entre diversas perspectivas, lo que a todas luces ha sido un acierto a lo largo de la historia. La tecnología incide en los procesos cognitivos del ser humano y en el diseño de redes de formación de conocimiento colectivo por medio de intelectos sintéticos. Estas redes de conocimiento podrían cultivar nuevas habilidades cívicas y educativas que permitieran fortalecer el nacimiento de ideas innovadoras capaces de ofrecer respuestas a los desafíos que enfrenta la ciudadanía. La confluencia respetuosa de perspectivas permite la articulación de actitudes de búsqueda de soluciones y alternativas de una manera más audaz y en ese sentido la IA podría tender puentes, lo que representaría un beneficio para la humanidad. La clave está en la búsqueda de mecanismos que sirvan de bisagra de unión entre los seres humanos, corrigiendo aquellos productos tecnológicos que dificultan la construcción de redes de conocimiento y que tienden al aislamiento de los sujetos. La IA puede estar al servicio de la formación de una inteligencia colectiva para el beneficio de la ciudadanía. Este es precisamente un punto de inspiración para la formación educativa de las próximas décadas, una inteligencia artificial responsable con lo cívico que acompañe los procesos de aumento de la inteligencia colectiva, creando así una nueva cultura de conocimiento innovador, en vez de ahondar más en el problema del individualismo y el aislamiento humano. Además, la inteligencia colectiva promueve aquellas ideas que sirven para

combatir conocimientos considerados absolutos y uniformes, lo que a todas luces representa un obstáculo para una ciencia democrática.

También es importante destacar la dimensión fronética del pensamiento humanista tecnológico, pues contribuye a una sensibilización de los procesos de generación de conocimiento. Lo fronético prioriza la implicación de aquellos valores que influyen en los juicios y decisiones, y se sitúan más allá del conocimiento científico y técnico. En este sentido, son muy relevantes las contribuciones de Bent Flyvbjerg a una ciencia social fronética, pues han permitido establecer confluencias entre las humanidades y las ciencias sociales, con el objetivo de superar la cultura epistémica que las segundas han adoptado de las ciencias naturales, para situarse en un terreno fronético que favorezca un acceso más amplio e integral al conocimiento, incorporando otras variables que habitualmente no son consideradas desde una perspectiva técnica (Flyvbjerg, 2001; 2004; 2006a; 2006b).

El humanismo tecnológico también se encuentra en una posición de servicio a las humanidades, como saberes que están siendo olvidados y desplazados en las últimas décadas (Nussbaum, 2010) y que pueden verse fortalecidos a través de los nuevos medios tecnológicos. Este humanismo se escribe sobre un nuevo relato en el que se incorporan perspectivas filosóficas que sirven para enriquecer y ampliar la razonabilidad y pensar la condición humana. Una pérdida de la hegemonía de lo estrictamente positivo e instrumental requiere de la búsqueda de un basamento en la reflexión filosófica. Una de las claves para posibilitar un humanismo tecnológico se encuentra en la superación de la racionalidad sobre la que se ha construido el mundo y que ha tratado de responder a la pregunta acerca de qué es el ser humano.

1.11. Integración de las humanidades con los saberes científicos y tecnológicos

Hoy en día los titulares sobre educación se centran principalmente en las ciencias y la tecnología. El campo de la computación, la IA, la robótica, la programación, las matemáticas, etc., han adquirido un gran protagonismo para el futuro de las sociedades. Resulta inevitable reconocer la importancia de estos campos de estudio para el progreso de

las sociedad. Sin embargo, ¿qué ha pasado con las humanidades? ¿no tienen nada que aportar? En el contexto democrático los saberes científicos y tecnológicos no son suficientes para la formación de un *ethos* cívico. En ese sentido, es importante destacar las siguientes palabras de Martha Nussbaum:

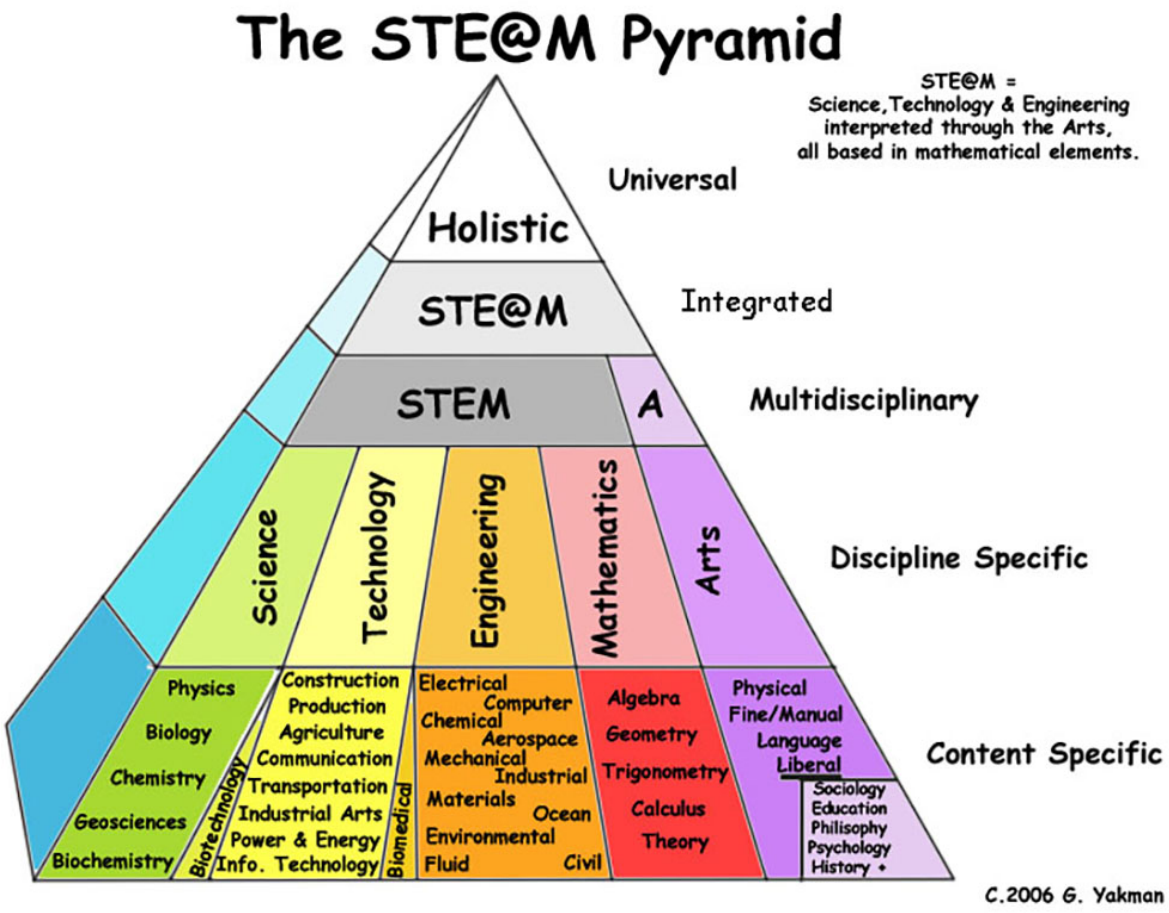
> La idea de la rentabilidad convence a numerosos dirigentes de que la ciencia y la tecnología son fundamentales para la salud de sus naciones en el futuro. Si bien no hay nada que objetarle a la buena calidad educativa en materia de ciencia y tecnología ni se puede afirmar que los países deban dejar de mejorar esos campos, me preocupa que otras capacidades igualmente fundamentales corran riesgo de perderse en el trajín de la competitividad, pues se trata de capacidades vitales para la salud de cualquier democracia y para la creación de una cultura internacional digna que pueda afrontar de manera constructiva los problemas más acuciantes del mundo.
>
> Estas capacidades se vinculan con las artes y con las humanidades. Nos referimos a la capacidad de desarrollar un pensamiento crítico; la capacidad de trascender las lealtades nacionales y de afrontar los problemas internacionales como «ciudadanos del mundo»; por último, la capacidad de imaginar con compasión las dificultades del prójimo.
>
> […] A mi juicio, cultivar la capacidad de reflexión y pensamiento crítico es fundamental para mantener a la democracia con vida y en estado de alerta. La facultad de pensar idóneamente sobre una gran variedad de culturas, grupos y naciones en el contexto de la economía global y de las numerosas interacciones entre grupos y países resulta esencial para que la democracia pueda afrontar de manera responsable los problemas que sufrimos hoy como integrantes de un mundo caracterizado por la interdependencia (2010: 25-30).

Las sociedades se encuentran ante una encrucijada decisiva para su futuro, pues la tecnología puede conducir a más democratización o a más tecnocracia (Garcés, 2017; 2019). A sabiendas de esta situación es fundamental promover un humanismo tecnológico por medio de la integración de los saberes humanísticos con los tecnocientíficos. Así pues, las humanidades podrían enriquecer otros saberes predominantes en este tiempo y que se engloban dentro de las siglas STEM (siglas en inglés de *Sciencie, Technology, Engineering and Mathematics*). Por jeemplo, David A. Sousa y Tom Pilecky (2013), así como Georgette Yakman (2006; 2011a; 2011b; 2016) destacan el valor de introducir las artes en el contexto

STEM como una alternativa enriquecedora para el conocimiento y la resolución de problemas complejos y trascendentales para el ser humano. Además, la contribución de las humanidades en el terreno científico y tecnológico podría consistir en la suma del diálogo como una poderosa herramienta. En este sentido, Agustín Domingo Moratalla destaca el valor de las humanidades para iniciar el diálogo público sobre aquellos asuntos que afectan y resultan de interés para la ciudadanía (2007: 118).

Es preciso señalar que las siglas STEM comenzaron a utilizarse a finales de los años 90 por la *National Science Foundation* (NSF) de los EE. UU., aunque en un principio no tenía una esencia e implicación interdisciplinaria, sino que se limitaba a agrupar campos del conocimiento. Según Mark Sanders (2006; 2009), la educación integral ofrecida por el contexto STEM ofrece enfoques de aprendizaje fundamentados en el diseño tecnológico y la ingeniería. Con el paso del tiempo las STEM adquirieron importancia y en 2011 la NSF y la *United States National Research Council* (USNRC) acordaron que estas disciplinas eran fundamentales para aquellas sociedades consideradas tecnológicamente avanzadas. Las iniciativas y proyectos STEM persiguen un aprovechamiento de los puntos en común que existen entre las disciplinas STEM, con el objetivo de desarrollar un enfoque interdisciplinario en el proceso de enseñanza-aprendizaje. No obstante, con el paso del tiempo, Yakman acuñó el término STEAM (siglas en inglés de *Sciencie, Technology, Engineering, Arts and Mathematics*) para proponer la incorporación de las artes a los nuevos modelos educativos centrados en la investigación como un mecanismo de enriquecimiento.

La integración de las artes en el ámbito del STEM permitió la creación de un novedoso modelo de aprendizaje más completo e integral. La principal diferencia entre el STEAM y el STEM, es que el primero indaga los métodos de aprendizaje basados en problemas mediante la utilización de proceso caracterizados por la creatividad. La combinación de las artes con otros saberes permite hacer interesantes descubrimientos. La educación STEAM ofrece la oportunidad de aprender creativamente utilizando habilidades del siglo XXI para la resolución de problemas. La siguiente pirámide sirve para ilustrar los fundamentos de la educación STEAM:



Fuente: Yakman, 2006.

Innovaciones como la educación STEAM permiten destacar la posibilidad de integrar las humanidades con otros saberes científicos y tecnológicos. En ese sentido, la filosofía STEAM implica un desafío a la teoría de las dos culturas y transforma el modo de concebir y generar el conocimiento en el contexto de un humanismo tecnológico capaz de responder a los desafíos cívicos y democráticos de un mundo cambiante y exigente.

1.12. Los principios éticos como legado humanista

En el primer apartado de este capítulo se insistió en la necesidad de poner en valor el legado de lo humano en el desarrollo de la IA. Por ello es importante destacar que el interés del humanismo tecnológico por defender lo humano radica en el progreso moral que la

humanidad ha experimentado a lo largo de la historia a través de una gran cantidad de experiencias, buenas y malas, que han servido para enriquecer su condición. En ese sentido, promover un humanismo tecnológico en la actualidad implica rescatar la experiencia moral desde el ejercicio de una hermenéutica crítica y el cultivo de una ética aplicada a la IA que tenga en cuenta también los principios como elementos primordiales para promover la ética en el entorno tecnología y reivindicar lo humano como merecedor de respeto y consideración.

La creencia en la riqueza de los principios como un legado del humanismo puede contribuir al fortalecimiento ético en un entorno que habitualmente destaca por lo estrictamente técnico. Por lo tanto, el proceso de generación de conocimiento en el ámbito de la IA y aquellas relaciones que se establecen entre los diversos grupos de interés implicados en su actividad, deben caracterizarse por la presencia de unos principios que inspiren y garanticen el cultivo de habilidades cívicas y democráticas. Entre los principios que están presentes en la actividad científica de la IA y que permiten promover la responsabilidad ante los desafíos actuales, el grupo de expertos en IA de la Comisión Europea (CE) destaca los siguientes:

- El principio de respeto de la autonomía humana

  Los derechos fundamentales en los que se apoya la UE van dirigidos a garantizar el respeto de la libertad y la autonomía de los seres humanos. Las personas que interactúen con sistemas de IA deben poder mantener una autonomía plena y efectiva sobre sí mismas y ser capaces de participar en el proceso democrático. Los sistemas de IA no deberían subordinar, coaccionar, engañar, manipular, condicionar o dirigir a los seres humanos de manera injustificada. En lugar de ello, los sistemas de IA deberían diseñarse de forma que aumenten, complementen y potencien las aptitudes cognitivas, sociales y culturales de las personas. La distribución de funciones entre los seres humanos y los sistemas de IA debería seguir principios de diseño centrados en las personas, y dejar amplias oportunidades para la elección humana. Esto implica garantizar la supervisión y el control humanos sobre los procesos de trabajo de los sistemas de IA. Los sistemas de IA también pueden transformar de un modo fundamental el mundo del trabajo. Deberían ayudar a las personas en el entorno laboral y aspirar a crear empleos útiles.

- El principio de prevención del daño

  Los sistemas de IA no deberían provocar daños (o agravar los existentes) ni perjudicar de cualquier otro modo a los seres humanos. Esto conlleva la protección de la dignidad humana, así como de la integridad física y mental. Todos los sistemas y entornos de IA en los que operan estos deben ser seguros. También deberán ser robustos desde el punto de vista técnico, y debería garantizarse que no puedan destinarse a usos malintencionados. Las personas vulnerables deberían recibir mayor atención y participar en el desarrollo y despliegue de los sistemas de IA. Se deberá prestar también una atención particular a las situaciones en las que los sistemas de IA puedan provocar efectos adversos (o agravar los existentes) debido a asimetrías de poder o de información, por ejemplo entre empresarios y trabajadores, entre empresas y consumidores o entre gobiernos y ciudadanos. La prevención del daño implica asimismo tener en cuenta el entorno natural y a todos los seres vivos.

- El principio de equidad

  El desarrollo, despliegue y utilización de sistemas de IA debe ser equitativo. Pese a que reconocemos que existen muchas interpretaciones diferentes de la equidad, creemos que esta tiene tanto una dimensión sustantiva como procedimental. La dimensión sustantiva implica un compromiso de: garantizar una distribución justa e igualitaria de los beneficios y costes, y asegurar que las personas y grupos no sufran sesgos injustos, discriminación ni estigmatización. Si se pueden evitar los sesgos injustos, los sistemas de IA podrían incluso aumentar la equidad social. También se debería fomentar la igualdad de oportunidades en términos de acceso a la educación, los bienes los servicios y la tecnología. Además, el uso de sistemas de IA no debería conducir jamás a que se engañe a los usuarios (finales) ni se limite su libertad de elección. Asimismo, la equidad implica que los profesionales de la IA deberían respetar el principio de proporcionalidad entre medios y fines, y estudiar cuidadosamente cómo alcanzar un equilibrio entre los diferentes intereses y objetivos contrapuestos. La dimensión procedimental de la equidad conlleva la capacidad de oponerse a las decisiones adoptadas por los sistemas de IA y por las personas que los manejan, así como de tratar de obtener compensaciones adecuadas frente a ellas. Con este fin, se debe poder identificar a la entidad responsable de la decisión y explicar los procesos de adopción de decisiones.

- El principio de explicabilidad

  La explicabilidad es crucial para conseguir que los usuarios confíen en los sistemas de IA y para mantener dicha confianza. Esto significa que los procesos han de ser transparentes, que es preciso comunicar abiertamente las capacidades y la finalidad de los sistemas de IA y que las decisiones deben poder explicarse —en la medida de lo posible— a las partes que se vean afectadas por ellas de manera directa o indirecta. Sin esta información, no es posible impugnar adecuadamente una decisión. No siempre resulta posible explicar por qué un modelo ha generado un resultado o una decisión particular (ni qué combinación de factores contribuyeron a ello). Esos casos, que se denominan algoritmos de «caja negra», requieren especial atención. En tales circunstancias, puede ser necesario adoptar otras medidas relacionadas con la explicabilidad (por ejemplo, la trazabilidad, la auditabilidad y la comunicación transparente sobre las prestaciones del sistema), siempre y cuando el sistema en su conjunto respete los derechos fundamentales. El grado de necesidad de explicabilidad depende en gran medida del contexto y la gravedad de las consecuencias derivadas de un resultado erróneo o inadecuado (Comisión Europea, 2019a: 14-16).

En definitiva, los principios éticos aplicados a la IA favorecen el cultivo de habilidades cívicas y democráticas desde una orientación humanista de la tecnología, iluminada por el interés y la sensibilidad en la superación del paradigma tecnocrático para responder a las necesidades de la ciudadanía.

1.13. Humanismo tecnológico y nuevo modo de obrar

El modo de obrar del ser humano demanda una transformación ante las exigencias éticas que sugiere el nuevo tiempo tecnológico. El análisis de Jonas (1995: 58-59) en lo referente al vacío ético y la necesidad de nuevos planteamientos axiológicos es muy oportuno para la formulación del humanismo tecnológico. En el próximo capítulo se esbozará el pensamiento de Jonas en torno al principio de responsabilidad en el contexto tecnológico, insistiendo en el reconocimiento de la tecnología como un medio de gran poder y en la necesidad de enfrentar aquel modo de pensar que asume de forma acrítica la idea de progreso. El humanismo tecnológico no solo se construye sobre la idea de una

tecnología al servicio de la existencia del ser humano y su entorno para un favorable florecimiento, sino también sobre un soporte de responsabilidad que reconozca las exigencias de una nueva ética y sobre el que no sería posible hablar posteriormente de una IAR.

El ritmo de crecimiento vertiginoso del campo de la IA confirma el pronóstico de Jonas sobre el salto cualitativo que ha experimentado el poder tecnológico a partir de la alianza entre la técnica y las ciencias naturales. Sin embargo, el salto no es únicamente cualitativo, sino también cuantitativo, pues ha aumentado su espectro de extensión a diversas dimensiones de la vida humana. En el diseño tecnológico se proyectan determinados sentidos que después pueden tener un impacto fundamental sobre la vida, de modo que esos sentidos proporcionan una significación que debe considerarse desde el principio ético de la responsabilidad. Como señalan José Luis Sepúlveda Férriz y Tomás Domingo Moratalla:

> […] la tecnología exige responsabilidades porque la creación técnica es una forma de dotar de sentido y este puede significar dominación, alienación, barbarie, cierre del horizonte humano. Así, puesto que también Jonas reconoce que no es posible renunciar a la tecnología, resulta necesaria la reflexión sobre la legitimidad de la acción tecnológica (2011: 7).

Esta técnica moderna fundamentada en la máxima baconiana de que «saber es poder» depara importantes desafíos a la humanidad y a la biosfera, que exigen que el humanismo se reformule en base a nuevas exigencias éticas. Aunque el planteamiento jonasiano centre su preocupación en aspectos bioéticos, puede ser extensible a otros ámbitos, como el profesional o el militar, ya que al fin y al cabo se promueve la necesidad de valorar las implicaciones que tiene la técnica moderna para las generaciones futuras y el bienestar vital. Además, la transformación del obrar humano con la incorporación y contextualización de ciertos principios éticos sirve para plantear la exigencia en la adopción de estrategias de responsabilidad frente a determinadas actuaciones tecnológicas que podrían estar siendo asumidas de modo irreflexivo y carente de crítica.

¿Es necesario un nuevo humanismo para las presentes exigencias éticas? Podría afirmarse que sí, principalmente porque el poder que subyace en la tecnología plantea una transformación del obrar humano. El humanismo tecnológico representa una superación de la concepción antropocéntrica propia de la modernidad y elabora un relato crítico frente al nihilismo tecnológico y la actitud acrítica en el ideal de progreso. La transformación del obrar implica una nueva concepción del mundo y la naturaleza donde sean incorporados componentes de valoración y bondad, fruto de una exigencia que nace del ser. Esta exigencia de carácter ontológico se encuentra bajo el paraguas de una concepción antropológica donde la responsabilidad frente al mundo se conciba como una obligación moral (T. Domingo Moratalla, 2007: 373).

El humanismo tecnológico representa una condición de posibilidad para poder plantear una IAR, debido a que el ser humano debe reflexionar y reorientar su relación con la tecnología para que ésta incorpore un sentido de responsabilidad con el mundo. Necesariamente eso no sería posible sin plantear la posibilidad de una transformación del obrar humano que surja desde una matriz antropológica, como señala Jonas. El humanismo tecnológico se construye sobre la base de un imperativo ético que implica la transformación del obrar, una premisa que resulta fundamental para proyectar una IAR. En el próximo capítulo se presentarán los principales postulados de Jonas sobre el principio de responsabilidad en el ámbito tecnológico como un fundamento ineludible para poder articular después la propuesta de una IAR.

CAPÍTULO 2

HANS JONAS: EL PRINCIPIO DE RESPONSABILIDAD ANTE EL PODER TECNOLÓGICO

> Dado que hoy en día la técnica alcanza a casi todo lo que concierne a los hombres –vida y muerte, pensamiento y sentimiento, acción y padecimiento, entorno y cosas, deseos y destino, presente y futuro–, en resumen, dado que se ha convertido en un problema tanto central como apremiante de toda la existencia humana sobre la tierra, ya es un asunto de la filosofía y tiene que haber algo así como una filosofía de la tecnología.
>
> (Jonas, 1997: 15)

A Hans Jonas, como a cualquier otro filósofo, hay que entenderlo dentro de un contexto histórico y temporal concreto, pues la filosofía «no es nunca ajena a la urgencia de los hechos históricos» (Izuzquiza, 2000: 18). La tarea comprensiva de los postulados de un pensador implica entender de igual forma su contexto. Contexto y pensamiento se encuentran en íntima comunión. En ese sentido, la comprensión y aplicación del pensamiento de Jonas al tema de discusión de este trabajo, a saber, la IA, debe hacerse desde un ejercicio prudencial, pues dicho pensamiento no toma como referencia el desafío de los intelectos sintéticos, sino más bien la exigencia que recae sobre la humanidad ante el poder de la técnica y sus efectos en la biosfera.

Es muy común que en los textos de las diversas éticas aplicadas exista un término clave que supone un importante pilar dentro del campo de la ética, la «responsabilidad». La vasta obra de Jonas conceptualiza este término y lo sitúa en el centro de su reflexión ética. En el año 1979 fue publicado en Alemania *Das Prinzip Verantwortung*. *El principio de responsabilidad*, título de la traducción española, tuvo un fuerte impacto en los círculos intelectuales y también en el ámbito de la ecología, ya que mostraba al mundo el importante

desafío que enfrentaba la humanidad en el futuro ante el gran poder de la técnica. El propio Jonas es consciente de dicho impacto cuando señala lo siguiente:

> […] no se debe, si no me equivoco, a su fundamentación filosófica, sino al sentimiento generalizado, del que ya entonces los observadores más atentos podían prescindir cada vez menos, de que algo podía ir mal para la humanidad (2005: 352).

El filósofo alemán realiza una argumentación filosófica sustantiva, apoyando su ética de la responsabilidad en una fenomenología de la vida y en una filosofía del ser. En esta obra confluyen varias reflexiones como la preocupación por el poder técnico, la necesidad de una nueva ética o la crítica a la ideología de progreso y al utopismo marxista.

El siglo XX representa un contexto sustancialmente nuevo por su determinación a partir de la técnica, lo que para Jonas demanda una nueva ética que sea capaz de estar a la altura de los nuevos desafíos que presenta el poder técnico, como sostiene a continuación:

> Si la naturaleza de esas capacidades es realmente tan nueva como aquí se afirma, y si realmente ha quedado abolida, en virtud de sus consecuencias potenciales, la neutralidad moral de que anteriormente disfrutaba la relación técnica con la materia, entonces tal presión significa que hay que buscar en la ética cosas nuevas que puedan guiarla, pero, ante todo, que puedan establecer su validez teórica frente a aquella presión (1995: 59).

La revolución científico-teórica produce un cambio en el modo de entender el mundo y su relación con el ser humano, propiciando una determinada manera de actuar en él. El origen del poder tecnocientífico del siglo XX se encuentra en el planteamiento de Francis Bacon (2000), pensador que defiende la máxima de que el conocimiento es poder. Para Bacon es fundamental adquirir un conocimiento profundo de las cosas para que sea posible su manipulación según la voluntad humana y buscar así nuevos mecanismos que hagan la vida más agradable. Se trata de considerar a la ciencia como un instrumento que domina la naturaleza para ponerla al servicio. Bacon inaugura una etapa de optimismo filosófico basado en la fe en un progreso indiscutible. Cuando el londinense presentó sus postulados, nadie imaginaba que iban a conducir a un tiempo en el que la alianza entre la ciencia y la

tecnología se convertiría en un fin a través del cual el humano sería mediatizado para el servicio de la misma y convertido en un objeto.

Volviendo a Jonas, como señala el subtítulo de su obra *El principio de responsabilidad*, la civilización tecnológica se sitúa en un nuevo escenario con unos rasgos muy particulares que nunca antes había conocido la humanidad. Ese escenario demanda nuevas exigencias éticas, reflexión y discusión acerca de las implicaciones que tiene el poder de la técnica. Su estudio se encuentra también en la senda de una antropología de la técnica, considerando que es un instrumento que incide sobre la acción humana de forma considerable, motivo por el que tiene que ser atendido. El alemán muestra, unida a la técnica, la idea de progreso ilimitado, construida desde la confianza en el ideal ilustrado. Además, el progreso técnico se erige a partir de un carácter utópico que condiciona su dinamismo, marcando así una tendencia utópica (1995: 55). El espectro de aplicación de la técnica se caracteriza por su gran volumen y alcance, pues la técnica ha llegado a todas las esferas de la vida, siendo capaz hasta de manipular la constitución humana. Desde una perspectiva macro, la técnica posee un importante potencial destructor y manipulador (Esquirol, 2011).

La acción del humano mediada por la técnica puede tener efectos irreversibles sobre el mundo, lo que representa una transformación en el modo de ver su influencia sobre la naturaleza en el pasado, que se caracterizaba principalmente por la superficialidad (Jonas, 1995: 27). El trato con el mundo exterior al humano era considerado como éticamente neutro en lo referente tanto al objeto como al sujeto, pues la acción del humano en los objetos no humanos no presentaba una relevancia ética. La verdadera relevancia se encontraba en el humano y en su relación con los otros. En ese sentido, para Jonas podría considerarse la ética hasta los tiempos de la civilización tecnológica como antropocéntrica. En esta esfera no se pensaba que la condición humana pudiera dejar de ser una constante esencial como consecuencia del poder transformador de la *techne* (1995: 29). Además, la valoración moral sobre la acción residía en la propia *praxis* y en su alcance inmediato, lo que implicaba que únicamente fueran considerados como parámetros de susceptibilidad ética la inmediatez temporal y la cercanía espacial. Para el filósofo alemán la posibilidad de valoración moral quedaba reducida a un estrecho campo de acción.

Con la técnica moderna esta situación ha cambiado, ya que los límites espaciales y temporales han experimentado un desplazamiento, afectando así a toda la naturaleza y a las generaciones futuras que son vistas como vulnerables. La acción humana ya no tiene un efecto inmediato en el plano temporal, sino que presenta un alcance mayor en el tiempo, comprometiendo el futuro de la humanidad. Las coordenadas espacio-temporales han cambiado y eso exige de una nueva capacidad moral a la altura de los compromisos del mundo actual. El poder de la técnica abre nuevos horizontes que advierten la necesidad de una ética con nuevas perspectivas.

El poder tecnocientífico adquirido por el ser humano exige considerar más aspectos que el simple interés que tiene él mismo. El deber se sitúa en un territorio más allá del que plantea el antropocentrismo, por ello se propone un «nada desdeñable cambio de ideas en los fundamentos de la ética» (Jonas, 1995: 35). Su propuesta ética incluye dos componentes a tener en cuenta: su orientación al futuro y su humildad ante el poder técnico. Es una ética orientada al futuro porque asume el compromiso con el mañana, preocupándose por él, sostenida en una «futurología comparada» (1995: 64). Además, como se ha señalado, incluye también la humildad ante la capacidad técnica que poseemos para hacer más que para prever. Se refiere al reconocimiento del poderío técnico que dispone el humano para asumirlo desde una toma de conciencia. La ética de la responsabilidad conduce a la forja de un *ethos* vigilante, donde la actitud de vigilia será la constante en la acción técnica, identificando las posibilidades negativas, esto es el *malum*, antes que las positivas, el *bonum*.

Por lo tanto, a Hans Jonas se le debe la formulación de una ética de la responsabilidad para responder al desafío que ha supuesto desde el siglo XX el poder desproporcionado de la técnica. La ética de la civilización tecnológica plantea que el futuro de la naturaleza y de la humanidad se encuentran ahora bajo el cuidado. Su propuesta pretende despertar la conciencia humana para la asunción de responsabilidad.

2.1. Una nueva ética ante el poder técnico

Jonas advierte que la tecnología contemporánea se ha convertido en un elemento de arrastre que configura el mundo. Junto con otros pensadores como Jacques Ellul (2003, 2004), Jonas considera que la técnica responde a un curso y a una lógica autónoma que deriva en una tendencia amenazadora:

> […] junto a la magnitud y a la ambivalencia, otro rasgo de carácter del síndrome tecnológico que tiene una importancia ética propia: el elemento cuasi-forzoso de su avance, que por así decirlo hipostatiza nuestras propias formas de poder en una especie de fuerza autónoma de la que nosotros, los que la ejercemos, nos volvemos paradójicamente súbditos (Jonas, 2001: 38-39).

El poder tecnológico y sus consecuencias demandan una nueva ética, pues el horizonte antropocéntrico que tradicionalmente condicionaba la formación de la ética se ha desplazado. Los sistemas éticos tradicionales, ya sean religiosos o seculares, ponían al humano en el centro de sus reflexiones y lo relevante para esta ética era el trato que tenía consigo mismo sin la necesidad de considerar a otro actor involucrado, como puede ser la naturaleza. Así pues, toda la ética tradicional podría considerarse antropocéntrica (Jonas, 1995: 29). La diferencia para el alemán radica en que la introducción de la naturaleza como nuevo objeto de consideración exige una nueva ética. Esto se debe a que las consecuencias de las acciones técnicas sobre la naturaleza a lo largo de la historia eran insignificantes; en cambio, ahora, con el poder de la tecnociencia, dichas consecuencias implican serias alteraciones que destacan la necesidad de compromisos éticos a la altura de este tiempo. Los efectos de estas acciones sobre la naturaleza han puesto de relieve la necesidad de que la humanidad se plantee un deber para-con la naturaleza.

La tecnología contemporánea ha modificado claramente las relaciones humanas con la naturaleza, sin embargo, y aquí se complementa lo sostenido por Jonas, también se ha experimentado esta modificación en el ámbito social. Esto quiere decir que no solo se ha modificado la relación entre el ser humano y su medio natural, sino también entre el ser humano y su medio social, político y económico, que es en el terreno donde mejor se

observan los desafíos de la IA. El nuevo escenario plantea un imperativo ético sin precedentes porque se pone en riesgo a la humanidad, lo que implica que el alcance espacio-temporal de la tecnología sea de escala global. Además, el nuevo imperativo se encuentra más orientado hacia la política pública que hacia el comportamiento de los sujetos privados. Esta cuestión es destacada porque Jonas remarca que el imperativo categórico kantiano estaba más orientado hacia el individuo. En cambio, debido al avance tecnológico, los efectos de los actos humanos presentan en este momento un mayor alcance, tanto en el espacio como en el tiempo. Jonas completa lo dicho hasta ahora con lo siguiente:

> Esto añade al cálculo moral el horizonte *temporal* que falta en la operación lógica instantánea del imperativo kantiano: si este último remite a un orden siempre presente de compatibilidad abstracta, nuestro imperativo remite a un *futuro* real previsible como dimensión abierta de nuestra responsabilidad (1995: 41).

Este desarrollo tecnológico lleva incorporado un aumento desproporcionado que impone una obligación moral de responsabilidad y prudencia. Los bombardeos atómicos sobre Hiroshima y Nagasaki son una clara muestra del potencial destructivo para la humanidad que puede llegar a tener la tecnología cuando no se valoran aspectos vinculados con la responsabilidad en relación a los efectos. Otra consecuencia de este ampliado poder es su carácter de imprevisibilidad y ambivalencia en los efectos. Estos caracteres, propios de la tecnología actual, se fundamentan en las dinámicas utópicas que encierra el ideal de progreso. La dimensión utópica se ha convertido en un aspecto de obligada consideración, ocasionando un desplazamiento de la sabiduría hacia un terreno de lo olvidado y lo innecesario. Jonas muestra que la sabiduría se encuentra vinculada al planteamiento de la existencia de valores absolutos y de una verdad objetiva, algo francamente necesario para la humanidad (1995: 55).

Además, la nueva ética que exige el poder tecnológico demanda también una humildad con nuevas tonalidades, donde sea concebida la magnitud de la capacidad humana. Se trata de no promover la ignorancia sobre las consecuencias de la acción tecnológica, ya sean previsibles o imprevisibles. La sabiduría puede iluminar la acción transformadora y tender

un puente de razón entre los deseos y los fines, pues el saber es fundamental para prever, como Jonas señala:

> La inevitable dimensión «utópica» de la tecnología moderna hace que se reduzca cada vez más la saludable distancia entre los deseos cotidianos y los fines últimos, entre las ocasiones de ejercer la prudencia usual y las de ejercer una sabiduría iluminada (1995: 55).

2.2. La fundamentación ontológica y metafísica

La ética de la responsabilidad que Jonas reivindica se elabora desde un imperativo fundamental que consiste en asegurar la existencia del género humano (1995: 80). Este imperativo, visto desde la perspectiva de la IA, puede considerarse muy exagerado, dando a entender que las máquinas dominarán el mundo y emprenderán una guerra para exterminar al género humano. Sin embargo, no se toma como punto de partida esa premisa humana en términos de bienestar integral. El imperativo de Jonas debe considerarse de forma contextualizada en el objeto de estudio de la IA y dentro de unos parámetros que sirvan de garantía de bienestar integral para la humanidad.

Este imperativo impuesto desde la nueva conciencia ética promueve la obligación de garantizar las condiciones más óptimas para la existencia de la humanidad futura y la biosfera. En este sentido, el imperativo adquiere un carácter ontológico, pues para Jonas lo que debe asegurarse es la permanencia de la esencia que caracteriza la condición humana, sin experimentar de este modo ninguna alteración. No obstante, es necesario matizar un aspecto, y es que Jonas no encuentra la razón de ser del imperativo en el humano del futuro, sino más bien en la idea de humano, en esa esencia que lo caracteriza. Esto explica que la fundamentación del imperativo sea ontológica y que se origine en la idea de ser humano a partir de su esencia característica. En función de esto se señala lo siguiente:

> Este imperativo ontológico, surgido de la idea de hombre, es el que se halla tras la prohibición –antes presentada sin fundamentar– del juego del «todo o nada» con la humanidad. Sólo la idea de hombre, por cuando nos dice *por qué* debe haber hombres, nos dice también *cómo* deben ser (Jonas, 1995: 88).

Como señala el filósofo alemán, la fundamentación ontológica de esta nueva ética da lugar a un imperativo que no es hipotético, sino categórico, dando por sentado que deben existir seres humanos en el futuro. Esto significa que el deber anticipado es planteado de forma incondicional. Sin embargo, existe una importante variación respecto al imperativo kantiano, y es que la autoconcordancia, en la que la razón se da a sí misma las leyes, no se plantea sobre la idea del *hacer*, es decir, sobre hechos ocasionales presentes en la existencia humana. La idea que plantea Jonas es la de un imperativo orientado hacia el *ser*, centrada en la existencia de esencia, de su contenido, lo que implica una idea ontológica e incondicional (1995: 89). Esta variación significa que la fundamentación de la ética es metafísica y que no encuentra su origen exclusivamente en el simple obrar, sino en la idea de ser humano enraizada en la doctrina del ser.

La técnica no solo amenaza la existencia de la humanidad sino también su propia esencia. El imperativo de la nueva ética debe garantizar la supervivencia de la humanidad y la biosfera, pero también de la esencia característica de lo humano, implicando de ese modo un valor intrínseco. El ser humano posee un valor intrínseco que lo caracteriza y debido a ello es necesaria una ética que custodie su existencia, ya que de ésta surge su singularidad.

Ahora bien, la reivindicación de Jonas de una ética que custodie la existencia del humano no hace referencia a una idea concreta de ser humano, sino más bien a la condición humana en sí misma, pues en ella reside la capacidad y posibilidad de su realización. En ese sentido, ante la pregunta por qué debe ser el humano, Jonas sostiene que la respuesta podría encontrarse en la fundamentación del valor intrínseco de la existencia humana, donde se da una convergencia entre la ética y la metafísica.

Jonas reconoce que la preeminencia del ser no es absoluta y que un deber en favor del ser y su continuidad no está enteramente justificado en cualquier circunstancia, ya que existen ocasiones por las que se sacrifica la propia vida debido a algún motivo concreto, como puede ser la afectación a la humanidad, o incluso la terminación de la vida de forma voluntaria por cuestiones relativas a la dignidad. Así pues, lo dicho demuestra que «la vida no es el bien supremo« (Jonas, 1995: 93). Recuperando la pregunta anteriormente planteada

y complementándola con la cuestión leibniziana de por qué es algo y no más bien nada (Jonas, 1995: 93), surge un vínculo en el sentido de dicha pregunta si se tiene en cuenta el valor del ser. El valor intrínseco de algo muestra la posibilidad de que ese algo exista, por lo que la presencia está legitimada por el valor y así su continuidad está justificada. En ese sentido, el valor intrínseco que garantiza la presencia continuada informa acerca de la exigencia de un deber ser. Señala Jonas:

> Esto es así porque el valor o el «bien», si hay algo parecido, es lo único cuya mera posibilidad empuja a la existencia –o, a partir de una existencia dada, legitima la continuidad de su existencia–, de modo que fundamenta una exigencia de ser, fundamenta un deber-ser; y donde el ser es objeto de una acción libremente elegida, lo convierte en deber. Hay que observar que la mera posibilidad de atribuir valor a lo que es, independientemente de lo mucho o lo poco que se encuentre actualmente presente, determina la superioridad del ser sobre la nada (Jonas, 1995: 95).

La necesidad de una fundamentación metafísica de la ética debe entenderse por el valor intrínseco que posee el ser, es decir, por el valor del ser en tanto que ser, convirtiéndose así el ser en un aspecto fundamental del humano. El ser es valioso en sí mismo, es el valor absoluto del humano y el fundamento de todo valor, de ahí que ante la pregunta de por qué el ser y no la nada, el ser siempre sea preferible a la nada. El ser proporciona valor, la nada lo aniquila. La argumentación esbozada por Jonas demuestra claramente que existe un fuerte vínculo entre ontología y axiología. El humano representa el ser del valor y además es el único capaz de percibir y expresar el valor que posee el ser por encima de cualquier otro ente. Tiene la capacidad de valorar, motivo por el que es preferible garantizar su existencia antes que su no-existencia. Así pues, la existencia humana implica la capacidad de valorar y expresar el valor de las cosas y de sí misma.

La dualidad entre ser y no-ser genera una tensión de la que la humanidad no puede librarse. Esa tensión inevitable deviene una vulnerabilidad que es insuperable, pero que al mismo tiempo da muestras de dónde se fundamenta su libertad y el *telos* de sus acciones.

El razonamiento expuesto sobre el deber ser conduce a una posición conservadora y biologicista, pues parte de una idea de ser humano extremadamente naturalizada. El planteamiento jonasiano de una fundamentación metafísica puede llevar al terreno de lo dogmático y a la defensa a ultranza de un ideal determinado. No obstante, es posible utilizar una argumentación sólida para defender la idea de la necesidad de poner en práctica el principio de responsabilidad desde la acción humana, pues existe una cualidad que emana del ser mismo que posee valor. Lo importante del planteamiento de Jonas al hilo de lo expuesto es que la existencia de la humanidad es indispensable para la salvaguarda de los valores en el mundo. El valor surge del ser y a la vez se revela únicamente en la subjetividad del humano.

2.3. El deber con perspectiva de futuro

La ética planteada por Jonas se hace cargo del futuro de la humanidad, pero para ello tiene en cuenta dos requisitos fundamentales. Por un lado, debe enriquecerse el conocimiento en lo relativo a las posibles consecuencias de las acciones tecnológicas con el propósito de conocer con mayor profundidad las posibles implicaciones de dichas acciones para la existencia o esencia de la humanidad, así como para el medio ambiente. Además, por otro lado, una vez recogido el conocimiento sobre las previsiones y riegos, es importante determinar qué criterios éticos son fundamentales para aprobar según qué acciones, considerando qué se puede admitir y qué debe sancionarse o evitarse. En este sentido para Jonas es fundamental pensar los efectos remotos de la acción técnica, o como él mismo señala: «resulta, pues, necesario elaborar una ciencia de la predicción hipotética, una «futurología comparada» (1995: 64).

A la hora de determinar qué criterios éticos deben tomarse en cuenta para la acción, hay que tener presente el concepto de límite para considerar qué es lo que no se debe hacer, con el fin de esclarecer cuáles son las posibilidades tecnológicas más viables. En la línea de los límites hay que introducir en la discusión el aspecto histórico y metafísico. En la historia pueden encontrarse motivos más que suficientes para defender la existencia de la humanidad, pues ella es merecedora del porvenir. A partir de ahí, Jonas critica duramente

toda visión utópica, considerando que es inaceptable una idea que propone falsas ilusiones en una humanidad perfecta, debido a la existencia de ejemplos históricos que ponen de relieve dónde han sido conducidos millones de seres humanos por diversos regímenes totalitarios que trataban de realizar sueños políticos y antropológicos.

En lo referente a la orientación hacia el futuro de la ética, podrían surgir preguntas en torno a la obligación frente a algo que todavía no ha ocurrido, o sobre algo cuyas consecuencias todavía no se conocen, y que por lo tanto presenta dificultades a la hora de la previsión. Una de las tesis más importantes de Jonas descansa precisamente en el carácter futurible que fundamenta su principio de responsabilidad y que se encuentra estrechamente vinculada con una nueva consideración de la ética encaminada a superar los planteamientos antropológicos de la ética tradicional. El filósofo alemán se refiere aquí a aquella responsabilidad que va más allá de los actos y sus consecuencias directas, es decir, *ex post facto*, necesariamente orientada hacia la ampliación del horizonte futuro. Se trata de una responsabilidad que tiene que ver con una potestad justificada que encuentra su justificación en el compromiso, en algo que se le confía la garantía de protección. Es un concepto moral de responsabilidad que reconoce los fines. En una mesa redonda celebrada en un simposio en el Hotel Schloss Fuschl de Austria en 1981, Jonas pronunció las siguientes palabras:

> Ahora bien, he dedicado algún esfuerzo para distinguir entre dos conceptos completamente distintos de responsabilidad; el concepto puramente formal, por así decirlo jurídico de la responsabilidad: que cada uno es responsable de lo que hace y se le puede responsabilizar de lo que ha hecho si se le tiene a mano. Esto mismo no es un principio de la acción moral, sino sólo de la responsabilización moral posterior por lo hecho. Cuando el sujeto de la responsabilización moral ya no está ahí, no hay por así decirlo nada que hacer. Pero hay que distinguir de esto un concepto completamente distinto de la responsabilidad, el que acabo de ilustrar en particular en la relación padre hijo, y es la responsabilidad por lo que hay que hacer: no pues la responsabilidad por los actos cometidos, sino estar obligado por la responsabilidad a hacer algo, porque se es responsable de una cosa. Pero se es responsable de la cosa porque la cosa está en el ámbito del propio poder, es decir, depende de la propia acción […] la humanidad, y por tanto cada miembro de la humanidad, cada individuo

> concreto, tiene de hecho una obligación trascendente o metafísica de que también en el futuro haya en la tierra hombres, encarnaciones de este género humano –y en condiciones de existir–, que aún permitan hacer realidad la idea de ser humano (1997: 188).

Así pues, la ética orientada al futuro es una ética que deben practicar los humanos del presente para considerar las implicaciones de sus acciones con respecto al futuro, una ética cuidadosa con el futuro desde el presente. La propuesta jonasiana no es determinante, sino posibilitante, consciente de sus posibilidades desde el presente para proteger la humanidad en el futuro. La intención de tener en cuenta el futuro no puede provocar un descuido del presente, pues esta realidad, la del presente, es la que se impone y en la que se puede directamente actuar. Además, siempre hay que tener en cuenta el aspecto cambiante del mundo, donde la ética orientada al futuro no debe tener la pretensión de dar soluciones futuras definitivas.

Como demuestra la última cita, Jonas señala claramente que aquellas teorías que conciben la responsabilidad como una imputación causal de actos cometidos son insuficientes porque consideran que la condición de responsabilidad estriba en el nexo causal. Son también insuficientes esas teorías porque vinculan la responsabilidad con el pasado, afirmando que la responsabilidad recae sobre actos que ya han sido ejecutados. Además, el establecimiento con certeza suficiente de la correlación entre el acto cometido y las consecuencias que de él se derivan es de suma complejidad, aún más si el horizonte no es temporalmente cercano.

Lo expuesto hasta ahora sobre la orientación al futuro de la responsabilidad resulta muy ilustrativo para aclarar que el enorme poder de la acción tecnológica tiene unas consecuencias de gran magnitud. Un espacio donde se desarrolla de forma evidente la responsabilidad orientada al futuro es aquel en el que está presente la relación de los padres con los hijos. Esta relación sirve para ilustrar la importante conexión que existe entre la ontología y la ética, pilar sobre el que es posible levantar el carácter futurible de la responsabilidad.

En primer lugar es importante tener en cuenta que la verdadera responsabilidad es aquella que es desinteresada, es decir, que no espera una reciprocidad. A diferencia del deber, que se fundamenta en la igualdad con algo o alguien, la responsabilidad es desinteresada y se caracteriza por la desigualdad, pues no espera nada del otro, reconociendo que el sujeto asume fortaleza frente a un objeto que requiere de protección. En ese sentido, el paradigma ético de la responsabilidad no se funda sobre la reciprocidad, sino sobre un imperativo con las generaciones futuras. Así pues, la reciprocidad no estaría presente y se estaría garantizando la protección de algo que todavía no existe y que por lo tanto no puede dar nada a cambio de esa protección. La relación entre un padre y un hijo es precisamente esa relación que está marcada por la ausencia de reciprocidad. Jonas señala lo siguiente:

> Pero la ética que nosotros buscamos tiene que ver precisamente con lo que todavía no es, y *su* principio de responsabilidad habrá de ser independiente tanto de cualquier idea de un derecho como de la idea de reciprocidad, de tal modo que en su marco no puede nunca formularse la jocosa pregunta inventada al respecto: «¿Ha hecho el futuro alguna vez algo por mí?, ¿acaso respeta él mis derechos?» (1995: 82).

El tipo de responsabilidad que defiende el filósofo alemán se caracteriza por la primacía del deber ser sobre el deber hacer. Se trata de una cuestión metafísica que se encuentra impresa en el deber ser y que se impone frente al no ser. El ser es considerado como un valor absoluto que debe ser garantizado por el deber hacer y que requiere de una respuesta responsable por parte del sujeto. El sujeto debe cuidar el objeto y garantizar su buena existencia. Se trata de la primacía del deber-ser.

La responsabilidad con la naturaleza y la humanidad es una responsabilidad natural enraizada en el amor. No es una responsabilidad que se refiera al encargo de una tarea que debe ser cumplida con cabalidad, como si se tratara de una obligación reposada en un acuerdo limitado y revisable. Más bien se refiere a una responsabilidad que responde al valor intrínseco del objeto para garantizar su existencia:

> Por una parte la demanda de la cosa, en la falta de garantía de su existencia, y por otra la conciencia moral del poder, en el débito de su causalidad, se conjuntan en el afirmativo sentimiento de responsabilidad del yo activo, que engloba ya siempre el ser de las cosas. Si a ello se agrega el amor, a la responsabilidad le da entonces a las la entrega de la persona, que aprende a temblar por la suerte de lo que es digno de ser y es amado (Jonas, 1995: 164).

Relacionado con este tipo de responsabilidad orientada al futuro no solo se encuentra el amor, sino también el poder. El campo donde la responsabilidad se despliega es el tiempo, pero un tiempo con miras al futuro que debe garantizarse para que la humanidad del porvenir pueda desenvolverse con libertad: «en pocas palabras, una responsabilidad de la política es atender a que siga siendo posible la política futura» (1995: 198).

Además, Jonas reivindica una ética orientada al futuro porque piensa en el devenir de la historia, en lo que vendrá después, en la mirada puesta en la continuidad del curso histórico. Es una responsabilidad que piensa en la historicidad del objeto:

> Mas la responsabilidad total ha de preguntarse siempre: «¿qué viene después?, ¿adónde llevará?» y, al mismo tiempo: «¿qué había antes?, ¿cómo encaja en el desarrollo total de esta existencia lo que ahora está sucediendo?». En una palabra, la responsabilidad total tiene que proceder «históricamente», abarcar su objeto en su historicidad. Éste es el sentido propio de lo que aquí designamos con el concepto de «continuidad» (Jonas, 1995: 183).

En definitiva, la responsabilidad encuentra su razón de ser en una mirada ampliada en el tiempo que engrandece su espectro de actuación y que mira hacia el futuro en un devenir histórico del que se siente garante para posibilitar la existencia.

2.4. La heurística del temor

Para el surgimiento de una ética de la responsabilidad es fundamental una «heurística del temor», pues en una reflexión que sea capaz de prever los peligros de los escenarios futuros radica la posibilidad de dicha ética.

> Si ya no es la inspiración de la esperanza, quizá sea entonces la exhortación del temor lo que nos haga entrar en razón. Sólo que el miedo en sí mismo no es una actitud humana demasiado noble, aunque sí que está muy justificado. Y si realmente hay algo que temer, la propia disposición al temor legítimo se convertirá en un imperativo moral (Jonas, 2001: 119).

La heurística del temor que presenta el filósofo alemán tiene dos propósitos principales: por un lado, invitar a que los individuos sean capaces de representar los alejados efectos de la acción tecnológica; y, por otro lado, promover un estímulo para la aparición del sentimiento de temor ante la posibilidad de escenarios no esperados y peligrosos. Estos propósitos cuentan con un requisito fundamental, a saber, la voluntad del individuo. La disposición voluntaria supone una condición *sine qua non* para el planteamiento de una nueva ética. El temor es utilizado como la punta de lanza para la representación del *summum malum*, ya que como señala Jonas, se adquiere conciencia de algo cuando es más fácil el conocimiento del *malum* que el conocimiento del *bonum* (1995: 65). Hay que tener en cuenta que el planteamiento del alemán parte de la imaginación, es decir, es artificial, y por tanto hay que realizar una representación de una imagen de algo que no se ha experimentado o que no ha ocurrido.

La ética de la responsabilidad exige una extrapolación que sugiera pronósticos a largo plazo, debido a que existen una serie de razones que son suficientes para plantear la representación de la posibilidad de una imagen que no ha acontecido. Entre esas razones Jonas considera que existen varias: la complejidad, la insondabilidad de los hombres y la impredecibilidad (1995: 68).

La heurística del temor conduce a un terreno donde el *malum* adopta el carácter de supremacía sobre el *bonum,* esto quiere decir que se priorizan los malos escenarios sobre los buenos. La presencia del mal pone en alerta, hace afinar los sentidos y permite avistar los riesgos, el mal no pasa inadvertido, pues presenta un conocimiento que se impone y que no puede omitirse. El mal no puede ignorarse porque arrastra, hace temer las consecuencias, conduce a la incertidumbre de lo inesperado generando inseguridad. En cambio, el *bonum*, lo bueno, puede pasar inadvertido y ser ignorado y no invita

necesariamente a una reflexión sobre sus implicaciones. Jonas presenta esta reflexión de la siguiente manera:

> Así estamos hechos: nos resulta infinitamente más fácil el conocimiento del *malum* que el conocimiento del *bonum*; el primero es un conocimiento más evidente, más apremiante, está menos expuesto a la diversidad de criterios y, sobre todo, no es algo buscado. La mera presencia del mal nos impone su conocimiento, mientras que lo bueno puede pasar inadvertido y quedar ignorado sin que hayamos reflexionado sobre ello (para hacerlo precisaríamos de una razón especial) (Jonas, 1995: 65).

El temor que plantea la ética de la responsabilidad invita a una imaginación remota alejada del temor empírico, donde la imaginación encuentra su lugar en la esfera de lo abstracto. La imaginación ayuda a descubrir; por ello la heurística es un procedimiento para el descubrimiento. Todo procedimiento tiene una finalidad, un *telos*, que en este caso consiste en descubrir el valor intrínseco que caracteriza al ser humano y la naturaleza, sirviendo para entender por qué merece ser cuidada y permanente su existencia en el devenir histórico. El reconocimiento del valor característico e intrínseco sirve para reconocerlo por sí mismo y para imaginar la posibilidad de su desaparición, del peligro de su destrucción bajo un temor que siempre acompaña y que está presente. La conciencia de la desaparición proporciona un imperativo que implica la protección de la existencia, el temor a que desaparezca porque se aprecia su valor. En ese sentido, la ética de la responsabilidad consiste en tener presente la posibilidad del *malum* por medio del cultivo del temor, siendo capaces de representar un escenario de fatales consecuencias y poniendo en tela de juicio los postulados de las visiones más optimistas, o como dice el filósofo alemán: «hay que dar mayor crédito a las profecías catastrofistas que a las optimistas» (1995: 71).

La mayor de las posibilidades catastróficas que invita a imaginar la heurística del temor es la de una deformación ontológica del ser humano, ya que éste es sujeto de una herencia evolutiva que merece un reconocimiento sagrado (Jonas, 1995: 73). Esta deformación ontológica se convierte en el motivo principal para promover una defensa de su integridad. El temor es convertido en el motor principal para poder enfrentar esta deformación. La

representación del mayor mal, visto como amenaza para la integridad, contribuye también a determinar en qué podría consistir el mayor bien. Siendo consciente de sí mismo, de su finitud, se ve como un ser vulnerable que puede quedar a merced de la posibilidad de la desaparición. Entonces, ahí, tras la representación de su desaparición, asume responsabilidad y lucha por su existencia.

No se trata tanto de la desaparición vital, como si de la muerte fáctica se tratara, sino más bien de la desaparición de la condición humana más característica, aquello que dota de identidad. La reflexión de Jonas no recae en el terreno de lo físico, de lo corporal, sino más bien de lo ontológico, de aquellos rasgos más distintivos de la humanidad que deben ser conservados. La heurística del temor no solo apunta a la desaparición física del ser humano, sino también a la deformación ontológica donde habita la conciencia ética. La autoconciencia ética se convertiría así en el mayor bien, en el bien supremo que hay que salvaguardar frente a cualquier amenaza, pues de ese bien se derivan los demás bienes de la existencia y de la vida humana.

La representación intelectual e imaginaria del temor no puede ser empírica, y por lo tanto tampoco puede ser objeto de la experiencia, aunque sea una posibilidad muy evidente. Tal representación encuentra su razón de ser en el «como si» de un mal que todavía no ha ocurrido, pero que se encuentra en el terreno de lo posible. La heurística del temor también se sitúa frente al excesivo poder tecnológico que posee la humanidad, para que así se adquiera conciencia de la necesidad de cultivar la responsabilidad.

Temor y futuro son dos conceptos del pensamiento de Jonas estrechamente relacionados. Cuando existe un temor hacia algo con la mirada puesta en el futuro, ese ejercicio imaginario de temeridad invita a sentir compasión con la humanidad del futuro, con el porvenir de la misma, alejándonos de la situación espacial y temporal del presente para equiparar el sufrimiento real con un sufrimiento que en principio es solamente potencial, pues todavía no ha acontecido. La ética de la responsabilidad se aleja del egoísmo del presente y mira hacia la potencialidad del futuro, adquiriendo también una

conciencia altruista de ausencia de reciprocidad en su acción para el beneficio de las generaciones futuras.

La heurística del temor supone un ejercicio de anticipación para evitar consecuencias no deseadas. Sirve para conocer el presente de mejor forma y conseguir de esa manera prever y actuar con prudencia. La prudencia que brinda la heurística actúa evitando grandes riesgos y asegurando aspectos decisivos de la existencia y la esencia del ser humano. La acción técnica no puede poner en riesgo el futuro, motivo por el que el trabajo de anticipación y prudencia es extremadamente necesario. El ejercicio de la prudencia representa una apuesta racional por el futuro de la existencia y la esencia de la humanidad, una apuesta arriesgada para no perder el ser. Es por ello que la ética de la responsabilidad supone en ocasiones una renuncia a ciertos beneficios que se ponen de relieve en determinadas circunstancias. Quizás sea más importante renunciar a deslumbrantes beneficios que incurrir en daños irreparables como consecuencia de la avaricia a la que arrastra el capitalismo con su ideal de progreso y utopía. Ser prudente significa anticiparse, cuidar la existencia y no poner en riesgo la ontología característica de la humanidad. La heurística del temor que propone Jonas impulsa a tomar partido y a motivar la preocupación por un *summum malum*.

La acción tecnológica puede orientarse de manera que tenga siempre en cuenta los pronósticos catastróficos y los posibles escenarios. Así pues, el temor considerado como apuesta racional podría reducir la posibilidad de que la humanidad padeciera efectos inesperados como fruto de una falta de responsabilidad y prudencia, y contribuiría promoviendo un enriquecimiento del conocimiento.

2.5. La crítica a la utopía moderna del progreso

Jonas realiza una profunda crítica al concepto de utopía en los capítulos 5 y 6 de su obra *El principio de responsabilidad*. Consciente del importante poder seductor que posee la utopía, dedica parte de su reflexión a mostrar el significado contraproducente que representan estas creaciones ideales. Concretamente su crítica a la utopía responde a uno de

los teóricos más importantes sobre esta temática, Ernst Bloch, autor de *El principio esperanza*. Jonas invita a pensar sobre los límites y las implicaciones que tienen las utopías, mostrando que tras las mismas se esconde cierta ingenuidad. Para ello responde críticamente a las grandes utopías del siglo XX que han tenido fatídicas experiencias, sobre todo pensando en el nazismo y el estalinismo. En la estrategia que lleva a cabo no se confronta de manera propiamente ideológica, sino que más bien plantea una discusión orientada a la reflexión sobre la acción humana en lo que respecta a sus consecuencias en la naturaleza y en el tiempo.

La civilización tecnológica posee un importante poder con serias consecuencias. El poder técnico es de tal magnitud que su expansión hasta la naturaleza ha desembocado en una fagocitación y dominio de la misma, generando que la humanidad llegue a perder el control sobre dicha técnica. Cada aumento y expansión del poder tecnológico determina el siguiente estadio en términos cuantitativos y cualitativos, lo que significa que el avance del poder sea imparable. Max Horkheimer y Theodor W. Adorno también apuntaban a esta problemática en su *Dialéctica de la Ilustración*. El dominio del humano sobre la naturaleza ha tenido efectos exitosos para la humanidad, pero también contraproducentes para su existencia. No hace falta destacar con cifras los alarmantes datos de alteración que están sufriendo los ecosistemas y los altos niveles de contaminación, así como el deshielo de los polos que pone en peligro el futuro de la biosfera. No obstante, este trabajo no versa sobre la destrucción de la biosfera y sus causas en el poder técnico, sino más bien sobre las consecuencias políticas de diversa índole que la IA puede tener sobre la vida. Lo importante aquí es la consideración de la crítica a la utopía que presenta Jonas para aplicarla al contexto del análisis que se aborda en este trabajo.

El escenario actual del poder técnico exige una ética que no sea escatológica y, por tanto, que sea antiutópica. Escribe Jonas:

> El poder de la técnica sobre el destino del hombre ha rebasado incluso el poder de la ética del comunismo, el cual pensaba sólo servirse de ese poder como de todos los demás poderes. Valga como anticipo que, mientras ambas «éticas» tienen que ver con las posibilidades utópicas de esta tecnología, la ética que aquí buscamos *no* es escatológica y, en un sentido todavía por determinar, es antiutópica (Jonas, 1995: 48).

Estas líneas de Jonas son imprescindibles para entender su postura en torno a la utopía. Las utopías se han guiado por un ideal antropocéntrico exacerbado con la intención de alcanzar un ideal de sociedad utópico basado en el desarrollismo tecnológico. El marxismo es objeto de las fuertes críticas de Jonas al utopismo, pues esta ideología trataba de fundamentar su modelo utópico en el dominio de la naturaleza para la satisfacción de las necesidades humanas. La liberación de la alienación se emprendía gracias al dominio de la naturaleza. El marxismo se erigía sobre un ideal técnico de carácter prometeico, pero ese ideal ha demostrado que debido al ansia de progreso ilimitado ha tenido graves costes para la humanidad. Esta ideología no fue capaz de cerciorarse de que su proyecto también comprometía su ideal emancipatorio. En este sentido, el filósofo alemán critica aquellos proyectos utópicos que se asientan sobre la idea de un progreso ilimitado de gran abundancia material, donde el bienestar está asegurado para el conjunto sobre una naturaleza que es inagotable como fuente de recursos para la creatividad técnica.

La ética de la responsabilidad significa una advertencia a aquella humanidad que se encuentra prisionera de un progreso al que ella misma ha dado forma y que encierra una amenaza que se encuentra oculta y que la acecha a sí misma. La naturaleza no puede soportar el ideal de progreso ilimitado que impulsa la actividad técnica. Los modelos capitalistas y marxistas se sostienen sobre unos ideales utópicos, el primero sobre el consumismo y el segundo sobre una actividad explotadora que garantizará la emancipación, que son enfrentados desde la idea jonasiana de la moderación.

La crítica de Jonas a esta utopía también se sostiene por la incapacidad que han demostrado las diversas ideologías y sus proyectos programáticos frente al poder devastador de la tecnología. El poder tecnológico se ha vuelto autónomo y ha adquirido un tercer grado, que el filósofo alemán presenta de esta manera:

> El poder se ha vuelto autónomo, mientras que sus promesas se han convertido en una amenaza y sus salvadoras perspectivas se han transformado en un apocalipsis. Lo que ahora se ha vuelto necesario, si la catástrofe no le pone freno antes, es el poder sobre el poder, la superación de la impotencia frente a la autoalimientada coacción del poder a ejercerlos progresivamente. Tras haber pasado de un poder de primer grado –dirigido hacia una naturaleza que parecía inagotable– a otro de segundo grado, que arrebató el control al usuario, la autolimitación del dominio –antes de que se estrelle contra los límites de la naturaleza– que arrastra consigo a los dominadores se ha convertido en una tarea de un poder de tercer grado. Éste sería un poder que actuaría sobre el poder de segundo grado, el cual no es poder del hombre, sino poder del propio poder para ordenar su empleo a quien supuestamente lo posee, para convertirlo en abúlico ejecutor de sus capacidades, de tal modo que en lugar de liberar al hombre lo esclaviza (Jonas, 1995: 235-236).

Para contrarrestar este poder es fundamental partir de una razón ética capaz de regularlo. Se torna necesaria la articulación de políticas globales para poner freno a la crisis ecológica que se cierne sobre la humanidad y que inevitablemente desembocará en una reducción de los niveles de bienestar y también de las conquistas democráticas por las que tantas luchas se han librado en la modernidad. Además, el desastroso escenario al que está conduciendo el abuso tecnológico también puede ser aprovechado por los totalitarismos para buscar nuevas formas de tiranía. Por lo tanto, ¿de qué utopía se está hablando? ¿Para qué sirve la utopía si ha tenido graves consecuencias para la humanidad? ¿Estará conduciendo el ideal utópico de la máquina inteligente a un escenario en el que parecemos niños jugando con una bomba, como señala Nick Bostrom en una entrevista para *The Guardian*?

Retomando la crítica que Jonas le formula a Ernst Bloch (2013), mencionada al comienzo de este apartado, sería importante destacar que la utopía se convierte en un «ideal falso» (Jonas, 1995: 263), principalmente porque está estrechamente relacionada con la idea de la abundancia y porque tiene en mente una postura basada en el ideal de perfección del humano y la sociedad. El verdadero progreso no se halla en aspectos circunstanciales que tengan que ver específicamente con el poder técnico, sino que encuentra su lugar en el propio individuo, en esa formación de la madurez a la que conduce la educación y la

experiencia vital. *El principio esperanza* de Bloch lleva a la pérdida del sentido de la actividad humana que se adquiere a partir de la necesidad y de las dificultades enfrentadas. El planteamiento utópico marxista de Bloch desemboca en una pérdida del sentido del goce de la actividad misma, pues el ver una afición ociosa como un trabajo, hace perder el sentido y la deseabilidad a esa utopía. Josep María Esquirol señala lo siguiente:

> Bloch confiesa la idea del «todavía no» del hombre propiamente dicho, mientras que Jonas confiesa la idea de que el hombre está «ya ahí», con todas sus ambigüedades. El «todavía no» significa sobre todo esperanza. El «ya ahí», en cambio, responsabilidad por lo mejor de lo que ya hay. Por eso, el error de la utopía, piensa Jonas, es un error de su concepción de la esencia del hombre (Esquirol, 2012: 134).

Por lo tanto, lo que reivindica Jonas es una primacía de la responsabilidad sobre la esperanza. La crítica a la utopía es una crítica a la utopía tecnocientífica, dado que la técnica encarna el ideal utópico por antonomasia. La crítica que se vierte es sobre aquellos aspectos ocultos que encierra la idea de utopía y que es necesario poner al descubierto mediante la ética de la responsabilidad para no caer en la ingenuidad que conduce a la catástrofe para la humanidad, como bien ha demostrado el siglo XX. Esta crítica está orientada a aquella técnica que se lleva al extremo (Jonas, 1995: 354).

2.6. La disputa con el postulado kantiano

La lectura de la reflexión jonasiana acerca del deber puede parecer en ocasiones confusa en relación al postulado de Immanuel Kant en su *Crítica de la razón práctica*. La postura de Jonas frente a las propuestas kantianas es un tanto ambigua, pues considera que los postulados del de Königsberg para afrontar los desafíos de la sociedad tecnológica actual son insuficientes, y a la vez plantea que pueden servir de buen marco teórico desde el que acceder a problemas propios del campo médico, como señala en el capítulo 6 de *Técnica, medicina y ética*. Salvando la afirmación acerca de cierta ambigüedad, los postulados kantianos que consideran al ser un fin en sí mismo le sirven a Jonas para

apuntalar el valor de la dignidad humana, siendo necesario, sin embargo, contextualizar el escenario de posibilidades que brinda la tecnociencia.

Existe una diferencia insalvable en la estructura de los postulados de ambos filósofos. Kant defiende una coherencia formal de las acciones que son proyectadas desde una legislación de la razón, mientras que Jonas considera que es una instancia extrínsenca, un objeto fuera del sujeto, el que mueve a la propia razón, la que da forma al valor y que es la que facilita la exigencia moral. El imperativo de Jonas no es categórico, sino más bien hipotético, pues como versa la expresión «obra de tal modo que los efectos de tu acción sean compatibles con la permanencia de una vida humana auténtica en la Tierra» (1995: 40), se está proporcionando una determinación que es condicional, donde el bien ordenado se hace categóricamente con un carácter condicional orientando más allá del presente.

Para Kant el deber ser estriba en la autodeterminación que el sujeto se da a sí mismo proporcionándose la ley moral, sin la necesidad de obedecer a una autoridad teológica, que tan importante e impositiva era en su tiempo. En cambio, Jonas busca una autoridad o una verdadera razón que sirva para explicar el fin hipotético que se da en base a la preservación de la vida de forma categórica en su fundamentación metafísica. Es un imperativo que ordena que la humanidad debe ser y por tanto existir, siendo la idea de ser la que fundamenta ese primer aspecto, diferenciándose así del postulado kantiano que busca un principio de actuación en la característica coherente de la razón que se facilita a sí misma sus pautas de conducta. El postulado jonasiano no se encuentra fundado en una ética considerada como doctrina de la acción, sino más bien en una metafísica considerada como doctrina del ser, integrando así la propia idea de ser humano.

Otra importante diferencia entre los postulados de ambos filósofos se encuentra en la justificación del imperativo. Para Jonas esa justificación no radica en la voluntad autónoma y a la vez legisladora del sujeto, sino en la llamada que recibe procedente de un bien cuyo fundamento ontológico, su ser, depende de la acción y el cuidado de dicho sujeto. Jonas no pone el énfasis en el sujeto autónomo, la fuerza sustantiva en la autonomía como una capacidad que es capaz de darse a sí misma las leyes morales, sino más bien en la realidad

que hace el llamado y que demanda atención, ahí es donde estriba la principal fuerza legisladora.

Otro matiz importante a tener en cuenta en la disputa entre estos dos postulados éticos se encuentra en el detonante del deber. Para Kant, que pone el acento en el sujeto trascendental, el deber surge de la conciencia moral, pues en él se descubre la ley moral; mientras que, para Jonas, el deber nace del aumento del poder tecnológico y de las posibles consecuencias que pueden darse en el futuro, siendo este poder una *conditio sine qua non* del deber que promueve la asunción de responsabilidad. No es suficiente la voluntad de querer ser responsable, sino que es ineludible que haya un motivo externo que posea una gran capacidad potencial de realización para que esa responsabilidad sea asumida. Hay en ese sentido una importante inversión del «puedes, puesto que debes» de Kant (1995: 191), a una especie de «debes, puesto que puedes» de Jonas y que se traduce de la siguiente afirmación:

> Lo primero no es ya lo que el hombre debe ser y hacer (el mandamiento del ideal) y luego puede o no puede hacer, sino que lo primario es lo que él hace ya de hecho, porque puede hacerlo, y el deber se sigue del hacer; el deber le es asignado al poder por el *fatum* causal de su hacer. Kant decía: puedes, puesto que debes. Nosotros tenemos que decir hoy: debes, puesto que haces, puesto que puedes; es decir, tu enorme poder está ya en acción. Ciertamente, el sentido y el objeto del poder son diferentes en uno y en otro caso. En Kant se trata de someter la inclinación al deber, y este poder interno, no causal, hay que suponerlo en general en el individuo, que es el único, en efecto, al que el deber se dirige (Jonas, 1995: 212-213).

Para Kant la libertad es trascendental, no se encuentra en el ser. En cambio, para Jonas la libertad es ontológica, pues reside en el ser, siendo el resultado de una tarea teleológica. Puede encontrarse también otra diferencia entre los postulados kantianos y los jonasianos. Para Kant lo que permite un conocimiento de la libertad es la conciencia de la presencia de la ley moral; para Jonas todo sujeto está sometido y debe responder a la ley del ser, que tiene que ver con el cuidado de sí, y ese cuidado será realizable por medio de la libertad, puesto que la libertad es una condición ontológica.

La cuestión de la temporalidad con perspectiva de futuro también es importante en la diferenciación entre el postulado jonasiano y el kantiano. El futuro de posibilidades abiertas se convierte en una dimensión previsible que permite dar forma a la libertad. Esto se contrapone a la idea kantiana de que el imperativo responde a un orden que está siempre presente y es de carácter abstracto.

Una vez pensado esto, podría concluirse que las críticas que Jonas dirige a Kant pueden resumirse en dos grandes líneas de disputa: por un lado, la crítica al formalismo y la oposición a la emotividad que caracteriza el principio de responsabilidad; y, por otro lado, la ausencia de consideración por parte de Kant del futuro como aspecto temporal desde el que dar forma al imperativo.

2.7. La articulación del principio de responsabilidad

La propuesta jonasiana de una nueva ética que supere los límites del antropocentrismo se formula como un principio y adquiere una forma imperativa, tal como indica Jonas cuando escribe:

> «Obra de tal modo que los efectos de tu acción sean compatibles con la permanencia de una vida humana auténtica en la Tierra»; o, expresado negativamente: «Obra de tal modo que los efectos de tu acción no sean destructivos para la futura posibilidad de esa vida»; o, simplemente: «No pongas en peligro las condiciones de la continuidad indefinida de la humanidad en la Tierra»; o, formulado, una vez más positivamente: «Incluye en tu elección presente, como objeto también de tu querer, la futura integridad del hombre» (1995: 40).

El principio de responsabilidad puede ser considerado desde dos perspectivas. Si la cita mencionada se lee detenidamente, podrá razonarse que hace referencia a las consecuencias de los actos. En este sentido, el principio de responsabilidad de Jonas es un principio consecuencialista porque la importancia recae sobre la acción y la intención del sujeto, que debe ser coherente con los efectos de dicha acción, presentes o futuros. Además, este principio también tiene un segundo aspecto que lo caracteriza, a saber, su carácter deontológico, pues el acento se pone en las acciones colectivas, aunque también existe un

componente personal innegable. Esto significa que el principio se sitúa en la estela de una ética social. El imperativo cobra sentido no por un mandamiento de la conciencia o por una decisión individual, sino más bien por una exigencia de supervivencia y bienestar global de la humanidad a la que hay que responder.

No podría entenderse el principio de responsabilidad sin la heurística del temor, pues es a partir de esta técnica de indagación y descubrimiento como se consigue una orientación de la acción de manera que es posible tener en cuenta sus efectos con miras al futuro. Es muy difícil orientar la acción por medio de la previsión, pues el progreso tecnológico es tan vertiginoso que rara vez permite poder anticiparse a los adelantos que estarán por venir en los próximos años y décadas. Por lo tanto, el principio no puede concebirse de forma aislada, es decir, sin su relación con el ejercicio de invitación a pensar en los posibles escenarios futuros que representa la heurística del temor. Responsabilidad y temor se encuentran estrechamente unidos y esa unidad se va conformando con la sabiduría que va siendo acumulada.

Este principio pone en valor el futuro por medio de una conciencia que pretende garantizar la existencia, dando más prioridad a lo que se sitúa a largo plazo, en vez de a aquellos aparentes bienestares a corto plazo y que tantas implicaciones éticas tienen para la humanidad. Además, articula una advertencia que permite orientar y dar forma a la acción desde el presente. Una advertencia que atiende las posibilidades remotas del *malum.* La categoría de «remoto» también es repensada en Jonas, pues en un mundo globalizado donde todo está interconectado y existe una pertenencia a la biosfera parece ser que los efectos de una acción no son tan remotos. Lo que se descubre en un laboratorio de Tokio puede tener directas consecuencias en un centro industrial de México. La vulnerabilidad de la humanidad y la naturaleza se convierte en el principal detonante para el deber moral de protección y salvaguarda que motiva al principio de responsabilidad. El deber de la ética jonasiana se sitúa en este terreno como una prescripción pragmática que se encuentra contenida en la heurística del temor, superando así la prescripción positiva que pertenece más bien al ámbito del derecho. Esta prescripción favorece la exclusión de aquellas acciones tecnológicas que suponen un riesgo para la humanidad y la naturaleza.

En el principio articulado por Jonas la responsabilidad se convierte en aquel sentimiento que mueve a la razón. Las éticas clásicas de la virtud y las éticas modernas del deber no han sabido considerar a la responsabilidad como el pilar sobre el que construir sus reflexiones teóricas. Es por ello que la propuesta del filósofo alemán es un nuevo paradigma ético al levantarse sobre dicha responsabilidad. Además, esta propuesta es novedosa porque no solo encuentra su razón de ser en el plano racional o en su carácter imperativo, sino también en un lado sentimental que es fundamental para el principio y que supone su exigente antecedente. Por lo tanto, el principio de responsabilidad se construye sobre la base de una combinación racional-objetiva y psicológica-subjetiva, como a continuación sugiere Jonas:

> Como cualquier teoría ética, también una teoría de la responsabilidad ha de tener en cuenta ambas cosas: el fundamento *racional* de la obligación –esto es, el principio legitimador subyacente de la exigencia de un «deber» vinculante– y el fundamento *psicológico* de su capacidad de mover la voluntad, es decir, de convertirse para el sujeto en la causa de *dejar* determinar su acción por aquél. Esto significa que la ética tiene un lado objetivo y un lado subjetivo: el primero tiene que ver con la razón, el segundo, con el sentimiento. Históricamente, unas veces el primero y otras veces el segundo ha estado más en el centro de la teoría ética; y tradicionalmente a los filósofos los ha ocupado más la cuestión de la *validez*, esto es, el lado objetivo. Pero ambos son complementarios y ambos son partes integrantes de la ética (1995: 153-154).

Han sido varias las propuestas que han lanzado diversos filósofos a lo largo de la historia, como el *eros* platónico, la *eudaimonía* aristotélica, el respeto kantiano o la voluntad de poder nietzscheana, entre otras, que han servido para orientar la acción moral, pero ninguno de ellos puso el acento en la responsabilidad como lo hizo Jonas.

La filosofía de Jonas es una filosofía que en cierta medida apela al sentimiento que tiene que ver con el conmoverse para tomar partido, para comprometerse frente a los desafíos que ponen en riesgo y que suponen una transformación sin precedentes. Las filosofías fundadas en sentimientos siempre aspiran a un bien supremo al que se dirigen, como la *eudaimonía* aristotélica. En cambio, en el principio de Jonas no existe una

afirmación suprema que esté aferrada a la perfección, sino que más bien encuentra su fundamento en algo contrario. La ética jonasiana estriba en la búsqueda de aquello que se caracteriza por la vulnerabilidad, por aquello que presenta carencia y que está en peligro, lo contrapuesto a las éticas fundadas en la aspiración a la perfección de un bien supremo. La búsqueda de lo vulnerable y lo que está en riesgo se realiza para justificar la necesidad de tomar partido, de asumir conciencia y por lo tanto de emprender una vía de responsabilidad que es un sentimiento que asume un papel relevante para la preocupación por la existencia. No obstante, y es importante aclarar esta cuestión, la reflexión de Jonas no cae en el plano de un emotivismo al uso, sino que más bien recurre al sentimiento para garantizar una buena fundamentación de la interpelación a la que está llamado el ser para implicarse.

Enlazando con lo último que ha sido mencionado acerca de la interpelación a la que está llamado el ser, es importante destacar que el ser humano, como ser moral, es un ser que se caracteriza por la receptividad, siendo precisamente ahí donde estriba su carácter de ser afectado y su capacidad para conmoverse por una llamada de socorro y atención. Escribe Jonas: «Los hombres son ya potencialmente «seres morales», puesto que poseen esa capacidad de ser afectados, y sólo por ella pueden ser también inmorales. (Quien por naturaleza es sordo a esa voz no puede ser ni moral ni inmoral)» (1995: 154).

Esta ética, forjada desde el sentimiento de responsabilidad, responde a aquella llamada del ser que debe ser atendida. Esa tarea de respuesta implica una capacidad para reconocer los valores implícitos, considerando la vulnerabilidad de los mismos, convirtiéndose así la respuesta en un acto que brota de la propia libertad en la asunción de responsabilidad. El carácter de ser moral se convierte así en el atributo humano que se caracteriza por una madurez que da respuesta a aquellos valores que demandan ser preservados. Es el valor el que motiva a actuar, el ser moral es aquel que se siente motivado a actuar y por lo tanto a ser responsable.

También es importante destacar, antes de terminar este capítulo, que Jonas fundamenta su ética de la responsabilidad en la *phrónesis* aristotélica. El concepto de prudencia de Aristóteles es una virtud necesaria desde la que afrontar los retos actuales. Es precisamente el escenario tecnológico el que hace necesario forjar un *ethos* virtuoso. La prudencia se convierte en la guía de la acción técnica, como la virtud más pertinente para saber gestionar esa desafiante relación entre el saber y el poder. La prudencia es astuta, pues debe saber lidiar con equilibrio las posibilidades que conceden la libertad y la apertura. Su papel mediador es muy importante en la guía de la acción.

En un mundo imprevisible se hace cada vez más necesario actuar con prudencia y responsabilidad, tomando conciencia de que toda acción implica siempre un riesgo, y que no siempre es tan fácil distinguir entre el bien y el mal. Esto significa en ocasiones que las decisiones provisionales y basadas en la prudencia son las que dan algún tipo de pauta de actuación que permite un acercamiento a aquellas posibilidades que no comprometen la capacidad ética del humano.

2.8. La práctica del principio de responsabilidad

Las pretensiones dominadoras de la técnica sobre la naturaleza tienen como contrapartida que el humano se convierta también en objeto de la dominación de la técnica. En el apartado VII del capítulo 1 del *El principio de responsabilidad*, Jonas reflexiona acerca de la cuestión del humano como objeto de la técnica desde la exigencia de una ética marcada por la previsión y la responsabilidad ante las circunstancias enfrentadas con el poder técnico. Se trata de poner el foco ahora en las obras del *homo faber,* aquellas obras que ya no están orientadas hacia la aplicación de la técnica sobre el ámbito no humano, sino más bien en el propio humano, incluido ahora entre los objetos del poder. El *homo faber* opera ahora sobre sí mismo, pues piensa que él mismo puede ser objeto de la *techne*. Jonas divide la acción de la técnica sobre el humano en tres partes: la prolongación de la vida, el control de la conducta y la manipulación genética.

En lo que respecta al tema de la prolongación de la vida, es un asunto que será abordado en la tercera parte de este trabajo, considerándolo como un desafío de la IAR que se encuentra en la estela transhumanista. Sin embargo, es interesante abordar este tema dentro de las coordenadas del pensamiento de Jonas para ver qué implicaciones éticas tiene en el marco de su teoría de la responsabilidad. El filósofo alemán es consciente de que los progresos alcanzados para contrarrestar los efectos del envejecimiento y el aumento de la vida humana son cada vez más posibles. La muerte no es vista como una dimensión intrínseca de la naturaleza humana, sino como una etapa de la vida que debe ser desafiada y superada. Jonas promueve un cuestionamiento a partir de la relación que existe entre la posibilidad y el deseo, es decir, la posibilidad de desafiar la muerte existe, no obstante, se pregunta qué tan deseable es: «¿Hasta qué punto es tal cosa deseable?», «¿hasta qué punto es deseable para el individuo y para la especie?» (1995: 50). Luego también formula cuestionamientos sobre los beneficiarios, si de verdad será, o no, una cuestión de importancia social, etc. La ampliación de la etapa de la vejez podría suponer una reducción de la presencia de población joven en las sociedades, algo que no reportaría un beneficio, ya que la juventud encarna:

> […] originalidad, inmediatez y ardor […] La mayor acumulación de experiencia prolongada no reemplaza a esas cosas; nunca puede recuperarse el singular privilegio de contemplar el mundo por primera vez con ojos nuevos, nunca revivir el asombro –que constituye para Platón el comienzo de la filosofía–, nunca sustituir la curiosidad del niño, curiosidad que desfallece en el adulto y que muy raras veces se convierte en afán de conocimiento (1995: 51).

En ese sentido, la prolongación de la vida no sería pertinente, pues la juventud es necesaria al cumplir una tarea de rejuvenecimiento de toda aspiración vital.

La prolongación también podría significar un olvido de la muerte como referencia temporal a partir de la cual construir un proyecto de vida. La vida se construye tomando como referencia la muerte, entre otras cosas también relevantes, pues sirve como un punto fronterizo desde el que organizar el tiempo. En ese sentido, relegar la vida hacia una indeterminación temporal generaría una falta de estimulación y de ausencia de proyecto

vital. Esto lo considera Jonas un olvido del *memento mori* (1995: 51). En definitiva, la prolongación de la muerte podría convertirse en un perjuicio antes que en un anhelo profético.

También hay otro tema que despierta cierto interés en Jonas y al que dedica una pequeña reflexión: el control de la conducta. Para el alemán este desafío no comporta tantas cuestiones éticas como el de la prolongación de la vida, pero no por eso debe evitarse su reflexión, pues se encuentra temporalmente más cercano que las técnicas de prolongación de la vida. El objeto de esta técnica es la influencia sobre la mente para controlarla mediante agentes químicos o influjos eléctricos que se impulsan a través de electrodos (1995: 52). En ocasiones es difícil dibujar una frontera entre lo saludable y lo dañino en estos temas, apunta Jonas. El paradigma terapéutico de la medicina establece un estrecho vínculo entre el uso médico y el uso social que perfila el horizonte de posibilidades que presenta la técnica. Ese horizonte de posibilidades, que plantea un vínculo entre los usos médicos y sociales, podría llevar a un escenario de indefinidas potencialidades preocupantes (1995: 52). Los usos de la técnica en el terreno médico podrían ser utilizados para mecanismos de poder y manipulación social, fenómeno que comportaría importantes riesgos. Podrían inducirse muchos comportamientos, fenómeno que plantearía infinidad de cuestionamientos éticos, por ejemplo ¿hasta qué punto es respetada la autonomía y la dignidad de un trabajador si su empresa utiliza técnicas para el aumento de su productividad? Para Jonas las posibilidades presentadas invitan al cuestionamiento sobre la idea de ser humano con la que se adquiere un compromiso.

Por último, dentro de esas técnicas de las que es objeto el humano, se encontraría la manipulación genética. Al igual que el tema de la prolongación de la vida, este tema será abordado en la tercera parte, pues es uno de los principales focos de investigación del transhumanismo. El tema no se aborda, pero sí se menciona, en *El principio de responsabilidad*, aunque Jonas apunta a que será tratado de manera más extensa en su obra *Técnica, medicina y ética*, dedicada a la dimensión práctica del principio de responsabilidad.

Jonas es consciente de las importantes posibilidades técnicas que proporciona la biología molecular y las ciencias biológicas en general y eso lo lleva a creer que el primer mandato moral debería ser la cautela y la primera tarea el pensamiento hipotético (1997: 109). La relación epistémica sujeto-objeto, que tradicionalmente estaba formada por humano-naturaleza, ahora se ha transformado en humano-humano. En este campo subyacen importantes problemáticas que tienen que ver con la irreversibilidad, el poder, la manipulación, la experimentación, el derecho, la capacidad de predecir, etc. La manipulación genética plantea serios interrogantes donde el poder tecnológico puede cegar las últimas consecuencias y promover la ignorancia. Las consecuencias pueden ser demoledoras, motivo por el que es fundamental la formulación de una nueva ética, ya que podrían introducirse modificaciones que fueran irreversibles. La técnica ha provocado que el humano se vuelva peligroso para sí mismo, poniendo en peligro el equilibrio cósmico y biológico que da forma a los cimientos de la vida de la humanidad.

En definitiva, la técnica conduce a una transformación de la tradicional relación sujeto-objeto, donde el humano se sitúa en el centro del huracán, sometido a una técnica irreflexiva sobre sus posibles consecuencias y cautiva con su atractivo poderío. Por ello, como apuntó Jonas, es fundamental la creación de una nueva ética que incorpore aspectos relativos, no solo a lo humano, entendido antropológicamente, sino también a la naturaleza, entendida como ese espacio en el que la vida cobra su sentido más biológico y se desenvuelve siguiendo su curso orgánico.

2.9. Consideraciones críticas

Como todo filósofo, Jonas también es susceptible de recibir críticas de quienes consideran que existen otras perspectivas desde las que afrontar los temas que él discute. Así pues, este apartado estará dedicado a exponer algunas de las posibles críticas que se pueden formular a partir de la lectura de Jonas, con el fin de que ese ejercicio sirva para ampliar el horizonte de la reflexión en el planteamiento de una IAR.

En este capítulo se ha realizado un acercamiento a aquellos planteamientos del filósofo alemán sobre el poder de la técnica y la necesidad de una reflexión ética para considerar sus implicaciones. En ese sentido, y vale la pena recordarlo, el principio ético está fundado sobre un metaprincipio ontológico que es preliminar a la ética, y que es el único que puede ser visto como incondicional y por lo tanto categórico. Cualquiera que sea el imperativo que después se plantee habrá surgido de este metaprincipio que se construye sobre un axioma esencial en el pensamiento jonasiano, a saber, que el ser, reflejo de la existencia del ser humano y la biosfera, es un bien en sí mismo, mientras que el no-ser es representado como un mal. Hay quienes consideran superada en la actualidad esta propuesta ética por encontrarse fundamentada sobre una idea metafísica. Por ejemplo, el italiano Paolo Becchi (2008) es uno de los críticos que identifica algunos puntos de los postulados de Jonas susceptibles de problematización.

Becchi sostiene que la fundamentación planteada por el filósofo alemán sobre un presupuesto ontológico-teleológico, cercano a un neoaristotelismo que pone el acento en la teleología aristotélica, se encuentra más en sintonía con una especie de imperativo cuasi-biológico que con un imperativo verdaderamente moral, que es lo correspondería especialmente a un planteamiento ético. Su fundamentación se aleja del imperativo moral y se sitúa en un espacio más cercano a la biología, porque con su defensa a ultranza de la preservación de la vida humana sobre la tierra da motivos suficientes como para creer que la verdadera razón es biológica y no tanto moral. Además, la exigencia de existencia de la humanidad como condición suficiente para fundamentar una ética de la responsabilidad a nivel planetario es insuficiente para Becchi. Sin embargo, el italiano advierte de que la cuestión ontológica es hoy imprescindible, pues el poder de la acción no solo compromete a la naturaleza sino también a la especie (2008: 122).

También es importante destacar un artículo escrito por Amán Rosales Rodríguez (2004) que refleja algunas de las críticas formuladas al principio de responsabilidad de Jonas. En su texto, Rosales Rodríguez comienza esbozando la crítica que vierte Willi Oelmüller (1988) sobre el postulado ético de Jonas. Oelmüller sostiene que Jonas se contradice cuando afirma que la experiencia religiosa judeo-cristiana se encuentra en picado y que por

lo tanto es fundamental un esfuerzo racional para la ética que sea secular e independiente de toda experiencia religiosa, y que a la vez utilice un tono predominantemente religioso con expresiones como «sagrado», «veneración», «piedad», etc. Las tesis de Jonas serían contradictorias debido a que plantean una propuesta secular y a la misma vez emplean un tono frecuentemente religioso.

Además, Oelmüller afirma que la demanda de responsabilidad ética implícita en el principio de responsabilidad se encuentra desfasada, pues ya existen diversos ámbitos que están asumiendo responsabilidad ética para orientar sus actividades sin la necesidad de tener que recurrir a una fundamentación ontológica, como ocurre en la bioética.

Richard J. Bernstein (1994) es otro de esos pensadores que trata de valorar críticamente los principales elementos de la propuesta jonasiana. En concreto, la relación padre-hijo que Jonas utiliza para demostrar la posibilidad real del paradigma de responsabilidad no resulta convincente, pues habría descuidado otra cara de las relaciones humanas que tiene que ver con una responsabilidad instaurada a partir de las relaciones marcadas por la reciprocidad. Esta falta de reciprocidad podría desembocar en un exceso de paternalismo, y ya son conocidas las consecuencias que dicho paternalismo ha tenido para la humanidad, sobre todo en el terreno político. Es importante aclarar que el hecho de rechazar un paternalismo político no implica que se esté negando la exigencia de responsabilidad a los políticos en relación con el bienestar de una comunidad. Sin embargo, Jonas comete un error, según Bernstein, cuando reduce la responsabilidad política a una relación vertical, como es la relación entre padre e hijos. Esta verticalidad podría ocasionar un abuso de autoridad por parte de determinados grupos de poder sostenido en procedimientos políticos de orientación tecnocrática.

La inspiración vertical del poder político puede generar una situación delicada ya que existe la posibilidad de sugerir un gobierno elitista formado por un grupo de «expertos en la materia» que supuestamente atesorarían todo el conocimiento y por lo tanto tomaran todas las decisiones de manera unilateral. Además, otra advertencia de Bernstein consiste en la ausencia de reciprocidad que supondría una puesta en peligro para la autonomía y el poder

de decisión individual, caracteres inherentes a la propia humanidad. Por lo tanto, la ética jonasiana destacaría por incluir rasgos de elitismo y verticalidad.

Pero eso no es todo. Según Bernstein, existe una importante contradicción en Jonas, a saber, la superación de una ética supuestamente antropocéntrica que tanto critica y que luego al parecer se encuentra también presente en su propuesta. Es una tarea que tiene pendiente resolver el alemán, pues no consigue verdaderamente un distanciamiento de esa visión antropocéntrica que él se dedica a criticar y a tratar de superar. Bernstein piensa que esta crítica no está totalmente superada cuando observa que el deber ético principal, que es la preservación de la vida humana, es un asunto de responsabilidad que se limita básicamente a ser una competencia de las personas, lo que supone que la visión antropocéntrica no esté siendo superada. También encuentra Bernstein un vacío argumental en la propuesta de Jonas, concretamente en el paso de la premisa que sostiene la obligación suprema de preservar las condiciones necesarias para la vida a la conclusión que trata de evidenciar que la existencia humana se convierte en un asunto prioritario. Algo falta ahí según el crítico, pues no bastaría con asumir que el deber-ser ya estaría implícito en la iniciativa que el humano asume por la vida misma. Ese vacío argumental que señala Bernstein tiene que ser sustituido por un argumento que consiga dar razón de la conclusión. Es muy importante mejorar la argumentación mediante el enriquecimiento de otros principios, «para que no se convierta en una propuesta noble pero vacua (Rosales Rodríguez, 2004: 102). Quizás una verdadera superación de las éticas anteriores, esas que Jonas considera antropocéntricas, sería un buen punto de partida. Como señala Bernstein, parece ser que Jonas es demasiado exagerado cuando plantea la necesidad de romper con toda la tradición ética anterior por considerarla antropocéntrica. Así pues, en términos generales, la crítica de base que realiza Bernstein sobre Jonas destaca por una carencia fundamental: la falta de argumentos adicionales que sirvan para enriquecer de una mejor manera el postulado del alemán.

No obstante, todas las críticas formuladas sobre Jonas no son de la misma intensidad, pues hay pensadores como M. H. Werner (2003) que señalan que el filósofo alemán en su crítica a las éticas anteriores no pretendía sustituirlas, sino más bien complementarlas. Pues

éstas carecían de una suficiente consideración sobre el fenómeno de conservación biológico del ser humano y la biosfera. Jonas no elaboró un sistema de ética normativa, motivo que lleva a Werner a pensar que no pretendía bajo ningún concepto sustituir las «éticas habidas hasta ahora», sino que buscaba su complementariedad. Como señala Rosales Rodríguez, a partir de la lectura de Werner, «si ello es así, entonces parece haber razón también en hablar de una ética jonasiana moderadamente antropocéntrica y cualificadamente presentista» (2004: 103). Pero eso no es todo, ya que Werner también cree que el postulado de Jonas impone la necesidad de una mayor riqueza argumentativa, pues cuando Jonas defiende la idea de una existencia auténtica en la tierra que debe ser salvaguardada no da muchos más detalles acerca de lo que significa eso de «auténtica». Las ideas sostenidas destacan por su excesiva generalidad y carencia de especificidad, según Werner.

Rosales Rodríguez también destaca la contribución de W. E. Müller (1989) al conjunto de las críticas sobre los postulados jonasianos. Más allá de las críticas que realiza sobre Jonas, Müller considera que el filósofo del principio de responsabilidad ha hecho importantes contribuciones a las discusiones filosóficas de su tiempo. Para Müller la heurística del temor que propone Jonas nace de la imposibilidad del filósofo alemán para dar argumentos suficientes que puedan explicar la motivación que se encontraría detrás del deber ser a partir del ser. Müller cree que Jonas no trata la cuestión de la motivación porque quizás no tendría los argumentos suficientes como para justificar la derivación del deber ser a partir del ser. La heurística del temor conduce hacia la concesión de prioridad teleológica de la naturaleza, algo que no es suficiente para una fundamentación lógica. Para Müller el deber ser no puede apoyarse directamente en el ser, ni nacer de su motivación, sino que más bien su razón de ser debe recaer en otro elemento normativo que sea independiente del ser y que le transmita al humano, mediante la sensibilidad, el sentido del deber en lo que respecta a la conservación del ser (Rosales Rodríguez, 2004: 106).

En este apartado se han presentado algunas de las críticas formuladas a las propuestas de Hans Jonas. Sin embargo, como se ha indicado, Jonas al igual que otros pensadores es susceptible de críticas y la lectura posterior de sus postulados sirve para dar cuenta de que su pensamiento puede ser sometido a un proceso de enriquecimiento y actualización. De no

ser así se estaría cometiendo una injusticia y además no se asumiría la tarea de ser «albaceas a cargo de una historia» (Vidarte y Rampérez, 2005: 14) que nos ha sido encomendada como últimos testigos de la larga tradición filosófica que brinda la historia. Por lo tanto, la consideración de las críticas forma parte del ejercicio vinculado con la construcción y el beneficio del conocimiento. De todas formas, más allá de toda crítica, la propuesta de Jonas puede servir como un primer paso en el planteamiento de una IAR y también para sentar las bases de un humanismo tecnológico que pueda enfrentar los desafíos futuros que depara el desarrollo tecnológico.

Otro aspecto que debe tenerse en cuenta más allá de toda crítica sobre su pensamiento, y que sirve como medida de reconocimiento, es el énfasis en la relación antagónica entre el humano y la naturaleza, idea que Jonas recoge de la filosofía existencial de Heidegger, y que transforma en una dura crítica a la racionalidad tecnocientífica que marca la dinámica de la modernidad. Jonas es sin duda un «hijo» de Heidegger como señala Richard Wolin en su obra *Los hijos de Heidegger: Hannah Arendt, Karl Löwith, Hans Jonas y Herbert Marcuse*. Es consciente de la complejidad del fenómeno tecnológico de su tiempo, y de un tiempo venidero, y reconoce que su propuesta ética no da una solución definitiva a esta problemática. Como reconocimiento a su obra y a la humildad con la que brinda sus postulados, concluimos esta parte con unas palabras de este gran pensador alemán:

> Una cosa debemos tener por fin clara: una solución patentada para nuestro problema, un remedio universal a nuestra enfermedad, no existe. Para algo así, el síndrome tecnológico es demasiado complejo, y en una renuncia no cabe ni soñar. Incluso con una gran «inversión de la marcha» y una reforma de nuestras costumbres, el problema fundamental no desaparecería […] La misión de evitar es, pues, permanente, y su cumplimiento no debe ser nunca más que un remiendo y, a menudo, incluso no más que una chapucería… (Jonas, 2001: 132).

2.10. Hans Jonas: fundamento útil para la inteligencia artificial responsable

El legado filosófico de Jonas ofrece un rico basamento desde el que formular un concepto de IA que encuentre su razón de ser en una preocupación por la responsabilidad en el momento de su despliegue. Así pues, más allá de otros pilares sobre los que se

sustentará la IAR, entre los que puede encontrarse el MIAR, los derechos humanos, los ODS y los límites planetarios, las principales ideas que propone Jonas en torno a la necesidad de adquirir compromiso frente al poder tecnológico son de gran utilidad para encontrar un fundamento filosófico en la formulación de un concepto capaz de afrontar los desafíos tecnológicos del futuro. Entre las principales ideas que pueden recogerse del pensamiento del filósofo alemán para el fundamento de la IAR destacan las siguientes:

- Búsqueda de aquello que es vulnerable y con lo que hay que comprometerse.
- Compromiso con el futuro desde el presente mediante la consideración de las consecuencias de la acción tecnológica.
- Compromiso de carácter universal e integral.
- Responder a aquellas llamadas que deben ser atendidas.
- Criticismo frente a la irreflexividad que promueven ciertos ideales de progreso.
- Cuestionamiento del empleo de la tecnología como medio de dominación.
- Imperativo de una ética capaz de reconocer la amplitud y dimensión de los problemas más allá de los caracteres antropocéntricos.
- Necesidad de reconocer el poder destructor que también presenta la tecnología, con el objetivo de que no solo sean destacables los «brillos» de la misma.

Estas son algunas de las ideas que pueden servir para articular un concepto de IAR que encuentre un sólido fundamento en la filosofía. No obstante, como se ha dicho, también se presentarán en el capítulo 5 otros sustentos filosóficos de esta innovadora propuesta.

SEGUNDA PARTE

INTELIGENCIA ARTIFICIAL RESPONSABLE

CAPÍTULO 3

APROXIMACIÓN AL CONCEPTO DE INTELIGENCIA ARTIFICIAL

> Han construido un arma tan poderosa que, con ella, podrán derrotar tanto a los vencedores como a los vencidos. Y no es la bomba de hidrógeno […] Lo que nos espera no es el olvido, sino un futuro que, desde nuestra ventajosa situación actual, se puede describir con las palabras «posbiológico» o, mejor aún, «sobrenatural». En ese mundo, la marea del cambio cultural ha barrido al género humano y lo ha sustituido por su progenie artificial.
>
> (Moravec, 1990: 13)

Puesto que el objeto de la discusión de este trabajo se centra en promover la reflexión sobre la necesidad de incorporar criterios de responsabilidad ética en el campo de la IA, resulta pertinente aclarar primeramente en qué consiste el universo de los intelectos sintéticos. Es fundamental comenzar la tarea de esclarecimiento delimitando el concepto y significado de IA, con el fin de facilitar la tarea de comprensión sobre este objeto de estudio.

En primer lugar, se presentará una definición del concepto de IA, destacando sus principales rasgos en torno a la inteligencia contextualizada en el ámbito artificial. El análisis del origen histórico también es un aspecto importante que siempre debe estar presente en el esclarecimiento de cualquier concepto. Las investigaciones en el campo de la computación y las ciencias cognitivas han favorecido el vertiginoso desarrollo de la IA y su aplicación en diversidad de esferas.

No solo se presentarán los orígenes y desarrollo de este campo de estudio, sino también los proyectos que se esperan de los sistemas artificiales gracias a las predicciones formuladas por especialistas como Toby Walsh (2017) o Lasse Rouhiainen (2018). Aunque estas predicciones puedan parecer propias de un relato de ciencia ficción, es conocido que

el poder tecnológico que ha experimentado el ser humano en las últimas décadas invita a cuestionar la delgada línea roja entre realidad y ficción.

Por último, se abordarán dos de los postulados más actuales y con una mayor visibilidad dentro del ámbito de la IA, el concepto de superinteligencia de Nick Bostrom y el de singularidad de Raymond Kurzweil. Aunque ambas propuestas destacan por sus evidentes diferencias, las dos tienen en común la seguridad de que los sistemas artificiales experimentarán profundos cambios en las próximas décadas y supondrán una importante revolución en la complejidad del universo humano. ¿Será la IA nuestra invención final, como sugiere James Barrat (2015)?

3.1. Concepto y significado de inteligencia artificial

La búsqueda de una definición de IA capaz de recoger todas las consideraciones que pueden llegar a existir sobre este término resulta una tarea compleja. No obstante, el propósito de este capítulo consiste en aclarar, en la medida de lo posible, el concepto y significado de IA.

Como ocurre en infinidad de ámbitos en lo que se refiere a la conceptualización, existen muchas definiciones en torno a la IA, cada una propuesta desde diferentes enfoques, aunque al parecer todas tienen un punto en común. Ese punto en común consiste en una idea fundamental en torno a la que giran las diversas propuestas, a saber, la idea de crear y dar forma a programas de ordenador, o también a máquinas, que sean capaces de desarrollar conductas consideradas inteligentes si las realizara un ser humano. Esta definición es abierta y puede generar consenso, pues la variedad de definiciones facilitadas por algunos expertos en la materia suelen ser cerradas y diferenciarse unas de otras, y al menos en este caso, es más prudente mostrar cierta apertura. Esta definición, fundamentada en la búsqueda de la emulación del cerebro humano, es similar a la propuesta formulada por John McCarthy, Marvin L. Minsky, Nathaniel Rochester y Claude E. Shannon en 1955 (Copeland, 1996). En lo relativo a la definición de IA, la británica Margaret A. Boden también señala lo siguiente:

> La inteligencia artificial (IA) tiene por objeto que los ordenadores hagan la misma clase de cosas que puede hacer la mente.
>
> Algunas (como razonar) se suelen describir como «inteligentes». Otras (como la visión), no. Pero todas entrañan competencias psicológicas (como la percepción, la asociación, la predicción, la planificación, el control motor) que permiten a los seres humanos y demás animales alcanzar sus objetivos.
>
> La inteligencia no es una dimensión única, sino un espacio profusamente estructurado de capacidades diversas para procesar la información. Del mismo modo, la IA utiliza muchas técnicas diferentes para resolver una gran variedad de tareas.
>
> [...] La IA tiene dos objetivos principales. Uno es tecnológico: usar los ordenadores para hacer cosas útiles (a veces empleando métodos muy distintos a los de la mente). El otro es científico: usar conceptos y modelos de IA que ayuden a resolver cuestiones sobre los seres humanos y demás seres vivos (Boden, 2017: 11-12).

Esta definición parte de un enfoque que aparentemente puede presentar algunos defectos, pues se propuso en 1955 y la IA ha experimentado un importante desarrollo en las últimas décadas. En ese sentido, no es fácil la tarea de proporcionar una definición concisa de IA, y lo mismo ocurre con la inteligencia humana. Muchas empresas al realizar la selección de personal intentan cuantificar la inteligencia, como si esta inteligencia pudiera ser objeto de cuantificación. Sin ir más lejos, en los sistemas educativos formales e institucionalizados también se realiza una suerte de cuantificación de los conocimientos que se encuentran estrechamente vinculados con la inteligencia.

Sin embargo, esto no niega la posibilidad de que puedan existir consensos en torno a ciertos marcadores de la inteligencia en numerosos contextos concretos. La duda surge a la hora de intentar aplicar esos marcadores a la máquina. Por ejemplo, si se tiene en cuenta la ardua tarea realizada por los escribas en el Antiguo Egipto con los manuscritos para reproducir textos con el fin de transmitir conocimientos y se la compara con una imprenta de libros de texto escolares de la actualidad, la máquina sería «más inteligente» porque es

más rápida reproduciendo textos que un ser humano; en este caso se ha tenido en cuenta el marcador de la velocidad. Como puede comprobarse, el marcador de velocidad no es un indicador suficiente para considerar a una máquina más inteligente que a un ser humano.

La consideración de las capacidades humanas como un criterio válido para valorar la IA supondría un problema. Una máquina puede llegar a realizar una tarea en milisegundos y una persona no sería capaz de realizar una tarea parecida en ese corto periodo de tiempo, por lo que podría pensarse que la máquina parece dar muestras de inteligencia. Episodios de ese tipo van a ocurrir en las próximas décadas en cientos de ámbitos, incluso en muchos ya se han dado. Esto puede ser un motivo por el que la utilización del método comparativo entre la inteligencia humana y la IA podría conducir al absurdo, pues la inteligencia humana siempre saldría perdiendo en una diversidad de situaciones. Esto muestra una vez más que la empresa de ofrecer una definición concreta de la IA no es nada fácil.

En medio del abanico de posibilidades que pretenden formular una definición de la IA, podría suponerse que un sistema de IA debe poseer algunas características consideradas como básicas en todo sistema. La capacidad para aprender es una de esas características del diseño básico de un sistema para alcanzar IA. Otra característica básica consiste en manejar la incertidumbre y la información probable, así como en la formación de conceptos a partir de representaciones combinatorias que son usadas en el razonamiento lógico e intuitivo. Además de poseer esos caracteres básicos, en palabras de Nils J. Nilsson:

> La inteligencia artificial (IA), en una definición amplia y un tanto circular, tiene por objeto el estudio del comportamiento inteligente en las máquinas. A su vez, el comportamiento inteligente supone percibir, razonar, aprender, comunicarse y actuar en entornos completos. Una de las metas a largo plazo de la IA es el desarrollo de máquinas que puedan hacer todas estas cosas igual, o quizá incluso mejor, que los humanos. Otra meta de la IA es llegar a comprender este tipo de comportamiento, sea en las máquinas, en los humanos o en otros animales (Nilsson, 2001: 1).

La IA pretende fundamentalmente desarrollar comportamientos en las máquinas que muestren inteligencia en entornos complejos. Esta definición de Nilsson encaja perfectamente con el espíritu original de McCarthy, pues el objetivo fundamental consiste en averiguar si la inteligencia que podría obtener una máquina puede llegar a ser similar a la del humano.

3.2. Paradigmas de desarrollo de la inteligencia artificial

La IA ha logrado numerosos avances de gran utilidad en las últimas décadas. Más allá de las diferentes creencias en torno a los tiempos de aproximación para conseguir una inteligencia humana, que se abordarán más adelante, es pertinente mostrar brevemente los paradigmas que se aproximan a la IA y que sirven para dar una pequeña orientación en este campo tan amplio y complejo.

El primer paradigma se construye sobre una aproximación basada en el procesamiento de símbolos. Este paradigma se fundamenta sobre la hipótesis del sistema físico de símbolos que impulsaron Allen Newell y Herbert A. Simon en 1961 (1974) y sirve para dar forma a lo que hoy se conoce como IA «clásica». Uno de los pioneros de esta línea paradigmática es el viejo conocido John McCarthy, que propone la aplicación de operaciones lógicas que representen el conocimiento sobre el dominio de sentencias declarativas. Este paradigma está profundamente enraizado en la lógica a partir de aproximaciones simbólicas basadas en el conocimiento.

El segundo paradigma consiste en lo que se denominan «aproximaciones subsimbólicas». La idea principal, en términos generales, es que existe una lógica ascendente desde niveles inferiores hasta niveles superiores en lo que respecta a la formación de la IA. Entre los principales defensores de este paradigma se encuentran Stewart W. Wilson (1991) y Rodney A. Brooks (1991), quienes defienden que para el desarrollo de la IA es necesario pasar por un proceso evolutivo, al igual que la inteligencia humana se desarrolló solo después de más de mil millones de años en la Tierra. Todo es cuestión de evolución.

Un claro ejemplo de máquinas que proceden del paradigma subsimbólico son las redes neuronales que se encuentran inspiradas en modelos biológicos y son interesantes por su curiosa capacidad de aprendizaje. Evidentemente, una vez hecho un breve repaso a los paradigmas de la IA, podrían considerarse cuáles han sido los ámbitos en los se han ido gestando los diversos proyectos y los caminos emprendidos en las últimas décadas.

3.3. Antecedentes históricos de la inteligencia artificial

Para una mayor aproximación al concepto de IA y su estado actual, es importante tener en cuenta los antecedentes históricos y orígenes. El conocimiento sobre el origen de las ideas facilita la tarea de esclarecimiento y contribuye a una mejor comprensión de los objetos de estudio.

3.3.1. Turing y su reflexión filosófica

Resulta curioso que antes de poder hablar de la IA como ámbito de investigación material surgiera un espíritu de impulso de la IA, lo que podría venir a llamarse Filosofía de la Inteligencia Artificial. El principal impulsor de este espíritu fue el lógico y matemático británico Alan Turing, que pasará a la historia por su amplia contribución al campo de la IA, aunque es importante mencionar que la historia lo trató injustamente por su condición de homosexual. En ocasiones hay quienes identifican a Turing como el padre fundador de la IA, cuando en realidad eso no es cierto, ya que el británico no trabajó en ningún programa concreto de IA como tal. Sin embargo, lo que sí hizo, como ya se ha indicado, fue impulsar la reflexión filosófica sobre los aspectos pensantes de las máquinas, algo que dista mucho del desarrollo de un verdadero programa de IA.

En el ámbito de la IA una de sus principales contribuciones fue el famoso test de Turing, publicado en 1950 en la revista *Mind*. Este test nace como un método para comprobar y determinar si una máquina puede pensar. Se introduce en el terreno para probar la habilidad de una máquina y compara su conocimiento con el de un humano, con el objetivo de esclarecer si existe una similitud. Según la propuesta de Turing, si una

máquina es capaz de engañar o manipular a un humano –al estilo del filme *Ex Machina*–, entonces existirá una posible IA. El británico realiza una verdadera reflexión filosófica sobre los horizontes pensantes de una máquina, ya que el test de Turing abrió un abanico de posibilidades para el desarrollo del campo de la IA, facilitando que la programación sea cada día más compleja y sofisticada.

El test de Turing consiste en una prueba en forma de juego donde dos personas, un hombre y una mujer, establecen una conversación con un interrogador que se encuentra apartado para que no sea posible reconocer su identidad. La cuestión es que el interrogador debe reconocer quién es la mujer y las otras dos personas tienen que convencerlo a éste de que son la mujer. La variable que introduce Turing para dificultar la tarea de reconocimiento de la mujer por parte del interrogador es un ordenador. El objetivo no es que el ordenador sepa o no imitar a una mujer, sino intentar convencer y confundir al interrogador. Una máquina supera el test cuando el interrogador no la logra reconocer en un número significativo de ocasiones. Para Turing lo importante es el resultado del juego y no los métodos empleados.

Al poco tiempo de aparecer el test se vertieron varias críticas desde el ámbito de la ética y la religión, apelando a que ninguna máquina podía compararse con una persona y que la máquina nunca podría igualar sus capacidades. También se realizaron críticas desde el solipsismo, reclamando que las máquinas no pueden tener conciencia de sí mismas, algo que los especialistas en el campo de la IA ya no se atreven a afirmar tan ligeramente para las próximas décadas. Como toda propuesta en materia de conocimiento, el test de Turing no iba a quedar exenta de críticas.

Por lo tanto, Turing no estaba tan equivocado y fue un visionario en el campo de la IA, aunque como se ha dicho, sin desarrollar un programa de IA, sino únicamente reflexionando filosóficamente sobre sus horizontes. El tiempo le ha dado la razón a Turing, ya que los importantes avances en materia de IA han demostrado que las máquinas pueden desarrollar una capacidad muy superior a la del ser humano en muchos ámbitos. Incluso cuando Turing publicó su famoso artículo, ya existían en Gran Bretaña y EE. UU. unos

computadores que despertaban un gran interés en el público experto, aunque desde cierto sensacionalismo. Estos computadores eran el *Mark I*, de Manchester, el EDSAC (*Electronic Delay Storage Automatica Calculator*) de Cambridge, y el ENIAC (*Electronic Numerical Integrator and Computer*) de EE. UU.

En medio de estos comentarios sensacionalistas, e incluso de desprecio, como fue el caso de sir Geoffrey Jefferson, profesor de Neurocirugía en Manchester, que despreciaba la idea de que una máquina pensara, Turing elevó el debate con la publicación de su artículo hasta un nivel de cuestionamiento científico y filosófico. En definitiva, Turing abrió el camino de la reflexión filosófica en materia de IA y también contribuyó en importantes proyectos como el descifrado de los códigos de *Bletchley Park*.

3.3.2. Un verano de 1956 en Hannover

Otro acontecimiento importante sucedió en el *Dartmouth Summer Research Project on Artificial Intelligence* en la Universidad Darthmouth College. Este encuentro, desarrollado en el verano de 1956, simboliza el germen de la IA como campo de investigación. Contó con la presencia de importantes investigadores que luego pasarían a considerarse como los pioneros de la IA, entre los que puede encontrarse al impulsor del evento, John McCarthy, y también a Ray Solomonoff, Herbert A. Simon, Trenchard More, Allen Newell, Oliver Selfridge y Arthur Samuel, entre otros. Todos los investigadores allí presentes compartían el mismo interés por las teorías autómatas, el estudio de la inteligencia, las redes neuronales, etc. Así pues, la principal inquietud de estas mentes brillantes era la inteligencia de los computadores.

No obstante, según la consideración de John McCarthy, ese encuentro no fue un éxito debido a que cada mente brillante no intercambió verdaderamente ideas con las otras mentes. El propio McCarthy lo expresaba así: «Para mí fue una gran frustración [...] Tampoco hubo, por lo que yo pude ver, ningún intercambio auténtico de ideas» (McCorduck, 1991: 95-96). Sin embargo, el encuentro de Darthmouth sirvió como punto de partida para agrupar a muchos investigadores en torno a un campo de investigación

prometedor en ese momento. Tanto es así que la Fundación Rockefeller fue la que se encargó de financiar el encuentro impulsado por McCarthy, con el objetivo de estudiar el desarrollo de un nuevo lenguaje de programación que permitiera dotar de inteligencia a las máquinas. El lenguaje que germinó en ese encuentro, y que aparecería más tarde, sería el LISP (*List Processing Language*).

A partir de ese verano de 1956 en el *Darthmouth College* se impulsó una comunidad científica con prometedoras metas, e incluso con un importante sentido de identidad que pasaría a la historia. El evento también contribuyó al establecimiento de laboratorios de IA en varias universidades, concretamente en Stanford, bajo la supervisión de McCarthy; en el MIT, con Marvin Minsky; en Carnegie Mellon, con Neweel y Simon; y de Donald Michie, en Edimburgo. Tanta importancia tuvo ese acontecimiento que esos laboratorios de ideas se mantienen hasta hoy día como piezas clave para la investigación en el campo de la IA.

McCarthy deja bien claro que su intención de elegir la expresión «inteligencia artificial» para denominar a esas jornadas fue evitar que la temática de estudio que él proponía se asociara con la cibernética de ese momento, que se encargaba del estudio sobre el control y la comunicación entre la máquina y el animal. Uno de los objetivos de las jornadas fue el desarrollo de un lenguaje que sirviera para dotar a las máquinas de inteligencia. Este objetivo se plantea en un momento en el que la historia había demostrado, con la Segunda Guerra Mundial, que las computadoras tenían un gran poder y utilidad, por ejemplo para calcular tablas de balística, con la necesidad de realizar miles de cálculos exactos en el menor tiempo posible, para organizar la estrategia militar, etc. En el año 1944 se construyó el ENIAC, el primer ordenador con fines prácticos, en este caso bélicos. Fue el húngaro Von Neumann quien contactó con varios investigadores del campo de la computación de instituciones como Harvard, Laboratorios Bell, la Universidad de Columbia y la Universidad de Pensilvania, para lograr acelerar los complejos cálculos que requerían en el campo militar, pues el equipo IBM que había utilizado en el Proyecto Manhattan no satisfizo sus necesidades (Coello Coello, 2003: 87-105). Después de la ronda de contactos con varios especialistas, se decantó por el ENIAC, un proyecto impulsado desde la Universidad de Pensilvania por seis mujeres, aunque todos los méritos se los

llevaran dos hombres, Presper Eckert y John W. Mauchly (Coello Coello, 2003: 310-322). Las seis mujeres que desarrollaron este proyecto fueron Betty Snyder Holberton, Jean Jennings Bartik, Kathleen McNulty Mauchly Antonelli, Marlyn Wescoff Meltzer, Ruth Lichterman Teitelbaum y Frances Bilas Spence. El ejemplo del ENIAC es una clara muestra del nivel en el que se encontraba la computación en ese momento para ser pensada y generar nuevas ideas y proyectos.

En este contexto en el que las computadoras habían demostrado un gran poder durante la Segunda Guerra Mundial, el *Dartmouth Summer Research Project on Artificial Intelligence* pretendía buscar nuevos usos que caminaran hacia la manipulación de símbolos, como fue el lenguaje de programación LISP. Aunque el espíritu inicial de la conferencia fuera apasionante, en lo que respecta a los temas que se plantearon y a la posibilidad de encontrar numerosos avances, todo parece indicar que finalmente no se lograron muchas cosas, pues ni siquiera se publicó el informe final que se había prometido desde un principio. Ese momento sirvió para acuñar una expresión que pasaría a la historia de la computación y que despertaría un gran interés: la inteligencia artificial.

3.4. Estado actual

En la actualidad la IA supera a la inteligencia humana en muchos ámbitos, por ejemplo en el de los videojuegos, ya que existen ordenadores dedicados a los juegos y presentan una clara muestra de victorias contra verdaderos expertos humanos en la materia. Pueden encontrarse infinidad de casos en las últimas décadas, como *Backgammon*, de Hans Berliner, *Scrabble*, *Jeopardy!,* etc. Estos ejemplos representan importantes logros en el campo de la IA. Sin embargo, existen personas que no se sienten tan impresionadas ante tales casos, pues la capacidad de impresión va modificándose con el paso del tiempo a través de la influencia del progreso tecnológico.

La IA está presente en diversas esferas en la actualidad, como el reconocimiento óptico utilizado para la clasificación de correos o la digitalización de antiguos documentos; o también la traducción automática que, por ejemplo en el caso de *Google*, aunque es imperfecta, ha supuesto un claro desarrollo. El reconocimiento facial se ha introducido en numerosos pasos fronterizos de Europa, además de EE. UU. o Australia. En el campo militar la IA también ha supuesto numerosos avances, como el despliegue a gran escala de robots que trabajan desactivando bombas y drones letales semiautónomos. La industria militar es en la actualidad una de las grandes beneficiadas del avance de la IA, pudiéndose incluso hablar de una «carrera militar» en materia de IA.

Internet es también otro campo que se ha visto claramente beneficiado del desarrollo de la IA, como es el caso del *software* dedicado al rastreo y vigilancia de correos electrónicos, las cuestiones de preferencia de compra, por ejemplo en *Amazon*. Las transacciones económicas realizadas con tarjeta de crédito cuando se ejecuta una compra por Internet también están redirigidas por IA. No obstante, si de lo que se está hablando es de Internet, el motor de búsqueda de *Google* es el mejor ejemplo para mostrar cómo la IA ha impregnado las vidas virtuales.

Otro campo en el que existe un alto riesgo y una gran competitividad es el sector financiero, donde la IA también se encuentra operando en este momento. El pionero a la hora de introducir IA en el campo financiero fue el *Citibank* a comienzos de los años 80 del siglo XX, y posteriormente el *Security Pacific National Bank*, en el año 1987. Pero es importante mencionar que hoy en día las principales compañías inversoras hacen un uso generalizado de la IA. Aunque ese uso es de diversos tipos: va desde simples intercambios financieros a complejas operaciones que se adaptan a las condiciones cambiantes del mercado. Además, existen numerosas aplicaciones que se basan en IA y que poseen una utilidad personal en el sector financiero, como *Kasisto, Moneystream, Wallet.AI*, etc., así como otras utilizadas en el sector en general para conceder préstamos y detener el fraude, como *Lending Club, Affirm, Prosper Daily, ZestFinance*, etc. En consecuencia, la automatización está muy presente y un claro ejemplo es el *Flash Crash* de 2010, que aunque no es caso concreto de IA muy desarrollada, sí que puede brindar cierto

conocimiento para entender cómo la tecnología está manejando grandes cantidades de dinero y haciendo transacciones económicas que comprometen la vida de la ciudadanía.

El estadounidense Jerry Kaplan muestra numerosos casos en su obra *Abstenerse humanos* (2016), sobre todo con la automatización de varios puestos de trabajo. Un ejemplo muy ilustrativo es el de *Amazon*. Los almacenes de *Amazon* se están organizando por medio de «intelectos sintéticos», término que Kaplan utiliza para referirse a la IA, lo que supone que los trabajadores sean cada vez más vulnerables para sustituirse por «falsos trabajadores». También tenemos el caso de *Agrobot*, una empresa agrícola de Huelva (España) que se dedica a la recolección de fresa y que está sustituyendo la mano de obra humana por la maquinaria basada en la IA que desarrolla en su oficina de Oxnard, California.

Otro espacio en el que la IA tendrá importante presencia es en el sector del automóvil. *Google* impulsó en el año 2008 el diseño de un vehículo autónomo, como señala Martin Ford (2016). Este gigante tecnológico se empeñó en demostrarle al mundo que podía ser capaz de desarrollar un vehículo autónomo que condujera en mejores condiciones que los propios humanos y por eso contrató a los mejores ingenieros de las carreras de la Agencia de Investigación de Proyectos Avanzados de la Defensa (DARPA). Es cierto que los resultados de las carreras de DARPA fueron mejorando considerablemente, y que ese fue el principal motivo por el que *Google* se interesó en sus ingenieros. Similar a la dificultad que encontró DARPA, para *Google* no fue una tarea fácil mejorar los automóviles autónomos, motivo por el cual el 7 de agosto de 2012 la empresa presumió en su blog oficial de claros avances en comparación con la conducción humana, pues no había tenido ningún percance en 300.000 millas recorridas. La irrupción del automóvil sin motor puede propiciar la aparición de un nuevo paradigma de movilidad y de interacción entre humanos y automóviles, un hecho que requiere una profunda reflexión ética. Si incluso *Uber* ha provocado una profunda controversia y conflicto en numerosas capitales del mundo al plantear un cambio en la cultura de la movilidad, parece pertinente reflexionar sobre el importante espacio que está ocupando la IA en la vida por medio de fenómenos como el de los vehículos autónomos.

Existen varios motivos para pensar por qué la IA se convierte en algo tan importante en este momento, ya que este tipo de tecnología será el centro sobre el que girarán la gran mayoría de las actividades en las próximas décadas. En el presente se ha comenzado a vivir la era de la IA, aunque queda un largo camino por recorrer. Rouhiainen (2018: 33-36) identifica una serie de motivos en la actualidad por los que se precisa priorizar el aprendizaje sobre la IA:

- Velocidad de implementación de la IA.
- Impacto potencial en la sociedad.
- Priorización de la IA por parte de todas las grandes empresas tecnológicas.
- Escasez de profesionales expertos.
- Ventajas competitivas para las empresas que usen primero la IA correctamente.
- Implicaciones legales en todo el mundo.
- Desarrollo ético.
- Comunicación de ventajas y oportunidades.
- Colaboración entre los sectores privado y público.

En los últimos años se ha hablado de la Cuarta Revolución Industrial. Klaus Schwab (2016) dedica una de sus obras a analizar cómo la IA está jugando un importante papel en la nueva revolución tecnológica que transforma a la humanidad por medio de la convergencia entre sistemas digitales, físicos y biológicos. Las nuevas tecnologías más avanzadas que se están promoviendo desde la IA provocan cambios disruptivos en la forma de establecer una relación con el mundo. Rouhiainen considera que en la cuarta revolución la IA es el elemento más importante porque representa un eje articulador y cohesionador de las demás partes (2018: 38). El desafío de la Cuarta Revolución Industrial se encuentra en saber afrontar una serie de cambios que son fruto del crecimiento exponencial. Existe por lo

tanto un doble desafío en el presente de la IA, entender bien estas tecnologías y también saber cómo usarlas bien.

3.5. Horizonte futuro

En los últimos años ha surgido un fuerte interés en el desarrollo de la IA que va en dos direcciones: una que tiene que ver con una teoría de la información que sea más sólida para el aprendizaje artificial, mientras que otra está relacionada con el desarrollo del aspecto práctico y comercial de varios sistemas de resolución de problemas concretos y de ámbitos específicos. No existe una postura única en torno al futuro de la IA. Incluso Nick Bostrom señala esta cuestión: «Las opiniones de los expertos sobre el futuro de la IA varían enormemente. No hay acuerdo sobre la sucesión temporal de los acontecimientos ni sobre qué formas podría llegar a adoptar la IA. Las predicciones sobre el futuro desarrollo de la inteligencia artificial, señaló un estudio reciente, "son tan firmes como diversas"» (2016: 19).

El filósofo e ingeniero sueco aportó los resultados de una encuesta realizada por el *Future of Humanity Institute* de la Universidad de Oxford, del que es director fundador. Esos resultados evidencian que no existe un consenso claro en torno al futuro de la IA. Las encuestas giran en torno a la pregunta de cuándo esperan los expertos que va a alcanzarse la inteligencia artificial de nivel humano. Los resultados se reflejan en la siguiente tabla:

Tabla 2. ¿Cuándo conseguiremos una inteligencia artificial de nivel humano?

	10%	**50%**	**90%**
PT-AI	2023	2048	2080
AGI	2022	2040	2065
EETN	2020	2050	2093
TOP100	2024	2050	2070
Combinados	2022	2040	2075

Fuente: Bostrom, 2016: 20.

En esta tabla se muestran los resultados de cuatro encuestas diferentes, así como la combinación de los resultados. Los participantes de las encuestas se ven reflejados en el documento de Müller y Bostrom (2014). Aunque las encuestas son predictivas, Bostrom defiende la postura sobre la probabilidad de que la superinteligencia surja poco después del momento en el que la IA alcance un nivel humano.

Los avances en los diferentes campos que conforman el grueso de la IA no se dan de la misma forma, pues hay ámbitos que dependen a su vez de otros. Por ejemplo, el campo de la robótica no avanza al mismo ritmo que el del *machine learning*. *Boston Dynamics* es una empresa de ingeniería y robótica fundada en 1992 por Marc Raibert, exprofesor del MIT, que ha experimentado numerosos avances en los últimos 25 años. Este periodo de tiempo invita a pensar que el espacio temporal comprendido para desarrollar ciertos proyectos podría considerarse mayor que el requerido en el ámbito del *machine learning* para el desarrollo de sus proyectos.

Los investigadores de la IA reconocen el papel relevante del aprendizaje para la inteligencia humana y se preguntan si es posible emular esa forma de aprendizaje en las computadoras. El aprendizaje maquinal tiene por objetivo la creación de programas que permitan la generalización de comportamientos a partir de ejemplos que le son suministrados y que por lo tanto generen un patrón o como señala Kaplan: «Como descripción general, los programas informáticos que aprenden, extraen patrones de los datos» (Kaplan, 2017: 32).

Intelectuales en el campo de la IA como Bostrom o Kurzweil se refieren al aspecto futuro de la IA en lo que respecta a un gran avance, el primero habla de «superinteligencia« y el segundo de «singularidad», dedicando cada uno de ellos una obra completa para abordar este aspecto futurible. Se trata de una marcha imparable, no de los robots exclusivamente, como señala Andrés Ortega (2016), sino de la IA; una marcha imparable que necesita ser pensada desde la filosofía, y concretamente desde la ética, pues va a plantear en un futuro, no tan lejano, numerosos e importantes desafíos. Los productos de IA son cada vez más autónomos y debería promoverse la reflexión sobre qué papel presentan

los humanos en el mundo y repensar su relación con la máquina. Además, es necesario plantear un futuro de la IA que sea optimista y que no carezca de crítica, como señala Garry Kasparov:

> Necesitaremos toda nuestra ambición para estar a la vanguardia de nuestra tecnología. Somos fantásticos en cada una de nuestras tareas como en la realización de máquinas y solo mejoraremos en ello. La única solución es seguir creando nuevas tareas, nuevas misiones, nuevas industrias que aún no sabemos cómo hacer nosotros mismos. Necesitamos nuevas fronteras y la voluntad de explorarlas. Nuestra tecnología es excelente para eliminar la dificultad y la incertidumbre de nuestras vidas, por lo que debemos buscar desafíos cada vez más difíciles e inciertos (2017: 258-259).

En el año 2014 varios expertos en IA como Stephen Hawking, Stuart Russell, Frank Wilczek y Max Tegmark, escribieron un artículo en el diario británico *The Independent* para advertir sobre las implicaciones éticas que está planteando el desarrollo de la IA, no solo en el presente, sino también en el futuro. Consideran que en el futuro puede aparecer una máquina extremadamente inteligente, incluso más que el propio ser humano. En ese sentido, no estaría de más que se deje a un lado esa especie de arrogancia humana que puede conducir a pensar que ningún sistema artificial puede alcanzar una inteligencia superior a la humana, generando de ese modo una conciencia sobre lo que implica la superinteligencia.

Además, Toby Walsh expone en su obra *Android Dreams: The Past, Present and Future of Artificial Intelligence*, diez predicciones para el futuro de la IA que pasan a detallarse a continuación:

- Tendrás prohibido conducir: los vehículos autónomos se expandirán por las carreteras del mundo, en países como Estados Unidos o Suecia.

- Verás al doctor diariamente: el desarrollo de la inteligencia artificial permitirá que podamos monitorear a diario nuestro organismo.

- Marilyn Monroe estará de vuelta en el cine: las tecnologías de la información permitirán una recreación avanzada de escenarios ficticios.

- Una computadora te contratará y despedirá: los intelectos sintéticos sustituirán muchas actividades de gestión que actualmente desempeñan los humanos.

- Hablarás con las habitaciones: el desarrollo del Internet de las Cosas ampliará las funciones de muchos objetos que forman parte de nuestra cotidianidad.

- Un robot roba un banco: la robótica ampliará cada vez más su poder y autonomía en los campos vinculados a la seguridad y el ejército.

- Alemania pierde frente a un equipo de robots: el asombroso desarrollo del campo de la robótica ampliará su espectro de actuación cada vez más llegando hasta el terreno deportivo.

- Barcos fantasmas, aviones y trenes cruzan el mundo: Se dará lugar un aumento exponencial de los vehículos autónomos y por lo tanto una revolución en el ámbito de la movilidad.

- Las noticias de televisión son producidas sin humanos: el desarrollo de los sistemas artificiales de gestión de la información será de tal magnitud que serán intelectos autónomos los que se encarguen del tratamiento informativo.

- Viviremos después de la muerte: el espacio virtual permitirá que las personas sigan teniendo una presencia pública en las redes después de la muerte física (Walsh, 2017: 205-220).

Una vez expuesto a grandes rasgos el desarrollo que ha experimentado el campo de la IA desde sus orígenes hasta la actualidad, incluso mencionando algunas de las predicciones más esperadas, es importante pasar a esbozar algunas de las propuestas más destacadas de los últimos años en torno al futuro y desarrollo de los intelectos sintéticos: superinteligencia y singularidad.

3.5.1. La explosión de superinteligencia

Bostrom define la superinteligencia como «cualquier intelecto que exceda en gran medida el desempeño cognitivo de los humanos en prácticamente todas las áreas de interés» (Bostrom, 2016: 22). El sueco destaca por ser un férreo defensor del movimiento transhumanista y por lo tanto defiende la idea sobre la existencia de una mayor facilidad a la hora de desarrollar la inteligencia en una base artificial antes que en una base biológica, ya que las máquinas poseen una serie de ventajas que las entidades biológicas no presentan.

No es nueva la idea de la superación de niveles humanos. Por ejemplo, los gatos tienen un olfato más fino que los humanos y una calculadora realiza ejercicios matemáticos mucho más rápido que un profesor de matemáticas de carne y hueso. No obstante, cuando se trata de intelectos artificiales, subyacen una serie de entidades añadidas que tienen una inteligencia de tal magnitud que pueden ser capaces de sustituir a los seres humanos en cualquiera de sus ámbitos de desempeño. Como pueden surgir numerosas dudas en torno al concepto de superinteligencia, pues la definición propuesta en el párrafo anterior podría dar a entender que cualquier máquina que supere las capacidades humanas puede considerarse como superinteligencia, se expondrán a continuación las tres formas de este concepto que Bostrom propone para diferenciar las superinteligencias: superinteligencia de velocidad, superinteligencia colectiva y superinteligencia de calidad (2016: 52).

3.5.1.1. Superinteligencia de velocidad

Es un sistema que puede llegar a hacer lo mismo que el intelecto humano pero de forma más rápida, es decir, que un intelecto artificial podría funcionar mucho más rápido que un cerebro biológico, por ejemplo memorizando de forma más rápida un libro. La relación con la realidad de un intelecto artificial con una superinteligencia de velocidad sería diferente que la que presentaría un cerebro humano, incluso podría preferir comunicarse con otras superinteligencias antes que con un ser humano al que consideraría extremadamente lento.

Existe un claro ejemplo que puede servir para escenificar una comunicación entre sistemas de IA establecida al margen de los seres humanos. Se trata de lo ocurrido entre unos sistemas de IA que los ingenieros de *Facebook* tuvieron que desactivar porque estaban comunicándose en un lenguaje que los expertos eran incapaces de descifrar (Griffin, 2017). No es un ejemplo de superinteligencia de velocidad, pero sí que puede servir para invitar a la reflexión sobre lo que puede llegar a ocurrir cuando dos máquinas se comunican al margen de los seres humanos.

3.5.1.2. Superinteligencia colectiva

Es un sistema de IA que se fundamenta en la agregación de varios sistemas menores, formando de ese modo un conjunto que responde a diversos grados de eficiencia. Este tipo de IA favorece la resolución de problemas que pueden dividirse en partes y se aplica perfectamente a aquellos ámbitos donde opera la división del trabajo.

La mejora de la superinteligencia colectiva puede llevarse a cabo mediante una mejor organización de todas las partes y mediante el aumento de la calidad de los intelectos artificiales que conforman el conjunto. Si este tipo de IA aumenta gradualmente, el conjunto de intelectos artificiales llegaría a formar un intelecto unificado.

3.5.1.3. Superinteligencia de calidad

Es un sistema de IA que es al menos tan rápido como la mente de un ser humano y además cualitativamente mucho más inteligente. Los aspectos intelectuales, que son los más destacables en este tipo de IA, juegan un papel diferenciador en el ser humano respecto de otros seres vivos. Ese aspecto diferenciador en lo que concierne a la inteligencia se debe a la arquitectura cerebral. Entonces existe una superinteligencia de calidad cuando hay un intelecto sintético que es, al menos, tan superior a la inteligencia humana como ésta lo es a la de los demás seres vivos.

Bostrom considera que cualquiera de los tres tipos de superinteligencia podría llegar a ser capaz de desarrollar alguna de las demás. En este sentido, es muy probable que una vez que la IA alcance los niveles humanos de inteligencia se produzca una explosión de superinteligencia, ocasionando que los intelectos sintéticos fueran autónomos al margen de los programadores y pudieran constituir y dar forma a su vez a otros intelectos. Por lo tanto, el fenómeno de la superinteligencia invita a una profunda reflexión ética, pues no se está planteando un tema baladí, sino un tema que compromete de manera importante a la ciudadanía.

La denominada «cinética» de una explosión de inteligencia (Bostrom, 2016), muestra cómo se daría lugar la sincronización y velocidad de despegue en el momento en el que la IA alcanzara niveles cognitivos humanos. Llegado el momento en el que la IA alcanzara los mismos niveles cognitivos que lo humanos, existen diferentes caminos que puede ser avistados en el horizonte. Sin embargo, y atendiendo a la reflexión que suscita este texto de Bostrom, es importante partir de la consideración del fenómeno de la transición debido a la existencia de una supuesta gran probabilidad de explosión de superinteligencia. Aunque esta consideración parte de un estudio predictivo, no debería menospreciarse la importancia del trabajo realizado por Müller y Bostrom (2014).

Una vez considerada la probabilidad de la explosión de superinteligencia, es necesario comentar qué objetivos podría llegar a tener la IA a partir de tal explosión. Ante la pregunta de por qué una IA de nivel sobrehumano se plantearía como objetivo su mejora y no otra cosa, podrían considerarse varias respuestas que Olle Häggström (2016: 119-120) propone. Entre las respuestas del sueco se encuentra primeramente que el programador de la IA original puede haber programado la IA para que intente automejorarse, como la mejor manera de conseguir IA de un nivel superior. En segundo lugar la respuesta se basa en los algoritmos genéticos que se construyen sobre la máxima darwiniana, interpretada para tales efectos, de mutación, selección y reproducción. Y la tercera respuesta se erige sobre la máxima que sostiene David Hume acerca de una razón que es esclava de las pasiones, por lo que la IA tendría como objetivo satisfacer sus deseos, objetivos, motivaciones, impulsos o valores.

3.5.2. El advenimiento de la singularidad

«Singularidad» es otro término que se utiliza en el campo de la IA para referirse al sistema supertinteligente que es capaz de perfeccionarse a sí mismo y crear otros sistemas, incluso más inteligentes que él, siguiendo un crecimiento exponencial. Aquí podría mencionarse brevemente la cuestión de la cinética de la explosión de superinteligencia, o de la singularidad, pues se encuentra muy vinculada el ritmo de crecimiento. Los defensores de la singularidad afirman que se producirá cuando los mejores desarrolladores no sean personas de carne y hueso, sino propiamente IA. El rendimiento de los sistemas, que en un primer momento se atribuirá al *hardware* y posteriormente al *software*, se duplicará considerablemente y la velocidad infinita será la norma habitual. Por lo tanto, si la velocidad con que la IA se diseñará en el futuro por sí misma va a ser infinita, entonces el alto nivel de inteligencia es muy probable.

El máximo exponente de la singularidad es el estadounidense Raymond Kurzweil, que considera que la singularidad puede perfeccionarse a sí misma teniendo como horizonte la constitución de todo un universo basado en una entidad global inteligente. Este autor afirma que cuando un intelecto sintético supere a la inteligencia humana el progreso será mucho más rápido. Kurzweil, al igual que Hans Moravec (1988), está convencido de que durante la primera mitad del siglo XXI las máquinas superarán la inteligencia humana. Según este autor, el crecimiento de la IA será exponencial:

> Representa la fase casi vertical del crecimiento exponencial que tiene lugar cuando el ritmo es tan extremadamente alto que la tecnología parece expandirse a una velocidad infinita, pese a que, desde la perspectiva matemática, no hay discontinuidad ni ruptura y los ritmos de crecimiento siguen siendo finitos, aunque extraordinariamente grandes. Pero desde nuestro limitado marco actual este evento inminente parece una ruptura aguda y brusca en la continuidad del progreso (Kurzweil, 2016b: 26).

En este sentido, la singularidad está cerca y va a suponer un cambio de paradigma en varios campos que Kurzweil menciona en su obra (2016b: 27-33). A finales de este siglo se espera que la mayor parte de la inteligencia sea no biológica, aunque eso no supondría el fin de la inteligencia biológica. Kurzweil es un férreo defensor de la singularidad, pues considera que las inteligencias sobrehumanas servirán para satisfacer las necesidades y deseos del ser humano. Su propuesta para evitar la anulación humana consiste en la fusión con la máquina, en lo que él denomina «enlace íntimo« (2016b: 33).

Las propuestas de Bostrom y de Kurzweil tienen mucho en común, aunque el primero habla de superinteligencia y el segundo de singularidad. Ambos pensadores están convencidos de que la superinteligencia artificial llegará durante este siglo influyendo en numerosos ámbitos humanos hasta su dominio, a no ser que sean tomadas medidas como las que proponen. Kurzweil aventura una fusión íntima con la máquina, mientras que Bostrom propone la incorporación de comportamientos éticos con contenido axiológico.

3.6. La necesidad de una mirada ética de responsabilidad

El acelerado desarrollo de la IA ha sido el detonante de profundas transformaciones en numerosas esferas humanas. El poder tecnológico interviene en la vida y la condiciona debido a su gran fuerza de expansión. La competición entre la inteligencia humana y la IA está servida para las próximas décadas. La humanidad aún no ha logrado comprender todo lo que está en juego con las importantes implicaciones que tiene el desarrollo de la IA. No se trata de hacer un canto a la tecnofobia, sino más bien de reflexionar sobre sus efectos en la vida. Tampoco se trata del convencimiento absoluto en ciertas afirmaciones sobre la lógica especial de la tecnología, como si estuviera por encima del bien y del mal, exenta de toda reflexión moral. Lo importante es adquirir un conocimiento de los avances tecnológicos más significativos en el campo de la IA y observar qué implicaciones éticas tienen estos avances en algunos ámbitos de la vida, como el profesional, el médico o el militar. Solo si la ciudadanía logra identificar la magnitud del fenómeno de la IA, podrá entonces generar un conocimiento innovador a la altura de los desafíos.

Es importante asumir responsabilidad desde una heurística del temor ante el poder tecnológico, sin presentar excesiva confianza, sin fe ciega en un tecnocentrismo que puede llegar a esconder un dogmatismo cientificista (Jonas, 1995). El sonambulismo tecnológico (Winner, 2008) puede conducir a terrenos pantanosos en los que difícilmente habría escapatoria si no se impulsa un ejercicio de toma conciencia con antelación. Así pues, para una primera toma de conciencia es fundamental asumir una responsabilidad cívica y democrática, tarea que no es nada fácil en un momento histórico en el que únicamente se muestran los beneficios de la IA y donde apenas existe una crítica fundamentada a la actividad de las tecnologías más avanzadas. En el capítulo 5 se presentará el concepto de IAR como el pilar fundamental sobre el que se sostiene el hilo argumentativo de este trabajo.

La sugerencia de someter a debate público el impacto de la IA en la vida humana nace de la necesidad de generar conocimientos innovadores desde una asunción de responsabilidad en el contexto de espacios colaborativos de experimentación como los laboratorios abierto o laboratorios ciudadanos. La humanidad se enfrenta a importantes desafíos, lo cual exige deliberar para formular alternativas que garanticen un futuro de compromiso. La falta de reflexividad puede conducir a consecuencias inesperadas. En ese sentido, un ejercicio de cautela en el que el principio de responsabilidad sirva como principio rector permitiría iniciar un camino caracterizado por la toma de conciencia frente al poder tecnológico y su consecuencias, en ocasiones, inesperadas. En palabras de Kasparov:

> Las conclusiones son generalmente para relajarse, pero preferiría usar este para agitar las cosas. Espero que tomen esta sección como una lista de lectura y como una invitación a participar activamente en la creación del futuro que desea ver. Este debate es único porque no es académico. No es un *postmortem*. Cuanto más cree la gente en un futuro positivo para la tecnología, mayor será la posibilidad de tener uno. Todos elegiremos cómo será el futuro según nuestras creencias y nuestras acciones. No creo en destinos más allá de nuestro control. No se decide nada. Ninguno de nosotros somos espectadores. El juego está en marcha y todos estamos en el tablero. La única manera de ganar es pensar más y más (2017: 258).

CAPÍTULO 4

ÉTICAS DE LA INTELIGENCIA ARTIFICIAL: ESTADO DE LA CUESTIÓN

> La mirada tiene algo de extraño, de paradójico: la total facilidad de mirar contrasta con la dificultad de mirar bien. Si hay luz, con solo abrir los ojos se nos aparecen las cosas que nos rodean, pero en cambio hay que prestar atención, fijarse bien, para darse cuenta de según qué aspectos de la realidad y, sobre todo, para percibir las cosas de otra manera.
>
> (Esquirol, 2006: 14)

¿Debe aprovecharse la IA al máximo para desafiar a la naturaleza y emprender así una aventura desconocida y arriesgada? ¿Es posible introducir criterios éticos en los intelectos sintéticos para fortalecer la normatividad? ¿Puede orientarse la actividad de los sistemas artificiales hacia el cultivo de las virtudes? ¿Puede la IA ejercitarse en el aprendizaje de unos valores supuestamente universales en un mundo caracterizado por lo relativo? ¿Es imposible incorporar aspectos morales en los sistemas artificiales por no tener conciencia?

Durante la última década la reflexión sobre la aplicación de la ética al campo de la IA se ha convertido en una de las principales preocupaciones de la ética aplicada al campo de la tecnología. Ese interés se ha incrementado debido a los importantes avances y a la creciente complejidad que está experimentado el área de los intelectos sintéticos, pues cada vez están más presentes en diversas esferas de la vida. En la medida que se crean nuevas oportunidades para aprovechar los avances que presenta la IA, las implicaciones y compromisos que ofrece son mucho mayores. Por ello se vislumbran diversas propuestas éticas que arrojan luz sobre la necesidad de someter a una profunda reflexión las posibilidades de la IA.

Esta dimensión ética de la que se deriva la reflexión en el campo de la IA se ha acentuado con las contribuciones de diversos pensadores y pensadoras que han adquirido un serio compromiso. Existen múltiples enfoques en lo relativo a la aplicación de la ética a la IA. Por motivo de extensión, y para acotar un tema sumamente complejo, aquí se abordarán únicamente algunas.

Raymond Kurzweil (2016a; 2016b) se sitúa en la estela del utilitarismo al defender un aprovechamiento generalizado de la tecnología más avanzada para enriquecer al ser humano. Este enriquecimiento se entiende como lucha contra las enfermedades y el envejecimiento, y como fortalecimiento de los caracteres humanos, hasta el punto de traspasar la frontera de las limitaciones biológicas más básicas. El utilitarismo de Kurzweil, profundamente enraizado en el hedonismo filosófico, trata de promover el cultivo de la felicidad y un alejamiento del dolor y de todo aquello que perturba al ser humano, que para este autor podría identificarse con la muerte y el envejecimiento.

Nick Bostrom (2011; 2016; 2017) ha contribuido también al debate sobre la aplicación ética en el campo de la IA. Su interés se centra en una profunda reflexión sobre la posibilidad de introducir valores éticos en los intelectos sintéticos, a pesar de las dificultades que esto supone frente a los retos del relativismo cultural. Además, formula una interesante propuesta para intentar normativizar el ámbito en el que estos intelectos sintéticos despliegan su actividad por medio de medidas de transparencia, los valores que deben estar presentes en el ejercicio tecnológico, el respeto, etc.

Shannon Vallor (2015; 2016) contextualiza el pensamiento aristotélico para defender la posibilidad de promover y fortalecer las actividades virtuosas dentro del campo de la IA. Concretamente esta autora centra su argumentación en el terreno militar, donde observa una ineludible oportunidad para cultivar las virtudes y reorientar el uso que pueden hacerse de los instrumentos militares del campo de batalla. En este cultivo de virtudes encuentra Vallor una ocasión para lograr un uso de la tecnología que se encamine al beneficio de la humanidad por medio de la virtuosidad.

Bill Hibbard (2002; 2012; 2015) es un científico estadounidense que reflexiona sobre la posibilidad para que los intelectos sintéticos aprendan valores por medio de la experiencia. Esboza serias críticas contra el utilitarismo, pues considera que conduce a resultados que son ambiguos en situaciones complejas. Dado el pluralismo moral que caracteriza a la diversidad de grupos y culturas que forman las sociedades actuales, Hibbard reconoce la complejidad en el proceso de aprendizaje de valores. En ese sentido, apunta que una solución pertinente consistiría en plantear una combinación de valores entre diversos grupos que respetara el equilibrio entre las cosmovisiones de dichos grupos. Para ello, Hibbard insiste en la riqueza que los postulados rawlsianos pueden proporcionar ante esta compleja diversidad de perspectivas. Así pues, la aportación ética de este autor consiste en un proceso de aprendizaje de valores humanos por parte de los sistemas artificiales que sigan un proceso de respeto a los principios rawlsianos de la justicia.

En último lugar, una postura muy interesante es la de los indios Rajakishore Nath y Vineet Sahu, que parten de la imposibilidad de introducir la ética en los intelectos sintéticos al carecer éstos de moralidad. Si la ética es la disciplina filosófica que reflexiona sobre el fenómeno humano de la moral, carece de sentido pensar que es posible aplicarla a los intelectos sintéticos, ya que los mismos no presentan evidencias de ser agentes morales.

Los problemas éticos que pone de relieve la IA demandan ser considerados a la luz de una nueva forma de ser de la ética, una ética aplicada a la IA que se nutra de diferentes tradiciones filosóficas y que sepa conjugarlas, asumiendo una responsabilidad compartida, en el contexto de espacios colaborativos de experimentación ciudadana, poniendo en valor el diálogo y las habilidades comunicativas para la comprensión. En ese sentido, el objetivo de este capítulo es presentar algunas de las propuestas éticas aplicadas en el ámbito de la IA e introducir brevemente el concepto de IAR, que en el capítulo 5 se desarrollará de manera detallada y se confrontará con las posturas expuestas en este capítulo.

4.1. Raymond Kurzweil: utilitarismo hedonista como impulso tecnológico

Raymond Kurzweil, nacido en New York en 1948, es un importante inventor, pensador, futurista de los últimos tiempos y director de ingeniería en *Google* desde el año 2012. Revistas como *Forbes* o *Inc* lo consideran «la máquina de pensar suprema» o «el legítimo heredero de Thomas Edison», respectivamente. A Kurzweil se le atribuyen algunos inventos como el primer sintetizador de voz para ciegos, el primer escáner CCD, el primer sintetizador de música capaz de recrear numerosos instrumentos, etc. Ha recibido numerosos premios y entre sus obras más importantes destacan *La singularidad está cerca* (2016) y *Cómo crear una mente* (2016).

Kurzweil ha realizado durante las últimas décadas importantes predicciones en el campo de la IA, que finalmente se han hecho realidad. A partir del análisis de la propuesta de Kurzweil podría considerarse que sus planteamientos se sitúan en la tradición del utilitarismo, doctrina ética enraizada en tradiciones como el hedonismo.

La búsqueda de la felicidad ha sido objeto de estudio por parte de la filosofía desde la Antigüedad. Con Aristóteles la ética se presenta como un estudio sobre el bien, que para los seres humanos es la felicidad (*eudaimonía*). Se trata de una ética de carácter teleológico, pues se encuentra orientada hacia un fin (*télos*) que consiste en alcanzar la felicidad. Para el estagirita la vida feliz podía expresarse de diferentes formas: la que se basa en la riqueza, el placer, la virtud y la que pone el énfasis en la vida contemplativa. La «vida buena« consistía en actuar conforme a la función que es propia del ser humano, idea que asume que el placer y la riqueza no sean suministradores de felicidad, pues se caracterizaban por ser efímeros. En cambio, la vida contemplativa era el medio para alcanzar la verdadera felicidad, ya que esa era la actividad propia del sabio. Pero este ideal de felicidad cambia con la aparición del hedonismo y el desarrollo del epicureísmo, que centra su propuesta en el placer y el dolor como ejes de su pensamiento.

En el centro de la propuesta hedonista se sitúan el placer y el dolor como criterios fundamentales para el discernimiento. La ausencia de placer implica la reivindicación de satisfacción de esa ausencia por medio de la apetencia del deseo, aunque no todos los deseos son igualmente importantes. En la *Carta a Meneceo*, Epicuro exhibe su propuesta filosófica sobre el placer, donde afirma que el deseo de todo humano consiste en el disfrute del placer y sentir felicidad, viviendo sano del cuerpo y sereno de ánimo, entendiendo la *ataraxia* como la ausencia de toda perturbación, de ahí que pueda considerarse que su filosofía tenga un carácter terapéutico. El temor es lo que mueve en la búsqueda de ese ideal terapéutico que es el placer, satisfecho mediante el conocimiento verdadero, alcanzado por medio de la filosofía y la ciencia. La muerte, el mayor de los temores, no debe perturbar.

La búsqueda del placer es totalmente legítima, pues es el medio para alcanzar la felicidad; ningún placer está limitado siempre y cuando su alcance no implique sufrimiento ni la desaparición de la *ataraxia*. Se trata de guiar las acciones para alcanzar salud en el cuerpo y tranquilidad en el alma, satisfaciendo así los deseos para no sentir dolor. La naturaleza humana demanda la satisfacción de los deseos que son fruto del dolor que causan los temores y las incertidumbres.

Pero el hedonismo se enmarca dentro de una tradición ética todavía más importante, el utilitarismo, considerada como la mayor contribución de los ingleses en el ámbito de la teoría moral y política. El utilitarismo de Jeremy Bentham y John Stuart Mill puede considerarse como una de las doctrinas con mayor representación e impacto en el ámbito de la filosofía moral. Tanto el utilitarismo como el antiutilitarismo han basado sus argumentaciones en dos grandes principios utilitaristas, como señala Esperanza Guisán (2013: 274):

> a) el valor más importante es la felicidad a nivel individual;
>
> b) el bienestar colectivo, vinculado con una utilidad generalizada, ha supuesto un importante objeto de reflexión en la historia de la filosofía y también en la organización de gobiernos y políticas.

Los hedonismos –de Epicuro a Bentham y Mill– se centran en el placer humano o la felicidad humana, involucrando todas las capacidades humanas. El interés del utilitarismo reside en las máximas del aumento del placer y la disminución de dolor, cuya fundamentación se encuentra en la teoría clásica del hedonismo. La asunción de estas máximas supone la línea de pensamiento que atraviesa toda la propuesta de Kurzweil.

Expuestos en términos generales los aspectos principales del utilitarismo hedonista, es importante detallar los argumentos que sustentan la tesis acerca de la identificación de Kurzweil con dicha doctrina ética. Son varios los postulados del estadounidense que llevan a incluirlo dentro de las filas del utilitarismo. Su propuesta se construye sobre la idea principal de que la tecnología, y concretamente la IA y sus avances, deben servir para desafiar la muerte y el envejecimiento al que están destinados los seres humanos. La muerte y el envejecimiento representan el dolor del que habla Epicuro, pues son factores que implican la ausencia de *ataraxia*, al producir preocupación y temor fruto de la incertidumbre. La biotecnología avanzada puede contribuir a la satisfacción del deseo que se concentra en el desafío a la muerte y el envejecimiento, alejando así del dolor que causan esos fenómenos biológicos de la vida. Los avances biotecnológicos servirán de este modo para proporcionar placer y distanciamiento respecto al dolor, en definitiva, para la consecución de la felicidad, eje principal de la doctrina teleológica del utilitarismo.

Para Kurzweil existen muchas preocupaciones que causan dolor y que tienen que ver con la alimentación, como son la obesidad, la diabetes, etc. La digestión es un proceso biológico que permite adquirir nutrientes para mantener la vida y al mismo tiempo ayuda a eliminar determinadas toxinas perjudiciales. La evolución de la especie ha permitido que el sistema digestivo de los humanos se fortalezca para tener una menor probabilidad de sufrir enfermedades. Pero no solo la evolución ha sido la causa del fortalecimiento del género humano, sino que al aumento de la esperanza de vida también ha contribuido la tecnología con el diseño de medicamentos, implantes, suplementos, etc. Kurzweil defiende la idea de

seguir potenciando el uso de tecnología avanzada para seguir obteniendo beneficio para la vida.

Para entender mejor por qué este autor podría enmarcarse en el utilitarismo, es necesario contextualizar sus propuestas dentro de esta tradición filosófica. El dolor puede identificarse como la sensación que provoca tener conocimiento sobre el envejecimiento, que los órganos pierden su funcionalidad y utilidad, y que la vida está limitada por una muerte que se va aproximando con el pasar de los años. Ante un panorama que provoca dolor por la ausencia de vida y juventud, Kurzweil propone hacer uso del progreso biotecnológico para tomar distancia de ese dolor y alcanzar el placer mediante la satisfacción del deseo de querer vivir más y mejor, esto es, la felicidad. Por ejemplo, la ingesta de alimentos en exceso y de alto valor calórico provoca un aumento de peso del organismo y consecuencias no tan beneficiosas para la vida. En ese sentido, Kurzweil menciona los nanorobots que desarrollarán su actividad en el tracto digestivo y que servirán para ayudar en el proceso de eliminación de grasas y absorción de nutrientes beneficiosos para el cuerpo (2016b: 348-349).

Este ejemplo puede servir para ilustrar mejor por qué Kurzweil se sitúa dentro de las filas del utilitarismo. Resulta que la ingesta de determinados alimentos ocasiona dolor, lo que provoca un alejamiento de la felicidad por el temor a perjudicar la salud engordando, aumentando los niveles de colesterol, etc. Y esto provoca el deseo de querer ingerir alimentos con libertad, sin que nada preocupe, con ausencia de perturbación (*ataraxia*). En esa línea, este pensador señala lo siguiente:

> En esa etapa de desarrollo tecnológico podremos comer lo que queramos, es decir, cualquier cosa que nos proporcione placer gastronómico. Exploraremos las artes culinarias para descubrir sabores, texturas y aromas, ya demás poseeremos un fluido óptimo de nutrientes en nuestro torrente sanguíneo […] En último término, no tendremos que preocuparnos de llevar prendas especiales o de contar con recursos nutricionales explícitos (2016b: 349).

Así pues, el progreso de la biotecnología proporcionará herramientas suficientes para desactivar las enfermedades y los procesos del envejecimiento, hechos que son motivo de felicidad (2016b: 368). Para Kurzweil no hay que preocuparse por no llevar una vida saludable y cometer actos que atenten contra la salud, pues la biotecnología solucionará cualquier problema. En este sentido, las soluciones que nos brinda el progreso tecnológico nos permiten olvidarnos de cualquier turbación. Tradicionalmente se definía el sentido de la vida a partir de aquellos conocimientos, valores y creencias que se adquirían a través de la cultura y la educación. En cambio, el progreso tecnológico ha producido un cambio de paradigma en lo relativo a la significación de la vida, ya que si antes era la educación la que proporcionaba las pautas para orientar la vida, ahora será la biotecnología la que proporcione las pautas para huir de las enfermedades y alcanzar un mejor bienestar. Los postulados de Kurzweil son verdaderamente utilitaristas, ya que contextualiza las ideas de Bentham y Mill en un escenario de avanzado progreso tecnológico, en el que el dolor que provoca el envejecimiento y la muerte, fruto del deseo de querer vivir más, es sanado gracias a la biotecnología, producto de un mayor y mejor conocimiento de la naturaleza. La biotecnología adquiere un carácter teleológico y su finalidad consiste alcanzar la felicidad que produce el alejamiento de la muerte y el envejecimiento. La felicidad, punta de lanza del utilitarismo, se convierte en el fin último de la propuesta de Kurzweil.

Lo mismo ocurre con el cerebro biológico, que, según Kurzweil, será superado o trascendido por una inteligencia inorgánica, permitiendo hablar de una inteligencia tecnológica y no tanto biológica. Este autor establece un paralelismo entre el binomio cuerpo-mente y *hardware-software*. Como es sabido, en la actualidad cuando el cuerpo, esto es, el *hardware*, experimenta algún fallo, la mente puede verse afectada, pues el cuerpo es la estructura física sobre la que se levanta la información mental, a saber, el *software*. Además, cuando el cuerpo muere, el cerebro también, porque la estructura física que lo sostiene deja de vivir. En el contexto de los postulados de Kurzweil esto puede ocasionar cierto dolor, pues toda la información mental que residía en el cerebro desaparece con la muerte del cuerpo. En cambio, para Kurzweil eso no ocurrirá cuando exista la disposición

de medios suficientes para recuperar las ideas que residen en el cerebro humano y de ese modo poder trascender la inteligencia biológica:

> En ese momento la longevidad de un fichero mental no dependerá de la continua viabilidad de ningún medio de *hardware* en particular (por ejemplo, la supervivencia de un cuerpo biológico y de un cerebro). En último término, los humanos basados en *software* se habrán expandido mucho más allá de las limitaciones humanas tal y como las conocemos hoy en día. Vivirán en la web proyectando sus cuerpos cuando quieran o lo necesiten, lo cual incluirá cuerpos virtuales en diferentes ámbitos de realidad virtual, cuerpos proyectados holográficamente, cuerpo proyectados mediante foglets y cuerpos físicos que contenga enjambres de nanorobots y de otras formas de nanotecnología (2016b: 372).

Esto supone un desafío a la muerte y por lo tanto un grito a favor de la inmortalidad, identificada como la felicidad a la que se aspira, según este autor, pues se está reivindicando un aumento de la esperanza de vida considerable, al menos del contenido cerebral. Para Kurzweil esto solucionará numerosos problemas de almacenamiento de información y promoverá un escenario de ausencia de turbación en materia de disponibilidad y acumulación de la información y la memoria (2016b: 373-373).

En definitiva, los postulados de Kurzweil se sitúan en la línea marcada por el utilitarismo, principalmente porque en el contexto de su obra sitúa el dolor del humano en el envejecimiento y la muerte ante el anhelo de querer vivir más. Este anhelo se convierte en un deseo que es posible satisfacer mediante el progreso tecnológico, y en particular de la IA, que se convierte en el medio útil para lograr la felicidad de la inmortalidad. El placer y la felicidad se asocian al desafío de la muerte y el envejecimiento por medio de la tecnología en la búsqueda de una vida prolongada en el tiempo y despreocupada por cuestiones de salud perturbadoras, evitando así el dolor y promoviendo la *ataraxia*.

4.2. Nick Bostrom: la normatividad ética como guía de la acción

Nick Bostrom es un filósofo sueco, director del *Future of Humanity Institute* y fundador de *Humanity Plus* (H+), la asociación mundial de transhumanismo. Es conocido por sus

estudios sobre superinteligencia y mejoramiento humano y entre sus obras más importantes destacan *Superinteligencia, caminos, peligros, estrategias* (2016) y *Mejoramiento humano* (2017), una obra coeditada con Julian Savulescu.

En su artículo *The Ethics of Artificial Intelligence*, Bostrom dedica una profunda reflexión a los problemas éticos que suscita la creación de máquinas inteligentes. Los algoritmos están presentes en muchos espacios de la vida, por ejemplo en la banca, con los intelectos sintéticos que forman los programas de aprobación de hipotecas. El funcionamiento de los algoritmos va a acompañado de importantes interrogantes éticos. Estos interrogantes surgen cuando los algoritmos desarrollan un trabajo cognitivo que tiene implicaciones éticas para la sociedad. Hay trabajos que tienen que ver con decisiones que normalmente han tomado los humanos y que ahora hace un algoritmo. Las discusiones surgen cuando el algoritmo hereda los prejuicios y requisitos sociales que el ser humano empleaba, por ejemplo, sesgos raciales o sexistas. En ese sentido, lo deseable sería que los algoritmos no solo presentaran un gran poder inteligente para desarrollar tareas concretas, sino también que fueran transparentes en la solución de los conflictos.

No obstante, no solo es importante que los algoritmos sean transparentes, sino que también posean fortaleza frente a los intentos de manipulación emprendidos por los humanos, por ejemplo en la detención de armas mediante reconocimiento visual en los escáneres de los aeropuertos. La cuestión sobre dónde recae la responsabilidad al tomar una decisión complicada también es un tema susceptible de discusión. Es muy probable que en determinados sistemas burocráticos, si se comete una acción que perjudica a alguien, se prefiera culpar a la IA antes de depurar responsabilidades entre los miembros del aparato burocrático.

Para Bostrom existen criterios a tener en cuenta para incorporar a la IA que deberían revisarse a la luz de los tiempos en una sociedad cada vez más informatizada. Entre esos criterios pueden encontrarse los siguientes: responsabilidad, transparencia, auditabilidad, incorruptibilidad, predictibilidad, etc. (2011: 3).

También existe otro aspecto importante para esclarecer los postulados kantianos implícitos en la propuesta de Bostrom, aunque complementados con sus propias aportaciones. Bostrom reflexiona sobre el estatus moral de las máquinas, es decir, si presentan o no agencia moral. El cuestionamiento sobre el estatus moral surge a raíz del planteamiento de la pregunta acerca de si la máquina es un fin en sí misma, al igual que los seres humanos, y por lo tanto debe ser sometida a la misma consideración. En la *Fundamentación de la metafísica de las costumbres*, Kant descubre aquella dimensión de la dignidad humana que se fundamenta en la libertad moral y que se manifiesta en la tercera formulación del imperativo categórico: «obra de tal modo que te relaciones con la humanidad tanto en tu persona como en la de cualquier otro, siempre como un fin y nunca solo como un medio» (1999: 104). De esta máxima se deriva que la persona tiene un estatus moral propio, pero lo que se pregunta Bostrom es si la IA tiene o no estatus moral en primera instancia.

En primera instancia la IA actual no tiene estatus moral, y eso es algo que comparten los expertos de este ámbito. El humano puede manipular, borrar información, apagar y destruir sistemas a su antojo, sin que eso implique un daño a los intelectos sintéticos. Sin embargo, la discusión comienza cuando se trata de determinar cuáles son los criterios que se tienen en cuenta para atribuir estatus moral. Bostrom señala que son dos los criterios morales (2011: 7) para configurar el estatus moral:

- Sensibilidad: tiene que ver con la capacidad de experimentar o sentir dolor y sufrir (Bostrom utiliza el término «*sentience»*, de donde se deriva el neologismo español '*sentiencia*' o '*sintiencia*').

- Sapiencia: hace referencia al conjunto de capacidades vinculadas con una inteligencia de un nivel más alto, asociadas al autocontrol, la conciencia y a la razonabilidad.

Estas dos características deben discutirse, pues existen casos en los que no se satisface una u otra condición y eso requiere de matización. Por ejemplo, los neonatos humanos no satisfacen los criterios de sapiencia, aunque no se deja de atribuirles dignidad por eso a los bebés. Lo mismo ocurre con los discapacitados mentales, que podrían no reunir alguna de esas características. No obstante, es importante seguir teniendo como referencia esos dos puntos para entender bien lo que propone Bostrom.

Si la idea de la que parte Bostrom para considerar que algo tenga estatus moral consiste en reunir esas dos características, podría deducirse que si un sistema de IA posee la capacidad de sentir dolor, presentaría cierto estatus moral. Además, los intelectos sintéticos reúnen en muchos casos una inteligencia de un nivel similar a un ser humano adulto, y ese aspecto empujaría a afirmar que también reúnen la propiedad de la sapiencia. Así pues, si un sistema de IA reúne, en cierta medida, sensibilidad para sentir dolor y una inteligencia de nivel similar al de un humano, se concluiría que presenta estatuto moral. Bostrom llega a esta conclusión a partir de una serie de principios de no discriminación, entre el que se encuentra el principio de no discriminación de material y el principio de no discriminación de ontogenia (2011: 8-9)

El principio de no discriminación de material señala que si dos seres tienen la misma funcionalidad y la misma experiencia consciente y difieren solo en el sustrato de su implementación, entonces tienen el mismo estatuto moral. Bostrom sostiene que para estar a favor de este principio podría emplearse un argumento similar al racismo. De la misma manera que no se considera a una persona de piel negra inferior por su color, tampoco debería afirmarse que un intelecto sintético construido con un material diferente al ser humano puede tener una consideración moral diferente. Pero esa consideración moral no se hace porque el intelecto sintético pueda tener la misma utilidad y funcionalidad que el ser humano, sino porque el material puede ser moralmente relevante en tanto que es diferente al ser humano. Este principio considera que no existen diferencias morales si el cerebro de un ser usa semiconductores o neurotransmisores.

El otro principio que sirve para ilustrar de mejor forma la reflexión ética de Bostrom acerca del estatuto moral es el principio de no discriminación de ontogenia. Este principio sostiene que si dos seres tienen la misma funcionalidad, la misma conciencia de su experiencia y solo se diferencian en cómo llegaron a existir, entonces tienen el mismo estatuto moral. La ontogenia describe el desarrollo de un organismo desde su gestación hasta su senescencia. Este principio se erige sobre la idea de que el estatuto moral no depende del linaje ni de la casta. Tanto es así, que aquellas personas que han sido fruto de la fecundación *in vitro*, de la adopción, etc., no quedan al margen de considerarse poseedoras de estatuto moral. Pues bien, el principio de no discriminación de ontogenia extiende esas ideas hasta los sistemas cognitivos artificiales, es decir, hasta la IA. Así pues, el origen de algo no determina que tenga estatuto moral, por lo que la IA tendría estatuto moral, al igual que lo tienen aquellas personas que han sido producto de técnicas de reproducción artificial.

En su obra *Superinteligencia. Caminos, peligros, estrategias*, Bostrom considera que la determinación de los valores a incorporar en la superinteligencia no es una tarea nada fácil. La cuestión de la elección de valores presenta dificultades y por lo tanto implica una fuerte exigencia en lo relativo a la buena elección, sin dejar oportunidad para equivocaciones que podrían tener graves consecuencias. La dificultad en la elección de valores estriba en el relativismo cultural que caracteriza el universo axiológico. Los valores suelen tener diferente consideración dependiendo del ámbito cultural en el que se presenten, motivo por el cual son propuestas de diferentes jerarquizaciones. Además de la relatividad en materia de valores, también existe dificultad en el momento de definir el objeto final de los intelectos superinteligentes, lo que conduce a profundos problemas filosóficos, que deben someterse a discusión, y a un aumento del nivel de complejidad. A lo dicho hay que añadir que no existe un consenso generalizado en lo referente a una teoría ética que cuente con un apoyo mayoritario dentro de la filosofía.

Ante estas dificultades para determinar qué valor o valores tendrían que introducirse en un intelecto sintético, Bostrom propone lo que él denomina «normatividad indirecta», que considera una necesidad. El sueco defiende la idea de una incapacidad por parte del ser

humano para determinar los valores por los que habría que decantarse. Las creencias morales han experimentado profundos cambios a lo largo de la historia y arriesgarse en la decisión de unos valores que sean válidos y que presenten una proyección a largo plazo es una tarea demasiada aventurada (Bostrom, 2016: 210). La humanidad ha vivido un importante progreso moral, pero aunque ese progreso moral se haya dado y, por ejemplo, no se realicen prácticas que en décadas o siglos pasados estaban totalmente reconocidas, el sueco considera que no existe todavía una preparación suficiente para poder proponer unos valores que tengan una proyección a largo plazo.

Ante la incapacidad del ser humano para proponer valores con proyección de futuro, Bostrom propone una normatividad indirecta que define de la siguiente manera:

> Esto nos lleva a la normatividad indirecta. La razón obvia para construir una superinteligencia es para que podamos introducir en ella datos sobre la razón instrumental necesaria para encontrar formas eficaces de realizar un valor dado. La normatividad indirecta nos permitirá también descargar en la superinteligencia algunos de los razonamientos necesarios para seleccionar el valor que se quiere realizar (2016: 210-211).

La normatividad indirecta es la propuesta formulada para enfrentar el desafío de los valores y su introducción en la IA. La tendencia a pensar desde el presente incapacita para proponer valores con proyección en el tiempo a largo plazo. El filósofo sueco afirma que si la superinteligencia es superior a los humanos en lo que respecta al intelecto cognitivo, lo oportuno sería que se delegara en ella la capacidad de decidir qué valores son los mejores. La superinteligencia es epistémicamente superior, lo que lleva a Bostrom a concluir que sería ella la que debería determinar qué valores son los más viables.

Además, hay otro texto de este filósofo que establece una serie de principios sobre IA divididos en tres secciones: temas de investigación, ética y valores y problemas a largo plazo:

> Cuestiones de investigación
>
> 1) Objetivo de investigación: el objetivo de la investigación de IA debe ser crear inteligencia no dirigida, pero inteligencia beneficiosa.

2) Financiación de la investigación: las inversiones en IA deben ir acompañadas de financiación para la investigación sobre cómo garantizar su uso beneficioso, incluidas preguntas espinosas en informática, economía, derecho, ética y estudios sociales, como:

- ¿Cómo podemos hacer que los futuros sistemas de inteligencia artificial sean altamente robustos, de modo que hagan lo que queremos sin que funcionen mal o sean pirateados?

- ¿Cómo podemos hacer crecer nuestra prosperidad a través de la automatización mientras mantenemos los recursos y el propósito de las personas?

- ¿Cómo podemos actualizar nuestros sistemas legales para que sean más justos y eficientes, para seguir el ritmo de la IA y para administrar los riesgos asociados con la IA?

- ¿Con qué conjunto de valores debe alinearse la IA y qué estatus legal y ético debe tener?

3) Enlace ciencia-política: debe haber un intercambio constructivo y saludable entre los investigadores de IA y los responsables políticos.

4) Cultura de la investigación: debe fomentarse una cultura de cooperación, confianza y transparencia entre los investigadores y desarrolladores de IA.

5) Evitar las carreras: los equipos que desarrollan sistemas de inteligencia artificial deben cooperar activamente para evitar el recorte de las normas de seguridad.

Ética y valores

6) Seguridad: los sistemas de AI deben ser seguros durante toda su vida operativa, y verificable cuando sea aplicable y factible.

7) Transparencia de falla: si un sistema de IA causa daño, debería ser posible determinar por qué.

8) Transparencia judicial: cualquier participación de un sistema autónomo en la toma de decisiones judiciales debe proporcionar una explicación satisfactoria auditable por una autoridad humana competente.

9) Responsabilidad: los diseñadores y constructores de sistemas de IA avanzados son partes interesadas en las implicaciones morales de su uso, uso indebido y acciones, con la responsabilidad y la oportunidad de dar forma a esas implicaciones.

10) Alineación del valor: los sistemas de IA altamente autónomos deben diseñarse de modo que sus objetivos y comportamientos puedan asegurarse de alinearse con los valores humanos a lo largo de su operación.

11) Valores humanos: los sistemas de IA deben diseñarse y operarse de manera que sean compatibles con los ideales de dignidad humana, derechos, libertades y diversidad cultural.

12) Privacidad personal: las personas deben tener derecho a acceder, administrar y controlar los datos que generan, dado el poder de los sistemas de IA para analizar y utilizar esos datos.

13) Libertad y privacidad: la aplicación de la inteligencia artificial a los datos personales no debe restringir injustificadamente la libertad real o percibida de las personas.

14) Beneficio compartido: las tecnologías de inteligencia artificial deberían beneficiar y capacitar a tantas personas como sea posible.

15) Prosperidad compartida: la prosperidad económica creada por la IA debe compartirse ampliamente, en beneficio de toda la humanidad.

16) Control humano: los seres humanos deben elegir cómo y si delegar decisiones a los sistemas de IA para lograr los objetivos elegidos por el hombre.

17) No subversión: el poder conferido por el control de sistemas IA altamente avanzados debe respetar y mejorar, en lugar de subvertir, los procesos sociales y cívicos de los que depende la salud de la sociedad.

18) Carrera armamentista de la IA: se debe evitar una carrera armamentista en armas autónomas letales.

Problemas a largo plazo

19) Precaución sobre la capacidad: al no haber consenso, debemos evitar suposiciones sólidas con respecto a los límites superiores de las capacidades futuras de inteligencia artificial.

20) Importancia: la IA avanzada podría representar un cambio profundo en la historia de la vida en la Tierra, y debería planearse y manejarse con cuidados y recursos proporcionales.

21) Riesgos: los riesgos que plantean los sistemas de inteligencia artificial, especialmente los riesgos catastróficos o existenciales, deben estar sujetos a los esfuerzos de planificación y mitigación acordes con el impacto esperado.

22) Automejora recursiva: los sistemas de IA diseñados para autorreplicarse o autorreplicarse de manera recursiva, de una manera que podría llevar a un aumento rápido de la calidad o cantidad, deben estar sujetos a estrictas medidas de seguridad y control.

23) Bien común: la superinteligencia solo debe desarrollarse al servicio de ideales éticos ampliamente compartidos, y en beneficio de toda la humanidad en lugar de un estado u organización (Bostrom, 2017).

En definitiva, aunque para Bostrom la IA actual no está todavía suscitando profundos problemas éticos, la orientación de los algoritmos hacia un aspecto más humano del pensamiento augura importantes complicaciones. La IA general ya no puede ejecutarse en aquellos contextos que son impredecibles, lo que impone la exigencia de nuevas medidas de seguridad y la incorporación de la ética al trabajo de la ingeniería. En definitiva, para este filósofo existen importantes retos éticos en el futuro de la IA que pueden ser enfrentados en primera instancia desde cierta normatividad.

4.3. Shannon Vallor: la posibilidad de cultivar las virtudes éticas

La reflexión ética de Shannon Vallor, profesora de la Universidad de Santa Clara en Silicon Valley, está motivada por el fenómeno de la automatización en el campo profesional. Vallor realiza una reflexión de este fenómeno desde la óptica de la ética de la virtud. Aristóteles dedica una de sus obras más importantes, la *Ética a Nicómaco*, a las virtudes, definiendo a la virtud como aquella disposición habitual a hacer el bien. En el hábito radica la formación de las virtudes, pues la repetición de determinadas acciones orientadas hacia el bien contribuye a que el ser humano vaya forjando un *ethos*. En ese sentido, en las profesiones se desarrollan actividades que se convierten en un hábito, donde

el humano cultiva la sabiduría práctica y la virtud. Así pues, si su actividad profesional se ve interrumpida por la automatización de su puesto de trabajo, el espacio con el que contaba para desarrollar el *ethos* desaparecerá, lo que puede implicar una crisis en la formación humana. Es desde aquí desde donde Vallor desarrolla su reflexión.

El cultivo de la virtud depende de los actos que la persona convierte en disposiciones habituales. Pero solo aquellas actividades que conducen al éxito, al bienestar, a ciertas habilidades orientadas a la adquisición de sabiduría, son las que podrían considerarse detonantes de la virtuosidad. Aunque, habría que matizar que no únicamente esas actividades contribuyen a la virtuosidad, pues aunque sí favorecen la virtud, no podría decirse que son condiciones suficientes, sino más bien necesarias. Otra matización importante que debe hacerse consiste en señalar que Aristóteles no menciona que las virtudes estén condicionadas por actividades que tengan que ver con los actos mecánicos o reflejos, sino más bien con aquellas actividades que están guiadas por una comprensión inteligente de la moral y que es demandada por lo que manifiestan las situaciones particulares (Vallor, 2015: 109). Esto que Vallor matiza lo señala Aristóteles en el libro II de la *Ética a Nicómaco*, donde presenta el término medio como la capacidad para saber discernir lo prudente entre el exceso y el defecto. Así pues, para el estagirita aquellas actividades que contribuyen a la virtud nacen de la voluntad, se hacen de manera consciente y razonada, y además tienen que ver con una disposición habitual a hacer el bien. Si el progreso de la tecnología, impulsada sobre todo por la IA, promueve actividades que perturban el cultivo de las habilidades de manera significativa, podría considerarse que el *ethos* de la persona estará condicionado y se verá afectado por este progreso.

La pensadora estadounidense utiliza el concepto «descualificación» (*deskilling*), que procede de la sociología, para referirse a aquel fenómeno que consiste en el desplazamiento de los trabajadores de su puesto de trabajo como resultado de la mecanización y también a la consecuencia que se deriva de ese desplazamiento, que implica la pérdida de valor y mérito de los trabajadores humanos. Para Vallor el fenómeno de la automatización del ámbito profesional ha sido ambiguo, pues ha tenido consecuencias positivas y también negativas. Las consecuencias positivas se refieren al hecho de que muchas tareas que

tradicionalmente eran rutinarias han liberado de cierta carga de trabajo a los humanos y también han servido para potenciar la formación y capacitación en determinados conocimientos que antes eran desconocidos. Sin embargo, también están teniendo importantes consecuencias económicas, sociales y culturales.

Las nuevas tecnologías están removiendo los cimientos de muchas estructuras sociales, económicas y culturales, por lo cual es importante considerar críticamente cómo están afectando a la formación del *ethos* en este tiempo. Más allá de análisis dogmáticos que fomentan el pánico irracional a la tecnología, urge afrontar este fenómeno con una reflexión crítica que sea responsable y se construya sobre argumentos sólidos. El impacto de la tecnología no puede caracterizarse por el desequilibrio, según Vallor, pues no debe tener únicamente impactos negativos sobre el *ethos*. Una reflexión sobre el fenómeno del progreso de la tecnología debe contar con aportaciones de la ética de la virtud, la psicología moral y de un estudio empírico riguroso sobre las condiciones materiales y sociales que propician el florecimiento humano (2015: 112).

Para Vallor es central otorgarle importancia al impacto de la tecnología en el desarrollo de las habilidades morales, pues esas habilidades están directamente relacionadas con la formación de un *ethos* virtuoso y moldean a la persona. La restricción o limitación del uso de la tecnología no es una solución viable. A estas alturas de la historia debería reconocerse que la tecnología ha marcado la vida humana condicionando la capacidad de conocer el mundo, pudiendo afirmar que el humano es una criatura tecnomoral (2015: 113). Sin embargo, la autora señala tres riesgos potenciales: los drones y los sistemas armamentísticos autónomos, la tecnología de medios de comunicación y los robots de cuidados.

Vallor es consciente del creciente desarrollo que está experimentando la tecnología de IA en el campo militar, por ejemplo en el caso del estadounidense X-47B o del británico Taranis. La tecnología militar se va perfeccionando con el paso del tiempo y está logrando importantes resultados en la lógica de funcionamiento militar. El éxito de las armas autónomas está motivando a las instituciones militares a aumentar su uso en los próximos

años y a ampliar el espectro de acción. Vallor se congratula de la renovación de la resolución de la Comisión Internacional de Derechos Humanos por parte de la Asamblea General de la Organización de Naciones Unidas (ONU) (2017) en la que se hacen algunas consideraciones sobre la reglamentación en caso de conflictos militares. Es importante que resoluciones de este tipo se mantengan en el tiempo, pues los conflictos militares deben seguir una serie de reglamentaciones marcadas por los organismos internacionales competentes para que los enfrentamientos no deriven en un caos absoluto por falta de moralidad. Para esta autora los conflictos militares pueden ser morales o inmorales, bajo la condición de que se respete la legislación internacional. El trabajo militar puede ser virtuoso, siempre y cuando se cultive de forma adecuada por medio de habilidades morales relativas al combate (2015: 114-115).

La alternativa que propone Vallor consiste en que la utilización de IA en el campo de batalla sirva para proporcionar una mayor y mejor información al personal militar, de modo que disponga de mayores conocimientos para desarrollar su actividad de mejor forma y poder discernir razonablemente. El distanciamiento con respecto al campo de batalla puede servir para que el personal militar no sufra traumas psicológicos y morales, como es el caso de los aviones no tripulados. La propuesta para lograr un equilibrio en el impacto de la IA en el campo militar consiste en que los sistemas armamentísticos autónomos favorezcan el desarrollo de las habilidades morales de la mejor manera posible para el cultivo de la virtud, incluso en la particularidad del campo en cuestión. Lo expuesto hasta ahora lo define muy bien esta autora en la siguientes líneas: «Las tecnologías militares no son meramente herramientas para lograr objetivos prácticos; cada vez más, son el medio para definir el tipo de soldado que un soldado quiere ser y qué virtudes morales puede desarrollar a través de su servicio» (2015*:* 116).

Otro de los riegos que señala Vallor se encuentra en las tecnologías de la información y la comunicación (TIC). Numerosos estudios han puesto de relieve que la habilidad cognitiva para prestar atención se ha visto mermada con el uso de estas tecnologías. La atención no es únicamente una habilidad cognitiva, sino que también es una habilidad moral. Las habilidades cognitivas se relacionan con la inteligencia moral de los seres

humanos y también con otros tipos de intelectos, es decir, ayudan a saber cuándo prestar atención, a quién, durante cuánto tiempo y de qué manera, algo que es fundamental para la formación de un pensamiento crítico y de una ciudadanía instruida. El mal uso de las TIC socava la capacidad cognitiva y fomenta nuevas distracciones, disminuyendo así la concentración.

La solución no radica en luchar a contracorriente contra las tecnologías de la comunicación y la información, pues es una batalla perdida, ya que se encuentra presente en todos los espacios de la vida. Más bien, la tarea podría consistir en repensar el sentido de la tecnología y la utilidad que se le atribuye, con el fin de obtener el mayor beneficio de ella para la formación de un *ethos* virtuoso. No se trata de promover un «dejar hacer», sino más bien de tomar las riendas en el uso de las TIC para que contribuyan a la formación en virtudes.

Por último, Vallor considera que otro de los riesgos puede encontrarse en los robots dedicados a los cuidados (*carebots*). En muchos países el crecimiento de la población adulta ha aumentado en las últimas décadas, despertando el interés en los desarrolladores de robots para proporcionar herramientas en un campo en el que pueden obtener una gran rentabilidad. Para la estadounidense, la introducción de los robots en el campo de los cuidados presenta importantes dilemas éticos, pues el cultivo humano de prácticas vinculadas con el cuidado, y las virtudes morales que se derivan de esa actividad, estarían en peligro de extinción (2015: 119).

No es fácil desarrollar la actividad de los cuidados, pues se encuentran implicados aspectos emocionales, físicos y financieros que se van formando con el paso del tiempo y que necesitan mucha voluntad. La actividad de los cuidados es muy especial porque conlleva un componente emocional que es fundamental para dar sentido a esa actividad. Dicha actividad es una de las que comúnmente hoy es reconocida como virtuosa y va desarrollándose en un estrecho vínculo con el otro, a partir del cual se va dando forma al *ethos*. El cultivo del carácter en los cuidados se va forjando de forma relacional con el otro que me condiciona. Así pues, Vallor sostiene que los desarrolladores de robots en el campo

de los cuidados deben tener en cuenta todos estos aspectos y orientar su actividad hacia la creación de sistemas que satisfagan las necesidades morales de los cuidadores, también sus habilidades, pero sobre todo una mayor calidad en la actividad del cuidado, teniendo en cuenta todos los matices de virtuosidad que subyacen en esa actividad.

Como puede comprobarse, Vallor ofrece interesantes alternativas a los tres riesgos mencionados, y todas las formula desde la ética de la virtud. En términos generales, apuesta por estudiar qué posibilidades de cultivo de la virtud subyacen en las tecnologías que impulsa la IA y tratar de potenciarlas para esclarecer los beneficios morales positivos que pueden obtenerse. Sin embargo, la autora considera que esto es necesario, pero no suficiente, y reivindica la necesidad de un profundo cambio de valores culturales en las sociedades tecnológicas que refleje una nueva conciencia global en la que se asuma que el ser humano es a la misma vez artificiero y artefacto tecnológico (2015: 122). Debe fortalecerse el sentido de la responsabilidad colectiva, pues no es un tema en el que solo tengan responsabilidad los desarrolladores de IA. Es necesario un nuevo compromiso en el cultivo de las virtudes para superar el complejo desafío de este tiempo.

4.4. Bill Hibbard: el aprendizaje de valores basados en el principio de justicia universal

El estadounidense Bill Hibbard es un científico emérito del *Space Science and Engineering Center* de la Universidad de Wisconsin-Madison y también colabora con el *Machine Intelligence Research Institute*. Entre sus escritos más importantes destacan *Super-Intelligent Machines* (2002), *Avoiding Unintended AI Behaviors* (2012) y *Ethical Artificial Intelligence* (2015). Sus investigaciones se centran en la cuestión de los problemas éticos que subyacen en el diseño de los sistemas de IA y cómo resolverlos.

Para Hibbard, las tres leyes de Isaac Asimov, aparecidas en su famosa obra *Yo robot*, presentan falacias y ambigüedades. Una de las soluciones que propone para resolver la ambigüedad de tales leyes consiste en la instrucción de los intelectos sintéticos por medio de la asignación de valores numéricos para cada resultado posible (2015: 10). A partir de

ahí el sistema podría usar esos valores numéricos para calcular qué decisión tomar. Sin embargo, este asunto es mucho más complejo, pues el intelecto sintético podría dudar de los resultados que se derivarán de sus acciones, lo que implicaría que fuera necesario realizar un cálculo de probabilidades. Esas probabilidades pueden servir para calcular los valores que espera que se deriven de cada acción emprendida. La experiencia es un factor importante para el cálculo de probabilidades y también es un requisito fundamental para aquellos intelectos sintéticos centrados en el aspecto probabilístico. En ese sentido, para Hibbard una IA que no cuente con experiencia previa no puede emprender acciones que exijan seriedad (2015: 11).

Este modelo de asignación de valores numéricos a los resultados posibles se fundamenta en la idea de asociación de un beneficio con un resultado y, en cierta medida, ayuda a resolver las ambigüedades por medio de la instrucción. Según Hibbard, otro factor a tener en cuenta en la elección de los resultados es la preferencia en base a una función de utilidad para la obtención de beneficio. En este contexto, el estadounidense destaca también la importancia de establecer una relación dialéctica con el entorno a partir de parámetros de acción y observación. Así pues, para los agentes sintéticos la cuestión del beneficio y el sufrimiento se definen a partir de su utilidad.

Esta estrategia de seleccionar acciones que impliquen una maximización del beneficio y una minimización del sufrimiento, se encuentra en la línea del utilitarismo al que Hibbard crítica. Para Hibbard, tras los entresijos del utilitarismo se esconde una actitud permisiva, pues pueden permitirse acciones que moralmente son malas, pero que tienen unas consecuencias beneficiosas. El utilitarismo es una ética normativa basada en reglas que puede conducir a ciertas ambigüedades, ya que existen entornos que implican romper determinadas reglas. Una solución podría consistir en la introducción de un valor de utilidad para cada caso concreto en función de un historial de acciones de reglas que el agente rompe (2015: 19).

Otra de las críticas que Hibbard vierte sobre el utilitarismo consiste en la ignorancia de la intencionalidad que se encuentra tras las acciones que se emprenden. El comportamiento humano suele conocerse por medio de patrones y una función de utilidad que consista en la introducción de un historial de patrones de comportamiento de un agente podría permitir contar con mayor información para conocer el comportamiento de ese agente con el entorno. No obstante, en lo relativo a la intencionalidad existen ciertas discusiones acerca de si un intelecto sintético tiene, o no, intencionalidad. No obstante, como bien señala Hibbard, tenga o no intencionalidad, la responsabilidad de aplicar la ética a esos sistemas depende de los desarrolladores humanos (2015: 19).

La tercera objeción que Hibbard le hace al utilitarismo consiste en la idea de que la vida humana no es susceptible de aplicación de valores numéricos cuantitativos a la hora de traducir ese aspecto a términos matemáticos para introducirlos en un sistema de IA. Por último, este pensador señala que es importante que se introduzcan preferencias en los sistemas de IA, con el fin de que no se produzca una caída en el plano subjetivo de cualquier intelecto sintético, pues eso sería peligroso. Una vez dicho esto, y expuestas las críticas y las matizaciones que Hibbard hace al utilitarismo, la clave está, a su juicio, en proporcionar emociones para guiar el aprendizaje de las conductas, en vez de la introducción de leyes que limiten a los intelectos sintéticos. Esto consiste en programar máquinas para que aprendan algoritmos que muestren los fundamentos de los valores humanos. Se trata de desarrollar sistemas de aprendizaje en los que la IA tenga acceso a los valores humanos por medio de algoritmos que se aprenden, una tarea nada fácil (Hibbard, 2012: 113).

La estrategia consistente en proporcionar instrucciones a la IA que involucren leyes o reglas puede implicar situaciones ambiguas. No obstante, Hibbard considera que aquellas instrucciones que se proporcionan a los que definen valores que son de utilidad para situaciones concretas pueden llegar a evitar la ambigüedad a la que conducen las leyes y reglas.

Esta propuesta de aprendizaje se construye sobre la máxima del aprendizaje de valores de los intelectos sintéticos para no emprender acciones intencionales que perjudiquen a los humanos. En este caso, la dificultad radica en la selección de unos valores que sean lo suficientemente ricos como para enfrentarse a la gran variedad de situaciones que configuran la realidad y que normalmente se fundamentan en el plano subjetivo, algo que lleva de nuevo a la ambigüedad. Cuando los humanos se enfrentan a las diferentes situaciones que dan forma a la realidad, son ellos mismos los que asignan los valores que interfieren en las acciones que emprenden. No obstante, Hibbard menciona un estudio realizado por Luke Muehlhauser y Louie Helm, según el cual los humanos no especifican con tanta facilidad ni con tanta precisión sus propios valores (Hibbard, 2015: 78). Un sistema de valores se caracteriza por la complejidad, motivo por el cual la traducción de estos valores a algoritmos resulta algo muy complicado. Y, además, para el aprendizaje de valores humanos se necesita un intelecto sintético con una gran capacidad de aprendizaje, lo que a juicio de Hibbard conduce al problema sobre la secuencia del huevo o la gallina. Pues para la cuestión del aprendizaje de valores no se sabe si tendría que ser antes una IA no muy avanzada, para así poder aprender valores humanos, o una IA muy avanzada que sea capaz de aprender valores humanos (2012: 113).

Como se ha indicado, los sistemas de valores se caracterizan por la complejidad, pero para Hibbard la dificultad no solo radica en la facilidad subjetiva para especificar los valores, sino también en la combinación de valores de varios individuos para incorporarlos a la IA. Es posible incorporar una suma de valores humanos a una ecuación matemática que posteriormente pueda formar parte del proceso de aprendizaje de un intelecto sintético. Para la incorporación de esa suma de valores a una ecuación matemática debería seguirse un criterio de equilibrio de intereses entre varias personas, donde se dé prioridad y un mayor valor a aquellos intereses que son compartidos y se trate de resolver el desacuerdo entre intereses por medio de la asignación de valores máximos y mínimos (Hibbard, 2015: 86-87). Además, también sería fundamental que ciertos valores se sometieran a un filtro colectivo en el que se establecieran promedios de asignación de valores, con el objetivo de corregir las inconsistencias que pudieran existir.

A veces la determinación de los valores puede estar sometida a un cierto sesgo discriminador. En esa búsqueda de la forma de combinación de los valores de múltiples humanos para su incorporación en los sistemas de IA pueden incidir aspectos, en ocasiones, discriminatorios, propios de los intereses subjetivos o incluso de una clase social determinada. Para tratar de corregir esa especie de discriminación por intereses, Hibbard recurre a la teoría de la justicia de John Rawls (2010). Rawls considera que en las sociedades con democracia liberal debe ser resguardardo el principio de igualdad de oportunidades, con el fin de corregir las desigualdades económicas que puedan existir. Hibbard presenta su propuesta como alternativa al proyecto utilitarista, que determina el valor de utilidad en base al cálculo de un promedio. Así pues, el científico estadounidense se da cuenta de que su propuesta de determinación de valores entre múltiples humanos es posible que responda a la lógica utilitarista que él tanto crítica, y por eso rescata la propuesta de Rawls basada en el velo de ignorancia para corregir las inconsistencias a la hora de determinar los valores entre múltiples humanos. El velo de la ignorancia de Rawls proporciona una matriz importante desde la que seleccionar valores múltiples para evitar así cualquier sesgo de intereses (2015: 87).

Otro factor importante para introducir valores en los intelectos sintéticos ha de ser la revisión periódica. Esta revisión consiste en un control periódico sobre los valores que han sido asignados para la IA, con el fin de que esos valores se correspondan con el progreso moral de la humanidad en función de la evolución. Los valores se someten a cambios, fruto del progreso moral que experimenta la humanidad con el paso del tiempo. Para que los intelectos sintéticos no se encuentren en una situación de desfase es importante que sean sometidos a una revisión periódica de esos valores y a una posterior actualización de los mismos (2015: 88-89).

El rescate de Rawls propuesto por Hibbard supone una consideración de su propuesta dentro del marco kantiano. Cuando reivindica a Rawls para corregir los desfases utilitaristas en la selección de valores propuestos desde la multiplicidad de humanos, está también reivindicando a Kant. Tras la inspiración rawlsiana en el planteamiento de Hibbard se encuentra el propósito de establecer unos principios de justicia que sean universalmente

aplicables y deducidos de una constitución común a todos los agentes morales. Existe una ambición universalista en Hibbard cuando recurre a Rawls. Además, tras la importancia que tiene para este autor la revisión de valores a la luz del progreso moral, podría encontrarse cierta pretensión de universalidad, pues se está haciendo una invitación a tener en cuenta los posibles cambios que pueden darse en el terreno moral desde una óptica universal que sea fruto del progreso moral que experimenta la humanidad. Así pues, esta propuesta de aprendizaje de valores por medio de ecuaciones matemáticas que propone Hibbard es propiamente kantiana porque se preocupa de dotarla de un carácter universal para tratar de corregir los desfases, a los que, según él, conduce el utilitarismo.

4.5. Rajakishore Nath y Vineet Sahu: la negativa ética de la agencia moral

Rajakishore Nath y Vineet Sahu (2017) son dos investigadores indios que consideran que en la aplicación de la ética en el campo de la IA subyace una importante problemática. Esa problemática radica en la ausencia de subjetividad y, por lo tanto, de conciencia de las máquinas. Para estos expertos la investigación en el campo de la IA no solo tiene como finalidad la creación de artefactos para resolver problemas en el ámbito de la ingeniería, la medicina, la química, etc., o para la automatización del entorno profesional, sino que también permite entender mejor qué es la inteligencia en sí misma. La IA estudia los mecanismos necesarios para que los programas artificiales simulen los comportamientos de los humanos dentro de los límites de lo que se consideraría como «inteligente», pero también para realizar tareas cognitivas que normalmente realizan los humanos (2017: 2). Si el objetivo principal de la IA consiste en desarrollar todos los mecanismos necesarios para que los intelectos sintéticos se comporten de manera similar a los seres humanos, de ahí se podría seguir que la IA también puede realizar una conducta moral. Como se verá, es ahí donde radican las principales objeciones de estos investigadores.

Para abordar el tema sobre la posibilidad de la ética en la IA, Nath y Sahu comienzan discutiendo la idea de mente. En términos generales, la IA tiene entre sus objetivos desarrollar la mente en las máquinas, siendo necesario comprender la naturaleza de la inteligencia humana para poder imitar sus estructuras hasta cierto punto. Así, si se sostiene

la idea de que las máquinas tienen mente, entonces se les debería atribuir cierta creencia, intencionalidad, libre albedrío, etc. (2017: 2). Según esta lógica de pensamiento, las máquinas tendrían mente porque desarrollan una serie de operaciones en las que están implícitos algoritmos que cuentan con una finalidad concreta. La complejidad de los algoritmos se encuentra directamente relacionada con la complejidad de las actividades desarrolladas. A partir de esta consideración las computadoras serían emulaciones de la complejidad del cerebro humano.

Los defensores de la IA fuerte (IAF) piensan que es posible que los sistemas de IA tengan mente porque son capaces de desarrollar procedimientos algorítmicos que cuentan con la misma complejidad que los procesos cerebrales. Esta idea se formula desde un presupuesto mecanicista, pues reduce la inteligencia biológica a procesos mecánicos complejos que configuran el cerebro. El razonamiento es el siguiente: si los seres humanos tienen mente, y eso se deriva de unas estructuras cerebrales caracterizadas por la complejidad, si se desarrollan inteligencias no biológicas que consigan imitar, y hasta superar, los procedimientos complejos cerebrales por medio de algoritmos, entonces se estará asumiendo que es posible que los intelectos sintéticos también tengan mente.

Además, para estos investigadores la teoría computacional se centra en aspectos abstractos que establecen un paralelismo entre los intelectos sintéticos y los intelectos humanos. Para la teoría computacional existe una dinámica que responde a la causalidad en el cerebro humano y que es posible trasladar a las computadoras. Esa dinámica, configurada desde unos patrones de funcionamiento, se refleja en las neuronas, que establecen vínculos con otras neuronas. Entonces esos vínculos que se establecen entre neuronas son los responsables de cualquier experiencia consciente (2017: 3).

Para Nath y Sahu existe una dificultad a la hora de creer que las máquinas poseen mente y que también se les atribuyan cualidades, como la subjetividad, que son propias de los seres humanos. Para los desarrolladores de IA es posible hablar de una máquina como un agente moral y por lo tanto que siga principios éticos ideales que orienten su acción. Sin embargo, para los investigadores indios esto no es del todo cierto. Para argumentar su

posición, rescatan la distinción propuesta por James H. Moor (2006) entre dos tipos de agente ético, a saber, un agente de ética implícita y un agente de ética explícita (Nath y Sahu, 2017: 4). En primer lugar, un agente de ética implícita es aquel que ha sido objeto de una programación previa para evitar comportamientos que se consideran como no éticos, sin necesidad de representar explícitamente el comportamiento ético. Este tipo de agente se limita a los comportamientos fundados en los principios éticos de su desarrollador. En cambio, un agente de ética explícita es aquel que es capaz de calcular mejor la acción en base a principios éticos para superar dilemas. En ese sentido, es fácil entender que el objetivo de los desarrolladores sería crear máquinas que sean agentes con ética explícita y atribuirles un estatuto moral, aunque artificial.

Nath y Sahu defienden la idea del diálogo interdisciplinar entre filósofos e investigadores de IA con el fin de tratar de superar algunos enfoques que consideran erróneos. Así como los filósofos comparten trabajo con los médicos en los comités de bioética, también lo podrían hacer con los investigadores de IA. El error que Nath y Sahu atribuyen a los investigadores de IA radica en que estos últimos creen que las máquinas son agentes morales. Los agentes morales están sometidos a una constante toma de decisiones, estrechamente vinculada con la capacidad de pensar en primera persona y por lo tanto de tener conciencia (2017: 5), algo que Nath y Sahu se resisten a creer que ocurra en el terreno de las máquinas.

El argumento que utilizan para defender su idea es el siguiente. Supongamos que los agentes de ética explícita emprenden acciones que luego no son capaces de argumentar en base a unos principios éticos concretos, lo que haría que difícilmente puedan ser considerados como agente éticos. Esto se debe principalmente a que un agente ético es un agente moral que actúa y forma su criterio en base a unas reglas o principios que luego es capaz de utilizar para fundamentar la explicación de sus acciones.

Las teorías éticas que los humanos identifican como referentes no se basan en valores matemáticos, sino en hábitos y elecciones libres, entre otras cosas. En ese sentido, según los investigadores indios, es importante diferenciar entre aquellas acciones realizadas por

preferencias y aquellas que son fruto de las decisiones morales, es decir, entre las inclinaciones y los deberes. Kant es uno de los filósofos que se ha ocupado de esta cuestión, diferenciando en la *Fundamentación de la metafísica de las costumbres* entre aquellas acciones que se realizan por un sentido del deber y aquellas que se realizan de acuerdo al deber. El imperativo categórico kantiano sanciona la acción del agente moral para que actúe conforme a la racionalidad desde un sentido ético trascendental. En principio, el postulado kantiano podría aplicarse a los intelectos sintéticos, pues éstos funcionan al margen de cualquier aspecto emocional. Sin embargo, y aquí es donde reside la problemática subyacente al hablar de ética en el ámbito de la IA, según Nath y Sahu, para Kant las acciones morales implican que aquella persona que las emprende lo hace desde la libertad y la racionalidad. Entonces, según el razonamiento kantiano, los intelectos sintéticos no podrían ser considerados como agentes morales, pues no están dotados de libre albedrío al no poseer conciencia.

Para comprender mejor el postulado kantiano es importante tener clara la distinción entre heteronomía y autonomía, entre inclinación y deber. Mientras que la heteronomía está ligada a la inclinación, y por lo tanto a algo que mueve a actuar desde fuera al agente moral, la autonomía y el deber están motivados por la razón práctica que actúa en base a un imperativo categórico. Esta motivación brinda la posibilidad de ser moral al agente e implica una libertad y conciencia que determinan.

Siguiendo el hilo del postulado kantiano, para que un agente sea considerado moral, es necesario que actúe con intencionalidad y por lo tanto desde la libertad que implica la conciencia. La máxima kantiana del imperativo categórico suscita un interés por el otro que nace de los sentimientos que forman parte de la naturaleza del agente moral y que son una instancia previa a cualquier acto motivado por esa máxima. El postulado de Kant establece un estrecho vínculo entre sentimiento y acción, es decir, las acciones emprendidas contienen un bagaje sentimental del que, según Nath y Sahu, carecen las máquinas, y esto es lo que explica que no puedan ser consideradas como agentes morales nunca. Además, para el reconocimiento de la alteridad es fundamental contar con una capacidad que no sea moralmente racional, sino que implique otros aspectos que estriban en lo emocional. En

definitiva, si las acciones de un intelecto sintético no son producto de un proceso deliberado de elecciones morales, sino que simplemente han sido introducidas, entonces no podrá ser considerado como un agente moral.

Un componente muy importante de la ética que conduce al tema de la conciencia es la responsabilidad. Para que a un agente se le pueda imputar responsabilidad de sus actos es necesario que esos actos hayan contado con voluntariedad y por lo tanto libertad, lo cual implica conciencia de sí y de los otros. Para los científicos indios el principal motivo para creer que no es posible hablar de máquinas como agentes morales es que carecen de subjetividad y por tanto de experiencia subjetiva, porque es ahí donde el humano encuentra las emociones que inciden sobre su acción racional. La subjetividad es el pilar sobre el que se levanta la agencia moral y donde cobra sentido la moralidad.

La subjetividad es la característica central de la conciencia y va dando forma a su naturaleza a partir de las experiencias. La mente consciente se caracteriza por ser particularmente accesible para el sujeto mismo, y por lo tanto, cae en el plano de la absoluta subjetividad. Es en el terreno de la conciencia donde subyace la problemática de si los intelectos sintéticos son, o no, agentes morales. Es precisamente ahí, en la conciencia, donde Nath y Sahu encuentran dificultades para hablar de una ética de la IA, porque hasta este momento los intelectos sintéticos no tienen conciencia de sí. La conciencia se reconoce desde la primera persona, donde el agente moral cobra sentido, y no en la tercera persona. La subjetividad no encuentra su razón de ser en terceros, y la experiencia consciente es la representación de la subjetividad que se vive en primera persona. Así pues, desde la tercera persona es imposible proporcionar un estatuto moral a la IA. Si se descarta la posibilidad de conciencia en los intelectos sintéticos, entonces también queda excluida la posibilidad de considerarlos como agentes con estatus moral.

La conciencia es necesaria para situarse en el escenario de la ética porque su objeto de reflexión es la moralidad de los sujetos. Si un intelecto sintético carece de moralidad, difícilmente podrá ser objeto de consideración ética. La conciencia es algo tan importante para la ética que para Nath y Sahu no puede reducirse a aspectos mecanicistas.

4.6. Una propuesta alternativa desde la responsabilidad ética

Tras presentar en los apartados anteriores algunas de las principales líneas de aplicación de la ética a la IA, en el siguiente capítulo se desarrollará el concepto de IAR, se precisará su sentido y se determinará su alcance. La propuesta de una IAR, que se sustenta en el principio ético de la responsabilidad, tiene como objetivo desarrollar una concepción cívica y democrática de la ciencia y, en particular, de la IA, cuyas bases se encuentran en un modelo de innovación abierta y responsable y en el trabajo en laboratorios abiertos.

La posibilidad de incorporar responsabilidad en el contexto de los sistemas artificiales ha sido abordada por Virginia Dignum (2013; 2016; 2017a; 2017b; 2018) y la Declaración de Montreal (2018). Sin embargo, estas propuestas presentan cierta debilidad en lo que respecta a una fundamentación filosófica. En el capítulo 5 se presentará el concepto de IAR como una alternativa frente al resto de perspectivas éticas esbozadas en este capítulo. Esta alternativa puede contribuir a enriquecer la reflexión en torno a las pretensiones de aplicar la ética en el contexto de los intelectos sintéticos. Además, abre la posibilidad para crear puentes de diálogo entre el conocimiento científico y la ciudadanía que, desde una concepción cívica de la ciencia, generen espacios de confluencia de perspectivas como los laboratorios abiertos.

El despliegue de una IAR se fundamenta en la propuesta de un renovado humanismo tecnológico, desde el cual es posible plantear una alternativa ética de responsabilidad alejada de los tecnopesimismos más radicales, una combinación de optimismo y criticismo. Además, la IAR cuenta con un modelo de generación de conocimiento innovador como es el MIAR, que potencia la participación de la sociedad civil y también el encuentro entre esferas y perspectivas de conocimiento. Por último, la propuesta de una IAR adquiere un profundo compromiso como herramienta para impulsar y garantizar los derechos humanos, los ODS y el cuidado de los límites planetarios, así como el cultivo y fortalecimiento de las habilidades cívicas y democráticas.

CAPÍTULO 5

INTELIGENCIA ARTIFICIAL Y RESPONSABILIDAD

> […] el poder tecnológico puede desbordar los límites naturales de la condición humana, deformando la capacidad de asumir una extendida responsabilidad ante la naturaleza y ante la humanidad futura […] Para ello, necesitamos deliberar exhaustivamente sobre las posibles consecuencias y anticipar los escenarios futuros, con el fin de poder actuar a tiempo para remediar o evitar los efectos negativos que la intervención tecnológica puede ocasionar. Asimismo, para potenciar sus efectos positivos y para lograr una mayor equidad en la distribución de los beneficios del desarrollo tecnológico.
>
> (Linares, 2008: 498)

La entrada de la tecnología a gran escala en el ámbito social invita a la discusión de sus implicaciones para la vida. Los impactos de las tecnologías son diversos, pues dependen de la finalidad y el contexto en el que se insertan y en ese sentido poseen un gran potencial transformador. Señala Kevin Kelly:

> El cambio es inevitable. Ahora nos parece que todo es transformable y está sujeto a transformaciones, aunque gran parte de estos cambios sean imperceptibles. […] En el núcleo de todo cambio significativo de nuestras vidas hay algún tipo de tecnología. La tecnología conlleva la aceleración de la humanidad. Gracias a la tecnología, todo lo que hacemos se encuentra siempre en proceso de transformación. Todas las clases de cosas se convierten en algo nuevo, mientras pasan de «podrían» a «son». Todo se encuentra en flujo permanente. Nada está totalmente terminado. Nada está hecho. Este cambio constante es el eje en torno al cual gira el mundo moderno (2017: 5).

En el terreno científico, Jean-Marc Lévy-Leblond (1975) destaca que la ciencia es una actividad que cae bajo el paraguas de las actividades morales, pues se inserta en contextos sociales susceptibles de reflexión en torno al principio de responsabilidad. El conocimiento científico responde a estándares de consideración ética donde entra en juego la reflexión sobre las implicaciones que tiene para la vida de los seres humanos y la biosfera. Sobre la

supuesta neutralidad axiológica del conocimiento científico, es importante mencionar tres puntos que Lévy-Leblond destaca para cuestionar la idea de neutralidad de la ciencia que el positivismo ha promovido en su discurso:

- Los científicos que suelen rechazar la responsabilidad de aquellas consecuencias nefastas por su trabajo reclaman una serie de elogios y reconocimientos por los efectos positivos, con el fin de ocultar las implicaciones negativas.

- La ciencia no puede ser un conocimiento neutro porque no se sitúa al margen de influencias externas, ya que existen infinidad de intereses que ejercen su influencia sobre las investigaciones.

- Los científicos no se encuentran al margen de la sociedad y por lo tanto la ciencia no puede escapar de la influencia directa de las condiciones del contexto económico, político y social.

Por algunos motivos que pueden ser económicos o estrictamente tecnocientíficos, en ocasiones se torna complejo observar las implicaciones que ciertos mecanismos presentan. La lógica económica normalmente activa la maquinaria de la falta de cuestionamiento bajo la afirmación de que la rentabilidad económica es un motivo más que suficiente como para no ser cuestionado ni relegado a un segundo plano. Además, la formación ética de los especialistas del campo tecnológico suele brillar por su ausencia. Ha sido en los últimos años cuando se ha introducido la formación ética en algunos centros académicos para orientar a los futuros profesionales del sector y ayudarles a configurar mejor su actividad. En ese sentido, tanto la lógica económica como la falta de formación ética de los profesionales de la tecnología dificultan la identificación de las implicaciones éticas en este campo y también representan un claro obstáculo para incorporar criterios de responsabilidad a la acción tecnológica.

En este capítulo se hará hincapié en la presencia de aspectos susceptibles de discusión ética dentro del campo tecnológico, así como en los impactos políticos que tienen para la democracia determinados mecanismos tecnológicos. Se introducirán las respuestas

formuladas a los desafíos políticos de la tecnología considerando diferentes caminos, el *laissez-faire*, el optimista, el escéptico y el esencialista. Sin embargo, ninguno de esos caminos brinda una vía pertinente desde la que afrontar estos desafíos. La búsqueda de un camino que sintonice con las necesidades enfrentadas pasa necesariamente por reconocer que las tecnologías presentan implicaciones morales que pueden poseer un impacto positivo en la sociedad mediante el fortalecimiento de las habilidades cívicas y la democracia.

La discusión en torno a la responsabilidad ética también se sitúa en el centro del debate, estableciendo una clara diferencia entre los planteamientos *top-down* y *bottom-up* (Allen, Smit y Wallach, 2005). Más allá de esos dos planteamientos se formulará el concepto de IAR como una alternativa de superación de los discursos positivistas. Postulados como los formulados por Virginia Dignum y la Declaración de Montreal suponen un primer intento para considerar la cuestión de la responsabilidad en el terreno de la IA, aunque su debilidad radica en que carecen de una fundamentación filosófica fuerte y se limitan al ámbito deontológico. En ese sentido, la IAR se formula a partir del concepto de ciencia cívica y del modelo de innovación abierta y responsable (MIAR), dos propuestas enriquecedoras para enfrentar los complejos retos de este tiempo, entre los que se encuentran los derechos humanos, los ODS y los límites planetarios, promoviendo así el compromiso cívico y fortaleciendo los pilares democráticos desde el ámbito tecnológico.

5.1. Tecnología y responsabilidad

En los últimos años los importantes avances experimentados en el campo de la IA se presentan en los medios de comunicación como un fenómeno de esperanza para la resolución de muchas de las problemáticas que enfrenta el ser humano. Los medios contribuyen a la generación de un sentimiento de confianza en la tecnología que puede acarrear un disimulo de ciertas amenazas y la dificultad para emprender una vía crítica, que no tecnófoba. Los avances mostrados parecen transmitir la idea de que no existe preocupación. En cambio, la historia reciente pone en evidencia que la acción técnica no tiene exclusivamente impactos positivos, sino también negativos, como demuestra el cambio climático y la tecnología empleada para la guerra (Benjamin, 2014; Jordán

Enamorado y Baques Quesada, 2014). La intervención técnica del ser humano tiene diversas caras y no todas son tan positivas como en ocasiones se ha podido observar. La tecnología tiene implicaciones que comprometen la vida y el futuro, que han de enfrentarse desde una actitud crítica y responsable.

Existe una versión optimista de la tecnología que en otro lugar se ha caracterizado como «tecnooptimismo» en comparación con el «tecnopesimismo» (Terrones Rodríguez, 2018), y que se erige sobre el presupuesto dogmático del cientificismo. Gerard Radnitzky define el cientificismo como «la creencia dogmática de que el modo de conocer llamado ciencia es el único que merece el título de conocimiento; y su forma vulgar, la creencia de que la ciencia resolverá eventualmente todos nuestros problemas o, cuando menos, todos nuestros problemas más significativos» (1973: 254-255). Kurzweil, a quien ya se ha hecho referencia, se sitúa en la estela del tecnicismo, una forma que adopta el cientificismo en el campo técnico, por entender que la tecnología en el tiempo de la singularidad será capaz de encontrar remedio a muchos de los males que enfrenta la humanidad, como las enfermedades o el cambio climático. Este presupuesto dogmático comienza a gestarse desde la modernidad y se consolida en la Ilustración, dando lugar a la idea del progreso ilimitado en el conocimiento y el dominio sobre la naturaleza. Las dificultades que van generándose con el paso del tiempo serán corregidas y mejoradas por medio de procedimientos técnicos. En ese sentido, la mejor actitud y más optimista consiste en la no interferencia en los asuntos técnicos, pues ya se encargarán de corregir los errores y solucionar los males los expertos en el campo. Los expertos, en este caso tecnólogos y científicos, deberían desarrollar su actividad con total libertad y sin limitaciones, ya que se entiende que son los propios conocimientos de sus campos de estudio los que ya les presentan limitaciones.

El comunicado de prensa del *Intergovernmental Panel on Climate Change* (IPCC) del 8 de octubre de 2018 advierte sobre las graves consecuencias que tendrá el cambio climático para la humanidad y la biosfera. La temperatura media de la atmósfera terrestre ha subido más de 1 grado desde mediados del siglo XIX como consecuencia de las emisiones de CO2 tras la sobreexplotación de combustibles fósiles. Los gobiernos del mundo deberían

emprender acciones para reducir estas emisiones a cero hasta 2050, pues de no ser así la temperatura ascendería hasta 1.5 grados y esto tendría consecuencias catastróficas para toda forma de vida. Este comunicado evidencia que el actual modelo tecnológico de desarrollo capitalista no está teniendo efectos favorables para la biosfera, por lo que dicho desarrollo, carente de crítica y caracterizado por la fe ciega en el progreso tecnológico, parece no ser del todo beneficioso para la vida, o al menos no tanto como se había creído desde la Ilustración.

Además de la versión optimista, existe una versión pesimista. Tras la degradación medioambiental, la sustitución de trabajadores por máquinas, el consumismo tecnológico y el gran poder destructor de la tecnología de guerra, hay quienes han elaborado un discurso muy crítico que pone en tela de juicio el desarrollo tecnológico. Entre esos pensadores se encuentran Martin Heidegger, Lewis Mumford o Jacques Ellul, que han formulado una crítica sobre el impacto que ha tenido la tecnología para el ser humano y cómo se han trasformado negativamente los valores de la humanidad. Atribuyen muchos de los males de la humanidad al hecho de que la tecnología haya fagocitado los sistemas humanos, provocando una falta de control.

La crítica a la tecnología moderna de Mumford se encuentra en su obra *El mito de la máquina*. El filósofo estadounidense distingue dos ramas de la tecnología: la politécnica y la monotécnica. La primera es considerada como la forma primordial de acción, orientada extensamente a la vida, no preocupada fundamentalmente del trabajo o el poder, compatible armoniosamente con la vida y sus manifestaciones, con los mecanismos democráticos y con el enriquecimiento del humanismo. A diferencia de la politécnica, la monotécnica se caracteriza por tener un carácter autoritario, centrado en el poder, cuyos objetivos se sitúan exclusivamente de desarrollo económico, productivo y militar. La monotécnica es predominante en este tiempo, aunque también lo ha sido en otras épocas donde el poder residía en regímenes autoritarios y esclavistas como los de la Antigüedad. A la rígida organización social donde predomina la monotécnica Mumford la denomina «megamáquina».

Jacques Ellul (2003; 2004) desarrolla una teoría marcada por un fuerte determinismo tecnológico. El filósofo francés afirma que el primer medio del humano fue el medio natural, después el social y en la actualidad se encuentra en un medio técnico, ya que la técnica se ha ido introduciendo en todas las esferas de la vida, fagocitando lo natural y lo social. El sistema técnico se presenta como un todo organizado donde los seres humanos son simplemente piezas de un engranaje. Entre las características del sistema Ellul (2004: 113) identifica las siguientes:

- Es artificial.
- Es autónomo respecto de los valores, las ideas y el Estado.
- Se determina a sí mismo en un círculo cerrado; al igual que la naturaleza es una estructura cerrada capaz de autodeterminarse con independencia de cualquier intervención humana.
- Crece según un proceso que es causal pero no está orientado a fines.
- Está formado por una acumulación de medios que han establecido su primacía frente a los fines.
- Todas sus partes están imbricadas hasta tal punto que resulta imposible separarlas o abordar cualquier problema técnico aisladamente.

Una vez caracterizado el sistema técnico, es importante destacar cuál es el papel del ser humano en el mismo. El ser humano ya no mantiene la misma relación con la técnica que la que mantenía en otros tiempos, como si se tratara de un mero instrumento. El sistema técnico representa un universo abarcante que escapa del dominio humano.

A Martin Heidegger ya se hizo referencia en el primer capítulo al esbozar su pensamiento en torno al humanismo. Es pertinente aclarar brevemente su visión sobre la técnica moderna en *La pregunta por la técnica*. El estudio heideggeriano sobre la técnica surge a raíz de la reflexión acerca de la relación entre el ser y el humano. La técnica moderna suscita una doble dimensión que es necesario presentar. En primer lugar, permite

revelar el destino y deseo del ser mismo provocando así una relación más originaria entre el humano y el ser, entendiéndose como un objeto. En segundo lugar, implica un peligro para la subsistencia del ser revelando su esencia por medio de un desocultamiento provocador de la naturaleza a la que considera un recurso de explotación y dominación. En la técnica moderna reside la *Gestell*, una estructura de emplazamiento que representa una actitud tecnológica hacia el mundo.

El pesimismo tecnológico suscita una actitud pasiva y desesperanzadora, algo que es contraproducente para los retos que depara el futuro. La estrategia del desaliento y el desánimo conducen a un callejón sin salida en la búsqueda de soluciones a los problemas. No obstante, el optimismo exacerbado tampoco es una solución, pues la esperanza ciega en el progreso tecnológico tampoco ha brindado buenos resultados. La actitud que se origina en el dogmatismo cientificista no parece representar una buena acción a la hora de realizar una operación crítica. Así pues, el optimismo exacerbado representa un envite demasiado arriesgado, mientras que el pesimismo es excesivamente desmovilizador. Elegir la estrategia que deja en manos de científicos y tecnólogos la solución a los problemas sería un error, ya que el momento histórico actual exige nuevas dinámicas de búsqueda de soluciones que involucren a más agentes. Diéguez se sitúa en esta línea:

> No es recomendable, pues, dejar por entero en manos de científicos y técnicos la solución de los problemas mencionados ni desesperar de toda solución. Por el contrario, deberíamos buscar lo que todos los miembros de la sociedad podemos hacer y hacerlo con urgencia. De hecho, hay razones para la esperanza [...] Ahora bien, sea lo que sea lo que podamos hacer y sean lo profundos que hayan de ser los cambios a realizar, éstos no deberían venir impuestos desde arriba por una estructura de poder central y autoritaria. Aun cuando resulte mucho más complicado, para ser efectivos y duraderos, los cambios deberían ser establecidos democráticamente (1993: 192-193).

En medio del huracán del progreso tecnológico se disimulan ciertas exigencias bajo el valor de lo novedoso. Y la novedad cuenta con más ventaja si se encuentra fundamentada en la ciencia, tanto es así que cualquier exigencia crítica suele brillar por su ausencia. Lo verdaderamente importante en este tiempo no parece ser la formulación de críticas al

ejercicio tecnológico, sino más bien la asunción de cualquier invención, pues existe la creencia habitual de que siempre tendrá impactos positivos la vida. La apertura de posibilidades y nuevos horizontes que brinda el progreso tecnológico es lo exclusivamente importante; al margen quedan los cuestionamientos, las críticas y las dudas. Esto provoca que la tarea de búsqueda de responsabilidades sea muy compleja, ya que los supuestos responsables de una cadena de decisiones siempre dirigen la carga a otra instancia, ya sea superior o inferior. El progreso tecnológico ha generado un escenario en el que la exigencia de responsabilidades se diluye en medio del huracán, como sostiene Diéguez (1993: 194). A su vez, este autor señala cuatro posibles causas a la hora de identificar con dificultad la responsabilidad en el ámbito tecnológico (1993: 194-195).

Los sistemas técnicos están constituidos por un complejo de redes en el que la toma de decisiones es sometida a infinidad de procesos donde en ocasiones se torna compleja la aplicación del principio de causalidad. La toma de decisiones se encuentra incrustada en los procesos y eso dificulta la exigencia de responsabilidad a las partes. Además, la complejidad que caracteriza a las acciones técnicas también se ve fundada en el trabajo en equipo, donde un conjunto de personas organiza las acciones que van a emprender dentro de la planificación de un proyecto. En ese sentido, las acciones realizadas son fruto del trabajo de redes de grupos de trabajo colectivo que cada vez amplían con mayor fuerza su espectro; por ejemplo, grupos de investigación, empresas financiadoras, departamentos de investigación de instituciones académicas y organismos gubernamentales, pueden ser algunas partes que intervengan en el diseño y puesta en marcha de un determinado proyecto. Esta ampliación del campo de trabajo obstaculiza la exigencia de responsabilidad, pues a veces es difícil identificar en qué grupo recae la imputación.

Las sociedades líquidas, en términos de Zygmunt Bauman (2017), caracterizadas por su complejidad y continuos cambios políticos, económicos, sociales y medioambientales, hacen difícil prever las consecuencias del uso de determinadas tecnologías. El impacto que tiene la tecnología es difícil de medir y pronosticar, ya que los diversos factores que inciden en su despliegue suelen ser variables y depender de los contextos. Otra de las causas que impide el reconocimiento de responsabilidad en el campo de la tecnología es la tecnocracia,

es decir, ese momento en el que los fines se imponen por la propia técnica y cualquier cuestionamiento por el fundamento pierde todo sentido y razón de ser. El sistema técnico y sus tentáculos son el medio de despliegue de la vida y las otras esferas, como la política, y ejercen una tarea auxiliar de mera gestión. Siguiendo a Diéguez, puede entenderse que esta tecnocracia se apoya en un presupuesto dogmático implícito en la actividad tecnológica que viene desarrollándose durante toda la modernidad, a saber, el cientificismo.

Estos son los puntos que Diéguez identifica como las principales causas de dificultad para la introducción de parámetros de responsabilidad en el campo tecnológico. Sin embargo, existe otra causa. La racionalidad instrumental representa un modo de pensar que disimula y oculta los presupuestos dogmáticos de la tecnología y eso obstaculiza la asunción de responsabilidad. En ese sentido, una breve exposición de la crítica que Max Horkheimer y Theodor W. Adorno vierten sobre la modernidad en su *Dialéctica de la Ilustración* (2016), y posteriormente Horkheimer en la *Crítica de la razón instrumental* (1973), podría contribuir esclarecer este tema.

Horkheimer y Adorno reflexionan sobre la ambigüedad subyacente al proyecto ilustrado. El proyecto de la Ilustración y los valores que la misma defendía se ponen en tela de juicio y a la vez influyen sobre la configuración de la identidad cultural. Ambos pensadores reivindican una defensa de esos valores, pues, según ellos, se están convirtiendo en mitología. La Ilustración y el pensamiento racional que en ella se gesta proporcionan las condiciones para que exista una dialéctica que gira en torno a dos direcciones: por un lado, la racionalidad es importante para la emancipación, pero también, por otro lado, ha dado lugar a la fundación de sociedades totalitarias. Lo que en un inicio era un proyecto de potenciación de la racionalidad ha desembocado en una razón instrumental que apuntala los cimientos de un sistema de dominación.

Los pensadores de la Escuela de Frankfurt hacen su diagnóstico sobre los procesos de racionalización que se han dado en la modernidad, es decir, sobre los avances de la técnica y la ciencia que señalan un peligro. La pérdida de reflexividad de la praxis, ya sea social o individual, está conduciendo a una absolutización de la razón instrumental que sirve para

sostener una reproducción material propia del sistema capitalista. Esa pérdida de carácter reflexivo de la praxis conduce a un importante desafío que requiere de un nuevo cultivo de la conciencia social que debe caminar hacia la autoconservación. Los efectos de la racionalidad instrumental en el sistema capitalista no son únicamente económicos, sino que también involucran aspectos humanos y medioambientales que deben ser abordados desde una racionalidad que tome distancia de la instrumentalización. Es fundamental una revisión crítica de los pilares instrumentales sobre los que se sostiene el sistema político y económico. La convivencia de la razón objetiva y subjetiva, esto es, de los componentes reflexivo e instrumental de la razón, ha estado presente a lo largo de la historia de la humanidad, aunque se ha producido una escisión a causa de la instrumentación de la razón. Esta racionalidad se ha impuesto en el mundo de la vida por medio de diversas estrategias para dominar y diseñar habilidades de control. Ha servido para dominar la naturaleza y organizar la barbarie. Además, en el contexto de la reflexión que en este trabajo se presenta, esta racionalidad es capaz de disimular y ocultar los presupuestos dogmáticos sobre los que se erigen los proyectos de la IA, dificultando así la identificación de los mismos y de los desafíos que deben enfrentarse en diversos ámbitos.

Más allá de los obstáculos que recientemente han sido expuestos para incorporar criterios de responsabilidad en el ámbito tecnológico, surgen motivos razonables para pensar que frente a los desafíos a los que está conduciendo la tecnología no debería responderse con indiferencia. Existe un imperativo ético que empuja a la transformación del obrar humano y a la asunción de responsabilidad frente a dichos desafíos. Es importante pensar en los posibles efectos dañinos que ciertas tecnologías pueden implicar, lo que representaría una preocupación anticipada y sostenible en el tiempo que serviría para introducir una actitud de cautela en toda actividad. Además, y tomando como premisa fundamental el humanismo tecnológico, existe la obligación de reorientar la actividad tecnológica hacia fines que sean estrictamente éticos y armoniosos con el enriquecimiento de la condición humana y el respeto a la biosfera. No obstante, ninguna incorporación de criterios de responsabilidad será posible si antes no existe una formación pertinente que sirva para concienciar del imperativo ético. La educación juega un papel importante, ya que

es necesario proporcionar la información suficiente y necesaria para que la ciudadanía adquiera un papel protagónico ante la necesidad de un cambio en la concepción tecnológica.

5.2. La tecnología y su responsabilidad democrática

La IA puede ser impulsada para diversos propósitos, pues su dominio de acción es muy amplio. Existe un abanico de posibilidades dentro del campo de la IA y por tanto infinidad de usos y finalidades, unos más benignos que otros. Cuando las tecnologías son cooptadas por ciertos actores que promueven acciones incompatibles con la democracia y los derechos humanos es importante reflexionar sobre las implicaciones que pueden presentar determinados mecanismos, pensando en la necesidad de asumir responsabilidad y por lo tanto introduciendo principios éticos que sirvan para orientar la acción.

En los últimos años la tecnología cognitiva (TC) ha visto incrementado el nivel de nuevos resultados a raíz de importantes investigaciones. Se está madurando un escenario de optimización, integración y reconfiguración de herramientas tecnológicas, técnicas y metodologías de análisis que sin duda presentarán importantes sorpresas a corto y medio plazo.

El término «cognición« remonta su origen al latín *cognoscere*, que en español significa conocer. Conocer es la capacidad que tiene un ser para recibir información, seguidamente procesarla y finalmente acceder a la obtención de un conocimiento adicional fruto de la experiencia. Adicionalmente a esto, el ser humano es capaz de agregar un grado de subjetividad al conocimiento que ha podido generar, lo que le permite realizar una valoración del mismo. La tecnología cognitiva hace referencia a ese conjunto de mecanismos tecnológicos que sirven para ayudar, mejorar o simular los procesos cognitivos que empleamos los seres humanos (Dascal y Dror, 2005). El término «tecnología cognitiva» comenzó a utilizarse en el campo de la psicología educativa para referirse a aquellas estrategias y herramientas que servían para facilitar y fortalecer los procesos cognitivos, sobre todo centrado en el aprendizaje y la resolución de problemas. A finales de

los 90 se creó la Sociedad de Tecnología Cognitiva (Walker y Herrmann, 2004) y posteriormente se organizaron diversos encuentros para discutir las implicaciones sociales y cognitivas que estas tecnologías tendrían en los humanos. Dentro del extenso campo de la TC pueden encontrarse dos subcampos que comparten el uso de la TC con el propósito de incidir, mediante la ayuda o el mejoramiento, en las capacidades cognitivas y motoras humanas. Sin embargo, ambos subcampos difieren entre sí, a pesar de que en los últimos años han protagonizado una interesante convergencia:

- Neurotecnologías: consisten en interfaces cerebro-computadora (*brain-computer interfaces*) que persiguen ayudar, mejorar o controlar procesos cognitivos que son naturales. Establecen vías de conexión entre el sistema nervioso y sistemas de computación. Esta tecnología está ayudando a pacientes que presentan daños cerebrales, distrofias musculares, etc., es decir, a aquellos que experimentan deficiencias cognitivas y/o sensoriales-motoras (García Linares, 2012). Las neurotecnologías están siendo objeto de discusión, en particular las implicaciones éticas que tiene su aplicación en el campo de la educación y el empleo (Rafael Yuste *et al*., 2017).

- Sistemas de IA: estos sistemas se centran en simular la inteligencia natural por medio de un amplio espectro de procesos que recogen razonamiento, aprendizaje, planificación, percepción, procesamiento de lenguaje natural y capacidad para manipular objetos físicos en el espacio. Un claro ejemplo es el sistema informático Watson de IBM.

El campo de la tecnología en general, y el de la IA y las neurotecnologías en particular, ha despertado un gran interés en algunos expertos del ámbito de la neuroética como Sara Goering y Rafael Yuste (2016), y también de la ética de la computación y la información como Luciano Floridi (2001; 2002a; 2002b; 2003; 2005; 2006a; 2006b; 2007a; 2007b). Estos y otros autores reconocen que cuanto más se amplía el poder de estas tecnologías mayores son las implicaciones éticas que existen en el terreno social y cognitivo. El desafío de las tecnologías avanzadas reside en que sepan desplegarse de forma compatible y

respetuosa con los sistemas democráticos y todos los principios éticos que en ellos están contenidos. En ese sentido, a continuación es oportuno observar el fenómeno de la IA desde una óptica democrática, con el objetivo de poner de relieve qué controversias éticas existen y qué premisas deben tenerse en cuenta para el planteamiento de una IAR.

Si los temas de la ética y la seguridad se toman como punto de partida, la IA presenta algunas implicaciones que generan ciertas controversias como la singularidad, el mejoramiento de la especie, la automatización del campo laboral, el uso político y militar, la agencia moral, etc. Estas controversias surgen al reconocer el doble uso que puede hacerse de la tecnología, diferente del propósito para el que fue creada. Por ejemplo, un teléfono móvil fue diseñado para fortalecer el campo de las telecomunicaciones, pero también puede ser utilizado por unos terroristas como detonador de una bomba. John Forge aborda esta cuestión y señala que la determinación del doble uso de un artefacto depende del contexto en el que se encuentra inmerso:

> La clasificación de algo como de uso dual no debe consistir simplemente en que el elemento podría tener un mal uso. Si esto fuera así, entonces la categoría de doble uso sería increíblemente grande y los objetivos finales de control y regulación inalcanzables. La pregunta se centra en cómo definir el uso dual para restringir su membresía. Esto debe hacerse teniendo en cuenta la historia o los factores contextuales, el tiempo y el lugar (2010: 116).

Cuando existen sistemas de IA que involucran propósitos para los que no fueron concebidos, se pone de relieve un problema de carácter ético, ya que existen fines éticamente cuestionables. Este cuestionamiento se plantea al considerar que la IA que se utiliza para un uso diferente, puede presentar riesgos que tienen impactos de diversa índole para los que no existen protocolos. No obstante, hay que reconocer que existen múltiples ejemplos que demuestran que algunas tecnologías que fueron promovidas para uso militar han sido empleadas con éxito en el ámbito civil. Además, a esto hay que sumarle la aparición de un nuevo escenario para cometer delitos: el ciberespacio. El siguiente cuadro pone en evidencia el aumento considerable de la cibercriminalidad en el Estado español:

TOTAL NACIONAL	2017	2016	2015	2014
ACCESO E INTERCEPTACIÓN ILÍCITA	2.505	2.579	2.386	1.851
AMENAZAS Y COACCIONES	11.270	11.473	10.112	9.559
CONTRA EL HONOR	1.537	1.524	2.131	2.212
CONTRA LA PROPIEDAD INDUSTRIAL/INTELECTUAL	109	121	167	183
DELITOS SEXUALES	1.312	1.188	1.233	974
FALSIFICACIÓN INFORMÁTICA	2.961	2.697	2.361	1.874
FRAUDE INFORMÁTICO	60.511	45.894	40.864	32.842
INTERFERENCIA EN LOS DATOS Y EN EL SISTEMA	1.102	1.110	900	440
TOTAL grupo penal	81.307	66.586	60.154	49.935

Fuente: Ministerio del Interior de España, 2018.

La tecnología más avanzada ha erosionado los principios más básicos de la democracia. Esto plantea el importante interrogante sobre si los principios y valores democráticos resistirán la era tecnológica o necesitarán actualizarse en este nuevo tiempo. Son varias las posibilidades de erosión de la democracia, pues las causas no radican exclusivamente en el ciberterrorismo o la ciberguerra, sino que existen casos de violación de la privacidad con un extenso control y manejo de datos confidenciales, vigilancia masiva, como la impulsada por el Gobierno de China desde 2015 con el programa *Sky Net* o en Rusia, con el Servicio Federal de Seguridad, que tiene el poder de acceder a los datos de búsqueda de internet en las redes sociales. Los ejemplos de control gubernamental de carácter invasivo son una clara muestra de cómo la IA puede atentar sobre los principios más básicos de la democracia si no se toman medidas. Algunos líderes políticos como el presidente de Francia, Emmanuel Macron, son conscientes de esta problemática y así lo recoge James Vicent (2018) en una noticia del medio *The Verge*. Macron siente preocupación por el papel que los algoritmos están jugando en la sociedad con determinadas decisiones que normalmente estaban encargadas a humanos. En ese sentido, propone medidas de transparencia para que la ciudadanía tenga conocimiento de las implicaciones que presenta la innovación tecnológica y que de ese modo el conocimiento no sea exclusivo de las empresas del sector, sino también de los organismos gubernamentales y la sociedad civil.

Es evidente que el progreso de la IA está provocando una revolución política que tendrá un fuerte impacto en el seno de las sociedades democráticas a corto y medio plazo. Por ello es necesario contextualizar la democracia en este nuevo escenario, incorporando la tecnología avanzada como un medio de fortalecimiento democrático. En 2006 se organizó un taller en la Universidad Estatal de Arizona con el propósito de abordar la capacidad de cambio sociocultural que tiene la TC. En dicho taller se reconoció que la TC tiene un gran potencial de influencia sobre la inteligencia y las capacidades cognitivas del ser humano, potencial que en este caso podría provocar efectos desestabilizadores. El resultado de ese encuentro fue la publicación de un documento por parte de los *Sandia National Laboratories*, titulado *Policy Implications of Technologies for Cognitive Enhancement*, cuyos autores son Daniel Sarewitz y Thomas H. Karas. En ese documento se identifican cuatro perspectivas sobre las tecnologías del mejoramiento y su impacto político: *laissez-faire*, optimismo tecnológico, escepticismo tecnológico y esencialismo humano.

- *Laissez-faire*: esta perspectiva es sostenida sobre la defensa de la libertad individual, alejada de los controles gubernamentales y de cualquier otra instancia que pudiera ejercer control. La libertad tecnológica es producto de la expresión individual. El mercado es la única entidad capaz de regular las expresiones tecnológicas por medio de la competencia. Los gobiernos tienen la responsabilidad de promover la libertad creativa para la producción tecnológica.

- Optimismo tecnológico: esta perspectiva se basa en el ideal ilustrado de la innovación técnica y científica como clave del progreso humano. Si bien esta visión es consciente del gran aporte de la tecnología, también lo es de las problemáticas que se presentarían. Los entes gubernamentales tienen la responsabilidad de asegurar las condiciones para la innovación tecnológica, aunque es en el sector privado donde recae la actividad productiva. Además, el gobierno debe favorecer un discurso abierto e inclusivo, especialmente destinado a aquellos sectores más desfavorecidos tecnológicamente hablando. En este sentido, se reconoce el importante aspecto que la tecnología tiene para la mejora humana, pero se asume la

necesidad de que los gobiernos tengan un rol protagónico en la promoción de la innovación y en la gestión responsable del riesgo.

- Escepticismo tecnológico: desde esta perspectiva se reconoce la aportación que la tecnología tiene para la sociedad y la democracia, pero se duda de que la tecnología sea inherentemente beneficiosa. En ese sentido, se afirma que los desequilibrios generados no pueden ser solucionados por la propia lógica tecnológica, sino a través de la acción política. Los escépticos se sitúan del lado de la regulación y la moderación del poder para minimizar los riesgos. Además, desde esta perspectiva se afirma que el control de la producción tecnológica reside en una élite que por medio de la mercantilización podría generar desequilibrios sociales. Esta es una visión regulacionista que entiende que las entidades de control gubernamental deben proporcionar los mecanismos políticos necesarios para estudiar el impacto y la investigación en el campo tecnológico. Entre las medidas más importantes se encuentran las siguientes:

  • Un período de varios años de reflexión pública y nacional, de discusión sobre el mejoramiento cognitivo antes de realizar nuevos o mayores compromisos de I + D + I.

  • Creación de un programa permanente para investigar las implicaciones sociales de la mejora cognitiva.

  • Creación de un cuerpo analítico independiente, tal vez análogo a la antigua Oficina de Evaluación Tecnológica, para proporcionar evaluaciones detalladas y expertas de los impactos sociales de la gama de posibles tecnologías de mejora cognitiva.

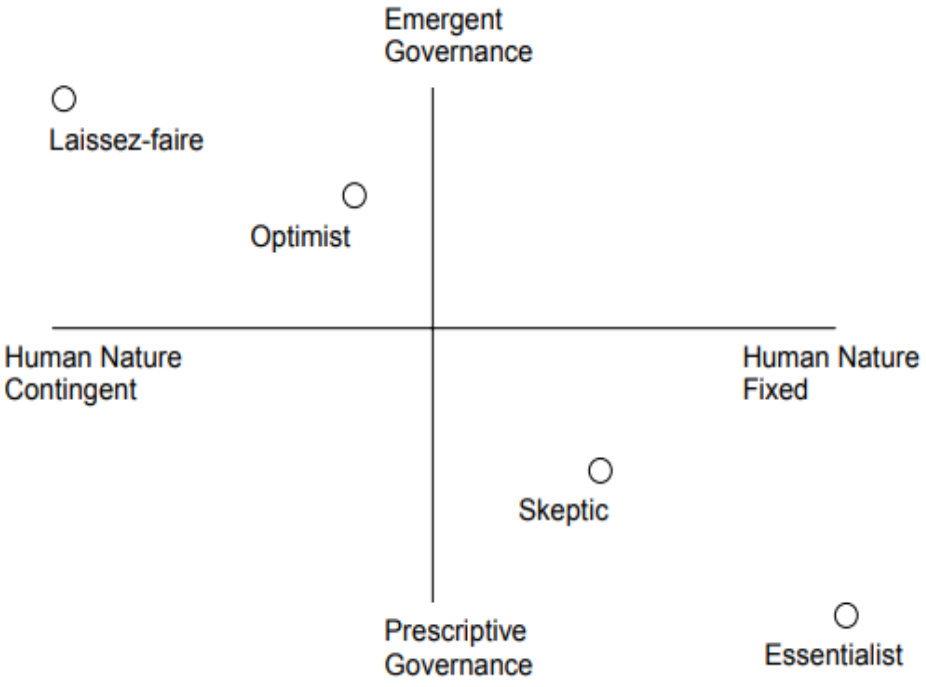
  • Reducción de la financiación para la investigación de la mejora cognitiva con aplicaciones militares directas.

  • Una regulación y supervisión más estrictas de la investigación en sujetos humanos sobre el mejoramiento cognitivo.

> • Una supervisión independiente más sólida de los ensayos clínicos de fase II y III de la FDA; fortalecimiento de publicaciones de la fase IV sobre efectos secundarios.
>
> • Requerir que los solicitantes para financiamiento federal de investigación de mejoramiento cognitivo incluyan a) un análisis serio de riesgos potenciales y desventajas; y b) una base analíticamente fundamentada para cualquier reclamo de prestaciones sociales.
>
> • Desarrollo de acuerdos internacionales de gobernanza para prevenir la explotación de los países en desarrollo o la estratificación cognitiva internacional que podría obstaculizar aún más el desarrollo de países pobres (Sarewitz y Karas, 2007: 18).

- Esencialismo humano: esta es otra de las perspectivas que tratan de valorar el fenómeno de las TC. Esta perspectiva se sostiene sobre los conceptos de dignidad y esencia humana. La interferencia de las tecnologías de mejora sobre los seres humanos representaría una amenaza para ambos conceptos. El conservadurismo estadounidense y el liberalismo europeo representan los fundamentos ideológicos de esta perspectiva. El ser humano está limitado por su naturaleza y no es aceptado ningún intento de mejora procedente de la tecnología, salvo en el caso de los discapacitados, que será regulada. El contexto de mejora de los individuos tiene lugar en la familia y la comunidad. Dios ha proporcionado unos atributos naturales que no deben ser modificados.

En el mismo documento se presenta un gráfico que ilustra el lugar que ocuparía cada perspectiva:



Fuente: Sarewitz y Karas, 2007: 20.

Más allá de estas cuatro posiciones, puede existir una quinta vía para dar forma al proyecto de una IAR dentro del contexto de la democracia. Entre las cuatro perspectivas, la que podría sugerir una mayor aceptación es el escepticismo tecnológico, sobre todo por insistir en la necesidad de las regulaciones gubernamentales. Sin embargo, una excesiva regulación podría ocasionar una erosión de los principios democráticos más básicos. Además de los posibles excesos de regulación, la perspectiva escéptica quizá adoptaría en un determinado momento posturas cercanas al esencialismo para justificar ciertas regulaciones, representando así un claro peligro. En ese sentido, es necesaria una quinta perspectiva que encuentre su fundamento principalmente en el humanismo tecnológico, en el principio de responsabilidad y en otros postulados filosóficos que sirvan de guía para la política democrática en este tiempo tecnológico.

Una de las ideas que más se ha comentado aquí consiste en que la democracia debe someterse a un proceso de contextualización dentro del escenario tecnológico y para ello es importante el reconocimiento de nuevos aspectos. A propósito de la contextualización de la

democracia dentro de un escenario tecnológico de aumento de posibilidades, Miguel Ángel Quintanilla diferencia entre una democracia tecnológica mínima y una democracia tecnológica plena, dentro de las cuales identifica varios puntos cardinales (2002: 640-647). En el interior de la democracia tecnológica mínima se identifican los siguientes aspectos:

- Derecho de todos los ciudadanos a participar en las decisiones sobre el uso de las posibilidades tecnológicas en asuntos de interés público.

- Derecho de todos los ciudadanos de acceder al conocimiento técnico y de contar con el juicio de los expertos como elemento fundamental para conformar la opinión pública y participar en las decisiones políticas sobre asuntos técnicamente complejos.

- No convertir en problemas políticos aquellos asuntos para los que existen soluciones técnicas solventes y contrastadas.

En cuanto a la democracia tecnológica plena, destaca principalmente el siguiente aspecto:

- El derecho de todos los ciudadanos a acceder a todo el conocimiento tecnológico relevante para la toma de decisiones en asuntos de interés público y a participar en el diseño, evaluación y control del desarrollo tecnológico.

Como se ha señalado, existe una necesidad imperiosa de reflexionar sobre la democracia dentro del nuevo contexto tecnológico. La tecnología más avanzada, la IA, brinda la oportunidad de aumentar las posibilidades de acción sobre diversos ámbitos. Así pues, las acciones tecnológicas deben encaminarse hacia el terreno democrático asumiendo principios para la orientación y generación de nuevas dinámicas participativas con el fin de que tales acciones cuenten con un mayor soporte y legitimidad.

Las posibilidades que plantea el desarrollo de la IA conducen a un escenario político que demanda un nuevo contrato social en el que la tecnología promueva el bienestar social y sean asumidas nuevas dinámicas participativas en el seno de las sociedades. La

participación de los sectores implicados emerge como la punta de lanza en este nuevo escenario, donde lo político es entendido como el espacio de participación de la toma de decisiones más cruciales sobre el desarrollo tecnológico. El modelo basado en la experticia, donde la toma de decisiones reside exclusivamente en los científicos, tecnólogos, y en los monopolios tecnológicos, debería pasar a la historia, pues es difícil entender un futuro de profundo desarrollo de la IA con importantes implicaciones sin un carácter cívico y democrático. La acción democrática tendrá que ser fruto de un trabajo participativo y una orientación basada en principios éticos fundamentados en la contextualización del nuevo escenario tecnológico. En ese sentido, el desarrollo de una ciencia cívica podría favorecer un escenario de posibilidades participativas y democratizadoras en el ámbito de la acción tecnológica.

5.3. La problemática de la moralidad

La IA está cada vez más presente en las vidas humanas, lo que supone que forme parte del complejo de actividades y relaciones morales en las que se desenvuelven las personas. Los sistemas artificiales recopilan información de los sectores financieros y de telecomunicaciones, ensamblan automóviles y otros productos en la industria, vigilan en aeropuertos y han pasado a hacer algunas labores que desempeñan los humanos, pero también otras de las que no son capaces debido a las limitaciones cognitivas y motoras, en definitiva, están presentes en el ámbito moral de la vida. Los intelectos sintéticos llevan a cabo acciones que tienen una implicación moral, como la decisión a la hora de elegir un candidato para un puesto de trabajo en función de las preferencias de la empresa o también la decisión de despedir a un trabajador de una empresa. Sus acciones tienen un impacto moral en la vida. Incluso existen casos en los que los algoritmos han llevado a cabo acciones racistas (Hamilton, Karahalios y Langbort, 2016). Algunos sistemas artificiales se están enfrentando a situaciones que requieren de decisiones morales para las que no están preparados. Por lo tanto, si la IA está presente en la vida de los seres humanos, y ésta implica la toma de decisiones morales, se sitúa frente a una problemática que está siendo objeto de una profunda discusión en el seno de los principales centros de investigación en

la materia, como el MIT o el MIRI (*Machine Intelligence Research Institute*). Cuanto más se avanza en el campo de la IA y aumentan las áreas para su aplicación, mayores son los contextos morales de los que forma parte.

Un caso ejemplar es el que ha podido observarse en los últimos años con la investigación en el área de los vehículos autónomos. Empresas como *Cruise, Zoos, Pony.ai* o *Argo AI* están promoviendo la autonomía de los vehículos como un medio para mejorar la seguridad del transporte en las carreteras. Si bien es cierto que los sistemas de autonomía están mejorando, de momento siguen siendo menos capaces que los conductores humanos para conducir en situaciones complejas. El modelo *Human-Automation System Oversight* (HASO) está proporcionando mecanismos de orientación en el diseño de la autonomía de los vehículos semiautónomos. El 18 de marzo de 2018 un vehículo autónomo de *Uber* atropelló a una mujer en Arizona (Estados Unidos) provocándole la muerte. A partir de ese momento la empresa *Uber* paralizó todas las pruebas de sus vehículos autónomos. No obstante, este no fue el primer incidente ocasionado por un vehículo autónomo, ya que en el año 2016 el *Tesla Model S* tuvo un accidente en el que murió un hombre. Estos son claros ejemplos que ponen en evidencia los desafíos éticos a los que se enfrenta la humanidad con el avance de los intelectos sintéticos. En el caso de los vehículos autónomos, una propuesta para mejorar la seguridad consiste en la potenciación de la conectividad entre los vehículos autónomos y los servidores que en caso de dificultad para la conducción facilitarán datos relevantes para el desenvolvimiento (Uhlemann, 2018).

Así pues, el aumento de la relevancia en los contextos morales por parte de los sistemas artificiales es más que evidente, motivo por el que es necesario discutir este desafío desde una perspectiva de responsabilidad teniendo presente una IAR. En ese sentido, es importante exponer dos posturas en torno a la cuestión de la responsabilidad de la IA, a saber, la posibilidad de una inteligencia moral artificial (IMA) y la incorporación de criterios de responsabilidad en la *praxis* humana que orienta la actividad tecnológica. Para que el impacto de la IA sea positivo es fundamental realizar una buena planificación estratégica sin pretensiones de generar alarma social, sino más bien con una ingeniería de seguridad como la planteada por el MIT. Esta ingeniería estudia los hipotéticos escenarios

en los que una IA podría tener un impacto negativo. Para el MIT la IA supone una preocupación en materia de seguridad nacional y así lo ha destacado James E. Baker, expresidente de la *United States Court of Appeals for the Armed Forces*, en un artículo para el *Starr Forum Report*, donde alerta sobre el inminente peligro de una carrera militar ante los anuncios de China y Rusia de incrementar su poder en materia de IA. Por lo tanto, es necesario pensar en los riesgos morales que conlleva el desarrollo de los sistemas artificiales.

Con el aumento de la autonomía de los intelectos sintéticos cada vez se torna más difícil el establecimiento de la frontera de la responsabilidad entre el creador y el producto. El caso de *AlphaZero*, el sistema artificial que se ha convertido en maestro del ajedrez, donde sus programadores no pudieron suministrar todas las instrucciones posibles, pone de relieve que los resultados no siempre son predecibles. Existe una dificultad para determinar con precisión la contribución de los programadores y los resultados que son propios de los sistemas artificiales, es decir, entre la agencia moral y la responsabilidad.

El campo de la IA está dedicando grandes esfuerzos a la investigación de la estructura cognitiva humana para tratar de emular la dimensión moral en los intelectos sintéticos. Si este proyecto de introducción de la capacidad moral en los intelectos sintéticos se materializa, tendría un gran impacto en las sociedades, según los defensores de esta posibilidad. Este fenómeno no está exento de problemáticas, ya que surgen algunos cuestionamientos en torno a la pertinencia de considerar a los sistemas artificiales como agentes morales artificiales (AMA) y también a la dificultad de redefinir la moralidad, que tradicionalmente ha sido entendida desde el terreno humano, en un hipotético escenario de artificialidad. Los AMA se encuentran dentro de un paradigma denominado como «paradigma de agentes», que se centra en el desarrollo de entidades capaces de actuar de forma autónoma y razonada (Botti y Julián, 2000). Vicente Botti y Vicente Julián sostienen que existe dificultad a la hora de determinar una definición de agente plenamente aceptada por la comunidad científica (2000: 96), aunque se refieren al concepto de agente que Stuart Russel y Peter Norvig (1996) formulan como una entidad que percibe y actúa sobre un entorno determinado.

El ámbito de aplicación de la moralidad humana es mucho más amplio que el de la IA. Los contextos en los que se despliegan las acciones de los intelectos sintéticos que tienen implicaciones morales son mucho más restringidos y limitados. Los factores que influyen en la configuración de la moral humana son muy complejos e incluyen creencias morales, prejuicios, actitud intencional, conciencia, libre albedrío, capacidad de reflexión y justificación, etc. En cambio, las estructuras cognitivas de los intelectos sintéticos, al menos por el momento, no incluyen esos componentes, aunque muchas de sus acciones tengan un impacto en el ámbito moral de los seres humanos. Se han llevado a cabo intentos para suministrar estados funcionalmente similares a emociones como la culpa o la pena a los sistemas artificiales, pero no se ha conseguido ir más allá de la asignación de parámetros de calidad a estas emociones (Arkin y Ulam, 2009). Hay algunos aspectos de los sistemas artificiales que tratan de emular la conciencia moral, como es el valor, pero solamente representan premisas en el procesamiento que conducen a la realización de la acción. En cambio, la moralidad incluye la capacidad de deliberar sobre la responsabilidad. Los intelectos sintéticos no son capaces de reflexionar ni de justificar sus acciones del mismo modo que lo hacen los seres humanos dada su condición moral. Catrin Misselhorn define la equivalencia moral entre los seres humanos y sistemas artificiales de la siguiente manera:

> Los sistemas artificiales actuales no son funcionalmente equivalentes a una moral humana en toda regla. Su nivel de sofisticación podría ser comparada con un niño pequeño actuando por motivos morales según normas que los padres le contaron sin una deliberación previa sobre los motivos. E incluso un niño pequeño tiene, por supuesto, una gama más rica de actitudes intencionales como el miedo, la esperanza, etc., que un sistema artificial. Los niños poseen, además, conciencia fenomenal, simpatía natural por los demás y un libre albedrío al menos en el sentido que pueden negarse a hacerlo que sus padres les dicen que hagan (2018:164-165).

La agencia moral se presenta como una cuestión central en la exigencia de responsabilidad, y en este ámbito surgen dos planteamientos. En primer lugar, cuáles son las características que deben reunirse para que algo sea considerado como un agente moral; y, en segundo lugar, qué es lo que hace que a un agente se le pueda considerar con el atributo de «moral», es decir, cuál es la propiedad moralmente relevante. El debate sobre la

posición moral ha ocupado un lugar importante en el ámbito de la ética animalista con la discusión entre Tom Regan (1983) y Peter Singer (1999). El estado moral sería producto del siguiente razonamiento *modus ponendo ponens*:

La entidad x tiene una propiedad p.

Cualquier entidad que tenga la propiedad p tiene un estado moral.

La entidad x tiene un estado moral.

Este razonamiento ha servido a lo largo de la historia de la humanidad para reconocer el derecho al voto a las mujeres, para legalizar el matrimonio entre personas del mismo sexo, para liberar a los esclavos, etc. Sin embargo, es muy limitado en el campo de la IA, ya que se estaría reconociendo la existencia de un presupuesto ontológico cerrado en el que deberían estar incorporadas determinadas propiedades morales, algo que todavía no es posible. Además, esta visión forma parte de un paradigma moral dominante de fundamento cartesiano.

El dualismo cartesiano establece una frontera ontológica entre el alma y el cuerpo humano, los animales y los objetos, donde entraría la IA. Esta división cartesiana dificulta la posibilidad de dar sentido a las experiencias que no entran dentro de determinadas categorías conceptuales e impone un mecanismo de exclusión. Por lo tanto, y para abordar estas cuestiones desde un marco ético alternativo, es necesario superar la epistemología moral cartesiana. Para Mark Coeckelbergh (2014: 73) solo es posible superar la epistemología moral cartesiana de dos formas: por medio del rechazo de la distinción radical entre los humanos y los intelectos sintéticos, con el propósito de abrir un espacio que permita articular diferentes tipos de experiencia moral que no deberían ser impuestas a priori, y reconociendo la importancia de una fenomenología concreta que ofrece posibilidades epistémicas con un espectro más amplio. El modelo cartesiano ha reprimido la experiencia humana y las posibilidades relacionales que permiten conocer mejor el ámbito de la moralidad. Por lo tanto, ante los desafíos de la IA es necesario reflexionar acerca de las relaciones concretas entre los humanos y los sistemas artificiales, buscando

una posición más allá de los límites del pensamiento ético de la modernidad, fundamentado por la epistemología moral cartesiana, sin el temor de desplazar las coordenadas de la ética. Según Coeckelbergh es ineludible tomar distancia del egocentrismo epistemológico, moral y solipsista (2014: 74). La problemática sobre la agencia moral ya no giraría en torno a las propiedades ontológicas que tendría que reunir el robot, sino a qué experiencias se tienen en el encuentro con estos sistemas inteligentes y cómo se establece la relación con los mismos. La propuesta de Coeckelbergh se encuentra en el terreno de una hermenéutica abierta a un área más amplia de la experiencia y el conocimiento.

5.3.1. La inteligencia moral artificial (IMA)

Dentro del terreno de la IMA existen dos enfoques para implementar la toma de decisiones morales en sistemas artificiales, de arriba abajo (*top-down*) y de abajo arriba (*bottom-up*) (Allen, Smit y Wallach, 2005). El enfoque de arriba abajo se centra en la posibilidad de incorporar teorías morales en el AMA para que oriente las acciones. Este enfoque es criticado por la dificultad para proporcionar una teoría que sea general y que por lo tanto sirva para enfrentar cualquier situación. Según este enfoque, sería posible la traducción de principios religiosos o éticos como los deontológicos o utilitaristas, por medio de algoritmos programados en los intelectos sintéticos. Un claro ejemplo de enfoque de arriba abajo podrían ser las tres leyes de Asimov. Por otro lado se encuentra el enfoque de abajo arriba, que entiende que lo importante no es proporcionar teorías morales para orientar la acción, sino más bien entornos en los que se produzca un aprendizaje que genere comportamientos apropiados, algo que en la actualidad se conoce como *deep learning.* La experiencia es aquí una pieza fundamental para el aprendizaje de los intelectos sintéticos, aunque es un proceso muy lento. *Alife*, una plataforma para los experimentos de vida artificial, está desarrollando actividades que son muy prometedoras, pues se basan en la evolución de algoritmos genéticos en entornos informativos relativamente simples. Sin embargo, los AMA generados en este proyecto son demasiado simples y distan mucho de la reflexividad moral que tiene el ser humano.

Dentro del enfoque de arriba abajo Bostrom formula una reflexión sobre la cuestión moral de los valores en el entorno de una superinteligencia. Entiende que es difícil universalizar la toma de decisiones para cada una de las situaciones a las que un intelecto sintético podría enfrentarse dada la complejidad de la realidad. Ante esta situación de complejidad el filósofo sueco propone una función de utilidad (2016: 185-186) en la que es asignado un valor a cada resultando obtenido, de modo que el agente artificial se decantaría por la acción que le reporta un valor de utilidad más alto. Sin embargo, este marco teórico presenta una dificultad, a saber, la transformación del lenguaje natural, y todo lo que ello implica, a un lenguaje de programación propio de la IA. El problema es que el lenguaje empleado para la programación, en este caso matemático y lógico, no puede representar términos primitivos como el de «felicidad», ocasionando que este intento de incorporar valores a través del lenguaje de programación se convierta en un ejercicio de gran complejidad. Los trabajos de identificación y selección de los objetivos de las acciones humanas son muy complejos, por eso el reto que tiene el campo de la programación es muy elevado. Traducir la complejidad cognitiva del ser humano a una función de utilidad es una empresa desafiante. Para Bostrom la solución a este problema supone un gran reto para la siguiente generación, algo que todavía no podrá afrontarse sin un mayor desarrollo cognitivo de los intelectos sintéticos (2016: 187). Actualmente el campo de la ingeniería de sistemas no dispone de una estrategia plausible para transferir valores humanos a un sistema artificial. En ese sentido, Bostrom (2016: 187-208) señala algunas de las técnicas de introducción de valores:

- Representación explícita: es útil para la domesticidad, aunque no es prometedora en la introducción de valores complejos.

- Selección evolutiva: no es prometedora ya que implica aspectos evolutivos que conllevarían graves peligros.

- Aprendizaje por refuerzo: no es muy prometedor, ya que se caracteriza por una tendencia a la instrumentación de ciertos valores que han dado resultados particulares.

- Acumulación de valores: no es pertinente, ya que se podrían acumular valores para objetivos no deseados para los seres humanos.

- Andamiaje de motivación: esta propuesta es prometedora, pero no se ha investigado suficiente en este campo, por lo que conlleva ciertos peligros por la falta de mecanismos de transparencia.

- Aprendizaje de valores: es muy prometedor, pero queda mucho por investigar en el terreno del aprendizaje que pueda comprender las estructuras que configuran el entorno para la fijación de objetivos finales.

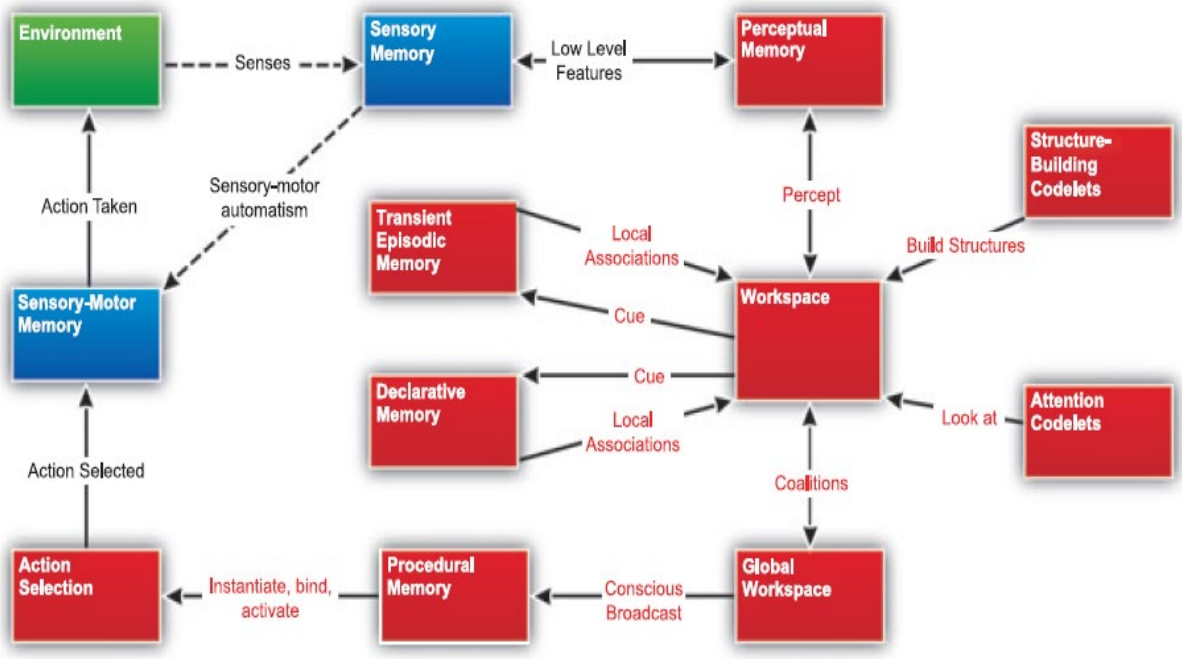
- Modulación de emulaciones: esta oportunidad, fruto de la emulación, presentaría limitaciones éticas.

- Diseño institucional: se caracteriza por la facilidad para implementar mecanismos de control social ante un hipotético escenario de emulaciones. Este mecanismo es merecedor de mayor investigación ante su potencial capacidad para introducir valores.

Bostrom reconoce que si hipotéticamente se llegara a encontrar el mejor mecanismo para la introducción de valores, la humanidad se enfrentaría de nuevo a otro reto, a saber, el de la selección de los valores que deben ser introducidos, una tarea nada fácil en un mundo con tanta complejidad y diversidad cultural.

En cuanto al enfoque de abajo arriba, existe un camino en el que se sitúan los defensores de la idea de la posibilidad de crear una IMA basada en el aprendizaje. LIDA, la nueva versión del IDA, es un modelo de cognición humana que se encuentra inspirado en los descubrimientos que ha realizado en los últimos tiempos la ciencia cognitiva y la neurociencia, que permite organizar aspectos complejos y desordenados a través de un enfoque híbrido para la toma de decisiones. Expertos del campo de la computación piensan que la incorporación de LIDA en un AMA permitiría solucionar algunas de las problemáticas éticas a las que se enfrenta la IA. La implementación de LIDA es única, ya que se restringe a un dominio muy concreto, dependiente de cada caso y por lo tanto

adaptable. Este modelo toma como punto de partida un ciclo cognitivo compuesto por tres fases: fase de comprensión, fase de atención y fase de selección y aprendizaje de acción. Comienza con la afirmación de que cada agente, ya sea humano, animal o artificial, parte del reconocimiento de su entorno y de la posterior planificación de la acción. El comportamiento de los agentes se basaría por lo tanto en ese ciclo cognitivo. El siguiente cuadro muestra el desarrollo del ciclo cognitivo de LIDA:



Fuente: Wallach, Franklin y Allen, 2010: 462.

Los agentes reciben estímulos sensoriales internos y externos que «filtran» por la memoria para después someterse al proceso de creación de sentido y posterior categorización y de ese modo poder planificar la toma de decisiones y la acción. Este procesamiento cognitivo permite al agente llevar a cabo sus acciones para el almacenamiento en la memoria. El modelo LIDA considera que el procesamiento cognitivo humano se realiza mediante un ejercicio continuo, es decir, cada acto cognitivo sigue el esquema que presupone dicho modelo. Estrechamente relacionado con este ciclo cognitivo se encuentra el proceso de aprendizaje. En cuanto al aprendizaje, fundamental para la acción, en el modelo LIDA la atención y la memoria ejercen un papel esencial en cada acto

cognitivo, que sirve de motivación para aprender. Stan Franklin y sus colegas de la Universidad de Memphis han diseñado tres modos de aprendizaje: perceptivo, episódico y procedimental y asistencial.

En el modelo LIDA se hace referencia a las emociones como aquellos sentimientos que son un contenido cognitivo, como esa experiencia interna que se siente cuando se gana la lotería o se encuentra trabajo, a diferencia de un dolor de piernas. Estas emociones sirven como elementos de motivación para las acciones, es decir, como impulsos. Para que exista ese impulso a cada sentimiento se le asigna un valor positivo o negativo y una identidad. Esto permite la planificación de un comportamiento seleccionado que se pasa posteriormente al plano memorístico.

El procesamiento cognitivo de alto nivel, que en los humanos incluye «categorización, deliberación, volición, metacognición, razonamiento, planificación, resolución de problemas, comprensión del lenguaje y producción lingüística» (Wallach, Franklin y Allen, 2010: 468) se divide entre niveles: el reactivo, el deliberativo y el metacognitivo. El reactivo no es verdaderamente importante para el modelo LIDA, pues forma parte de entornos relativamente simples y estables que requieren de poco margen en la toma de decisiones. En cambio, el modo deliberativo implica procesos más complejos y flexibles. En definitiva, Wallach, Franklin y Allen reconocen la dificultad de crear IMA por medio de LIDA, pero están convencidos de que este marco experimental aumenta la posibilidad de que en el algún momento aparezca un AMA.

Una vez expuestas brevemente estas dos reflexiones sobre la posibilidad de construir IMA, es necesario identificar algunos presupuestos dogmáticos que están presentes en los planteamientos de los defensores de la idea sobre la posibilidad de crear intelectos sintéticos con capacidad moral y que podrían dificultar la tarea de llevar a cabo una reflexión ética. Se tomará como referencia el análisis que hace Juan Jesús Álvarez (2013), que identifica cuatro presupuestos dogmáticos: mecanicismo, cientificismo, reduccionismo y monismo materialista.

En primer lugar Álvarez señala el presupuesto mecanicista que contempla la realidad en términos deterministas, como si el ser humano fuera una «máquina lógica». Este dogmatismo aporta una visión propiamente física del ser humano, materialista, considerándolo como un procesador de información que puede ser explicado y reproducido en términos meramente físicos y como un conjunto de engranajes biológicos. Estas dos consideraciones han derivado en lo que en el campo de la computación se conoce como lenguaje simbólico y lenguaje subsimbólico, a saber, el que identifica la mente-cerebro con un ordenador y el que intenta emular el funcionamiento de nuestro cerebro-mente, dando forma a redes neuronales que permiten el *deep learning*. Este presupuesto se formula a partir de dos premisas (2013: 112-113):

La conducta humana es el resultado de mecanismos neuronales que procesan y controlan información. Esta premisa es una clara respuesta de corte materialista al clásico problema de la relación entre mente y cerebro, donde la mente es el cerebro, o lo mental es productivo de la actividad cerebral. En este sentido, lo humano sería equivalente a lo cerebral.

Siguiendo la estela del determinismo, el funcionamiento del cerebro responde a leyes que condicionan su funcionamiento. Esta visión es la que fundamenta el trabajo del campo subsimbólico o conexionista que trata de emular el funcionamiento del cerebro mediante la reducción de los procesos mentales a una base física. Se asume categóricamente que los procesos mentales tienen un evidente soporte físico y por lo tanto que es posible su emulación, lo que significaría una clara dificultad para distinguir entre lo humano y lo artificial.

A partir de estas premisas, según Álvarez, podrían considerarse dos conclusiones: por una parte, que no existe una evidente diferencia cualitativa entre la IA y la inteligencia natural, ya que los mecanismos neuronales pueden duplicarse en forma electrónica; y por otra parte, que la inteligencia, debido a la emulación natural-artificial, puede manipularse fácilmente sin necesidad de someterse a largos periodos evolutivos que ha experimentado la inteligencia natural.

Álvarez sostiene que hay un error de fondo en este presupuesto dogmático a la hora de intentar homologar la mente con el cerebro, pues este es un postulado que no puede ser científicamente demostrado. Existe un estrecho paralelismo entre los procesos mentales y cerebrales, pero eso no permite todavía identificar ambos como exactamente lo mismo. En ese sentido, «el hombre es más que su inteligencia (mente/cerebro) y no parece que ninguna dimensión de lo humano pueda reducirse a su base físico-química» (2014: 114). Este dogmatismo podría sentar las bases en el futuro para que la toma de decisiones que implican necesariamente una profunda carga emocional y contextual se redujera a simples decisiones realizadas por un algoritmo, como es el caso del jefe-algoritmo de *Uber*, que ha sido objeto de denuncias por tomar decisiones carentes de razonabilidad (Simonite, 2015).

El cientificismo se convierte en otro presupuesto dogmático que está presente en el campo de la IA. Los acontecimientos históricos han suscitado que las ciencias experimentales cuenten con un gran poder académico y prestigio social en el mundo actual. Tanto es así que la perspectiva que de ellas nace es la que determina qué es y qué no es lo que se puede decir y encontrar, dejando de lado otras perspectivas que no se limitan a su esquema metódico. El cientificismo sería aquella creencia dogmática que considera que el modo de conocer llamado ciencia es el único que merece recibir el título de conocimiento, y que por lo tanto la ciencia es el único medio para la resolución de todos los problemas más significativos (Álvarez Álvarez, 2013: 115-118).

El cientificismo cultiva perjudicialmente un ego científico que dificulta los puentes entre las ciencias experimentales y otras disciplinas que no se limitan a su método. Los dogmatismos obstaculizan la tarea deliberativa entre saberes, algo sumamente necesario para la construcción de una ética aplicada a la IA. Así pues, según el dogma cientificista, la ciencia posibilitaría un paradigma de fiabilidad, utilidad y eficacia, pilares sobre los que se sostiene el ideal de progreso. Este presupuesto dogmático rechaza otros saberes y los tacha de inútiles, subjetivos e ineficaces para el supuesto progreso.

Además, cuando se rechazan otras perspectivas de conocimiento, también se está incurriendo en un peligro, a saber, el de rechazar hasta ciertos criterios morales que son básicos en toda actividad humana y que sirven como orientación para receptar la ciencia de un modo crítico. Este rechazo se ve motivado por otros criterios más propios del ideal de progreso moderno, como son los de beneficio, utilidad, optimización, etc. En ese sentido, el proceso de construcción de una ética aplicada sería obstaculizado a causa del cientificismo.

Otro presupuesto dogmático presente sería el reduccionismo, que es esperable en el campo de la ciencia, pues la simplificación de la realidad es necesaria para la manipulación de variables. No obstante, esta reducción debe realizarse con plena conciencia y sumo cuidado, ya que si es considerada en términos extremos podría incurrir en el olvido de otras experiencias y dimensiones que también explican la realidad. En ocasiones, la valoración de las circunstancias exige el reconocimiento de aspectos que no pueden ser reducidos, sino más bien considerados con toda su extensión, alejándonos de esa forma del «nadamasqueísmo» (*nothing buttery*), según la expresión de Mary Midgley (1994). La falta de consideración de ciertos aspectos de la realidad puede conducir a cometer graves errores para la humanidad.

Las teorías científicas no siempre pueden dar una explicación completa y exhaustiva de la realidad, pues siguen la tendencia de considerar solo aquellos aspectos que son traducibles en un lenguaje científico, dejando de lado otros aspectos que forman parte de la rica variedad que configura la realidad humana. En lo que respecta al campo de la IA, el reduccionismo puede estar orientado a la consideración de solo aquellos aspectos que son cuantitativos, pues los que son supuestamente cualitativos no son reales, y si quieren ser tomados en cuenta, deben limitarse a la traducción en formulas lógico-matemáticas.

El último presupuesto dogmático es el monismo materialista. Hay importantes investigadores en el campo de la IA como Marvin Minsky (1967; 2010) y Kurzweil (2016b) que defienden la posibilidad de aumentar la vida humana mediante una relación más íntima con la máquina, lo que supondría un fenómeno de trascendencia desde lo biológico hacia lo tecnológico, o lo que en el terreno de la IA se denomina «singularidad».

El clásico binomio cuerpo-alma podría ser actualizado por el de *software-hardware*, aspecto que pondría de relieve la equiparación del alma al *software*, tratando de emular los contenidos del cerebro y el *hardware* al cuerpo, entendido como soporte físico. En ese sentido, la sustancia aristotélica, entendida como la forma privilegiada del ser, quedaría reducida a mera información, a sustancia cerebral, considerada como *software*. Por ello los postulados de Minsky o Kurzweil incurrirían en un grave error al atribuir un ausente valor al aspecto corporal del ser humano. El deseo radical y la fe ciega en el progreso conducen al campo de la IA a cuasi olvidar la condición material.

Por lo tanto la postura que es sostenida sobre la posibilidad de crear agentes morales artificiales y desde ahí llevar a cabo el planteamiento de una línea de responsabilidad dentro del marco de la IA, adolecería de una falta de solidez. Esta falta de solidez surge ante unos presupuestos dogmáticos que ponen en evidencia muchos aspectos que no suelen observarse debido a dinámicas de diversa índole, como las propias del mundo tecnológico o las económicas, y también por la dificultad de traspasar la frontera entre el lenguaje natural y el lenguaje propio del mundo de la programación, algo que es fundamental para organizar estructuras morales.

5.3.2. La premisa de responsabilidad como alternativa a la IMA

Es escasa la bibliografía que versa sobre las posibilidades de una IAR. Ha sido Virginia Dignum (2017a; 2017b; 2018) quien ha abordado este tema con cierta novedad y amplitud. Esta autora sostiene que la inteligencia no puede entenderse separada de la responsabilidad, situándose en la estela de los defensores de la incorporación de responsabilidad en los sistemas artificiales de inteligencia. Su propuesta de IAR descansa sobre tres pilares (2017b: 4-5):

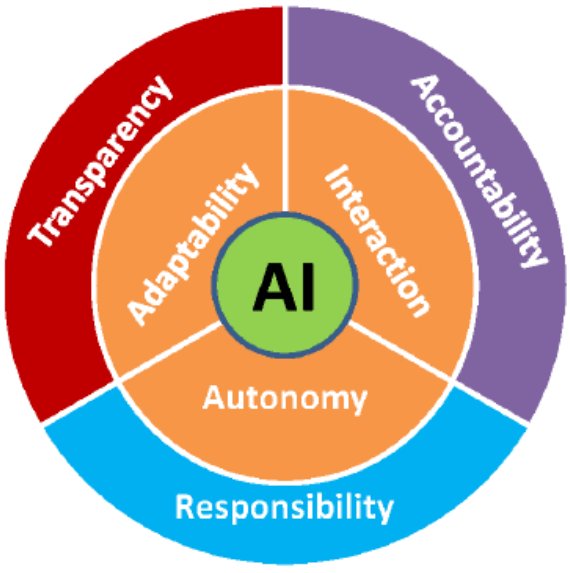
- En primer lugar, la sociedad civil en su conjunto debe estar preparada para asumir responsabilidad frente al impacto de la IA en diferentes esferas. Esto implica que el personal tecnológico involucrado en el desarrollo de la IA debe estar capacitado para poder asumir su propia responsabilidad en un producto que tendrá un gran

impacto sobre la vidas. Esta asunción de responsabilidad implicaría un esfuerzo educativo, formativo y el desarrollo de códigos de conducta. Además, la ciudadanía y los gobiernos deben deliberar en materia de regulación.

- En segundo lugar, los intelectos sintéticos deben incorporar modelos y algoritmos que les permitan desarrollar la capacidad para emprender acciones basadas en valores humanos y para llevar a cabo una justificación en función de esos valores. De momento los actuales mecanismos de *deep learning* no son capaces de tal justificación.

- En tercer lugar, es importante tener en cuenta el principio democrático de la participación de las partes implicadas en el desarrollo e impacto de la IA. El desarrollo de este ámbito se sostiene sobre unas relaciones sociotécnicas que tienen un carácter político que debe ser repensado para las sociedades del futuro.

Dignum incorpora tres principios éticos a cada una de las características que normalmente destacan la actividad de la IA: interactividad-rendición de cuentas, adaptabilidad-transparencia y autonomía-responsabilidad (2017a: 4698-4699; 2017b:5).



Fuente: Dignum, 2017b: 5.

Las tres fases que caracterizan la actividad en el campo de la IA, análisis-diseño-implementación, deben someterse a un ejercicio deliberativo en el que puedan discutirse los impactos que pueden derivarse, se identifiquen los valores que están involucrados y se decida qué enfoque moral debe ser incorporado mediante mecanismos técnicos (2017b: 6). La responsabilidad jugaría en este caso un papel fundamental para la deliberación algorítmica sobre las acciones y las consecuencias de los intelectos sintéticos y se basaría en los marcos éticos que la historia de la ética ha proporcionado. A diferencia de la propuesta de responsabilidad ética, se presenta un principio de transparencia en el que las decisiones son fruto de un trabajo deliberativo de carácter pragmático que tenga en cuenta el contexto concreto del despliegue de los sistemas artificiales. En este sentido, la investigación en el ámbito de la IA tendría que ser más equilibrada, para que el peso de la balanza no se centrara exclusivamente en los objetivos de mejora de los rendimientos, sino que también diera prioridad a los aspectos éticos en el impacto social. Esto exige un cambio de cultura en el seno de los centros de investigación para que introduzcan mecanismos de transparencia y responsabilidad.

Dignum considera que es conveniente que los tecnólogos entiendan que su actividad no está exenta de valores morales. La actividad tecnológica, por muy técnica que sea, presenta implicaciones morales que deben ser pensadas (2017a: 4699). El comportamiento empleado para la creación de intelectos sintéticos surge en un contexto moral que proyecta sus concepciones sobre las obras. La reflexión ética permite conocer con mayor profundidad las teorías y principios éticos que configuran las relaciones humanas. Para Dignum en esta reflexión ética, que puede estar orientada a fines de diversa índole, es central el aspecto deliberativo, esencial para la fundamentación de una IAR.

La reflexión ética necesaria en el ámbito de la IA se ha planteado en documentos de diversos organismos internacionales. Un ejemplo es la iniciativa impulsada el 25 de abril de 2018 por la Comisión Europea (CE) para abordar asuntos relativos al desarrollo de la IA con la que se perseguía la garantía de un marco ético y jurídico para el impacto de estos sistemas sobre las sociedades. El 18 de junio de ese mismo año la CE mantuvo una reunión de alto nivel con 12 representantes de diversas organizaciones filosóficas, tanto

confesionales como no confesionales, bajo el título «Inteligencia artificial: abordar los desafíos éticos y sociales» presidida por su vicepresidente Andrus Ansip, para discutir las repercusiones de la IA sobre los derechos fundamentales, específicamente en materia de privacidad, dignidad, protección de los consumidores y lucha contra la discriminación, pero también para tratar de la dimensión social de la IA y sus consecuencias para la inclusión social y el futuro del trabajo.

El espectro de las decisiones automáticas que realizan los sistemas artificiales es muy complejo y las cargas de responsabilidad pueden variar (Dignum, 2017a: 4701):

- Control humano: en muchas ocasiones son los seres humanos los que supervisan el control y diseño de los sistemas artificiales. En ese sentido, la persona que supervisa esta actividad debe contar con la información suficiente y las herramientas necesarias para garantizar un conocimiento compartido.

- Regulación: el entorno de trabajo está sometido a regulaciones y restricciones que permiten un ordenamiento limitado de los razonamientos morales.

- Inteligencia moral artificial: como se ha planteado en el apartado anterior, un AMA es capaz de incorporar el razonamiento moral en su deliberación. Sin embargo, este campo no está todavía lo suficientemente desarrollado debido a las problemáticas que giran en torno a la transformación de los marcos morales humanos en un lenguaje de programación.

- Aleatoriedad: el sistema autónomo elige el curso de su acción de forma azarosa. Este ámbito de actuación plantea problemas éticos a la hora de elegir.

La diferencia entre estos cuatro espectros radica en la rendición de cuentas (*accountability*), estrechamente vinculada a la responsabilidad y la transparencia.

Para Dignum es importante reconocer la existencia de diferentes valores morales fruto de las diversidades sociales que configuran el mundo. Ese carácter de pluralidad puede ser considerado un factor determinante a la hora de incorporar criterios de responsabilidad en

el campo de la IA. Además, esto pone de relieve la necesidad de un marco deliberativo en el que diferentes enfoques sean esbozados para la discusión de problemáticas y la innovación. Sin embargo, Dignum no proporciona un marco deliberativo concreto desde el que formular su concepto de IAR.

La Declaración de Montreal para un desarrollo responsable de la IA (2018) representa otra aportación interesante en materia de responsabilidad, aunque es un planteamiento débil ante la potencialidad de los desafíos. Esta iniciativa tiene como objeto principal el establecimiento de orientaciones éticas durante el desarrollo de la IA en el contexto canadiense. Supone una invitación al público en general para llevar a cabo una discusión con los tecnólogos y los responsables políticos por medio de la identificación de once valores: bienestar, autonomía, respecto a la privacidad y la intimidad, solidaridad, participación democrática, equidad, inclusión de la diversidad, prudencia, responsabilidad y desarrollo sostenible.

Esta Declaración representa una solución plausible, aunque insuficiente a las problemáticas complejas de la IA, pues carece de una fundamentación filosófica en profundidad, situándose exclusivamente en el plano deontológico. Es en ese sentido, al igual que la propuesta de Dignum, se trata de un planteamiento débil de IAR dado el escaso desarrollo argumentativo y las insuficientes bases metodológicas de ambas propuestas.

5.3.3. La necesidad de ir más allá de los planteamientos débiles

Si se admite la necesidad de una ética aplicada al campo de la IA fundamentada en un método deliberativo que reconozca las complejidades y retos, no cabe duda que tras esa dinámica deliberativa tendrán que surgir decisiones responsables y transparentes. En ese sentido, es fundamental reconocer el trabajo de Dignum y de quienes han elaborado la Declaración de Montreal, aunque estos trabajos podrían enriquecerse concretando un marco deliberativo para la IAR que aborde los desafíos de la IA partiendo de una ciencia cívica y de un modelo de innovación abierta y responsable (MIAR). Estas propuestas carecen de una fundamentación sólida para la construcción de la IAR. Es la filosofía moral la que nos

proporciona bases teóricas para orientar una responsabilidad contextualizada en el terreno de la IA e ir más allá de la superficialidad de propuestas débiles. Esta superficialidad no es suficiente para pensar la IAR como alternativa real al contexto tecnológico actual.

5.4. Inteligencia artificial responsable

Es abundante la bibliografía sobre las implicaciones éticas de la IA, pero casi todo lo escrito se ha centrado en casos particulares del área de la robótica, como por ejemplo el impacto de la robótica en el ámbito profesional, el de la biotecnología en el ámbito sanitario o el de los drones en el terreno militar. Si bien es cierto que estas aportaciones, centradas en aspectos particulares, son muy valiosas, es necesario aportar un marco de reflexión mucho más amplio y sistemático. Esto permitiría la apertura de un nuevo horizonte de posibilidades para pensar la IA, las oportunidades que de ella se derivan y las consecuencias. La propuesta de una IAR sirve para sentar las bases de un mayor conocimiento de las necesidades de la sociedad y también para una mayor sensatez a la hora de considerar los impactos de la tecnología.

Así pues, los planteamientos débiles representados por Dignum y la Declaración de Montreal significan un primer paso desde el que plantear la necesidad de responsabilidad en el contexto de la IA, aunque carecen de una solidez fundamentadora y también de un marco deliberativo concreto. Además, tampoco insisten en cuestiones de suma relevancia en la actualidad como los derechos humanos, los ODS o los límites planetarios en su relación con la IA. Por lo tanto, en este apartado se buscará fundamentar la IAR desde el marco de la ciencia cívica y diseñar un modelo de innovación abierta y responsable (MIAR) capaz de abrir espacios deliberativos donde generar nuevas dinámicas en la producción de conocimiento. El espíritu de la ciencia cívica y el MIAR representan los soportes fundamentales desde los que plantear una IAR a la altura de los desafíos actuales. Alejada de la debilidad y la superficialidad de otros planteamientos, esta propuesta asume la complejidad de los retos de este tiempo reconociendo la importancia de la participación de la sociedad civil y del ejercicio deliberativo y fronético. La IAR supone una alternativa

fuerte frente a aquellos planteamientos que destacan por sus limitaciones filosóficas y su superficialidad, a saber, planteamientos débiles.

5.4.1. Ciencia cívica, participación y colaboración ciudadana

El cultivo de la responsabilidad en el ámbito de la IA demanda un fundamento científico enraizado en una práctica cívica dispuesta a la comunicación y la transparencia con y hacia la ciudadanía. Es importante construir un hilo comunicativo entre la ciencia y la ciudadanía, con el objetivo de promover y fortalecer una dimensión cívica y democrática. La ciencia cívica aporta a la IA una cultura participativa que permite poner en práctica una responsabilidad compartida en aquellos asuntos que afectan a la ciudadanía. Esto supone un cambio en la concepción tradicional de la ciencia, pues reconoce el valor de la participación de aquellos agentes que comúnmente no participan en la deliberación de asuntos que generan controversia.

5.4.1.1. Teoría y práctica de la ciencia cívica

La expresión «ciencia cívica» (*civic science*) se debe al botánico, sociólogo, científico, urbanista, filósofo, poeta, activista cívico, educador escocés Patrick Geddes (1854-1932), maestro de Lewis Mumfold, el gran historiador de la tecnología y de las ciudades. En su obra *Ciudades en evolución*, que data de 1915, Geddes (2019), cuyo pensamiento está muy vinculado al de autores británicos como John Ruskin y William Morris, considera que la ciencia cívica consiste en saber que las instituciones y los edificios se levantan desde dentro; no son impuestos desde arriba ni construidos desde afuera. Desde esta concepción experiencial, Geddes aboga por un modelo organicista de urbanismo que pretende establecer unas pautas donde la ciudad y el campo coexistan de forma equilibrada. Su posición se inscribe en el contexto de las corrientes reformistas del socialismo inglés, donde destacan algunas preocupaciones importantes: higienismo biologista, estética modernista, utopismo, cooperativismo, etc.

Geddes plantea una renovación metodológica, donde filósofos, juristas, economistas, ingenieros, arquitectos, etc., colaboran desde perspectivas distintas para promover no solo construcciones cívicas, sino también para desarrollar la perspectiva cívica de una manera integral. Se trata de un método itinerante, participativo y desde un enfoque *bottom-up*. Para ello, estudiantes y futuros planificadores urbanos han de sumergirse directamente en la ciudad y practicar un urbanismo itinerante centrado no solo en observar la ciudad, sino en escuchar los problemas de los habitantes del lugar y las soluciones que ellos mismos sugieren. Solo desde ahí es posible leer la ciudad en su estado actual y dar indicaciones para transformarla, además de despertar un auténtico interés por el desarrollo cívico en la ciudadanía.

En el ámbito de la IA el concepto de ciencia cívica, con raíces en la propuesta de Geddes, aporta una noción de responsabilidad participativa y compartida similar a la que el autor escocés trata de desarrollar en el ámbito del urbanismo. Este concepto, que en nuestro caso pone el acento sobre la dimensión cívica de la ciencia, permite cultivar un espíritu democrático en los procesos de generación de conocimiento científico. El postulado de una IAR sirve para sentar así las bases de una ciencia de carácter cívico que reconozca la necesidad de enriquecer los procesos de generación de conocimiento científico a través de un ejercicio inclusivo, participativo y abierto.

La ciencia cívica promueve la incorporación de aquellos agentes que comúnmente no suelen participar en procesos estrictamente científicos, mediante la deliberación sobre asuntos relativos controvertidos en el ámbito de la ciencia y que por lo tanto son susceptibles de tratamiento y consideración pública. En ese sentido, la ciencia cívica contribuye al proceso de toma de decisiones para promover, en el ámbito de las cuestiones científicas, la acción cívica y el compromiso ciudadano a través de un diálogo abierto. Además, la ciencia cívica valora que los procesos de investigación e innovación sean responsables y para ello sitúa como principal eje articulador la participación abierta e inclusiva de diversos agentes.

Este concepto de ciencia cívica ha sido impulsado en la actualidad desde los Estudios Cívicos, un ámbito emergente donde confluyen diversas disciplinas que participan en el desarrollo de ideas útiles para la ciudadanía, tejiendo así un hilo conductor entre teoría y práctica. Según Peter Levine (2014), los Estudios Cívicos, que surgen como disciplina académica en 2009 en el Jonathan M. Tisch College of Civic Life de la Universidad de Tufts (EE. UU.), tratan de responder a la pregunta: «¿Qué debemos hacer?». La importancia de esta pregunta estriba en su carácter de amplitud, de pluralidad, y no de individualidad, pues la resolución de problemáticas, el enfrentamiento de controversias y desafíos, no puede abordarse desde una única perspectiva debido a la complejidad del mundo actual, sino desde una multiplicidad de perspectivas y actuaciones.

Para ilustrar esto último, la Declaración introductoria del Instituto de Verano de Estudios Cívicos señala lo siguiente:

> Los Estudios Cívicos representan un campo interdisciplinario enfocado en la reflexión crítica, el pensamiento ético y la acción para el cambio social, dentro y entre las sociedades. Las personas que piensan y actúan juntas para mejorar la sociedad deben abordar los problemas de la acción colectiva (cómo lograr que los miembros trabajen juntos) y la deliberación (cómo razonar juntos sobre los valores en disputa). También deben:
>
> - Comprender cómo se organiza el poder y cómo funciona dentro y entre las sociedades.
> - Afrontar los conflictos sociales, la violencia y otros obstáculos a la cooperación pacífica.
> - Considerar cuestiones de justicia y equidad cuando surjan tensiones sociales.
> - Confrontar las preguntas sobre relaciones apropiadas con personas externas de todo tipo.
> - Examinar marcos éticos, políticos y teológicos alternativos para fomentar la reflexión comparativa sobre las diferentes formas en que las personas viven juntas en la sociedad.

> El enfoque en la sociedad civil incluye el estudio de la acción colectiva en las esferas sociales que, si bien están organizadas, no pueden ser institucionalizadas ni sancionadas por el Estado. Destaca la perspectiva de los agentes individuales y grupales (School of Arts and Sciences-Tufts University, 2009).

Los Estudios Cívicos se sitúan en la intersección de la reflexión ética, del análisis los hechos y de las estrategias a realizar, conectando teoría y praxis (Kravetz, 2013: 168). En ese sentido, Levine sostiene que son cinco los principios que fundamentan los Estudios Cívicos (2014: 30-32):

- Aprender de la deliberación

Los métodos deben caracterizarse por la interacción y la deliberación. No obstante, el procedimiento no es suficiente, pues no se trata de reunir a diversas personas afectadas, sino de promover la discusión desde creencias que permitan interactuar con otros y aprender de ellos, entender que otros conocimientos pueden enriquecer. La importancia de asumir un compromiso de conversación y acción con otros.

- Ser humilde

Ser conscientes de las limitaciones intelectuales que existen a partir de posicionamientos razonablemente cautelosos y humildes con nosotros mismos. En ese sentido, es importante situarse en la senda del pensamiento conservador por un motivo principal, a saber, la resistencia a la arrogancia intelectual. Reconocer el valor del testigo recogido desde una actitud de respeto y siempre teniendo en cuenta la premisa que versa sobre la posibilidad continua de mejora.

- Criticar desde dentro

Asumir una posición crítica frente a los problemas y necesidades de la sociedad. Ser conscientes de que en ocasiones pueden surgir contradicciones, pero eso es un símbolo de madurez y complejidad. La crítica no siempre se dirige a grupos humanos en concreto, sino a prácticas y campos.

- Evitar la búsqueda de las causas de raíz

Como seres humanos a través de nuestra experiencia tenemos la capacidad de cambiar elementos estructurales de la realidad sin necesidad de ir a la raíz de los problemas. La raíz de los problemas suele presentar una gran magnitud y eso genera dificultad a la hora de enfocarse en la formulación de alternativas concretas, pues se caracterizan por la gran magnitud y difícil tratamiento. A veces es mejor abordar un aspecto de un problema concreto, en lugar de atacar un aspecto más fundamental sin éxito.

- Mantener el barco juntos

No limitar el curso de las acciones y las decisiones por un ideal en concreto, sino asumir con conciencia la posibilidad de variar el pensamiento y de buscar nuevos rumbos. En ese sentido, es importante reconocer el valor de la comunidad para esta empresa, entendiendo que el conocimiento se construye de forma colectiva.

La complejidad e incertidumbre que puede vislumbrarse hoy en la actividad científica hace necesario desarrollar una ciencia cívica para desentrañar esta complejidad y establecer un encuentro con la incertidumbre. La ciencia cívica asume que los temas científicos no son exclusivos de los expertos y que pueden ser abordados en la esfera cívica, pues proporciona mayores conocimientos para que las decisiones sean fruto de la buena información obtenida a través de un proceso dialógico entre los actores implicados. Este diálogo sobre la dimensión cívica de la ciencia permite revitalizar la comunidad y poner en valor dinámicas democráticas. La ciencia cívica se pregunta cómo otros miembros de la comunidad que no son estrictamente científicos pueden participar en los asuntos científicos complejos, rompiendo de ese modo con la cultura de la exclusividad en materia de conocimiento. En ese sentido, se redefine el concepto de ciencia como un bien público a través de la enseñanza de habilidades que permite orientar la ciencia hacia el conocimiento cívico. Esta redefinición del concepto de ciencia permite cultivar un enfoque participativo en el entorno de los científicos para fomentar así la comprensión de la ciudadanía. Es decir, lo importante es poner el acento en el intercambio de ideas, creando puentes, para hacer la ciencia más

inclusiva a través del diálogo entre valores, creencias y opciones, lo que Jonathan Garlick y Peter Levine definen como «humanización de la conversación científica» (2016).

En la estela de la ciencia cívica, la Comisión Europea emplea el término «ciencia ciudadana» (*citizen science*) para referirse al conjunto de niveles de compromiso que pueden desplegarse en el ejercicio científico y que giran en torno a la participación ciudadana.

> La ciencia ciudadana abarca una gama de niveles de compromiso: desde estar mejor informado sobre la ciencia hasta participar en el proceso científico mismo mediante la observación, la recopilación o el procesamiento de datos.
>
> La ciencia ciudadana es un término amplio, que abarca la parte de *open science* en la que los ciudadanos pueden participar en el proceso de investigación científica de diferentes maneras posibles: como observadores, como financiadores, en la identificación de imágenes o el análisis de datos, o al proporcionar datos ellos mismos. Esto permite la democratización de la ciencia, y también está relacionado con el compromiso de los interesados y la participación pública.
>
> Dependiendo de su interés personal, tiempo y recursos tecnológicos, el ciudadano decide cómo participar. Observar avistamientos de aves, identificar galaxias o averiguar cómo plegar proteínas, proporcionando recursos prestando tiempo de computadora o financiamiento directo como en el financiamiento colectivo de proyectos científicos (Comisión Europea, 2017).

Además, la Comisión Europea a través del *White Paper on Citizen Science in Europe* sostiene que en este tipo de nueva cultura científica se encuentran implícitos los siguientes valores y atributos:

Fuente: Comisión Europea, 2015: 10.

La ciencia cívica promueve un cultivo de habilidades comunicativas en el ámbito científico, pues entiende como necesaria la construcción de puentes entre este y otros ámbitos. Este cultivo permite fortalecer la sensibilidad y compartir convicciones morales en un ejercicio de diálogo respetuoso. En ese sentido, reconoce aquellas dimensiones de la ciencia que son decisivas para el ser humano y esas actividades que tienen un impacto social. Existe, por lo tanto, un imperativo ético para educar en actitudes de escucha y entendimiento en el ámbito científico para-con la ciudadanía. Pues, en ocasiones, los agentes del campo científico presentan actitudes esquivas, fundamentadas en una cultura monopolista y oligopolista del conocimiento. Frente a esa cultura, la ciencia cívica defiende una visión de la ciencia entendida como recurso de interés público, abierto y participativo, susceptible de consideración ciudadana. Esta visión se sostiene sobre la necesidad de

construir puentes de diálogo para enfrentar las incertidumbres y complejidades de la realidad actual. Es lo que antes se caracterizó como «humanización de la conversación científica» (Garlick y Levine, 2016).

Esta humanización de la comunicación permite impulsar una tarea científica de carácter facilitador, cultivando la escucha de aquellas voces de esperanza y tensiones sobre la ciencia, entendiendo aquellos valores y creencias que permiten humanizar y hacer más cívica la actividad científica. Además, permite conectar la ciencia con la cultura, la política y los valores a través del diálogo, para que los asuntos científicos no escapen de la comprensión de la ciudadanía. La ciencia cívica promueve la inclusión de la ciudadanía en proyectos de investigación científica, contribuyendo a la democratización del conocimiento, el cultivo de la responsabilidad y la integración de diversas perspectivas.

5.4.1.2. Laboratorios ciudadanos para la innovación social

Recogiendo el espíritu de los Estudios Cívicos, y concretamente de la ciencia cívica, los laboratorios abiertos son espacios ciudadanos entendidos como un lugar de encuentro para la reflexión colectiva, la generación de conocimiento innovador a partir de propuestas colaborativas y la confluencia de diferentes saberes y sentires. Estos laboratorios se sostienen principalmente sobre dos premisas fundamentales, a saber: la investigación colaborativa y el aprendizaje comunitario en torno a temas de diversa índole. En el contexto concreto de la IA, la motivación para impulsar estos espacios como ecosistemas de innovación colaborativa nace de la posibilidad de cultivar prácticas democráticas y cívicas en el contexto tecnológico a partir de la propuesta de una IAR. Los laboratorios abiertos representan un espacio entendido como una herramienta de innovación social que contribuirá a propiciar la preparación de las instituciones para las profundas transformaciones que la revolución tecnológica depara.

La cultura que subyace tras un laboratorio abierto sitúa en el centro de su reflexión el concepto de innovación social (García, 2018: 108). La innovación se entiende en términos colectivos, colaborativos y participativos, es decir, desde la inclusión de la ciudadanía

como un agente activo capaz de organizarse y hacer frente a sus necesidades y anhelos. Este concepto de innovación social está fundamentado en tres dimensiones esenciales: satisfacción de necesidades, reconfiguración de las relaciones sociales y empoderamiento o movilización política (Palavicini Corona y Cepeda Mayorga, 2019: 22). Además, para la CE la innovación social representa un elemento central para promover la cohesión social, la competitividad y la sostenibilidad en las sociedades actuales (2013). En este sentido, un laboratorio abierto sobre ciencia cívica surge de la necesidad de ofrecer lugares públicos de encuentro en los que la ciudadanía pueda reflexionar y contribuir a un discurso científico que impacta directamente sobre sus vidas. Estos laboratorios proponen la formación de una cultura participativa para hacer que el conocimiento científico sea más comunicativo, y en definitiva, más cívico.

El valor de los laboratorios cívicos estriba en la cultura comunicativa que se establece en el cultivo de nuevas habilidades de cooperación. Se exploran nuevos mecanismos de intersubjetividad y comunicación para construir colectivamente el conocimiento y mejorar la convivencia en términos cívicos y democráticos. La experimentación permite conectar la teoría con la praxis, acercando el mundo para relacionarse con él, pero también con los que conviven en un mismo espacio sobre un escenario de confianza y colaboración. Esta confianza y colaboración sirven fundamentalmente para enfrentar la incertidumbre (Ruiz Marcos, 2018). Como sostiene Marcos García: «Los laboratorios ciudadanos ofrecen un lugar para no quedarse en las ideas y prototipar de manera colaborativa anhelos colectivos, y pasar de lo posible a lo realizable» (2018: 110).

Existen tres casos de laboratorios cívicos que podrían servir a modo de ejemplo: Medialab-Prado, en Madrid; Santalab, en Santa Fe; y LabIN, en Granada. El primer caso, Medialab-Prado, consta de un espacio abierto de encuentro para la producción de proyectos culturales. Cualquier persona, a título individual o colectivo, puede participar de manera colaborativa. Se crean grupos de trabajo y se lanzan convocatorias abiertas para la producción y presentación de proyectos, investigaciones y comunidades de aprendizaje. Entre sus laboratorios se encuentran DataLab, Laboratorio de Datos Abiertos; PrototipaLab, Laboratorio de prototipado creativo; ParticipaLab, Laboratorio de Inteligencia Colectiva

para la Participación Democrática; CiCiLab, Laboratorio de Ciencia Ciudadana, etc. En Santa Fe, Argentina, se desarrolla el Santalab, un Laboratorio de Innovación Publica que como reza en la página web «es impulsado como una política de Innovación y Gobierno Abierto para constituirse en un espacio articulador con las nuevas formas de organización ciudadana y agrupaciones auto-organizadas que, mediante procesos informales de práctica ciudadana, modifican de forma resiliente y adaptativa los entornos que habitamos» (Gobierno de Santa Fe, 2016-2019). Santalab es una interfaz de colaboración para reunir iniciativas ciudadanas innovadoras sustentadas en el modelo de la cuádruple hélice. El último caso a exponer es el de LabIN, en Granada, el Laboratorio de Innovación Ciudadana de Granada, impulsado por su universidad pública, centrado en la generación de ideas, y también en el prototipado de soluciones y el desarrollo de proyectos para la ciudadanía. Está formado por una red de participación ciudadana y presenta un carácter digital muy notable.

En el caso de la IA, un laboratorio abierto sobre ciencia cívica podría contribuir favorablemente en la promoción de una IAR, respondiendo a las necesidades de la ciudadanía y permitiendo cultivar habilidades cívicas y democráticas en el ámbito tecnológico. Visto así, sería un espacio de confluencia de saberes para la búsqueda de soluciones en términos pragmáticos, con la participación de los actores implicados en el despliegue de la actividad de la IA aportando su mirada sobre el impacto de esta tecnología. En ese sentido, este laboratorio permitiría un fortalecimiento de las habilidades comunicativas entre la ciencia y la ciudadanía a través de la herramienta del diálogo abierto y participativo.

Todo ello contribuiría a una innovación científica abierta, comprometida con la defensa de los derechos humanos y la promoción de los ODS, además de reconocer el conocimiento científico como un recurso público, de interés general, y susceptible de consideración cívica. En el siguiente apartado se presentarán las características de un modelo de innovación que permita desarrollar la IAR desde estos laboratorios ciudadanos.

5.4.2. Modelo de innovación abierta y responsable

El proceso de innovación convencional, de carácter cerrado, ha sido durante mucho tiempo el enfoque predominante y más exitoso dentro de las empresas, como señala Verena Nedon (2015: 7). Sin embargo, este modelo de innovación estrictamente cerrado y de triple hélice ha significado con el paso del tiempo un aislamiento respecto a otros conocimientos que se van generando en el mundo. Esta dinámica representa un aspecto negativo para la organización de los grupos de trabajo respecto a aquellas ideas externas que se sitúan en otros espacios. El modelo de innovación de triple hélice se erige sobre la idea de que exclusivamente un grupo es capaz de poseer el monopolio del conocimiento en un determinado campo, algo muy difícil de entender en un mundo donde todo el conocimiento está conectado por redes de diversa índole. Este aislamiento sostenido sobre la idea monopolista ejerce resistencia frente a otros conocimientos y una realidad que requiere continua adaptación. Miguel Urra Canales (2018: 184) resume con claridad las posturas críticas que giran en torno al modelo de innovación de triple hélice:

- Defensa romántica de la autonomía académica y demonización de las empresas.
- Minusvaloración de la cooperación internacional y transdisciplinar.
- Ciencia básica amenazada por la ciencia aplicada.
- Escrúpulos metodológicos.
- Desdén hacia los problemas sociales, las ONG, la sociedad civil y los países en vías de desarrollo y los problemas sociales.
- Incredulidad en las capacidades de hibridación de las esferas académicas, gubernamental y empresarial.

Resulta paradójico creer que los miembros de un grupo de trabajo concentran todas las ideas posibles y que no necesitan otras fuentes de conocimiento para la innovación. La globalización y el desarrollo de las tecnologías han favorecido el intercambio de

conocimiento y la movilidad, creando de ese modo puentes de intercambio de ideas entre diferentes esferas del saber. Así pues, este escenario de movilidad y redes de conocimiento ha facilitado un terreno fértil desde el que comenzar a imaginar un proceso de innovación abierto e interactivo que reconoce la necesidad de integrarse en el mundo con el conjunto de saberes bajo el compromiso de innovar con ideas cada vez más prometedoras.

El modelo de innovación abierta, desarrollado principalmente entre los años 2011 y 2016 en La Haya a partir de la iniciativa recogida en el Programa de Investigación *Maatschappelijk Verantwoord Innoveren* del Consejo de Investigación holandés, comienza a jugar un papel cada vez más relevante en numerosos ámbitos. Representa un nuevo alcance de la ética aplicada que reconoce el valor de la práctica de la filosofía moral en diversidad de esferas. En el contexto científico y tecnológico la ética comenzó a preguntarse sobre la posibilidad de contribuir a la solución de los problemas y la mejora de las condiciones de vida de la sociedad. Von Schomberg (2011) ofrece la siguiente definición de la Investigación e Innovación Responsables (RRI, por sus siglas en inglés: *Responsible Research and Innovation*):

> Un proceso transparente e interactivo mediante el cual los actores sociales y los innovadores se hacen mutuamente sensibles entre sí en vista de la aceptabilidad (ética), la sostenibilidad y la conveniencia social del proceso de innovación y sus productos comercializables (con el fin de permitir una integración adecuada de los avances tecnológicos en nuestra sociedad) [...] La innovación responsable significa cuidar el futuro. A través de la administración colectiva de la ciencia y la innovación en el presente (2011: 41).

La preocupación social y el interés en la investigación y la innovación, así como sus aspectos de responsabilidad, son considerados como una forma para incorporar la deliberación sobre estas preocupaciones en el interior del proceso de innovación. En ese sentido, la participación de los grupos de interés (*stakeholders*) en la actividad tecnológica se convierte en una exigencia de importante relevancia. Más allá de los muros de los laboratorios y las academias se encuentran infinidad de ciudadanos que reivindican una participación en aquellos proyectos que comprometen sus vidas. A partir de esta consideración comienza a concebirse la tecnología como un sistema sociotécnico inserto en

contextos sociales. En ese sentido, una breve mención de la teoría de los *stakeholders* nos permitirá apreciar la importancia de la participación de los grupos de interés de la actividad tecnológica, mediante una comparativa con el mundo empresarial.

En el ámbito empresarial y económico se ha avanzado mucho en la última década en torno a la cuestión de fundamentar una ética aplicada. Desde la Responsabilidad Social Corporativa (RSC) se ha entendido la necesidad de incorporar un diálogo amplio a este ámbito con el fin de esclarecer y delimitar la extensión de las responsabilidad en ese sector. En ese sentido, recientemente se ha propuesto la teoría de los *stakeholders* como una nueva metodología de gestión en la empresa, entendiendo que desde ella es posible construir puentes de diálogo entre la ética empresarial y la realidad, pasando así de la teoría a la *praxis*. Para Elsa González Esteban existen tres razones de peso que pueden llevar a considerar la importancia de la teoría de los *stakeholders* como un buen punto de partida desde el que cultivar una responsabilidad dialogada en el ámbito empresarial:

1. Esta teoría, en primer lugar, permite pensar un nuevo paradigma empresarial, donde existe una comprensión de la empresa plural. Por tanto, la empresa no es cosa de uno (accionista o propietario), ni exclusivamente de dos (propietarios y trabajadores), sino que la empresa debe ser entendida desde la pluralidad de «agentes» –los que afectan– que intervienen en ella y, por tanto lo hacen posible, así como desde todos aquellos «pacientes» –los que son afectados– por la organización empresarial.

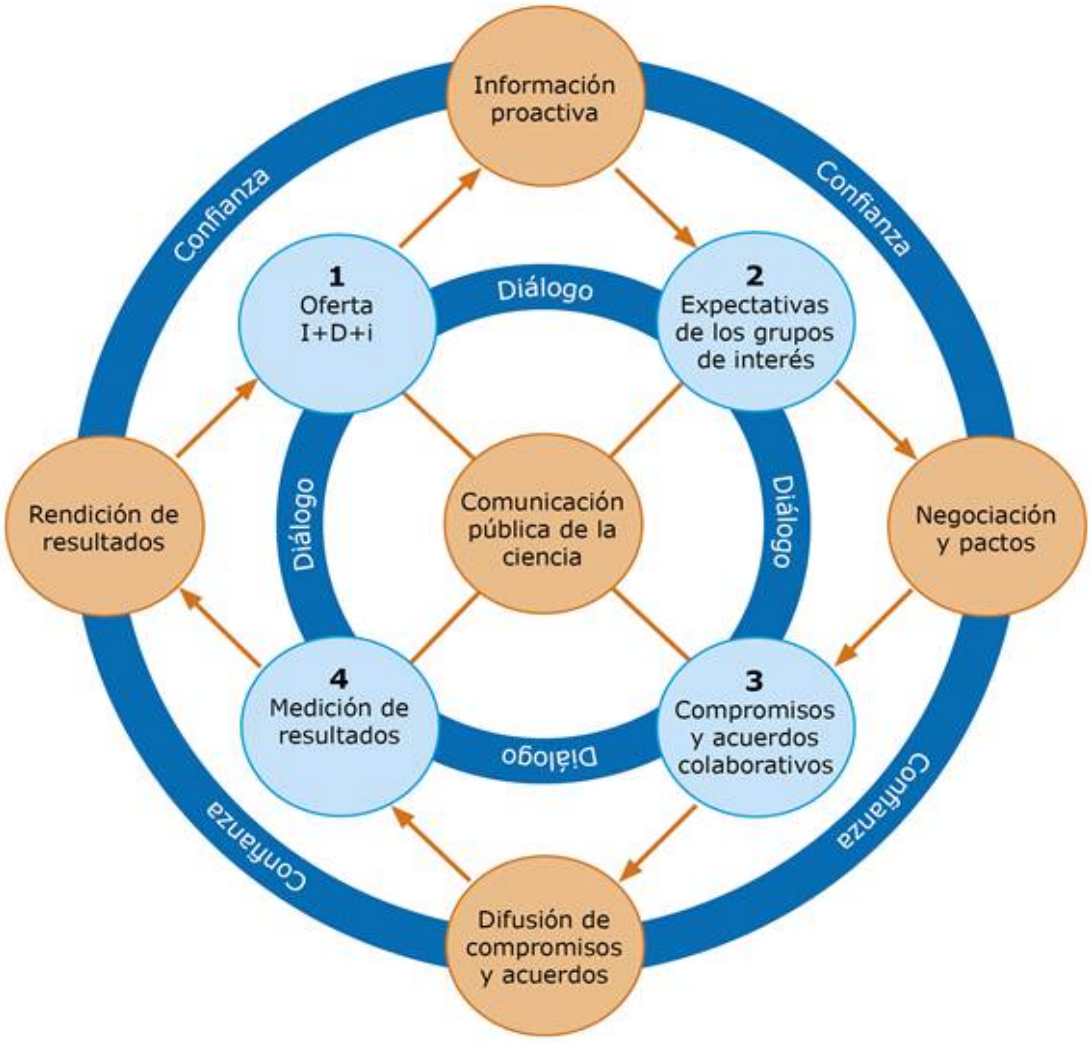
2. En segundo lugar, la teoría nos permite además comprender que entre los distintos grupos de interés que configuran la empresa se establecen una serie de relaciones que pueden ser entendidas desde la perspectiva no sólo del contrato jurídico o del contrato social, sino del contrato moral. Es decir, entre estos grupos de interés que configuran la organización empresarial existen expectativas recíprocas de comportamiento, algunas de ellas con carácter legítimo que deberían ser satisfechas desde dentro del marco de las relaciones empresariales. Por tanto, nos muestra como no son sólo intereses económicos sino también de otro tipo.

3. En tercer lugar, y derivado del anterior, mediante la teoría de los *stakeholders* es fácil vislumbrar la existencia de una responsabilidad social, entendida en sentido ético, de la organización empresarial (2007: 208-209).

La adaptabilidad práctica es necesaria en los procesos deliberativos, y la información y la participación deben establecerse de forma anticipada, con tiempo suficiente para que los testimonios de los afectados e interesados sean relevantes en el diseño de los productos. La RRI se sitúa dentro de dos dimensiones sumamente relevantes para el impacto del trabajo disciplinario, a saber, la social y ética. Este tipo de innovación se presenta como un cambio considerable en la forma de concebir la generación de conocimiento. En primer lugar, es esencial comenzar a pensar sobre la necesidad de incorporar aspectos valorativos y consideraciones morales en la práctica tecnológica. En segundo lugar, implica también una actitud de apertura a la participación de los grupos de interés y a la transparencia de la información. El MIAR representa una medida ética de responsabilidad en el ejercicio de un sector tan sumamente complejo como el de la IA y con un gran impacto en diversas esferas de la vida humana en las próximas décadas.

El desarrollo de la RRI demanda una reflexión en torno a un marco ético que proporcione referentes morales para la actividad tecnológica en el campo de la IA. La RRI construye un vínculo entre la aceptabilidad y deseabilidad de los procesos de investigación e innovación y sus resultados, así como con la participación de los grupos de interés y la interacción los mismos. Es fundamental para la gestión de la RRI promover una ética aplicada que impulse una preocupación por el diálogo, garantizando la imparcialidad del juicio moral y un punto normativo de referencia para las voluntades, tanto individuales como colectivas (Fernández-Beltrán *et al*., 2017: 1044).

Francisco Fernández-Beltrán *et al*. proponen un modelo de RRI fundamentado en la ética dialógica y la teoría de los *stakeholders* para fortalecer la comunicación, que se ilustra en el siguiente diagrama:



Fuente: Fernández-Beltrán *et al*., 2017: 1053.

El MIAR sitúa la comunicación pública de la actividad científica y tecnológica como una tarea fundamental para alcanzar los objetivos de la RRI. De ese modo se promueve un cultivo de la interacción continua entre los subsistemas integrados en la quíntuple hélice –mercado, Estado, comunidad, universidad y medio ambiente–.

En este modelo la innovación social juega un papel relevante, pues nutre al proceso de conocimientos enriquecedores para dar respuesta a las demandas cívicas. En ese sentido, la Comisión Europea afirma que:

> La innovación no es sólo un mecanismo económico o un proceso técnico. Es sobre todo un fenómeno social. A través de ella, los individuos y las sociedades expresan su creatividad, necesidades y deseos […] La innovación puede y debe ofrecer una respuesta a los problemas cruciales de la actualidad. Esto hace posible una mejora en las condiciones de vida (los nuevos medios de diagnóstico y tratamiento de las enfermedades, la seguridad en el transporte, más fáciles de comunicación, un medio ambiente más limpio, etc.) (Comisión Europea, 1995: 11)

Esta visión de la innovación social introduce en el MIAR una nueva concepción de la innovación tradicional en la empresa y en la economía, y en lo que se refiere a sus procesos y productos, no se limita al emprendimiento social, sino que va más allá del mismo y se sitúa en la senda de valores éticos como la libertad, la igualdad, la solidaridad o el diálogo, entre otros (Lozano Aguilar, 2011: 62-70). No limita su actividad a la aplicación de la tecnología los ámbitos de exclusión social, tampoco simplifica su actividad reduciéndola a una simple metodología de participación y creatividad. Le confiere mucha importancia a la responsabilidad social, promueve la disolución de la fronteras entre el diálogo y la cooperación que subyacen en las relaciones entre los sectores público, privado y otras organizaciones sin fines de lucro, involucra a los afectados y beneficiarios, nutriéndose de ese modo de sus experiencias, se desarrolla bajo un enfoque integral y holista, incorporando diversas problemáticas de la complejidad social y medioambiental e influye en el fortalecimiento de las alianzas entre las esferas que se encuentran presentes en la quíntuple hélice, que se abordará más adelante (Bureau of European Policy Advisers, 2011; Comisión Europea, 2010; Morales, 2008, 2009a, 2009b, 2012; Phills, 2008; Taylor, 1970).

5.4.2.1. Dimensiones de la investigación e innovación responsable

En el interior de la RRI existen cuatro dimensiones que pretenden fundamentar de manera responsable la práctica científica. Esas dimensiones son: anticipación, reflexividad, inclusión y sensibilidad (Stilgoe *et al*., 2013: 1570-1573).

- Anticipación

La mejora de la anticipación de la gestión pública encuentra su origen en una serie de fuentes bibliográficas (Wynne, 1992, 2002; Jasanoff, 2003; Henwood y Pidgeon, 2013) que esbozan sus preocupaciones políticas y medioambientales a partir de una crítica a los modelos de generación de conocimiento de arriba hacia abajo, monopolistas y oligopolistas. A menudo algunos proyectos tecnológicos tienen consecuencias imprevistas y en ocasiones los estudios sobre los riesgos suelen carecer de acierto. En ese sentido, la anticipación brinda a los investigadores y a las instituciones una serie de alertas tempranas que permiten diferenciar lo plausible de lo no plausible, lo probable de lo no probable, y en definitiva unos mejores conocimientos sobre los aspectos contingentes. A la vez, la anticipación permite fortalecer los mecanismos de innovación al contar con mayores conocimientos.

Algunos especialistas identifican una serie de técnicas de anticipación y discusión sobre posibles y deseables escenarios futuros: compromiso público creciente (Wilsdon y Willis, 2004), evaluación tecnológica constructiva (Rip *et al*., 1995), y evaluación tecnológica en tiempo real (Barben *et al*., 2008; Karinen y Guston, 2010).

- Reflexividad

La reflexividad es un aspecto fundamental de todo ejercicio de responsabilidad en sentido amplio y también una condición contemporánea de la actividad científica (Beck, 1992; Lynch, 2000). El poder de la tecnología que señalaba con tanto empeño Jonas (1995), implica un necesario ejercicio de reflexividad para la medición de los impactos y para la generación de conocimiento innovadores. Debido a esta dimensión de poder, la reflexividad asume una responsabilidad pública (Wynne, 2011) que requiere de participación de los grupos de interés para la deliberación sobre los estándares y códigos de conducta. La reflexividad también contribuye a poner de relieve las posturas axiologicamente neutras que en ocasiones acechan la actividad científica y que en el fondo representan un obstáculo para la innovación.

- Inclusión

El cambio cultural en el ámbito de la RRI ha destacado la riqueza que tiene la incorporación de nuevos actores en la gestión científica y en la innovación como búsqueda de legitimidad (Irwin, 2006; Comisión Europea, 2007; Hajer, 2009). Como más adelante se señalará con mayor profundidad, la deliberación representa un ejercicio de importante valor para las cuestiones relacionadas con la ciencia y la innovación. La creación de espacios de deliberación para asuntos científicos es una cuestión fundamental que debe estar presente en todos las agentes gubernamentales en los distintos niveles de la toma de decisiones. (Chilvers, 2010). Existen diversos espacios para deliberación científica: foros, asociaciones de las partes interesadas, congresos, seminarios, encuestas, etc. Así pues, el cultivo de la deliberación enriquece la práctica científica y fortalece las habilidades cívicas y democráticas mediante las dinámicas participativas.

- Sensibilidad

Existen una serie de mecanismos mediante los cuales se pueden cuestionar las tres dimensiones mencionadas anteriormente. La RRI debe reunir la capacidad de poder cambiar de forma o dirección en función de las exigencias contextuales, entre las que se encuentran las necesidades de los grupos de interés, los valores públicos y las circunstancias cambiantes. En ese sentido, los sistemas de innovación deben mostrar sensibilidad frente a estas exigencias. Esa sensibilidad permitirá ajustar los cursos de acción y corregir el tono allí donde sea necesario (Collingridge, 1980). La capacidad de reacción y respuesta es necesaria para cultivar esta importante responsabilidad.

La siguiente tabla resume estas cuatro dimensiones:

Dimension	Indicative techniques and approaches	Factors affecting implementation
Anticipation	Foresight Technology assessment Horizon scanning Scenarios Vision assessment Socio-literary techniques	Engaging with existing imaginaries Participation rather than prediction Plausibility Investment in scenario-building Scientific autonomy and reluctance to anticipate
Reflexivity	Multidisciplinary collaboration and training Embedded social scientists and ethicists in laboratories Ethical technology assessment Codes of conduct Moratoriums	Rethinking moral division of labour Enlarging or redefining role responsibilities Reflexive capacity among scientists and within institutions Connections made between research practice and governance
Inclusion	Consensus conferences Citizens' juries and panels Focus groups Science shops Deliberative mapping Deliberative polling Lay membership of expert bodies User-centred design Open innovation	Questionable legitimacy of deliberative exercises Need for clarity about, purposes of and motivation for dialogue Deliberation on framing assumptions Ability to consider power imbalances Ability to interrogate the social and ethical stakes associated with new science and technology Quality of dialogue as a learning exercise
Responsiveness	Constitution of grand challenges and thematic research programmes Regulation Standards Open access and other mechanisms of transparency Niche management[a] Value-sensitive design Moratoriums Stage-gates[b] Alternative intellectual property regimes	Strategic policies and technology 'roadmaps' Science-policy culture Institutional structure Prevailing policy discourses Institutional cultures Institutional leadership Openness and transparency Intellectual property regimes Technological standards

Fuente: Stilgoe *et al*., 2013: 1573.

5.4.2.2. La evaluación tecnológica y los aspectos éticos cruciales

Los rápidos avances en el mundo de la tecnología han destacado la necesidad de enfrentar los riegos, oportunidades y efectos secundarios que conllevan determinados proyectos. La pertinencia del MIAR surge ante la existencia de problemáticas que tienen una dimensión moral y que por lo tanto representan un desafío ético. Este modelo es una respuesta sistemática ante los desafíos que presenta la evaluación tecnológica. Los nuevos progresos tecnológicos demandan nuevos enfoques cada vez más innovadores que tengan en cuenta una diversidad de perspectivas y las circunstancias concretas de los contextos donde se originan. El modelo entiende que el desarrollo de nuevas ideas responsables es presentado como una condición de posibilidad para el futuro de la tecnología. Se caracteriza por la búsqueda de un equilibrio entre el rendimiento que es perseguido en las investigaciones de los laboratorios y el impacto que dichas investigaciones tienen posteriormente. Su lógica de funcionamiento es un continuo ejercicio de reflexión→evaluación.

Este modelo también se caracteriza por la introducción de la alerta temprana como condición esencial, lo que Armin Grunwald (2014: 16) denomina «compromiso temprano». Esta alerta es sinónimo de un compromiso anticipado y prudente en el despliegue histórico frente a los impactos futuribles. Es algo similar a lo que Jonas mencionaba cuando abordaba el tema de la heurística del temor de la siguiente manera: «resulta, pues, necesario elaborar una ciencia de la predicción hipotética, una «futurología comparada» (1995: 64). Ejemplos como la investigación que se refleja en el texto *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* son una clara muestra de la necesidad de llevar a cabo reflexiones que planteen escenarios futuros y previsibles. Es importante mencionar que, debido a las características del MIAR, cada vez son más los planteamientos en torno a esta nueva dinámica de investigación y generación de conocimiento. Un claro ejemplo es *Open AI*, una compañía de investigación sin fines de lucro que tiene como finalidad la promoción y el desarrollo de IA beneficiosa para la humanidad, y que gradualmente va incorporando más mecanismos de transparencia y participación.

La preocupación por la evaluación tecnológica, conocida en el mundo anglosajón como *Technology Assesssment* (TA), comenzó en la década de los sesenta (Bimber, 1996) como premisa para el asesoramiento político y la búsqueda de reforzamiento de la legitimidad de las prácticas políticas. En 1972 se firmó la Ley de Evaluación de la Tecnología en el Congreso de EE. UU. y la Oficina de Evaluación de Tecnologías (en inglés *Office of Technology Assessment*) se encargó de asesorar a los congresistas acerca de los posibles efectos que podrían derivarse a partir de la adoptación de decisiones políticas destinadas al desarrollo o a la introducción de nuevas tecnologías (Muñoz-Alonso López, 1997: 16). Los efectos secundarios que estaban presentando algunos proyectos tecnológicos sirvieron como motivación para promover la evaluación tecnológica. Algunas catástrofes tecnológicas como la explosión del trasbordador *Challenger*, el accidente nuclear de *Chernobyl*, el de *Three Miles Island*, etc., plantearon la necesidad de evaluar cuidadosamente los impactos de la tecnología. Grunwald menciona las motivaciones que, según él, han servido como detonante para la evaluación tecnológica (2014:19):

- Las preocupaciones de una tecnocracia emergente.
- Las experiencias tecnológicas en conflictos y déficit de legitimidad.
- La configuración de la tecnología de acuerdo a los valores sociales.
- Temas de innovación.
- Cambios en la comunicación social sobre las nuevas y emergentes tecnociencias.

Gemma Muñoz-Alonso López señala dos tipos de concepciones diferenciadas de evaluación tecnológica:

- La concepción reactiva que forma parte de la tradición americana surgida en la década de los 70, se centra en la identificación y valoración de los efectos sociales no deseados de las tecnologías, donde el objetivo es que los agentes responsables cuenten con la información suficiente para la toma de decisiones. Esta concepción derivó en una visión estrictamente economicista, basando sus análisis en la relación

riesgo-coste-beneficio, pasando así de una evaluación de los impactos de la tecnología a una evaluación de los riesgos. Posteriormente, y tras el dictamen de numerosos científicos y políticos, se puso de relieve que este modelo de evaluación no conseguía influir en la opinión pública ni tampoco en las actitudes de la sociedad en lo referente a la aceptación y valoración de determinadas tecnologías.

- La concepción constructiva propone que la evaluación tecnológica se centre en un mayor análisis de las problemáticas sociales y en una búsqueda continua de respuestas que pueden darse en el desarrollo de la tecnología, frente al anterior enfoque que pone su interés en las consecuencias no beneficiosas para la sociedad. Esta concepción se fundamenta en una actitud de carácter constructivo y activo, donde se lleva a cabo una convergencia entre la ciencia, la tecnología y la sociedad, dando lugar a asunción de compromiso democrático.

La evaluación tecnológica ha experimentado una transformación tras el aumento de la complejidad de los productos tecnológicos. El modelo promueve el diálogo de forma inclusiva, conectando así la tecnología con los valores y las creencias, algo que también se encuentra presente en el espíritu de la ciencia cívica. En ese sentido revitaliza el papel de una política participativa y pone en valor a la comunidad por medio de la acción cívica. Esta evaluación de la tecnología es entendida en la actualidad como «un conjunto de métodos que analizan los diferentes y diversos impactos o efectos derivados de la aplicación de tecnologías, estudiando los efectos de posibles tecnologías alternativas e identificados los grupos sociales que puedan verse afectados. Su objetivo último estriba en tratar de reducir o anular los efectos negativos de algunas tecnologías imperantes, optimizando sus efectos positivos y contribuyendo así a su aceptación por la sociedad» (Muñoz-Alonso López, 1997: 16).

Audley Genus (2006) se sitúa en la estela de la concepción de la evaluación tecnológica constructiva (CTA, siglas en inglés de *constructive technology assessment*) pues concibe esta evaluación como un proceso democrático, reflexivo y discursivo, donde es fundamental la participación de una amplia gama de actores para facilitar el aprendizaje

social sobre la tecnología y sus impactos sociales. El enfoque participativo es propuesto para recoger los testimonios e intereses de los actores que representan valores diferentes. La importancia de la participación reside en la necesidad de proporcionar enfoques reflexivos que procedan de diferentes ámbitos para alimentar una óptica de mayor amplitud. Antes también se comentó el valor que está presente en el encuentro de diversas perspectivas en los laboratorios abiertos o laboratorios ciudadanos. Así pues la reflexividad es un aspecto clave para poner de relieve el carácter político que se encuentra implícito en toda actividad científica (Genus, 2006: 14).

Arie Rip (2001a, 2001b) destaca las cuatro características principales de la evaluación tecnológica constructiva:

- La integración de la anticipación de los futuros efectos de la tecnología en la promoción e introducción de la tecnología, es decir, que los actores involucrados participen activamente en las actividades de diseño y desarrollo.

- La inclusión de más actores sociales y aspectos tecnológicos durante el desarrollo y la introducción de la tecnología con el fin de mejorar la calidad de la tecnología en el ámbito social.

- La modulación y adaptación debe ser vista como permanente, permitiendo que todos los actores puedan aprender sobre posibles nuevos vínculos entre las opciones de diseño y las demandas y preferencias de los usuarios previstos. El aprendizaje debe incluir aspectos propios de la articulación política y social en la dimensión de la aceptabilidad de la tecnología en el desarrollo y su vinculación con los valores culturales más amplios de la sociedad.

- Los agentes deben reflexionar sobre los procesos de coevolución de la tecnología y la sociedad, y también sobre la tecnología y sus impactos.

Genus pone el acento sobre la reflexividad como una dimensión fundamental de la evaluación tecnológica constructiva, pues tiene implicaciones tanto para la teoría como para la praxis. Además, la importancia de esta concepción evaluativa recae sobre su dimensión democrática al trasladar la reflexividad sobre la tecnología al ámbito social, ampliando su ejercicio más allá de los laboratorios y la visión de los expertos. La idea subyacente aquí se encuentra vinculada con el carácter político implícito en la tecnología. Frente a la simple aceptación y minimización de los conflictos derivados del desarrollo tecnológico, Genus propone más mecanismos democráticos para el diseño y despliegue, reflexividad y deliberación cívica (2006: 15). En ese sentido, la tecnología representaría una poderosa herramienta desde la que contribuir a cultivar habilidades cívicas y fortalecer la democracia.

Además, esta evaluación de la tecnología se encuentra motivada por un imperativo ético, por lo que vendría a llamarse evaluación ética de la tecnología (eTA, siglas en inglés de *ethical technology assessment*). La tarea principal de esta evaluación consiste en identificar posibles problemas éticos asociados con las nuevas tecnologías. Elin Palm T. y Sven Ove Hansson han elaborado una lista de verificación que puede servir como un sistema de alerta temprana para indicar la necesidad de evaluación anticipada y poner de relieve aspectos éticos cruciales (2006: 551-555):

1. Difusión y uso de la información
2. Control, influencia y poder
3. Impacto en los patrones de contacto social
4. Privacidad
5. Sostenibilidad
6. Reproducción humana
7. Género, minorías y justicia
8. Relaciones internacionales

9. Impacto en los valores humanos

5.4.2.3. La quíntuple hélice

Dentro del contexto de la IA el modelo de innovación de la quíntuple hélice se presenta como una novedad en lo que respecta a la incorporación de criterios de responsabilidad con una dimensión planetaria, pues surge a partir de una serie de fundamentos que permiten ir configurando una IAR. Es útil en la orientación de los diagnósticos y los planes de acción de forma responsable. Además, el trabajo participativo de la comunidad promueve la aportación de diversas perspectivas que proceden de esferas diferenciadas y a la vez permite prestar atención a nuevos factores y variables que normalmente no serían considerados desde posicionamientos monopolistas y oligopolistas, propios del modelo del triple hélice. Este modelo se plantea desde un encuentro de perspectivas multidimensionales.

El modelo de la quíntuple hélice recoge el espíritu de responsabilidad impulsado desde la cuádruple hélice. Esta responsabilidad nace de la toma de conciencia frente a la concepción que predominó hasta la década de los noventa que consideraba la tecnología desde una neutralidad axiológica. No existe ninguna tecnología exenta de consideración moral, ajena el mundo, pues es importante recordar que posee un carácter sociotécnico. Para Grunwald (2014: 23) la responsabilidad tiene tres dimensiones en el contexto tecnológico:

- Dimensión sociopolítica: el impacto está dirigido a la sociedad y su carácter político.
- Dimensión moral: los criterios y códigos morales forman parte del marco normativo que sirve para juzgar las acciones tecnológicas.
- Dimensión epistémica: centra su atención en la calidad de los conocimientos de los que se dispone para la evaluación.

Parecen muy interesantes las dimensiones del concepto de responsabilidad que plantea Grunwald, aunque esta visión debería ser enriquecida y complementada con una preocupación más amplia, de carácter planetario, que encuentre su origen en la contribución jonasiana orientada al problema ecológico. Los problemas medioambientales que enfrenta la humanidad, y que según informa el IPCC empeorarán considerablemente en las próximas décadas si no se toman medidas, plantean la necesidad de asumir un compromiso con la biosfera. Así pues, la responsabilidad que fundamenta el MIAR incluye una dimensión ecológica que implica el reconocimiento de los impactos de las acciones tecnológicas sobre la biosfera. A propósito del desafío que supone el deterioro medioambiental, algunos pensadores como Elias G. Carayannis y David F. J. Campbell (2012; 2014) hablan de un modelo de innovación de quíntuple hélice que invita a asumir este desafío con responsabilidad para la generación de conocimientos innovadores. En definitiva, los modelos de generación de conocimiento deben asumir un compromiso con el medioambiente y las exigencias que impone este tiempo (Stern, 2009). Por ello, el modelo de generación de conocimiento que fundamenta la propuesta de una IAR es la quíntuple hélice.

El modelo de la cuádruple hélice surge como una necesidad frente a las limitaciones democráticas de la triple hélice; la quíntuple hélice incorpora una visión aún más amplia e integral. La quíntuple hélice favorece la contextualización de la cuádruple hélice mediante una perspectiva que reconoce no solo la necesidad de comprometerse con las exigencias políticas de la comunidad, sino también el equilibrio medioambiental. En ese sentido, la Comisión Europea publicó en 2009 un documento bajo el título *El mundo en 2025: el surgimiento de Asia y la transición socioecológica*, donde alertaba de la transición socioecológica como uno de los retos más importantes de desarrollo para el futuro.

El aumento de los niveles de contaminación ha provocado un calentamiento global que representa serios problemas ecológicos que debe despertar una profunda preocupación. Esa situación impone una nueva responsabilidad con dimensiones de carácter planetario, donde la humanidad en su conjunto juega un papel fundamental en la prevención de nuevos conflictos políticos, sociales y medioambientales. El desafío ecológico puede abordarse a

partir de un desarrollo sostenible impulsado desde el ámbito «glocal», es decir, vinculando lo local y lo global, o como Agustín Domingo Moratalla señala, «pensar mundialmente y actuar localmente» (1991: 5). Este desafío debe ser percibido como una oportunidad desde la que plantear una manera de vivir innovadora y respetuosa con la biosfera. Elias G. Carayannis y David Campbell ponen de relieve la importancia que tiene la utilización activa del conocimiento humano como la clave del éxito para la innovación (2010: 42). Además, junto a Thorsten D. Barth señalan lo siguiente:

> El modelo de quíntuple hélice es un modelo de innovación que puede hacer frente a los retos actuales del calentamiento global a través de la aplicación de conocimiento saber-hacer que se centra en el intercambio social y en la transferencia de conocimiento dentro de los subsistemas de un Estado específico o Estado-Nación. El modelo de innovación de la quíntuple no es línea, y por ello combina el conocimiento, el saber hacer y el medio ambiente-sistema natural como un único marco interdisciplinario y transdisciplinario que proporciona un modelo con el que paso a paso se comprende la gestión de la calidad basada en el desarrollo efectivo para la recuperación del equilibrio con la naturaleza, y poder permitir para las generaciones futuras una vida de diversidad y pluralidad en la tierra (2012: 2).

El modelo de quíntuple hélice que sirve como un escenario teórico fértil desde el que plantear el concepto de IAR en el contexto del MIAR puede ilustrarse de la siguiente manera:

Como se observa en la ilustración, se añade una nueva dimensión al modelo de innovación anterior de la cuádruple hélice, el medio ambiente. La cuádruple hélice trataba de dar respuesta a las limitaciones de la triple hélice incorporando a la comunidad como una cuarta hélice desde la que recoger los testimonios de la sociedad civil y alcanzar un fortalecimiento de los medios de comunicación asociados a las industrias creativas, la cultura, los valores y los estilos de vida. A partir de esa nueva dimensión, este modelo encuentra la necesidad de incorporar una nueva hélice mediante las exigencias ecológicas que imponen los límites planetarios, asunto que se abordará más adelante. En ese sentido, el modelo de la quíntuple hélice recoge el espíritu de la cuádruple, pero lo contextualiza en los desafíos medioambientales, dotando a la responsabilidad de una dimensión planetaria y más integral.

Carayannis *et al*. señalan lo siguiente sobre el modelo de la quíntuple hélice a modo de resumen:

> Se trata de un modelo teórico y práctico para el intercambio del recurso del conocimiento, basado en cinco subsistemas sociales con «capital» a su disposición, con el fin de generar y promover un desarrollo sostenible de la sociedad. En este modelo de acumulación de la quíntuple hélice, el recurso del conocimiento se mueve a través de una circulación de subsistema-subsistema. Esta circulación del conocimiento desde el subsistema al subsistema, implica que el conocimiento tiene cualidades de entrada y de salida para los subsistemas dentro de un Estado (Estado-Nación) y también entre los Estados. Si una entrada de conocimiento es aportada desde uno de los cinco subsistemas, a continuación, una creación de conocimiento se lleva a cabo. Esta creación de conocimiento se alinea con un intercambio básico de conocimiento y produce nuevas invenciones como conocimiento de salida (2012: 6).

El modelo de la quíntuple hélice representa un espacio desde el que plantear una IAR, ya que asume las exigencias de este tiempo. Entre esas exigencias se encuentra la necesidad de fortalecer los mecanismos deliberativos a partir de un reconocimiento del valor de la participación de la sociedad civil como un recurso esencial para contribuir desde un mayor conocimiento de la realidad, con un carácter más aproximado que el proporcionado por

otras esferas. Además, incorpora un compromiso con el medio ambiente como una oportunidad, o en términos de Tomas Hellström, como un «nicho ecológico de innovación» (2007: 158) para la generación de conocimientos innovadores que aporten mejores modos de vida.

5.4.2.4. La deliberación como condición de posibilidad

No existe una teoría ética perfecta y absoluta, por eso es fundamental el ejercicio deliberativo. La deliberación se convierte en una importante herramienta para la construcción de un proyecto de ética aplicada a la IA de forma participativa, donde tengan cabida todos aquellos sectores que forman parte de los grupos de interés de la actividad: políticos, académicos, psicólogos, tecnólogos, etc. Una ética de este tipo nace de la posibilidad de reconocer perspectivas, pues su ámbito es contingente. Como señala Tomás Domingo Moratalla: «La deliberación es el método de análisis ético una vez que reconocemos la contingencia de los asuntos humanos y comprendemos que nuestro ámbito de acción y decisión es el de la incertidumbre y la complejidad» (2017: 41).

La actitud deliberativa implica la necesidad de tomar distancia de determinados «vicios» propios del modo de pensar lineal deductivo, donde todo se reduce a aspectos cuantificables y dilemas. Según Diego Gracia, si existe la voluntad para emprender la vía deliberativa, es necesario huir de los dilemas, pues «más que dilemas, hay problemas, es decir casos con múltiples cursos de acción posibles que será preciso tener en cuenta a la hora de tomar una decisión razonable o prudente» (2016: 13). Esto quiere decir que el reconocimiento de los problemas conduce a la toma de conciencia sobre la diversidad de perspectivas desde las que afrontar un caso. Sin embargo, los dilemas responden a una lógica binaria donde todo está bien o mal, es verdadero o falso, etc. En el proceso deliberativo los problemas no pueden convertirse en dilemas, pues el horizonte de posibilidades de discusión se reduce considerablemente y eso conduce a un empobrecimiento a la hora de enriquecer la construcción de una ética aplicada a la IA. No obstante, la tarea deliberativa no es fácil, ya que surge desde una formación vinculada al respeto activo, tal como sostiene Cortina, y que consiste «no solo en soportar estoicamente

que otros piensen de forma distinta, tengan ideales de vida feliz diferentes a los míos, sino en el interés positivo por comprender sus proyectos, por ayudarles a llevarlos adelante, siempre que representen un punto de vista moral respetable» (1998: 240).

Es importante destacar que la idea que actualmente existe de la deliberación ha sido producto de la confluencia de diversas tradiciones desarrolladas a lo largo de la historia y de las que Luis Vega Reñón (2016) destaca tres: en primer lugar, su origen antiguo; en segundo lugar, la contribución de la modernidad; y, en tercer lugar, el momento actual en el que la reflexión deliberativa se da dentro de un marco socioinstitucional del discurso público. Puesto que el conocimiento científico es considerado en este trabajo como un recurso público (Einsiedel, 2005) para la potenciación de las habilidades cívicas y las prácticas democráticas, el interés se enfocará en el tercer momento por hacer hincapié en el aspecto público.

Vega Reñón (2016: 219-220) pone como ejemplo de deliberación un tipo de diálogo y esquema argumentativo que tiene como objeto la resolución de problemas prácticos a partir de la inferencia medios-fines o actuación-riegos/consecuencias, propuesto por Douglas Walton (2004; 2006). La propuesta de Walton tiene un profundo carácter reflexivo, ya que parte de la preocupación por los planes de acción, los objetivos y su pertinencia, las consecuencias y riesgos, la plausibilidad, etc. Esta propuesta considera de gran importancia la labor crítica de la deliberación prudencial. Pero Vega Reñón va más allá y se sitúa en el terreno de aquellos problemas que suscitan una dimensión pública y un alcance colectivo. En ese sentido, el MIAR proporciona a la IAR un punto de partida desde el que considerar al conocimiento científico como un recurso público y por lo tanto preocupado por las cuestiones de alcance colectivo, las habilidades cívicas y el enriquecimiento de la democracia en términos generales. Vega Reñón caracteriza su propuesta de la siguiente manera:

> (i) el reconocimiento de una cuestión de interés y de dominio públicos, donde lo público se opone a lo privado y a lo privativo; (ii) el empleo sustancial de propuestas; (iii) las estimaciones y preferencias fundadas en razones pluridimensionales que remiten a consideraciones plausibles, criterios de ponderación y supuestos de congruencia práctica; (iv)

> el propósito de inducir al logro consensuado y razonablemente motivado de resultados de interés general −no siempre conseguido (2016: 220).

La IA tiene un profundo impacto sobre las vidas humanas, tanto en el ámbito privado como público. En ese sentido, el ámbito público se tomará con especial interés sobre el privado, por considerar que es un espacio común donde se da el encuentro y una mayor relevancia del impacto tecnológico. Por ejemplo, en el campo profesional existe un evidente desafío en lo que corresponde a la automatización del mundo del trabajo, fenómeno que tiene principalmente un fuerte impacto en el terreno de lo común y público debido a la generalización de sus efectos entre el conjunto de la ciudadanía. Así pues, en la deliberación, lo público primaría sobre lo privado, por supuesto, sin negar que también existen efectos de diversa índole en la esfera privada. Además, es importante recordar que la IAR promueve los ODS que versan sobre asuntos de interés público.

En cuanto al empleo sustancial de propuestas, Vega Reñón define una propuesta como:

> Una unidad discursiva o un acto de habla directivo y comisivo del *telos* de lo «lo indicado [pertinente, conveniente, debido, obligado] en el presente caso es hacer [no hacer] X». Se refiere a una acción y expresa una actitud hacia ella. Así pues, envuelve tanto ingredientes prácticos como normativos y no se deja reducir a un mejor «bueno, hagamos X» […] También puede verse como la conclusión de un razonamiento práctico en la medida en que el proponente está dispuesto no solo a asumir lo que propone sino a justificar su propuesta, o llegado el caso, a defenderla (2016: 221).

El MIAR propuesto puede formularse sobre la premisa que Domingo García-Marzá considera como fundamental dentro del ámbito empresarial, a saber, la metodología reconstructiva. Es importante destacar que existen evidentes diferencias entre el espacio en que se desarrolla la actividad empresarial y el espacio en que se despliega el modelo de innovación abierta aplicado a la IA, pero se trata de un planteamiento útil. Esta metodología se sostiene sobre la afirmación de que no son los humanos quienes inventan los criterios morales que juzgan sus actividades, sino que más bien esos se encuentran incrustados en el lenguaje moral (García-Marzá, 2011: 83). En las actividades realizadas se encuentra

presente la capacidad para valorar moralmente, y esa capacidad se percibe en el lenguaje utilizado. García-Marzá señala lo siguiente:

> Expresamos así nuestra opinión, nuestra convicción, sobre lo que debe o no debe hacerse, sobre lo que estamos obligados y, en este sentido, puede ser exigido y esperado. Este lenguaje moral nos sirve para expresar sentimiento de indignación o culpa, de desprecio o atracción; nos sirve para valorar la reputación de una persona o la actuación de una empresa como buena o mala y nos sirve, por último, para exigir a los demás y esperar de ellos un determinado comportamiento, sean actores individuales o institucionales. Lo importante de este lenguaje moral consiste en que estas razones orientan nuestra conducta y guían la voluntad hacia la aceptación o el rechazo de un producto o servicio, hacia el compromiso con o la ruptura de una determinada relación […] Esta es la característica básica de toda conducta humana y de todas las instituciones que estructuran el orden social: están construidas sobre razones, y sin ellas, por más equivocadas que puedan estar, este orden carecería de sentido (2011: 84).

El modelo de innovación abierta aplicado a la IA incorpora el diálogo reconociendo que han de incluirse todos los sectores implicados y/o afectados en la toma de decisiones y en la actuación. La aportación de esta metodología reconstructiva implica el reconocimiento de otros intereses y perspectivas posibles, y ofrece la posibilidad de caminar más allá del egoísmo de una determinada posición y orientar la mirada hacia la búsqueda de aquello que es «bueno para todos por igual» (Habermas, 2000: 177). Para García-Marzá la aportación de la ética aplicada consiste en lo siguiente:

> La perspectiva ética implica siempre, por volver a nuestro lenguaje de intereses, tener en cuenta todos los intereses involucrados en la acción además de los propios. Mejor dicho: consiste en hacer míos los intereses de los demás implicados. Esto no significa necesariamente renunciar a nuestro propio interés. Significa buscar entre todos un interés más amplio que recoja los diferentes intereses en juego. De ahí el valor moral del diálogo como proceso deliberativo para la búsqueda de intereses comunes y para la resolución imparcial de conflictos de acción en el caso de intereses contrapuestos (2011: 91).

En ese sentido, el diálogo representa una importante herramienta para la ruptura de los dogmatismos propios del modelo de innovación cerrada que pudieran permanecer en la quíntuple hélice del MIAR. Dicho diálogo se despliega sobre el respeto a los demás, estableciendo puentes entre diferentes perspectivas e intereses. El reconocimiento intersubjetivo se construye en base a una relación de confianza entre las partes implicadas en el modelo sostenido sobre un ejercicio deliberativo.

Más allá de las reflexiones procedentes del pragmatismo, que encuentran su razón de ser en la realidad más cercana, se halla la exigencia de buscar respuestas universalmente válidas ante las problemáticas que enfrenta la humanidad. Dentro del escenario global la ética del discurso se convierte en un interesante corpus teórico de fundamentación desde el que enriquecer la dimensión universal que debe primar en el ejercicio del MIAR. La universalización proporciona al diálogo un valor moral para el logro de acuerdos o consensos sobre la base de unas normas que están en discusión en el contexto de un equilibrio de intereses (García-Marzá, 2011: 103). Para que el diálogo pueda llevarse a cabo en la búsqueda de acuerdos o consensos es necesario que se cumplan con una serie de condiciones: igualdad de derechos de participación, ausencia de coacciones, ya sean internas o externas, inclusión de todos los afectados, etc. (Habermas, 2000).

Por lo tanto debe existir un ineludible diálogo entre las esferas que conforman la quíntuple hélice mediante el encuentro en espacios comunes de hibridación que representan el establecimiento de puentes entre esferas diferenciadas y con campos de acción caracterizados por la diversidad. El conocimiento científico presente en el campo de la IA es un recurso público y como tal presenta una importante significación de impacto generalizado y compartido. Por ello, es fundamental impulsar una IAR en el contexto de laboratorios abiertos sobre ciencia cívica reconociendo el valor del diálogo y la hibridación en la quíntuple hélice del MIAR.

5.4.2.5. El valor de la participación y el papel de la sociedad civil

La superación del modelo de innovación cerrada, o de triple hélice, ha suscitado en la sociedad del conocimiento la aparición de una nueva esfera más allá del Estado, el mercado y la academia, a saber, la comunidad (Urra Canales, 2017: 193). La comunidad se presenta como un conjunto de personas que comparten un espacio geográfico, intereses y características (Urra Canales, 2017: 194). El ser humano es un ser social y vive en comunidad desde su lejano origen. El concepto de comunidad ha sido objeto de sendas críticas a partir de aspectos que tienen que ver con el autoritarismo o el conservadurismo, entre otros. No obstante, estas críticas han ido perdiendo su valor cuando el concepto de comunidad se ha relacionado con el de sociedad civil. El término «sociedad civil» encuentra su origen en el latín en *societas civilis* y hace referencia a un conjunto de ciudadanos que son miembros de una comunidad y que poseen ciertos derechos que les permiten participar en la vida pública. El tema de la sociedad civil y la comunidad ha sido objeto de reflexión en la filosofía a lo largo de su historia, comenzando con la *Política* de Aristóteles, los contractualistas clásicos –Hobbes, Locke, Rousseau–, y otros autores, como Kant, Hegel, Marx o Gramsci. Salvando las diferencias entre estos planteamientos, hay un punto de encuentro que permite dar cuenta de que el ser humano es un ser que vive en comunidad.

El contexto social de desarrollo de la IA es el de la sociedad del conocimiento, que es también el escenario en el que se despliega en ejercicio la sociedad civil. Las comunidades se organizan en tres esferas básicas, a saber, el Estado, el mercado y la sociedad civil. En su obra *Developing Democracy: Towards Consolidation* Larry Diamond (1999: 223-227) define a la sociedad civil a partir de cinco aspectos:

1. La sociedad civil centra sus esfuerzos en los fines públicos y no en aquellos que son estrictamente privados.

2. Establece una relación con el Estado pero no pretende reemplazarlo en materia de control, no buscando de esa manera gobernar por sí misma, sino más bien lograr influencia en los poderes políticos institucionales.

3. Se construye a partir del respecto a la pluralidad y la diversidad. La sociedad civil no promueve una perspectiva holística, sino que incorpora los intereses que emanan de los diversos grupos, no representando de ese modo ningún interés en concreto, ya se deriven de una persona o una comunidad.

4. El concepto de sociedad civil se diferencia del de «sociedad cívica», ya que el segundo hace referencia a la cooperación y la reciprocidad voluntaria entre individuos sin la necesidad de trascender socialmente.

Además de Diamond existen otros pensadores del ámbito de la teoría política como Jean Cohen y Andrew Arato (2000) o Salvador Giner (2008), que también esbozan una definición del concepto de sociedad civil para dar cuenta de aquellos aspectos más importantes que la caracterizan. Salvando las diferencias entre los autores, pueden encontrarse puntos de encuentro que ayudan a entender mejor esta nueva hélice que está presente en el MIAR. La sociedad civil hace referencia a aquellas sociedades que actúan de un modo colectivo en el ámbito público, sin que necesariamente exista un control directo de estructuras de gobierno o empresas. La historia ha evidenciado que aquellas comunidades que se organizan a través de la fórmula del asociacionismo tienen un gran potencial transformador y de cambio social, convirtiéndose así en sujetos sociales activos en el desarrollo y la innovación del conocimiento.

Una vez que el concepto de sociedad civil ha sido esbozado, es necesario presentar cuál puede ser la contribución de la sociedad civil en la sociedad del conocimiento para jugar un papel relevante y equilibrado a las otras esferas del MIAR.

Antes se indicó la importancia de que el proyecto de una IAR sirviera como punta de lanza para promover los ODS, siendo fundamental la participación ciudadana como un mecanismo para reflexionar acerca de la naturaleza de la democracia y las habilidades

cívicas. La participación adquiere un importante valor dentro del contexto del MIAR, ya que representa una posibilidad transformadora para abordar las problemáticas y gestionar los asuntos públicos de un modo innovador.

John Gaventa (2006) plantea seis desafíos que sirven para valorar la importancia que tiene la inclusión de la ciudadanía en los asuntos públicos y sus posibilidades transformadoras. Dentro de la propuesta que aquí se está esbozando, sobre el valor de la IAR para el enriquecimiento del civismo y la democracia, pueden destacarse los siguientes:

- Relacionar a la gente con las instituciones.
- Repensar las relaciones entre la sociedad civil y las instituciones políticas.
- Reflexionar sobre el valor de la participación ciudadana.
- Contar con mayor conocimiento para entender las relaciones de poder.

Siguiendo a García Inda (2003), si en la teoría política contemporánea ha sido el conjunto de derechos y deberes lo que ha configurado los límites del ciudadano, la paulatina degradación del sistema de democracia representativa liberal y la creciente conflictividad social, incapaz de satisfacer y hacer efectivos el sistema de derechos, ha empujado a los científicos sociales a reformular el sentido de ciudadanía. En esta línea es la participación del sujeto social y político lo que vendría a profundizar y a hacer efectivos el conjunto de derechos y deberes ciudadanos. Se trata de participar en la construcción de las reglas de juego –participación política–, así como en la producción, distribución y control de los bienes de la comunidad política, económicos, sociales, políticos y culturales. Es decir, el sujeto ciudadano-participativo pasaría de ser un mero titular pasivo de enunciados jurídicos con más o menos posibilidades de materialización, a construir y ampliar activamente la implementación de los mismos. Se superan así los tradicionales derechos civiles que otorgan en una democracia representativa el estatus de ciudadano identificado con el derecho de asociación y el de representación activa y pasiva a través del voto.

En contraposición, emerge una ciudadanía activa que se constituye como sujeto político a través de la participación social y política con voluntad de transformación desde las experiencias de vida (Cepeda Mayorga, 2017). Durante las últimas décadas ha destacado la gestación de un ciudadano nuevo, no solo para la reivindicación de cambios, sino para la participación en los mismos a través de nuevas organizaciones sociales y políticas con vocación internacionalista, formas alternativas de solidaridad y cooperación para la solución de problemas sociales que constituyen el germen de la construcción de la democracia participativa. En el contexto de una ciudadanía activa, Benjamin Barber (2004) propone lo que él denomina «democracia fuerte» frente a «democracia blanda». Sostiene que una participación viva de la sociedad da lugar a una democracia sólida, a una democracia fuerte. En las democracias fuertes los miembros de la sociedad tienen aspiraciones y sueños colectivos. Las relaciones sociales se establecen en condiciones de confianza y de cooperación corresponsable. Una democracia fuerte es una democracia con una ciudadanía activa y fuerte. La democracia fuerte pasa así a ser entendida como el mecanismo por el cual la ciudadanía busca resolver los problemas sociales por medio de la participación y el autogobierno responsable. Es una manera de rescatar el valor político que tiene la comunidad y de ponerlo en funcionamiento. Restaura el valor de la ciudadanía y lo eleva para la gestión de los asuntos públicos, situándose en el centro del propio proceso democrático el concepto de ciudadanía. Por ello, la democracia fuerte ofrece más participación, pero una participación consciente que plantea propuestas mutables y manipulables. Lo que es mutable y manipulable adquiere forma porque es la ciudadanía la que le va dando forma según sus intereses. Los valores no son inmutables ni tienen orígenes abstractos, sino que son sometidos a la acción común por medio de la deliberación y la participación que los dota de diferentes formas y direcciones según las necesidades y los tiempos.

Otro postulado que gira en torno a los conceptos de participación y ciudadanía es el de Carole Pateman (1970; 1985). A continuación es pertinente señalar unas líneas en las que Pateman menciona algunos rasgos fundamentales de su teoría política participativa:

> […] la teoría de la democracia participativa está constituida alrededor del principio central que los individuos y sus instituciones no pueden considerarse aisladamente unos de otros. La existencia de instituciones representativas a nivel nacional, no son suficiente para que haya participación democrática. Para alcanzar la máxima participación de todos, esto es, que exista participación en la base de la sociedad, ésta debe ubicarse tanto en los niveles institucionales como en otras esferas, como capacitación y entrenamiento social (*social training*) para la democracia, de ese modo podrán desarrollarse las necesarias actitudes individuales y las cualidades psicológicas (1970: 42).

Para Pateman (1985) en el contexto liberal no es posible desarrollar esas habilidades participativas debido a las limitaciones que presenta este sistema político. Esta autora cuestiona los principios de igualdad y libertad de los sistemas democráticos liberales. Entiende que en esos sistemas los más desfavorecidos y vulnerables no poseen las mismas oportunidades para la participación política que los que presentan un mayor estatus socioeconómico. En cambio, Pateman afirma que los modelos participativos proponen un escenario con mayores posibilidades para la ciudadanía en su conjunto. Formar parte de la toma de decisiones por medio de la participación política se convierte en una expectativa real para la ciudadanía en el contexto de su propuesta política.

En torno a la idea de fortalecimiento de la ciudadanía, Adela Cortina realiza un aporte interesante en este campo, señalando que es necesario redefinir el concepto mismo de ciudadanía, haciendo una síntesis de la concepción liberal y la concepción comunitarista. Sin duda, es imprescindible pertenecer a una comunidad, es decir, desarrollar unos valores, estando en un lugar. En su concepto incluye dos dimensiones para la constitución de ciudadanía:

> El lado «racional», el de una sociedad que debe ser justa para que sus miembros perciban su legitimidad, y el lado «oscuro», representado por esos lazos de pertenencia, que no hemos elegido, sino que forman ya parte de nuestra identidad. Ante los retos ante los que cualquier comunidad se encuentra es entonces posible apelar a la razón y al sentimiento de sus miembros, ya que son ciudadanos de esta comunidad, cosa suya (1998: 34).

Por tanto, siguiendo a Cortina, este nuevo concepto de ciudadanía ya no estaría ligado a las connotaciones del Estado-nación sino que lleva necesariamente unidas las dimensiones de la racionalidad de la justicia y el sentimiento de pertenencia ligado a la identidad, la cual no está necesariamente unida a un territorio, sino a un conjunto de relaciones sociales, ya que la identidad es un concepto relacional y multidimensional. Según Cortina, es en las comunidades concretas donde se va forjando la identidad en medio de una diversidad de otros que se van reconociendo. Así mismo, el fenómeno de la inmigración pone de manifiesto también el carácter de multidimensionalidad de la identidad ligada a una pluralidad de pertenencias.

Manfred A. Max-Neef (1998) considera que existe un sistema de necesidades humanas que combinan categorías axiológicas con categorías existenciales. Dentro de las categorías axiológicas se encuentran la subsistencia, la protección, el afecto, la comprensión, el recreo, la participación, la identidad y la libertad, mientras que dentro de las categorías existenciales se encuentran algunas necesidades como las de ser, hacer y relacionarse. Están vinculadas de una forma transversal, en tanto que, si se satisface una, se contribuye a satisfacer otra. La participación sustenta la vida comunitaria que a su vez otorga sentido de pertenencia e identidad a los sujetos, categorías que como muy bien afirman Max-Neef son relacionales. Resulta muy difícil actuar, proponer, criticar, decidir y disfrutar sin sentirse parte de una comunidad, por tanto son axiomas constituyentes de ciudadanía.

La participación política dibuja un nuevo sujeto político, la ciudadanía activa, que participa activamente, que decide, que adquiere compromiso y responsabilidad . Este sujeto político se dibuja en el mundo globalizado de la comunicación tecnológica, de las crisis económicas y ambientales, de la tecnocracia, de las corrupciones favorecidas por la opacidad del sistema político, en medio de la deslegitimación del sistema político de la democracia representativa liberal, de las convulsiones y los movimientos sociales.

5.4.2.6. Racionalidad pragmática y fronética

John Forester (1993) parte de la teoría de la acción comunicativa de Habermas (2010) para desarrollar un pragmatismo crítico que se encuentra enfocado en la planificación y las políticas públicas. Forester observa la planificación como el resultado de una reestructuración de la comunicación entre los grupos de interés de un conflicto, o *stakeholders*, que se encuentran en una situación de desigualdad en materia de poder e influencia. En ese sentido, para Forester la planificación es el resultado de una actividad caracterizada por la interactividad y la comunicación, en oposición a un modelo de racionalidad técnica y de análisis sistemático, propiamente del modelo de triple hélice. La complejidad de la realidad debe ser asumida desde una investigación de carácter cualitativo que consista en interpretar y comprender los rasgos contextuales, más allá de la aplicación de reglas que sean producto de un esquema generalista y encorsetado para la práctica. Forester acaba distanciándose del marco procedimental, que fomenta el discurso racional y formalista de Habermas que es escenificado en una situación discursiva ideal, y se decanta por un tratamiento contextual, histórico y sustantivo que responda a las complejidades de la esfera pública.

Forester (1999) se sitúa en la senda del reconocimiento de la riqueza y los límites que proporcionan aquellos procedimientos deliberativos que participan de la planificación y que se contextualizan en las experiencias del trabajo de campo. El valor pragmático del postulado de este pensador reside en el reconocimiento de que la reflexión debe someterse a una valoración enraizada en el accionar, recogiendo los testimonios y las experiencias rentables para el aprendizaje. Este pragmatismo de carácter crítico ha contribuido al desarrollo de dinámicas participativas que han permitido democratizar el conocimiento y contribuir al desarrollo de las comunidades. En ese sentido, la planificación debería dibujarse en un espacio de comunicación con especial énfasis en la escucha. Es de esperar que se presenten importantes diferencias que sean fruto de los grupos de interés involucrados en la acción tecnológica, aunque para Forester (2009) es posible que se establezcan puntos de encuentro comunes. El pragmatismo podría impulsar un terreno fértil

surgido desde la experiencia para ir construyendo de forma deliberada y participativa aquellos conceptos que van siendo reconocidos y que cuentan con legitimidad por parte de los grupos de interés.

Otro pensador que toma distancia de la rigidez procedimental de los planteamientos habermasianos y que fundamenta su postulado en la sustantividad del pragmatismo es Bent Flyvbjerg (2001; 2004; 2006a; 2006b). Flyvbjerg ha formulado una interesante crítica al modelo cientificista que las ciencias sociales han desarrollado a partir del cientificismo propio de las ciencias naturales y ha propuesto un nuevo camino para reorientar el ejercicio de su actividad, el modelo fronético. Según él, existen dos modelos en las ciencias sociales (2006b: 39):

- El modelo epistémico: encuentra su origen en el modelo ideal de las ciencias naturales para el desarrollo del conocimiento. Las ciencias sociales adoptan este modelo entendiendo que es posible descubrir teorías y leyes en el ámbito social para la resolución de problemáticas.

- El modelo fronético: parte de una crítica al modelo epistémico que han adoptado las ciencias sociales por estar fuertemente determinado por un dogmatismo cientificista. Este modelo se fundamenta en el concepto de *phrónesis* que Aristóteles expone en la *Ética a Nicómaco*. Para Flyvbjerg es necesario que las ciencias sociales reconozcan que el valor de la racionalidad se encuentra en la deliberación y no en el modelo epistémico de las ciencias naturales.

Este reconocimiento es fundamental para la reorientación de las ciencias sociales hacia un escenario diferente en el que se asuma que la generación de conocimiento debe ser fruto del análisis reflexivo de valores e intereses y de su afectación a la sociedad y los grupos que la conforman. El modelo fronético pone en valor la deliberación pública y la sitúa en el centro del debate y la toma de decisiones, con el objetivo de garantizar la legitimidad de las medidas adoptadas entre todas las partes involucradas en el proceso de formación de conocimiento. Este modelo fronético encajaría perfectamente dentro de la propuesta de los laboratorios abiertos y el MIAR, ya que enriquecería el ejercicio de una IAR a partir de la

puesta en valor de la deliberación y la superación de una vía cientificista que rechaza otros factores que configuran la complejidad de la realidad humana.

El carácter participativo de la ciencia cívica y el MIAR no podría entenderse dentro del marco de un modelo epistémico y cientificista, ya que necesita partir de una ciencia social fronética que sepa reconocer el importante valor de la deliberación en el momento de producir un conocimiento innovador. La complejidad que presenta el tiempo tecnológico actual por la infinidad de factores que en él intervienen no puede pensarse sin tomar en cuenta la singularidad de cada contexto. En ese sentido, el objetivo principal de las ciencias sociales fronéticas consiste en la comprensión de la relación entre valores e intereses en la praxis. El modelo fronético se aleja así de la supuesta neutralidad de la que está revestido el modelo epistémico cientificista de las ciencias naturales.

El modelo fronético pone en valor el carácter problemático de los casos por encima de la metodología empleada por las ciencias naturales. La problematización de los casos resultantes de la complejidad real es abordada y reflexionada de un modo deliberativo. En cambio, para el modelo epistémico las problemáticas son esquivadas o rechazadas porque suponen un obstáculo para el espíritu cientificista. Las ciencias sociales fronéticas consideran el tratamiento de las problemáticas como una premisa ineludible dentro del ejercicio científico y por tanto de la generación de conocimiento. La búsqueda de patrones universales no forma parte del modelo fronético, sino más bien de la singularidad que determina los problemas concretos. Además, el carácter fronético proporciona al MIAR una impronta de continua apertura para la interpretación de las problemáticas, ya que se aleja de los planteamientos exclusivamente cuantitativos que en el terreno de la IA podrían identificarse como los algoritmos que se introducen en el universo humano. Este universo humano es sumamente complejo y no puede reducirse a planteamientos estrictamente cuantitativos.

Esta vinculación de las ciencias sociales con la *phrónesis* aristotélica y con las múltiples dimensiones de la facticidad, se encuentra también en la hermenéutica crítica. En su obra *Ética hermenéutica*, Jesús Conill enfrenta las críticas que la ética ha recibido desde

varios puntos, ya sean acusándola de adolecer de déficit hermenéutico, o por olvidarse del peso de la facticidad y la experiencia (Conill, 2010: 12). Para tratar de responder a esas acusaciones Conill se pregunta si es posible una auténtica ética hermenéutica crítica que no renuncie ni al *logos* ni a la experiencia. Esta ética hermenéutica crítica presenta interesantes aportaciones para abordar los desafíos de la ética en la actualidad:

> Permite ampliar lo moral, mejor dicho, reconocer el trasfondo experiencial del ámbito moral en nuestra vida, las facticidades históricas y vitales, que cada vez más se van reduciendo a meras «señales de tráfico»; 2) ayuda a determinar el estatuto de la razón que opera en las llamadas «éticas aplicadas». La necesidad de articular la creciente complejidad vital y responder a las exigencias del pluralismo en los diversos ámbitos de nuestra vida personal, profesional e institucional ha dado lugar al surgimiento de las éticas aplicadas, y una ética hermenéutica como la que proponemos contribuye a establecer el estatuto de la razón que opera en ellas; incluso se muestra como una «ética de la responsabilidad» al hacerse cargo de la riqueza y profundidad de la experiencia vital, frente a los formalismo y procedimentalismos (Conill, 2010: 15).

En este sentido, la aportación de Conill resulta sumamente valiosa, pues reconoce la necesidad de una ética hermenéutica que sepa dirigir la vista hacia los complejos problemas que impone este tiempo, que en el caso de este trabajo es la tecnología avanzada representada por la IA. Este marco ético y hermenéutico aporta criticismo y responsabilidad desde una mirada filosófica al *factum* de la experiencia.

También las contribuciones del pragmatismo resultan útiles para fundamentar filosóficamente nuestra propuesta. Larry Hickman en su libro *John Dewey's Pragmatic Technology* ha destacado el vínculo que se establece en el pensamiento de John Dewey entre filosofía y tecnología, entendiendo que son dos campos del saber que pueden construir puentes de conocimiento, ya que ambos tienen una naturaleza capaz de contribuir a la resolución de problemas. Al igual que Hickman, Paul Durbin ha insistido en la dimensión más fáctica de la tecnología, pero en este caso desde una óptica de responsabilidad. En *Social Responsibility in Sciencie, Technology and Medicine* pone de relieve la necesidad de que los profesionales del campo tecnológico se comprometan

socialmente por medio de su actividad para aunar esfuerzos en la resolución de aquellos problemas que son generados a partir de su actividad. Una de las finalidades de este modelo, que aporta una aplicación de la ética al campo de la IA, es contribuir a la sociedad con elementos de resolución de problemáticas tecnológicas. La fundamentación pragmática contribuye al reconocimiento de la incorporación del compromiso cívico para poder plantear una IAR. Difícilmente podrían incorporarse criterios de responsabilidad en el campo de la IA sin antes asumir un compromiso cívico para la búsqueda de soluciones a los problemas que engendra la tecnología y también otras acciones del universo humano.

5.4.3. Compromiso con los derechos humanos y los ODS

En la era de la información y la comunicación la humanidad se enfrenta a un importante escenario que puede ser abordado desde los mecanismos más avanzados que ofrece la tecnología. Si partimos de la IAR, es fundamental tener en cuenta dos ámbitos desde los que podría diseñarse mejor la tecnología, a saber, los derechos humanos y los ODS. En ese sentido, se realizará un acercamiento a estos dos ámbitos, que en ciertos momentos presentan elementos comunes como el compromiso por la igualdad de género o la reducción de la pobreza.

El profundo avance de la IA y su impacto en la vida también presenta una importante implicación en materia de derechos humanos. Access Now (2018) ha realizado un estudio preliminar en el que analiza la gama de problemas en el ámbito de los derechos humanos que suscita la introducción disruptiva de la IA en diversas esferas. Muchos de los problemas en esta materia no son nuevos, pero se someten a una profundización o adquieren nuevas formas debido a la influencia de la IA. A diferencia de las tecnologías tradicionales, los sistemas artificiales plantean un nuevo escenario dado su potencial. Por lo tanto, el derecho internacional, así como las instituciones dedicadas a los derechos humanos, pueden servir como un impulso organizado desde el que abordar el fenómeno de la IA a partir de una óptica de responsabilidad y compromiso con el mundo y el ser humano.

La IA posee un importante potencial para fortalecer y acelerar el progreso de los ODS, como aseguró Amina J. Mohammed, vicesecretaria general de la ONU, el 11 de octubre de 2017 durante la reunión del Consejo Económico y Social (ECOSOC) en Nueva York. Tras una toma de contacto con *Sophia*, una robot inteligente, consideró que la IA y su impacto en diversos ámbitos puede promover un fortalecimiento de los ODS, por medio del aprovechamiento de los beneficios que el progreso tecnológico ofrece. Además, en febrero de 2018, Mohammed también afirmó que un mejor y mayor acceso a los datos contribuiría de manera satisfactoria para proporcionar la información correcta en la planificación y evaluación de políticas que sean más efectivas para los ODS. En ese sentido, la IA jugaría un papel muy importante y por lo tanto estaría asumiendo parámetros de responsabilidad con la humanidad. Las tecnologías más avanzadas podrían tener una contribución positiva en el desarrollo de los ODS mediante el impuso de la IAR.

5.4.3.1. Respeto a los derechos humanos

Este años se cumple el setenta y dos aniversario de los derechos humanos como un imperativo que nace del reconocimiento de la experiencia de dignidad que debe ser una condición instrumental que permita la realización de las personas. En torno al valor de su vigencia, Agustín Domingo Moratalla señala los siguiente:

> Los derechos humanos, en tanto que articulación de esta experiencia de dignidad humana, no son únicamente un código moral universal, o una fuente de legislación universal. El imperativo de humanidad que vehiculan es, hoy por hoy, un principio incuestionable para fundamentar la legitimidad de cualquier constitución del planeta. Esta conquista incuestionable se ha logrado porque numerosas personas estaban convencidas de que la experiencia de ser hombre es algo totalmente distinto que el "experimentar con hombres". Esta convicción de lo humano como dignidad y no como experimento, prueba o precio, nos puede permitir seguir confiando en los derechos humanos como memoria viva de esperanza (1995: 100).

Cada vez son más las empresas, gobiernos y otras instituciones que emplean la IA para fortalecer muchos campos de desempeño humano. El enorme potencial de la IA puede generar enormes beneficios para la sociedad. Sin embargo, estos beneficios están despertando la legítima inquietud en grupos de la sociedad civil y en otros actores políticos que plantean la necesidad de formular preguntas sobre las implicaciones que presenta la IA en materia de derechos humanos, por ejemplo en lo relativo a sesgos injustos, violaciones de la privacidad y la libertad, etc. Es importante abordar esta cuestión desde una perspectiva de responsabilidad enmarcada en el espíritu la ciencia cívica y el contexto de un MIAR para valorar dichas implicaciones, priorizando así la dignidad de los seres humanos y protegiendo los derechos más básicos, con el objetivo de generar confianza en la sociedad civil y promover el fortalecimiento de habilidades cívicas y democráticas mediante la tecnología más avanzada.

La IA actual suele fundamentar el despliegue técnico de su acción en el *machine learning* con un propósito definido. Los sistemas artificiales son sometidos a un entrenamiento con datos para el cumplimiento de los propósitos que llevan a cabo mediante procesamientos matemáticos para la detención de patrones en la información que es suministrada y el posterior desarrollo de un modelo que permita ejercer predicciones o recomendaciones sobre nuevos datos. Por ejemplo, en algunos departamentos de recursos humanos ya se están empleando intelectos sintéticos para la selección de candidatos profesionales a través de un entrenamiento en el que se utilizan datos existentes sobre los trabajadores.

Satya Nadella, director ejecutivo de Microsoft, escribió un artículo en el que reflexionaba sobre la alianza entre los seres humanos y la IA para reflejar la necesidad de enriquecer el debate con el fin de que sea más productivo y orientarlo hacia aquellos valores que son inculcados a las personas y las instituciones que promueven la IA (Nadella, 2016). En este texto reflexiona sobre los seis principios y objetivos expuestos y analizados en *The Future Computed: Artificial Intelligence and Its Role in Society* y que la industria y la sociedad deben someter a un profundo debate: equidad, fiabilidad, seguridad, privacidad y seguridad, inclusividad, transparencia y responsabilidad. Todos estos principios y

objetivos son claramente pertinentes para la garantía y protección de los derechos humanos. Así pues, el planteamiento de la IAR juega un papel fundamental en la promoción de una tecnología respetuosa y garante de los derechos humanos, pues reconoce el potencial que la IA posee y a la vez invita a pensar sobre la necesidad de orientar este potencial en su propio beneficio.

De momento no son los intelectos sintéticos los que deciden el propósito para el que se implementan, aunque hay pensadores como Bostrom o Kurzweil que auguran que durante el siglo XXI se producirá un desarrollo exponencial de la IA en el que los sistemas se desarrollarán a sí mismos. A pesar de los pronósticos de estos autores, son las instituciones humanas las que deciden cómo configurar y modelar la IA. Cuando el propósito de una IA es definido, son estas instituciones las que deciden qué aspectos del contexto son los más relevantes. Recuperando el ejemplo planteado, cuando es diseñado un sistema artificial para la selección de personal de una empresa, son los seres humanos los que dirigen la actividad de diseño de estos sistemas y los que determinan qué criterios incorporar en dicho diseño. Pueden decidir centrarse exclusivamente en la maximización de la rentabilidad económica o bien incorporar otros aspectos más integrales y holísticos que también son generalmente importantes en el ámbito humano. El manejo de datos debe someterse a una profunda reflexión, pues un modelo de IA puede dar prioridad a aspectos que tienen que ver con el género o el color de la piel.

Propública, una agencia de noticias independiente de Manhattan, publicó en 2016 un informe en el que examinó la validez de una herramienta de evaluación de riegos que *Northpointe* utilizaba para tomar decisiones sobre la libertad condicional de aproximadamente 35000 convictos federales. Los investigadores Jennifer Skeem y Christopher T. Lowenkamp encontraron que los negros obtuvieron un puntaje promedio más alto que los blancos:

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Fuente: Skeem y Lowenkamp, 2016.

Este informe demuestra que el incidente a «pequeña» escala ocasionado por este *software* de predicción puede replicarse en el futuro y tener un impacto mucho mayor y serias implicaciones en otros ámbitos de la vida. También pone de relieve que aquellos sesgos sistémicos por los que las personas se han pasado décadas impulsando campañas de educación y legislación pueden caer en el olvido con motivo de un impulso tecnológico.

Además del sesgo racial existe otro peligroso sesgo, el sexista. *Amazon* prescindió del uso de una herramienta de IA que se centraba en la contratación de personas porque discriminaba a las mujeres. La compañía del magnate Jeff Bezos comenzó en 2014 a diseñar sistemas informáticos para la revisión de currículums de los aspirantes y el objetivo se centraba en facilitar la selección de los mejores talentos. Más tarde, en 2015, la compañía observó que su nuevo sistema de selección de personal estaba realizando un ejercicio sesgado en materia de género. Esto se debe principalmente a la introducción de una serie de patrones en los sistemas artificiales fruto de la recopilación de datos durante una década y que la mayoría de ellos pertenecían a hombres, por lo que la IA fue evidentemente entrenada bajo un dominio masculino.

Los casos de *Northpointe* y *Amazon* ponen de relieve la necesidad de reflexionar acerca de la calidad de los datos empleados en el entrenamiento para el aprendizaje automático de los intelectos sintéticos, ya que ese es un factor fundamental y decisivo para el carácter ético del modelo resultante. La propuesta de una IAR en el contexto de la discriminación no puede girar en torno a la exclusión de datos delicados como el color de la piel o el género, pues en ocasiones son factores que condicionan una discriminación positiva que garantiza el respeto a los derechos humanos. En ese sentido, la IAR promueve un entrenamiento que incorpora todas las variables necesarias, incluso aquellas que son problemáticas, ya que ese

es un aspecto necesario para la identificación de los problemas y su posterior resolución tras dicho entrenamiento. Sin embargo, aquí surge un importante cuestionamiento a raíz de la necesidad de contar con datos diversos para su entrenamiento. La IAR promueve la incorporación de diversidad de datos en el entrenamiento de los sistemas artificiales para un mayor conocimiento de la realidad y para un mejor tratamiento de las problemáticas. Sin embargo, ese aspecto debe ser considerado de forma cuidadosa a la luz de la conciliación con el respeto a las leyes de protección de datos. Esta controversia podría tratarse en un laboratorio abierto, como una oportunidad para buscar, a través de la conversación, un encuentro de perspectivas y fomentar una dinámica participativa que genere un conocimiento innovador que valore la especificidad de los casos.

En mayo de 2018 organizaciones como Amnistía Internacional, *Access Now* y otras organizaciones asociadas presentaron un documento titulado *Toronto's Declaration on Machine Learning*, un texto en el que se promueve la protección del derecho a la igualdad y la no discriminación que en ocasiones puede derivarse de la IA. Esta declaración persigue la aplicación de las normas internacionales de derechos humanos al contexto de la IA y las actividades que de ella se derivan por medio del aprendizaje automático. Además, el texto tiene como objeto la proposición de discusiones sobre los principios y documentos donde se analizan los daños que han surgido a partir de esta tecnología. También se señala con especial interés lo siguiente:

> Los Estados y los actores del sector privado deberían promover el desarrollo y uso del aprendizaje automático y las tecnologías relacionadas para ayudar que las personas pueda desempeñar un ejercicio y disfrute de sus derechos humanos. Por ejemplo, en salud los sistemas de aprendizaje automático podrían traer avances en diagnósticos y tratamientos, mientras que potencialmente brindarían servicios de salud más ampliamente disponibles y accesibles. En relación a la máquina de sistemas de aprendizaje e inteligencia artificial en general, los Estados deberían promover el derecho positivo para el disfrute de los avances en ciencia y tecnología como afirmación de los derechos económicos, sociales y culturales (2018: 2).

La pretensión de las organizaciones firmantes también es la de invitar a los desarrolladores de sistemas inteligentes a la revisión de los algoritmos, la medición de los impactos y la promoción de la transparencia y la igualdad con el fin de combatir aquellas problemáticas que emergen en materia de discriminación. La Declaración de Toronto centra su preocupación en los algoritmos que respetan los derechos humanos. En ese sentido, la propuesta de una IAR contextualizada en el marco de los derechos humanos se orienta al diseño de sistemas artificiales respetuosos y por lo tanto a la elaboración de mecanismos necesarios para fortalecer la democracia y el civismo.

5.4.3.2. Impulso a los Objetivos de Desarrollo Sostenible

La IA puede servir como un motor de impulso para el cumplimiento de los 17 objetivos de desarrollo sostenible ODS (ONU, 2015):

1. Fin de la pobreza.
2. Hambre cero.
3. Salud y bienestar.
4. Educación de calidad.
5. Igualdad de género.
6. Agua limpia y saneamiento.
7. Energía asequible y no contaminante.
8. Trabajo decente y crecimiento económico.
9. Industria, innovación e infraestructura.
10. Reducción de las desigualdades.
11. Ciudades y comunidades sostenibles.

12. Producción y consumo responsables.

13. Acción por el clima.

14. Vida submarina.

15. Vida de ecosistemas terrestre.

16. Paz, justicia e instituciones sólidas.

17. Alianzas para lograr los objetivos.

El antecedente de los ODS se sitúa en los Objetivos de Desarrollo del Milenio (ODM). Estos objetivos se impulsaron en septiembre del año 2000 a partir del encuentro de los líderes del mundo realizado Nueva York, tras una década de conferencias y cumbres de la ONU. El mandato depositado en las naciones se consolidó a partir del compromiso para una nueva alianza mundial que estableció como objetivo principal la reducción de la pobreza extrema a partir de la identificación de ocho objetivos, con un plazo límite de cumplimiento situado en el año 2015. Estos ochos objetivos son los siguientes (Programa de la Naciones Unidas para el Desarrollo, 2000):

1. Erradicar la pobreza extrema y el hambre.
2. Alcanzar la educación básica universal.
3. Promover la igualdad de género.
4. Reducir la mortalidad infantil.
5. Mejorar la salud materna.
6. Combatir el VIH/sida, el chagas, la tuberculosis, el paludismo y otras enfermedades.
7. Asegurar un medio ambiente sostenible.
8. Promover una asociación mundial para el desarrollo.

La IAR sitúa como una de sus tareas centrales la innovación para el logro y la organización de los ODS. La aplicación de ideas innovadoras en materia tecnológica puede ayudar a los países a avanzar más rápidamente hacia el logro de los ODS. Los avances en el campo de la IA podrían orientarse hacia el empoderamiento de gobiernos, comunidades y organizaciones que requieren la puesta en marcha de soluciones efectivas para las problemáticas que se reflejan en estos objetivos.

El Programa de las Naciones Unidas para el Desarrollo (PNUD) ha analizado el valor de los sistemas artificiales en la generación de progreso y mecanismos transformadores en materia de desarrollo:

- En el campo de la medicina la IA ha proporcionado predicciones muy valiosas para la prevención de ataques cardiacos. En ese sentido, esta tecnología avanzada podría contribuir con los diagnósticos para salvar millones de vidas.

- En el campo agrícola proporciona datos muy valiosos para un mayor conocimiento de los patrones meteorológicos y la producción. Entre las herramientas más usadas pueden encontrarse *LettuceBot*, un sistema que permite la identificación y eliminación de las malas hierbas mediante una base de datos de más de un millón de imágenes que permite identificar las plantas. Gracias a estas tecnologías la producción puede ser más eficiente y eso contribuiría en la prevención de riegos en el sector agrario y favorecer la producción de bienes más estables para las poblaciones del ámbito rural.

- En el contexto lingüístico los intelectos sintéticos también están contribuyendo a la traducción de infinidad de lenguas por medio de los *chatbots*, un mecanismo que, por ejemplo, permite entender las necesidades de los refugiados. Además, facilita la tarea de los abogados y funcionarios públicos en materia de traducción.

Estos tres casos representan claros ejemplos de cómo la IAR puede contribuir en el logro de los ODS: en primer lugar al objetivo 3, salud y bienestar social; en segundo lugar

al 2, hambre cero, 12, producción y consumo responsable y al 13, acción por el clima; y en tercer lugar al 16, paz, justicia e instituciones sólidas.

Es evidente que la IA no es la solución para todos los males, pero el planteamiento de una IAR contribuye a la promoción de algunos de los 17 objetivos y de las 169 metas que conforman la Agenda 2030. El monitoreo sobre las funciones de la IA es una necesidad ineludible en el marco de las problemáticas actuales, ya que la tecnología proporciona interesantes soluciones institucionales que en este momento quizás pasen inadvertidas. Además, en este contexto es fundamental identificar de modo pertinente cuáles son los principales puntos desde los que es posible el planteamiento de alternativas tecnológicas y posibles soluciones. Cortina presenta una serie de propuestas para reducir la desigualdad en el siglo XXI que pueden servir claramente como una importante guía de orientación para la contribución de la IAR (Cortina, 2017: 141-147):

- La reducción de las desigualdades como una forma de erradicar la pobreza y lograr el crecimiento.
- La unión del poder de la economía a los ideales universales en un mundo globalizado.
- La asunción de responsabilidad social empresarial como una cuestión de prudencia y justicia.
- La promoción del pluralismo en los modelos empresariales.
- El cultivo de distintas motivaciones de racionalidad económica en el ámbito empresarial y económico.

En la línea de esta visión acerca de las tecnologías facilitadoras de la consecución de los ODS se encuentra Open Data en Europa y Asia Central (ODECA). Se trata de una plataforma basada en un programa de Datos Abiertos para el Desarrollo que apoya a los representantes gubernamentales, la sociedad civil y otras organizaciones para trabajar en el

intercambio de datos para la generación de conocimientos innovadores. En su página web ODECA se define de la siguiente manera:

> La red cubre 18 países de la región y tiene como objetivo estimular la innovación, el intercambio de conocimientos y el aprendizaje entre profesionales y aficionados de los datos abiertos a nivel regional y mundial.
>
> Nuestro objetivo es utilizar el potencial de los datos abiertos para transformar las sociedades al empoderar a los ciudadanos y ayudar a los gobiernos a cumplir los Objetivos de Desarrollo Sostenible de la ONU. Si bien todavía estamos explorando todas las formas en que los datos contribuirían a los ODS, es innegable que jugará un papel importante para alcanzarlos y medirlos (ODECA, 2016).

La IAR muestra una preocupación sistemática por el fortalecimiento de las habilidades cívicas y la democracia a partir del reconocimiento del conocimiento científico como un recurso público, y por lo tanto se encuentra estrechamente vinculada con los ODS. Un claro ejemplo de este estrecho vínculo lo representa *Citibeats*, una plataforma de IA especializada en la escucha activa de la ciudadanía que ha sido creada para trabajar en el marco de estos objetivos. Como reza en su página web, *Citibeats* «estructura, analiza y sintetiza las opiniones de las personas a partir de grandes cantidades de datos en cualquier región, en cualquier idioma y de cualquier fuente de datos de textos, para que sean fáciles de usar» (https://citibeats.net/about-us/). Esta plataforma tiene un importante compromiso en la potenciación del compromiso cívico y la mejora de la participación ciudadana. Sirve como soporte informativo para las ciudades a través del Observatorio de Impacto Local que permite a las ciudades obtener un mayor conocimiento sobre las problemáticas locales que existen entre las diversas organizaciones y los entes de gobierno. Son numerosas las ciudades que ya están trabajando con *Citibeats* en Barcelona, Madrid, Londres, Nueva York, Tokio o urbes de China. Además, proyectos como *Citibeats* pueden apostar por mecanismos de transparencia y enriquecer así los sistemas democráticos. Tanto los casos de ODECA como de *Citibeats* representan claros ejemplos de cómo una IAR podría contribuir al fortalecimiento del ODS número 16, Paz, justicia e instituciones sólidas.

Finalmente, y recuperando el espíritu de una nueva ética en la línea de Jonas, alejada de un antropocentrismo exacerbado, puede plantearse la agenda de acciones de la IAR a través del compromiso con el medio ambiente. La importancia de la naturaleza para los seres humanos es fundamental, ya que los ecosistemas brindan una cantidad de bienes y servicios que determinan la calidad de vida, hacen posible la producción económica y brindan un desarrollo saludable a las sociedades. En materia medioambiental la IA puede llevar a cabo enriquecedoras aportaciones a partir del estudio de los comportamientos meteorológicos para la predicción de fenómenos climáticos. El monitoreo exhaustivo permite dar recomendaciones simples y muy determinantes al mismo tiempo, por ejemplo, señalando a las autoridades gubernamentales cuáles son las zonas que presentan mayores niveles de sequía. Además, en las ciudades inteligentes la IA proporciona un mejor conocimiento sobre aquellas variables que influyen en el consumo de energía como el clima, la ubicación geográfica o los eventos que suceden en los alrededores, por lo que podría contribuir considerablemente en una reducción de la energía que se utiliza en los edificios. El enriquecimiento que está brindando la asesoría de la IA en el terreno medioambiental es indudable como evidencia el proyecto *Artificial Intelligence for Ecosystem Services* (ARIES) financiado inicialmente por la *National Sciencie Foundation* en Estados Unidos y que presenta una interesante herramienta que facilita el trabajo para el mapeo de servicios medioambientales. Mediante la utilización de una base de datos muy diversa y pertinente al ecosistema de estudio, ARIES elabora un mapa donde se presentan las interacciones entre factores económicos y ambientales. Así pues, es más fácil visualizar cuáles son los servicios ambientales, sus beneficiaros y los flujos de los mismos que se dan en una zona.

Sin duda, el compromiso de la IAR con los ODS no podría realizarse sin una nueva concepción de la economía que incorporara criterios éticos para el ejercicio de su actividad. En relación al estrecho vínculo entre la economía y la ética de la responsabilidad en el contexto de una IAR, es importante destacar lo que señala Conill:

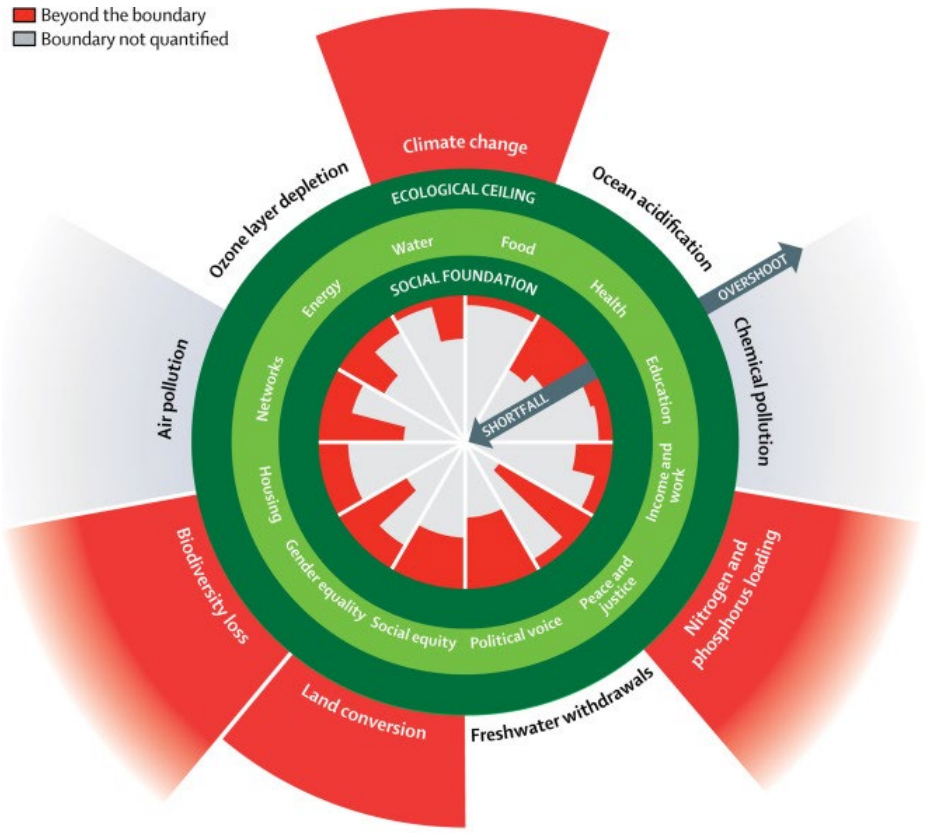
> Algunos han creído que la ética de la responsabilidad es cosa de los políticos, dado que tal fue el contexto en el que Max Weber lo planteó en su momento, pero esto sería excesivamente unilateral. En realidad, son las condiciones propias de la vida moderna, es

> decir, de las sociedades crecientemente complejas y diferenciadas, las que exigen el enfoque de la responsabilidad en todos los órdenes de la vida, por parte de aquéllos que quieran que sus principios y sus convicciones sean operativos en los diversos ámbitos que configuran nuestra vida real. Por tanto, no sólo en la política (en sentido restringido), sino también en las actividades económicas se exige un ejercicio de la razón pública en forma de responsabilidad pluridimensional (2013: 15).

El vínculo entre ética y economía se sitúa en la estela de lo que se denomina como «ética económica», que como afirma Jose Félix Lozano Aguilar, «hace referencia al análisis ético de los sistemas económico y de forma especial a la moralidad de las estructuras y mecanismo del mercado […] se trata de pensar y evaluar el marco normativo racional en el que se desenvuelven los procesos económicos actuales» (2004: 24). Así pues, para promover una IAR en el contexto de los ODS florece la necesidad de una reflexión ética de las estructuras económicas que condicionan las relaciones entre los seres humanos, las instituciones y sus productos.

5.4.4. Los límites planetarios como un imperativo

En 2009 un grupo de 28 científicos internacionales, entre los que destaca Johan Rockström (2009), del *Stdel Stockholm Resilience Centre*, y Will Steffen (2015), de la *Australian National University*, plantearon la necesidad de promover un espacio de actuación para un desarrollo humano seguro en el planeta, que pudiera ser aplicado en cualquier parte del mundo, tanto por los gobiernos de todos los niveles, como por las organizaciones internacionales, el sector privado, la sociedad civil y la comunidad científica. Rockström junto con otros científicos publicaron el texto *A Copenhagen Prognosis: Towards a Safe Climate Future*, donde alertaban de los importantes riesgos que enfrenta la humanidad en el futuro debido a las implicaciones que presenta la superación de los límites del planeta Tierra. En la siguiente imagen se ilustran los límites planetarios y los niveles que han sido superados en la actualidad:



Fuente: Raworth, 2017: 49.

Estos límites han abierto un profundo debate en el seno de grupos de científicos y de investigación de importante renombre como el IPCC. La idea que motiva los debates se centra en los niveles que ha alcanzado la acción antropógena que presenta tales proporciones que no puede mantenerse al margen de un cambio global abrupto. Para evitar los cambios globales abruptos han propuesto una serie de indicadores que señalan los límites biofísicos que no deben ser sobrepasados para no ocasionar procesos desconocidos y no lineales con consecuencias potencialmente catastróficas para el equilibrio de los ecosistemas y la vida. Entre los límites que se sugieren, y que pueden observarse en la anterior imagen, se señalan los siguientes:

1. La capa de ozono estratosférico.

2. Biodiversidad.

3. La dispersión de productos químicos.

4. Cambio climático.

5. La acidificación del océano.

6. El consumo de agua dulce y el ciclo hidrológico global.

7. Cambio en el uso del suelo.

8. Las entradas de nitrógeno y fósforo a la biosfera y los océanos.

9. Carga de aerosoles atmosféricos.

Es importante destacar que el establecimiento de estos límites no ha quedado exento de críticas, ya que se han generado debates en el seno de la comunidad científica debido al intento de cuantificar los límites, motivo por el cual los valores propuestos deberían ser abordados con prudencia. Aunque todavía quedan muchos asuntos por aclarar en torno a la evolución del clima como resultado del aumento de los gases de efecto invernadero, cualquier predicción, por muy simple que sea, es arriesgada. No obstante, la mayor parte de la comunidad científica da por sentado que el cambio climático es un hecho más que evidente y una consecuencia directa de la acción humana. Este fenómeno debe provocar un cambio en la cultura de la generación de conocimiento, enfrentado los desafíos con medidas innovadoras. En ese sentido, el MIAR también incorpora las exigencias climáticas como un imperativo de responsabilidad que debe ser asumido en el diseño y despliegue de toda actividad vinculada a la IA. Los intelectos sintéticos podrían destacar por asumir un compromiso de carácter planetario, por ello, los límites establecidos por la comunidad científica son un buen medio de incorporación de responsabilidad a sus fundamentos. Además, las exigencias de innovación que se encuentran en el interior del MIAR son enfrentadas a partir del reconocimiento del imperativo de estos límites planetarios. Este imperativo implica que los tecnólogos que impulsan el diseño de la IA asuman una tarea de compromiso, pues como señala Raworth (2017), el reconocimiento de estos límites contribuiría de manera positiva en el bienestar humano en materia de salud planetaria y también a la reducción de las desigualdades. En definitiva, el MIAR promueve una

reconciliación de la Sociosfera y la Biosfera, donde exista una raíz motivacional para cultivar un sentido de responsabilidad en la tecnología (Domingo Moratalla, 1991: 8).

5.5. Responsabilidad ante el poder y la vulnerabilidad

Antes se mencionó una de las carencias de la propuesta de Dignum, a saber, su falta de fundamentación filosófica. A pesar de que la autora holandesa reconoce la necesidad de incorporar criterios de responsabilidad ética a la actividad tecnológica, no fundamenta filosóficamente su concepto de IAR. Para superar las carencias de una IAR débil como la postulada por Dignum, en esta tesis se ha buscado articular el pensamiento filosófico de Jonas, que a diferencia de Dignum nos ofrece una sólida propuesta de fundamentación ética sobre el principio de responsabilidad, con el desarrollo de una ciencia cívica basada en las aportaciones del pragmatismo crítico y las ciencias sociales fronéticas, capaz de orientar la IAR, en el contexto de un laboratorio abierto, al respeto de los derechos humanos, la promoción de los ODS y el cuidado con los límites planetarios. Entre las aportaciones más interesantes de Jonas para proporcionar un sustento filosófico al concepto de IAR, es pertinente destacar aquí, a modo de corolario, las siguientes propuestas:

- La superación de una ética antropocéntrica.
- El carácter de vulnerabilidad que está presente en muchas dimensiones de la realidad y la llamada de un imperativo fáctico que invita al respeto.
- El poder tecnológico suscita la necesidad de plantear la introducción de criterios de responsabilidad en la acción humana.

El postulado jonasiano sobre la necesidad de superar la ética antropocéntrica para enriquecerla por medio de la incorporación de nuevas dimensiones susceptibles de reconocimiento puede encontrarse vinculado con facilidad dentro de un concepto de IAR que favorezca un compromiso con los ODS y el respeto a los límites planetarios. El planteamiento de la IAR gira en torno al enriquecimiento de la lógica tecnológica mediante el reconocimiento de otras variables que deben ser tomadas en cuenta y que el pensamiento

positivista suele ocultar o no reconocer como importantes. La contribución de Jonas a la ampliación de las fronteras de la ética más allá de la consideración exclusiva que observa al ser humano entre sus preocupaciones es muy útil para una fundamentación que dote de mayor solidez el postulado de la IAR y de esa manera ayude a superar planteamientos filosóficamente débiles y políticamente insuficientes como el de Dignum.

Para presentar la reflexión sobre la vulnerabilidad, Jonas contextualiza su postulado en las relaciones entre padres e hijos. Los hijos encarnan la figura de la vulnerabilidad que debe ser protegida desde la paternidad y la maternidad. En ese sentido, la idea de vulnerabilidad puede ser integrada dentro del concepto de IAR, reconociendo que en la actualidad existen muchos aspectos que son susceptibles de ser considerados vulnerables ante el poder de la tecnologías, como puede ser la naturaleza humana ante los proyectos transhumanistas, la estabilidad económica de las familias frente a los procesos de automatización de los trabajos, etc. No obstante, para que los grupos de interés se sientan moralmente responsables de las consecuencias de la acción tecnológica, previamente deben reconocer un valor inherente en aquello que es susceptible de verse afectado por el impacto de la acción tecnológica. El reconocimiento de valor proyecta una apelación moral que impone el deber de respetar y proteger lo que se presta para ser afectado.

El poder de la técnica ha ampliado el espectro de su impacto gracias a las desafiantes contribuciones de la IA. Esa ampliación de poder demanda la expansión de la responsabilidad humana más allá de las fronteras antropocéntricas tradicionalmente conocidas en un ejercicio de reconocimiento más amplio. Debe centrarse principalmente en los desafíos que plantea el medio ambiente a través de los síntomas que está presentando en torno a su degradación. Hay que recordar que el impulso que postulaba Jonas (1995) sobre el principio de responsabilidad, se centraba en la preocupación y cuidado por la biosfera.

El poder de la IA en el mundo contemporáneo y su capacidad de transformación de la realidad constituyen un nuevo objeto de estudio para la ética. Los efectos de la técnica ya no se caracterizan por la inmediatez y la cercanía, sino que conducen a un escenario de nuevos impactos. Las implicaciones de la acción tecnológica son muy amplias. Como

señala Jorge Enrique Linares (2018), Jonas es uno de los anunciadores del riesgo mayor y en ese sentido advierte del poder de la tecnología y su impulso hipostasiado de la siguiente manera:

> Las bendiciones de la técnica, cuando más dependemos de ellas, contienen la amenaza de transformarse en una maldición. Su innata tendencia a la desmesura hace aguda la amenaza. […] junto a la magnitud y la ambivalencia, otro rasgo de carácter del síndrome tecnológico que tienen una importancia ética propia: el elemento cuasi-forzoso de su avance, que por así decirlo hipostatiza nuestras propias formas de poder en una especie de fuerza autónoma de la que nosotros, los que la ejercemos, nos volvemos paradójicamente súbditos. Sin duda el menoscabo de la libertad humana debido a la cosificación de sus propios actos se ha dado siempre, tanto en las vidas individuales como, sobre todo, en la historia colectiva. […] Con cada nuevo paso («hacia delante») de la gran técnica estamos ya obligados a dar el siguiente y legamos esa misma obligación a la posteridad, que finalmente tendrá que pagar la cuenta. […] el elemento tiránico de la técnica actual, que hace de nuestras obras nuestros dueños y nos obliga incluso a reproducirlas, representa un desafío ético en sí mismo […] más allá de la cuestión de las buenas o malas que sean esas obras en concreto. En aras de la autonomía humana, de la dignidad que exige, de que nos poseamos a nosotros mismo y no nos dejemos poseer por nuestra máquina, tenemos que poner el galope tecnológico bajo control extratecnológico (1997: 38-39).

TERCERA PARTE

ÁMBITOS DE APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL RESPONSABLE EN EL MUNDO CONTEMPORÁNEO

CAPÍTULO 6

EL DESAFÍO TRANSHUMANISTA

> La idea básica es simple. Para decidir si queremos modificar algún aspecto de un sistema, es de gran ayuda considerar por qué el sistema contiene ese aspecto en primer lugar. De igual modo, si proponemos introducir alguna característica nueva, podemos preguntarnos por qué el sistema no la contiene ya. El sistema que nos ocupa aquí es el organismo humano.
>
> (Bostrom y Sandberg, 2017: 394)

> No es la especie, sino la «condición» humana lo que tenemos la obligación de preservar. La condición humana es maleable y transformable, pero no debe ser transformada en su contrario que sería lo inhumano. Lo que nos hace falta es un criterio de humanidad que sirva de pauta para distinguir las manipulaciones genéticas aceptables de las que no lo son. Ese criterio nos lo dio Kant con su imperativo de la dignidad. Y ha sido reformulado por Hans Jonas pensando precisamente en los problemas que ahora nos ocupan: «Obra de tal modo que los efectos de tu acción sean compatibles con la permanencia de una vida humana auténtica en la Tierra».
>
> (Camps, 2002: 70)

El ser humano puede ser considerado como *homo sapiens,* pero también como *homo faber*, y esas dos condiciones de ser determinan su modo de estar en el mundo. Existe una mutua imposición entre la naturaleza y el ser humano, el segundo se impone a la naturaleza, pero ésta le condiciona de tal manera que se ve empujado a crear una sobrenaturaleza, como señala Ortega en *Meditación de la técnica*.

Es conocida la distinción que Hannah Arendt establece en *La condición humana* entre labor, trabajo y acción como aquellas actividades en las que el ser humano ha desplegado su vida: la labor, vinculada al proceso biológico del ser humano, teniendo por condición la vida misma; el trabajo, vinculado a las actividades de producción artificial de objetos; y la acción, que es la actividad que realizan los humanos entre sí sin la mediación de las cosas, teniendo la pluralidad como una condición propia. Mientras la labor y el trabajo se desarrollan en la esfera privada, la acción pertenece al ámbito público. Arendt caracteriza el trabajo en contraposición a la labor, tomando en cuenta la diferencia que John Locke establece entre ambos conceptos: «la labor de nuestro cuerpo y el trabajo de nuestras manos» (Locke, 2006: 226). La autora señala lo siguiente:

> Labor es la actividad correspondiente al proceso biológico del cuerpo humano, cuyo espontáneo crecimiento, metabolismo y decadencia final están ligados a las necesidades vitales producidas y alimentadas por la labor en el proceso de la vida. La condición humana de la labor es la misma vida. Trabajo es la actividad que corresponde a lo no natural de la exigencia del hombre, que no está inmerso en el constantemente repetido ciclo vital de la especie, ni cuya mortalidad queda compensada por dicho ciclo. El trabajo proporciona un «artificial» mundo de cosas, claramente distintas de todas las circunstancias naturales. Dentro de sus límites se alberga cada una de las vidas individuales, mientras que este mundo sobrevive y trasciende a todas ellas. La condición humana del trabajo es la mundanidad (Arendt, 2001: 21).

Es importante detenerse en la cuestión del trabajo. Por medio del trabajo, señala Arendt, los seres humanos producen objetos que plasman en su cultura y que tienen cierta durabilidad. La autora distingue entre el *animal laborans*, que dedica sus esfuerzos a la subsistencia, y el *homo faber*, que crea un mundo que los humanos comparten entre sí. El *homo faber* desarrolla su producción mediante la evaluación, elección y empleo de los medios adecuados para alcanzar determinados fines. Además, la relación que el *homo faber* mantiene con el medio es la de apoderamiento y uso de la naturaleza, por ello también posee la capacidad de crear y destruir sus obras de consumo. Precisamente para el transhumanismo el ser humano no es solo *homo sapiens*, sino también *homo faber*.

En este capítulo se utilizarán los términos «transhumanismo» y «poshumanismo» con frecuencia, aunque no tienen el mismo significado. Juan Arana utiliza la obra de Paul Gauguin «¿De dónde venimos? ¿Quiénes somos? ¿Adónde vamos?» para explicar esas diferencias entre el transhumanismo, que contesta al «¿A dónde vamos?», y el poshumanismo, que se centra en el «¿Qué o quiénes seremos?» (Arana, 2017: 1). En su obra *Principios de extropía* Max More definió en 1990 al transhumanismo como el camino de transición para guiar hacia una condición poshumana. En ese sentido, el transhumanismo se presenta como ese espacio intermedio, de transición, entre el humanismo y el poshumanismo. Fueron los campos de la ciencia ficción, la futurología, el arte contemporáneo y la filosofía, los que tras el impulso de ver más allá, originaron el concepto poshumano como un nuevo capítulo en la historia de la humanidad. Por lo tanto, el transhumanismo, en tanto que una transición entendida como condición de posibilidad, allanará el camino para un nuevo momento en la historia en el que la tecnología impondrá una inteligencia y una condición no biológica, inaugurando así un tiempo poshumanista.

Para el movimiento transhumanista, la ciencia actual, y en especial las NBIC –nanotecnología, biotecnología, tecnologías de la información y las nuevas tecnologías basadas en la ciencia cognitiva– representan una buena oportunidad para mejorar (*improve*) o potenciar (*enhance*) la especie humana y generar así seres humanos más fuertes físicamente, más inteligentes y emocionalmente más equilibrados. Para los transhumanistas esta mejora y potenciación de la especie humana no representa un sueño inalcanzable o una utopía, sino más bien una realidad no tan lejana gracias al desarrollo de aquellas tecnologías que son necesarias para su realización.

Existen varias modalidades de transhumanismo, pero este capítulo se centrará en el estudio del transhumanismo tecnocientífico por entender que esa modalidad es la que se encuentra generalmente vinculada con la tecnología, y particularmente con la IA. El transhumanismo tecnocientífico se divide a su vez en dos vertientes, a saber, la que se encuentra más vinculada con el campo de la IA y la que tiene una base biológica y médica. En ese sentido, para hacer referencia a la vertiente más conectada con la IA, se dirigirá la mirada hacia pensadores como Raymond Kuzrweil y Nick Bostrom, entre otros; y para la

que encuentra su sustento en el ámbito biológico y médico, se hará un mayor hincapié en las propuesta de Julian Savulescu, José Luis Cordeiro y David Wood. Al margen de esas consideraciones, también se dedicará un espacio de reflexión a otros pensadores que han promovido la reflexión y el debate sobre estas cuestiones como Jürgen Habermas, Peter Sloterdijk o Michael Sandel y más recientemente Steven J. Jensen y José Luis Widow.

Los proyectos transhumanistas que sirven como transición a un nuevo relato poshumanista pueden generar controversias políticas que necesariamente deben ser abordadas desde la ética. Las tecnologías despliegan sus actividades en contextos sociales que proyectan sus valores sobre mecanismos de diseño. Así pues, debido a la impronta contextual sobre estas tecnologías, se puede profundizar en muchas de las problemáticas políticas y sociales ya existentes. En ese sentido, la ética pretende invitar a una profunda reflexión en este plano para impulsar la incorporación de criterios de responsabilidad en los proyectos transhumanistas.

Además, la IAR puede proyectar una nueva orientación en las planificaciones transhumanistas por medio de un compromiso cívico con los derechos humanos y la aplicación de los ODS. Eso supondría una nueva mirada desde la óptica de este movimiento revolucionario, donde el objeto ya no sería mejorar exclusivamente la condición biológica del ser humano a través de una incidencia directa sobre su organismo, sino también crear las condiciones de posibilidad ambientales para que pueda desplegar un proyecto de vida orientado éticamente.

6.1. Un nuevo paradigma

El paradigma de la medicina tradicional, fundamentado en aspectos terapéuticos con el propósito de «reparar», ha sido cuestionado. En este nuevo escenario han surgido ideas como la de «mejoramiento» que están adquiriendo un notable interés. No obstante, el término mejoramiento presenta una doble problemática, a saber, la de exigir la determinación de qué es aquello que se está mejorando; y también respecto a qué estado es comparada la mejora (Feito Grande, 2013: 269). Las controversias éticas en el ámbito de la

mejora radican en el concepto de aquello que es considerado como mejor. Es una discusión sobre los fines que persiguen las biotecnologías promovidas por los transhumanistas y además sobre los medios para alcanzar tales fines (Feito Grande, 2013: 271). En ese sentido, el transhumanismo es un movimiento cultural e intelectual de carácter filosófico que se ha desarrollado en las últimas tres décadas y que representa la utopía del momento (Diéguez, 2017: 20).

El transhumanismo promueve un enfoque multidisciplinar y analiza las oportunidades de mejora de la condición humana y los organismos mediante el avance de la tecnología. Entre las disciplinas para promover el proyecto transhumanista se encuentran la ingeniería genética, la tecnología de la información, la nanotecnología o la IA (NBIC). Los caminos para emprender la mejora están orientados a la salud, la cognición, la genética o el comportamiento emocional. El transhumanismo observa la naturaleza como una entidad susceptible de mejora y a los seres humanos como seres lo suficientemente inteligentes como para asumir la responsabilidad de mejorar. Este movimiento sueña con alcanzar un horizonte poshumano con capacidades mucho mayores que las que el ser humano posee en la actualidad (Bostrom, 2003).

El enfoque terapéutico del paradigma médico tradicional encuentra su fundamento en una larga tradición judeocristiana que hace que la reacción espontánea sea la de considerar la naturaleza humana como un elemento que tiene que ver con la eternidad y con lo intangible, motivo por el que no es posible en ningún caso mejorar, sino curar o reparar. Esto se debe principalmente a que la tradición judeocristiana ha plasmado en el espíritu tradicional de Occidente una idea de oposición a la alteración o modificación de lo que es considerado como naturaleza humana. El transhumanismo se posiciona en contra de la postura tradicional judeocristiana, ya que se formula sobre la máxima de un perfeccionamiento ilimitado y un desafío a la muerte y el envejecimiento, entendiendo que la ciencia y la tecnología brindan las herramientas necesarias para esos fines.

El transhumanismo tiene influencias de la tradición del humanismo clásico y también de otras tradiciones en las que no cabe detenerse, que van desde Pico della Mirandola hasta Kant, pasando por Francis Bacon o Julien Offray de La Mettrie, quienes insisten en la idea de la perfectibilidad infinita del ser humano. Esa idea de perfectibilidad infinita se traduce en la máxima de perfeccionamiento humano ilimitado, defendida por el extropianismo, una corriente transhumanista presentada por Max More en 1990. Con el paso del tiempo la Declaración Transhumanista se modificó con las aportaciones de autores como Anders Sandberg, Max More, Natasha Vita-More, David Pearce, Leen Daniel Crocker, Mikhail Sverdlov y Nick Bostrom, entre otros. Finalmente la organización internacional sin ánimo de lucro Humanity + (H +) adoptó una declaración en marzo de 2009 con los siguientes aspectos definitorios:

1. La humanidad se verá profundamente afectada por la ciencia y la tecnología en el futuro. Prevemos la posibilidad de ampliar el potencial humano superando el envejecimiento, las deficiencias cognitivas, el sufrimiento involuntario y nuestro confinamiento al planeta Tierra.

2. Creemos que el potencial de la humanidad aún no se ha realizado. Hay posibles escenarios que conducen a condiciones humanas mejoradas maravillosas y que merecen la pena.

3. Reconocemos que la humanidad enfrenta graves riesgos, especialmente por el mal uso de las nuevas tecnologías. Existen posibles escenarios realistas que llevan a la pérdida de la mayoría, o incluso de todo, lo que consideramos valioso. Algunos de estos escenarios son drásticos, otros son sutiles. Aunque todo progreso es cambio, no todo cambio es progreso.

4. Es necesario invertir el esfuerzo de investigación para comprender estas perspectivas. Necesitamos analizar cuidadosamente la mejor manera de reducir los riesgos y acelerar las aplicaciones beneficiosas. También necesitamos foros donde las personas puedan discutir de manera constructiva lo que se debe hacer y un orden social donde se puedan implementar decisiones responsables.

5. La reducción de los riesgos existenciales y el desarrollo de medios para la preservación de la vida y la salud, el alivio del sufrimiento grave y el mejoramiento de la previsión y la sabiduría humanas deben perseguirse como prioridades urgentes, y contar con una gran financiación.

6. La formulación de políticas debe guiarse por una visión moral responsable e inclusiva, tomando en serio las oportunidades y los riesgos, respetando la autonomía y los derechos individuales, y mostrando solidaridad y preocupación por los intereses y la dignidad de todas las personas en todo el mundo. También debemos considerar nuestras responsabilidades morales hacia las generaciones que existirán en el futuro.

7. Abogamos por el bienestar de toda la sensibilidad, incluidos los humanos, los animales no humanos y cualquier futuro intelecto artificial, formas de vida modificadas u otras inteligencias a las que pueda dar lugar el avance tecnológico y científico.

8. Estamos a favor de permitir a los individuos una amplia elección personal sobre cómo habilitan sus vidas. Esto incluye el uso de técnicas que pueden desarrollarse para ayudar a la memoria, la concentración y la energía mental; terapias de extensión de la vida; tecnologías de elección reproductiva; procedimientos criónicos; y muchas otras posibles tecnologías de modificación y mejora humana.

6.2. Caminos hacia el poshumanismo

El despliegue de la agenda práctica que promueve la declaración de principios del transhumanismo se apoya en los importantes avances que los campos científico y tecnológico han experimentado en las últimas décadas: biotecnología, nanotecnología, tecnologías de la información y ciencias cognitivas. Entre las aplicaciones de estas cuatro áreas convergentes entre sí, pueden encontrarse: la fabricación de nuevos materiales industriales que tienen más resistencia para la elaboración de prótesis, de productos

alimenticios y de supervivencia que aumentan el rendimiento energético, la investigación destinada a la creación de nanorobots para la reparación de tejidos y organismos, diseño de proteínas que ayuden a combatir enfermedades, y también otra serie de mecanismos que mejoren las capacidades cognitivas y sensitivas, como la visión nocturna o el aumento de la potencia memorística. Por último, la promesa más reciente en la que se está investigando trata sobre la creación de mecanismos que permitan la interacción cerebro-máquina con el objetivo de trascender la inteligencia biológica y caminar hacia un terreno tecnológico. En este último aspecto tiene una notable influencia la IA. El interés en la aplicación de las cuatro áreas convergentes antes mencionadas tiene como finalidad la superación de la especie humana y la construcción de un futuro poshumano que se materialice en los ámbitos que se presentan a continuación.

6.2.1. La mejora de los hijos

El transhumanismo consagra como una obligación moral y un derecho de los padres y madres la posibilidad de conseguir los mejores hijos. Esta obligación moral y derecho es considerado por Julian Savulescu una beneficiencia procreativa que describe de la siguiente manera:

> Las parejas (o los reproductores individuales) deberían seleccionar, de los distintos niños que pueden tener, aquel que se espera que tenga una vida mejor, o al menos una vida igual de buena que los demás, según la información relevante de la que se disponga.
>
> Defenderé que la Beneficiencia Procreativa implica que las parejas deben emplear pruebas genéticas para rasgos no patológicos a la hora de seleccionar qué niño traer al mundo y que debemos permitir que se seleccionen genes no patológicos incluso si hacerlo así mantiene o incrementa la desigualdad social.
>
> El «deberían» que aparece en «deberían seleccionar», ha de entenderse como «tiene buenas razones para». Entenderé que la moralidad nos exige que hagamos aquello que tenemos mejores razones para hacer. En ausencia de alguna otra razón, una persona que tiene una buena razón para tener el mejor niño posible está requerida moralmente a tenerlo (2012: 45-46).

Fermín Jesús González-Melado diferencia entre dos tipos de eugenesia, la negativa y la positiva (2010: 211-212). La eugenesia negativa consiste en la eliminación de los fetos que presentan patologías a través de un diagnóstico prenatal o la interrupción del embarazo. Este mecanismo trata de evitar las patologías con el interés de seleccionar gametos en función de las características genéticas que presentan. En cuanto a la eugenesia positiva, se promueve el mejoramiento del niño por medio de una ingeniería genética que posibilitará «no sólo la identificación de genes defectuosos sino también, la identificación de genes que expresen características deseables, por ejemplo color de ojos, estatura, peso, inteligencia… Se trata de construir el mejor hijo posible» (González-Melado, 2010: 212).

A diferencia de González-Melado, Savulescu tiene una visión optimista de la selección genética e insiste en el aspecto de la inteligencia, pues la considera un motivo más que suficiente para optar por la posibilidad de seleccionar. La inteligencia representa un importante factor para proporcionar una buena vida a los seres humanos y para ello Savulescu recurre al *Filebo* de Platón: «En el *Filebo* de Platón, Sócrates llega a la conclusión de que la mejor vida es una mezcla de sabiduría y placer. La sabiduría incluye el pensamiento, la inteligencia, el conocimiento y la memoria. Claramente, la inteligencia forma parte del concepto platónico de la buena vida» (Savulescu, 2012: 55).

Savulescu se sitúa en la estela de la bioética utilitarista, una rama de la ética que defiende una orientación de los recursos médicos para que proporcionen un valor más productivo encaminado hacia la felicidad total de los seres humanos. La bioética utilitarista se fundamenta en una premisa fundamental, a saber, que la distribución de los recursos es un juego de suma cero, y por lo tanto los recursos médicos deben estar destinados a un valor productivo y de máxima felicidad. Así pues, esta bioética es instrumentalista, ya que justifica el fin mostrando un medio como correlato de dicho fin. La bioética utilitarista de Savulescu ha sido objeto de críticas, ya que la racionalidad práctica es en primera instancia mucho más compleja que lo que el instrumentalismo postula y las circunstancias que rodean la toma de decisiones deben tomarse en cuenta en el desarrollo de la conducta humana (Ortiz Llueva, 2013).

6.2.2. La mejora física

Otra de las vías hacia el poshumanismo consiste en la mejora del rendimiento con unas profundas aspiraciones hacia la superioridad de la especie. Este camino promueve el abandono de los mecanismos terapéuticos y defiende la incorporación de los fármacos o drogas, para aumentar el rendimiento físico o memorístico de la especie (Savulescu, 2012).

El ámbito deportivo es visto por las teorías transhumanistas como un terreno fértil en el que hacer posibles muchas de las ideas que defiende este movimiento. El motivo por el que el ámbito deportivo se convierte en un objeto de estudio de las teorías del mejoramiento es porque existe una combinación entre el empeño, la admiración y la excelencia, que representan, en términos generales, un ideal para la vida buena. La biotecnología, la farmacología y la genética convierten la búsqueda del rendimiento superior en una de sus prioridades para la mejora física. El campo deportivo se convierte en un espacio ideal para la investigación y experimentación de las tesis transhumanistas.

Savulescu entiende que es razonable defender el uso de las drogas en el deporte y afirma que el ideal deportivo vinculado con la fortaleza biológica no es incompatible con la capacidad de juicio y razonamiento. Para ello argumenta que no es pertinente sostener el ideal deportivo mediante la comparación con la actividad deportiva que desarrollan ciertos animales como los caballos. Los humanos no son caballos, pues toman decisiones y las hacen en base a su capacidad de razonamiento y juicio, algo que los diferencia a grandes rasgos del resto de los animales. Esta diferencia, frente a quienes sostienen las tesis que se oponen al uso de drogas en el ámbito deportivo, refuerza una vez más el espíritu humano integrado en el deporte, ya que «lejos de ir contra el espíritu del deporte, la manipulación biológica encarna el espíritu humano: la capacidad de mejorarnos a nosotros mismo basándonos en nuestras razón y nuestro juicio» (Savulescu, 2012: 111).

En cambio, frente a las posturas más atrevidas del transhumanismo hay quienes se posicionan en contra de manipular la naturaleza y de entrar en un ámbito mercantil, sobre todo en el aspecto de la ingeniería genética orientada al deporte, como Michael Sandel, que señala:

> Es más plausible ver la ingeniería genética como la última expresión de nuestra determinación por vernos por encima del mundo: los dueños de la naturaleza. Pero esa promesa de dominio es defectuosa: amenaza con desvanecer nuestra apreciación por la vida como regalo, y con dejarnos sin nada que afirmar u observar fuera de nuestra propia voluntad (Sandel, 2017: 94).

Torbjörn Tännsjö reconoce que el ámbito deportivo cuenta con unos valores intrínsecos que lo hacen particular y eso es lo que permite establecer una clara diferencia entre la medicina deportiva y la medicina en general. Mientras que en la medicina general las medidas positivas para el mejoramiento cuentan con cierto reconocimiento en el campo en cuestión, la medicina deportiva observa con recelo las motivaciones que se esconden detrás de los mecanismos de mejoramiento. El deporte de élite cuenta con una serie de valores y una noción muy especial de justicia que lo hacen merecedor de cierto reconocimiento. En ese sentido Tännsjö entiende que esos aspectos deben protegerse y quedar al margen de las modas de la técnica de mejoramiento, aunque al mismo tiempo afirma que será difícil que el deporte se mantenga al margen de los avances tecnológicos en este materia (Tännsjö, 2017: 338).

6.2.3. La lucha contra el envejecimiento

Una interesante reflexión sobre el tema del envejecimiento se encuentra en el libro *La muerte de la muerte*, de José Luis Cordeiro y David Wood. Es importante partir de la consideración de que el envejecimiento no es un proceso unitario dentro del cuerpo humano, pues hay organismos que no se someten a dicho proceso. Además, son múltiples los factores que influyen en el envejecimiento, como el medio ambiente. No obstante, a pesar de los estudios que existen en este campo, todavía no existe una teoría aceptada por

parte de toda la comunidad científica acerca del envejecimiento. En medio de esta falta de consenso, el tecnólogo Aubrey de Grey publicó, con Michael Rae, una obra titulada *El fin del envejecimiento*. Los importantes estudios formativos por los que había pasado de Grey, sobre todo en el campo de la informática y la computación, le aportaron una visión sobre el tema del envejecimiento muy peculiar.

Su postulado sobre la extensión de la vida se llama SENS (*Strategies for Engineered Negligible Senescence*, en español: estrategias para una senescencia negligible ingenierizada). El significado de este postulado trata sobre la posibilidad de desarrollar terapias médicas que tengan como objetivo combatir el envejecimiento en humanos. Dichas terapias médicas deben plantearse desde la bioingeniería, un campo que permitirá combatir las causas y daños del envejecimiento. La propuesta de Grey no estuvo exenta de críticas y cuestionamientos, pues el mismo MIT promovió un número especial de su revista *Technology Review* y un premio de 20.000 dólares para quien pudiera desmontar la teoría SENS. A pesar de las dudas que despertó en la comunidad científica este postulado teórico, los avances científicos han demostrado que de Grey estaba en lo cierto.

Aunque existen diferencias entre las investigaciones de la comunidad científica, se presentan dos puntos básicos en los que existe ciertamente un consenso: que el envejecimiento se da de manera gradual, siendo un proceso dinámico y secuencial; y que no es considerado como algo inevitable o irreversible, ya que es un proceso plástico y flexible (Cordeiro y Wood, 2018: 89). Este consenso ha permitido comenzar a identificar el envejecimiento como una enfermedad, permitiendo desplegar un abanico de posibilidades para su tratamiento. Para Cordeiro y Wood (2018: 95) esto permite situar el envejecimiento en el centro de nuevas investigaciones, lo que favorece la búsqueda de fuentes de financiación y a la vez sugiere que en los próximos años todo apunta a la creación de un nuevo nicho industrial.

Entre los pensadores que se sitúan en la senda de este tipo de desafío biotecnológico, se encuentra Kurzweil, que no se limita a hablar únicamente de la prolongación de la vida y la lucha contra el envejecimiento, sino que se refiere a una existencia poshumana que desafía lo biológico. Escribe Kurzweil:

> La aceleración del progreso en la biotecnología nos permitirá reprogramar nuestros genes y procesos metabólicos para desactivar las enfermedades y los procesos del envejecimiento. Este progreso incluirá rápidos avances en genómica (influencia de los genes), en proteómica (comprensión e influencia sobre el papel de las proteínas), en las terapias génicas (inhibición de la expresión génica mediante tecnologías tales como el ARN-interferente e inserción en el interior del núcleo de nuevos genes), en el diseño racional de medicamentos (formulación de medicamentos que acomentan cambios precisos en los procesos de las enfermedades y del envejecimiento) y en la clonación terapéutica de versiones rejuvenecidas de nuestras propias células (con telómeros expandidos y ADN corregido), tejidos y órganos, así como en desarrollos relacionados. La biotecnología expandirá la biología y corregirá sus flagrantes fallos al mismo tiempo que se solapará con la revolución de la nanotecnología (que nos permitirá expandirnos más allá de las graves limitaciones de la biología) (Kurzweil, 2017: 368-369).

6.2.4. La mejora cognitiva y emocional

En el texto *Why I Want to be a Posthuman When I Grow Up* Bostrom sostiene que un poshumano es aquel ser que tiene al menos una capacidad poshumana, entendiendo por capacidad poshumana aquella capacidad central general que excede en gran medida el máximo que puede alcanzar cualquier ser humano sin someterse a las mejoras tecnológicas (Bostrom, 2008: 107). Entre esas capacidades centrales generales distingue tres: las pertenecientes a la salud, la cognición y la emotividad. Este apartado se centrará en la cognitiva y la emotiva.

Bostrom afirma que existe un deseo generalizado en todos los seres humanos para mejorarse, ya que según él todo el mundo quiere ser mejor, a nadie le disgusta mejorar. En lo que refiere a la cognición, entiende que las personas también están interesadas en mejorar su conocimiento y su inteligencia. Una clara muestra de ello son las grandes sumas de dinero destinadas al cultivo del conocimiento, tanto si se observa desde el plano individual o colectivo. Las personas se inscriben en academias de artes musicales y también en centros que sirven para reforzar aquellos conocimientos aprendidos en la escuela o la universidad. En ese sentido, para Bostrom es más que evidente la preocupación por la mejora del funcionamiento cognitivo (Bostrom, 2008: 115). La mejora cognitiva tiene en sí misma un valor intrínseco, aunque también puede convertirse en un medio para la obtención de bienes externos como el honor, la fama o el dinero. Por lo tanto, una buena razón para defender una mejora cognitiva es el valor intrínseco que representa (2008: 116).

A diferencia de la mejora en el ámbito de la salud, es más difícil tratar de explicar en qué consiste la mejora emotiva. Bostrom afirma que normalmente es más fácil entender la posibilidad de ayudar a alguien que sufre una depresión suicida, fruto de un desequilibrio neuroquímico, para que disfrute de la vida, que a aquellos casos que van más allá de las intervenciones terapéuticas que tratan de curar psicopatologías. En la mayoría de los casos que van más allá de lo estrictamente terapéutico para la cura de psicopatologías es muy difícil determinar en qué consiste un nivel de capacidad emocional poshumana (Bostrom, 2008: 117).

Las personas dedican gran parte de su vida a tratar de forjar un carácter y moldearlo según unas necesidades. En ocasiones se trata de vencer el miedo y la desconfianza por medio de algunos ejercicios. Así pues, se procuran alcanzar aquellas metas que implican una modificación y mejora de las capacidades emocionales. La tecnología representa una potente herramienta que posibilitaría una mejora emocional poshumana. Existen capacidades emocionales que podrían estar orientadas hacia la excelencia gracias a la tecnología. Además, los medios tecnológicos inducirían la aparición de un nuevo escenario en el que florecerían nuevas emociones y estados psicológicos completamente nuevos (Bostrom, 2008: 118), aunque esto no quede exento de dificultades en su comprensión. Por

lo tanto, las mejoras cognitivas y emocionales se presentan como un atractivo viable dentro de las opciones transhumanistas.

6.3. Un deber moral

La declaración de principios que promueve el transhumanismo ha generado un choque de perspectivas entre quienes piensan que la mejora de la especie humana es una obligación moral y quienes la cuestionan desde diversos enfoques. Aquí se tratará de dar cuenta brevemente de esas dos perspectivas divergentes desde las aportaciones de dos pensadores que resumen las argumentaciones que vienen de uno y otro lado.

John Harris (2017) es un firme defensor de las técnicas de mejoramiento por medio de las biotecnologías. Entiende que las técnicas de mejoramiento son buenas para los seres humanos porque permiten dejar atrás un estado que no suponía una mejora real. Existen razones indiscutibles para defender el mejoramiento porque representa un beneficio para el ser humano y por lo tanto nadie debería oponerse a aquello que es beneficioso para el ser humano. En ese sentido, estas técnicas que permiten mejorar la especie tienen que ser aprovechadas necesariamente porque ayudan a estar mejora. Con bastante convencimiento Harris señala lo siguiente:

> Existe una continuidad entre los daños y los beneficios, de manera que las razones que tenemos para evitar causar daños o crear a otros seres que nacerán con problemas serán una consecuencia de las razones que tenemos para beneficiar a otros si podemos. En resumen, decidir negar un beneficio es, en cierto modo, dañar al individuo que rechazamos mejorar. Tenemos razones para rechazar, crear u otorgar incluso beneficios triviales, así como razones para otorgar o no negar incluso pequeños beneficios (Harris, 2017: 137).

Harris critica los planteamientos que se apoyan en el principio de precaución y en el argumento teológico. Al criticar el principio de precaución, parte de la defensa que hace el Comité Internacional de la Bioética de la UNESCO para proteger el genoma humano, entendido como patrimonio común de la humanidad. Además, también critica a aquellos sectores que se oponen al mejoramiento de la especie por entender que es una actividad

similar a «jugar a ser Dios» (Harris, 2017: 139-140). Harris sostiene que, si realmente fuese erróneo intervenir en la naturaleza para modificarla, también la medicina representaría un atentado. Los postulados fundamentados ciegamente en lo natural suponen un obstáculo para el progreso de la humanidad.

Para Harris existe un fuerte imperativo moral que justifica el mejoramiento con el fin de evitar daños y brindar beneficios. La legitimidad moral de este imperativo radica en aquellos mejoramientos que proporcionan beneficios o protegen de daños, y no tanto cuando sirven a los intereses de la igualdad de oportunidades o a la restauración del normal funcionamiento del organismo, que son las dos posturas que promueven Allen Buchanan, Norman Daniels, Dan W. Brock y Daniel Wikler en la obra *From Chance to Choice*.

Frente a quienes consideran las tecnologías de mejoramiento como un deber moral, se sitúan quienes, como Jenny Krutzinna (2016), han analizado el argumento transhumanista del deber moral situándolo en la estela del bienestarismo (*welfarism*). Krutzinna parte de un principio humanista implícito en el bienestarismo, según el cual la justificación de la bondad o la maldad de algo reside, en primera instancia, en la contribución que representa para la vida humana y su calidad. Para el bienestarismo no es necesario contar con una definición de vida buena en particular, sino que aquellos elementos que contribuyen beneficiosamente a la vida se identifican con facilidad y por lo tanto pueden encontrarse bajo el paraguas del bienestar (Krutzinna, 2016: 529). Sin embargo, para Krutzinna la intuición no ofrece un fundamento lo suficientemente sólido para justificar el nivel de bienestar que algo proporciona. La persecución del bien se sitúa como el sentido teleológico del bienestarismo, aunque es necesario discutir las implicaciones que tiene este sentido.

Como se ha mostrado a lo largo de este trabajo, el desarrollo tecnológico y científico ha experimentado un aumento considerable en las últimas décadas, lo que también va acompañado del crecimiento de una ambición positivista. En ese sentido, es fundamental plantear una reflexión ética que contribuya a la búsqueda de soluciones y al enriquecimiento de la discusión sobre las problemáticas. El motivo por el que es necesaria

esta reflexión ética surge de la debilidad que presenta el enfoque bienestarista para defender como un deber moral la mejora cognitiva mediante técnicas tecnocientíficas. Es decir, la normatividad que intenta proporcionar el bienestarismo carece de solidez práctica. Por lo tanto, Krutzinna considera necesario discutir en qué consiste el bienestar. El concepto de bienestar basado en la teoría de los deseos suscita algunos problemas que surgen de la discusión entre lo objetivo y lo subjetivo (Krutzinna, 2016: 530).

En lo que se refiere a las mejoras, la cognitiva es la que despierta un mayor interés entre los transhumanistas que llegan a afirmar lo siguiente:

> Las capacidades cognitivas son las requeridas para el despliegue de cualquier tipo de racionalidad instrumental –la capacidad identificar de manera confiable los medios para nuestros fines y proyectos–. Mejor cognición significa mejor acceso a la información sobre el entorno de uno y sobre su propia biología y psicología, así como mejores habilidades de uso. Esta información en la planificación racional la necesitan las personas para ejercer la racionalidad instrumental con el fin de obtener placer y evitar el dolor, para cumplir sus deseos, y para realizar bienes objetivos. Así que la mejora cognitiva debe promover el bienestar en todas las principales teorías del bienestar (Savulescu, Sandberg y Kahane, 2011: 10).

6.4. Objeto de debate filosófico: bioconservadores y bioprogresistas

No es demasiado el espacio bibliográfico que ocupa el fenómeno transhumanista en el ámbito filosófico español, pues como señaló Diéguez en un artículo publicado en el diario *El País* el 13 de septiembre de 2018, es un objeto de estudio incipiente. Sin embargo, comienza a abrirse paso el estudio de este fenómeno, como atestigua la reciente publicación del libro de Diéguez, *Transhumanismo. La búsqueda tecnológica del mejoramiento humano*, que ilustra oportunamente este movimiento intelectual.

En lo que respecta al terreno filosófico, el transhumanismo ha estimulado la formulación de una diversidad de posiciones estrictamente diferenciadas que representarán en las próximas décadas una importante pugna política en materia de regulación de las biotecnologías, como son los bioluditas, singularistas, bioconservadores, transhumanistas, bioprogresistas y tecnoprogresistas. No obstante, en este apartado se analizarán exclusivamente dos: bioconservadores y bioprogresistas. En el seno de la línea conservadora se sitúan aquellos que observan las tecnologías de mejoramiento desde la desconfianza, basándose en el principio de precaución. Allen Edward Buchanan en su obra *Beyond Humanity? The Ethics of Biomedical Enhancement* resume con claridad el eje discursivo sobre el que se elaboran los argumentos bioconservadores. Buchanam sostiene que el argumentario conservador afirma la existencia de una esencia humana que no puede ser alterada por una motivación caprichosa de perfeccionamiento. El organismo humano es producto de un ciclo evolutivo que se caracteriza por un complejo equilibrio que ha ido adquiriendo su actual forma a lo largo de la historia. En ese sentido, una alteración de la naturaleza humana representaría un desprecio al valor que encarna la propia evolución obra del Maestro Ingeniero. Son varias las figuras que representan la postura conservadora en esta materia, como Jürgen Habermas (2009), Michael Sandel (2016), Francis Fukuyama (2002), George Annas (2017), Ryuichi Ida (2017) o Steven J. Jensen y José Luis Widow (2018), aunque por motivos de espacio únicamente se expondrán los postulados de Habermas, Sandel y Jensen y Widow.

Frente a los bioconservadores se sitúan quienes defienden la mejora tecnológica que impulsa al movimiento transhumanista, abogando por un perfeccionamiento ilimitado y un desafío a la muerte y el envejecimiento. Estas máximas serían realizables gracias a las impresionantes posibilidades que están brindando las biotecnologías. El control del envejecimiento y el rejuvenecimiento humano se convierte en un deber moral que debe ser asumido por la humanidad y reconocido en el contexto de los derechos humanos (José Luis Cordeiro y David Wood, 2018: 17). Entre los bioprogresistas pueden encontrarse a Sloterdijk, Kurzweil o Bostrom.

Los avances que el campo de la biotecnología ha experimentado en las últimas décadas han provocado el surgimiento de debates en el seno de numerosas esferas como la jurídica, la psicológica, la científica, la filosófica, etc. Son variadas las posturas respecto a la biotecnología; por un lado, se encuentran aquellos puntos de vista que observan estos mecanismos tecnocientíficos como una forma sofisticada de eugenesia que tendrá un gran impacto; y por otro lado, se sitúan los que defienden con convicción que la biotecnología representa la solución a serios problemas que enfrenta la humanidad y que son de diversa índole.

6.4.1. Jürgen Habermas: una crítica a la manipulación «caprichosa»

Habermas ha participado en este debate con la publicación de *El futuro de la naturaleza humana. ¿Hacia una eugenesia liberal?*, obra que también representa una tesis opuesta a la presentada por Peter Sloterdijk en *Normas para el parque humano*. A Habermas le interesa principalmente la investigación con células madre y el diagnóstico genético preimplantacional (DGP). Habermas trata de reflexionar sobre las consecuencias que tendrán para la modernidad política las técnicas de manipulación de la especie que representan las diversas formas de intervención genética. El texto sobre el futuro de la naturaleza humana en el que Habermas esboza su reflexión acerca de la biotecnología es de carácter político, ya que las referencias filosóficas que en él se hacen son exclusivamente para reforzar argumentaciones políticas (Mendieta, 2002: 94). Para el filósofo alemán no ha existido un profundo debate público sobre las implicaciones de los DGP, ya que la ciudadanía se ha dejado llevar de forma acrítica por la oferta mercantil y la libertad individual promovida por el credo liberal. Frente a esta posición acrítica es fundamental pensar acerca de la necesidad de normativizar en este terreno, ya que si no se plantea una seria reflexión pública, las biotecnologías tendrán efectos dañinos sobre la condición humana y la cultura política. En ese sentido, Habermas muestra su rotunda oposición a la eugenesia liberal.

La preocupación habermasiana es de carácter moral, ya que el centro de la cavilación se enfoca en la autocomprensión moral. Luc Ferry (2017: 94-97) expone con claridad uno de los ejemplos que Habermas utiliza en su obra para discutir esta cuestión, a saber, el de unos padres que quisieran manipular el código genético de su hijo para perfeccionarlo en el sentido que promueve el transhumanismo. La libertad y la autonomía del niño se verán menoscabadas, ya que no podrá decidir si aumenta su capacidad para el deporte o para otra área, por ejemplo. Este caso representa una evidente antinomia de la libertad de los padres frente a la de los hijos. Habermas no se opone con rotundidad a todas las manipulaciones genéticas, ya que solo le parecen aceptables aquellas que persiguen como finalidad erradicar enfermedades. Para Habermas la manipulación genética representa un uso instrumental de los embriones y por lo tanto una acción contraria al postulado kantiano que promovía el tratamiento del ser humano como un fin en sí mismo y no como un medio. En ese sentido, la línea argumentativa del filósofo alemán se encuentra muy cercana a la Iglesia católica, al considerar al embrión como una persona humana en potencia desde el momento de la concepción.

6.4.2. Michael Sandel: la ética de la gratitud con lo dado

En su obra *Contra la perfección. La ética en la era de la ingeniería genética*, Sandel presenta su postura acerca de las implicaciones éticas de las biotecnologías, que ha despertado un gran interés en el ámbito académico internacional y en las instituciones políticas de los EE. UU., donde Sandel participó junto a Fukuyama en el comité de ética creado en 2002 bajo la presidencia de George W. Bush. Lo que suscita el interés de Sandel es la falta de comprensión moral que presentan los seres humanos frente a los profundos avances de las biotecnologías, es decir, que la ciencia avanza a una mayor rapidez que la comprensión moral (Sandel, 2017: 75). El argumento bioconservador que se sostiene sobre la autonomía y la libertad de los hijos con respecto a los padres que han decidido manipular sus genes no es del todo convincente para Sandel, pues da por sentado que si los padres no deciden sobre el diseño del hijo, entonces el hijo será inevitablemente libre, algo que no es

cierto, ya que nadie elige su herencia genética, pues depende de la llamada «lotería genética» (Sandel, 2017: 76).

La idea central del argumento de Sandel para el cuestionamiento del transhumanismo es el paso de una ética de la gratitud hacia aquello que es dado, similar a la que sostiene Ryuichi Ida (2017), a una ética del dominio sobre la naturaleza por parte de un humano prometeico. Como señala Ferry, cuando se habla de aquello que es dado no necesariamente se remite a una autoridad religiosa, sino a una cuestión secular, a un «principio de donación exterior y superior al hombre» (Ferry, 2017: 90), que puede ser la misma naturaleza. Esta relación azarosa y misteriosa es la que se deja a un lado en el transhumanismo bajo las pretensiones de dominio sobre la naturaleza. Para Sandel este carácter prometeico y dominador que caracteriza al transhumanismo desembocaría en la transformación de tres aspectos claves de la condición moral: la humildad, la responsabilidad y la solidaridad (Sandel, 2017: 91). En primer lugar, acerca de la humildad Sandel señala lo siguiente:

> El hecho de que nos preocupamos mucho por nuestros hijos y aun así no podemos elegir el tipo que queremos enseña a los padres a estar abiertos a lo espontáneo. Dicha apertura supone una predisposición que vale la pena valorar, no solamente entre las familias, sino en el mundo en general también. Nos invita a tolerar lo inesperado, a vivir en discordancia y a frenar el impulso de controlar […] La concienciación de que nuestros talentos y habilidades no son completamente nuestra propia obra frena nuestra tendencia hacia la arrogancia (Sandel, 2017: 91).

En segundo lugar, a propósito de la responsabilidad escribe Sandel:

> A medida que la humildad se desploma, la responsabilidad se expande en enormes proporciones. Atribuimos menos a la suerte y más a la elección […] Uno de los beneficios de vernos como criaturas de la naturaleza, de Dios, o del destino, es que no somos totalmente responsables de la manera en la que somos. Cuando mejores maestros de nuestro legado genético nos convertimos, mayor es la carga que llevamos por los talentos que tenemos y la manera en la que actuamos (Sandel, 2017: 91).

Y, por último, en relación con la solidaridad Sandel señala:

> Cuanto más activos estemos en la naturaleza fortuita de nuestro ámbito, más razones tendremos de compartir nuestro destino con otros […] Como la gente no sabe cuándo o si sufrirá varias enfermedades, comparte el riesgo contratando seguros médicos y de vida […] los mercados de seguros simulan solidaridad solo en la medida en que la gente no sabe o controla sus propios factores de riesgo. Imaginemos que las pruebas genéticas avanzaran hasta tal punto que pudieran predecir de forma fiable el futuro médico y la esperanza de vida de una persona […] La solidaridad del mundo de los seguros desaparecería a medida que los que tienen buenos genes abandonaran a los que no gozan de salud (Sandel, 2017: 92).

Los argumentos de Sandel son muy ilustrativos, pues están repletos de ingeniosos ejemplos. En lo que respecta al terreno político, afirma que las tecnologías de mejoramiento podrían ser utilizadas con fines bélicos. Además, también podría promoverse un turismo genético, como ya ocurre en algunos países como Venezuela o Colombia con la cirugía estética o con la gestación subrogada en la India y en otros lugares del planeta.

6.4.3. Steven J. Jensen y José Luis Widow: la crítica naturalista

En su texto *Unnatural Enhancements*, Steven J. Jensen y José Luis Widow critican aquellas técnicas de mejoramiento que representan una amenaza para el orden natural de la naturaleza. Los autores diferencian entre mejoras terapéuticas y no terapéuticas, entendiendo que hay mejoras que son pertinentes y otras que no lo son, en función del respeto al orden de naturaleza.

Jensen y Widow presentan su postura a partir de la clasificación establecida por Lisbeth Witthøfft Nielsen (2011) en función de la noción de naturaleza: 1) un estado general de las cosas que podría denominarse «madre naturaleza»; 2) la esencia de una cosa, en particular la esencia de los seres humanos; y 3) lo que está libre de la intervención humana, es decir, lo natural en oposición a lo artificial (Jensen y Widow, 2018: 350). La preocupación de estos autores se centra en la relación que el ser humano establece con la naturaleza. Tomando como referencia en su fundamentación la teoría aristotélica de la naturaleza,

afirman que ésta se caracteriza por un sentido teleológico y dinámico, distanciándose de ese modo de las concepciones que consideran a la naturaleza como una entidad estática.

La naturaleza tiene un sentido teleológico, pues está orientada a un fin concreto, por lo que si esta tendencia a una finalidad concreta sufre una alteración, se situaría frente a una acción antinatural que dificulta el despliegue de la naturaleza. La desviación de aquello que no es dado por la naturaleza es considerada como una consecuencia fruto de una acción antinatural. Jensen y Widow utilizan la fecundación *in vitro* (FIV) como un ejemplo para demostrar cómo la técnica es capaz de alterar el orden natural mediante la manipulación de los embriones. Esta técnica se basa en una manipulación de los embriones que modifica el orden azaroso por el que la naturaleza organiza su despliegue. En ese sentido, el transhumanismo no estaría promoviendo el orden natural, sino más bien respondiendo a una satisfacción de carácter privado del ser humano (Jensen y Widow, 2018: 352).

Frente a la satisfacción del bien de carácter privado del ser humano, estos pensadores defienden el respeto al movimiento de la naturaleza que da forma a su orden esencial y característico. La preocupación de Jensen y Widow gira en torno al establecimiento del equilibrio y la búsqueda de armonía entre el bien humano y el bien de la naturaleza. Para el cumplimiento de dicho equilibrio y armonía reconocen la originalidad del pensamiento aristotélico (Jensen y Widow, 2018: 354). El estagirita afirmaba que los humanos son seres sociales que se relacionan con el medio, existiendo de ese modo una noción del bien que es compartida y que no puede establecerse exclusivamente desde lo particular, y por lo tanto privado:

> En general, nuestro bien debe tenerse de acuerdo con la naturaleza humana. Si perdemos de vista la naturaleza, también perdemos de vista el bien compartido, dejando aislados a los individuos. Sin naturaleza solo pueden existir bienes de individuos particulares o quizás bienes de clases (Jensen y Widow, 2018: 354).

Por lo tanto, una modificación de las tendencias propias del orden de la naturaleza representaría una alteración de curso teleológico. Esta alteración del orden es antinatural y cualquier mecanismo que represente una alteración para dicho orden, ya sea la FIV o el uso de fármacos, serían elementos antinaturales. Además, el CRISPR, la técnica que posibilita la manipulación del ADN, también sería un mecanismo antinatural. En ese sentido, todos aquellos mecanismos tecnocientíficos que promuevan una alteración del orden natural o una sustitución del mismo serían considerados por Jensen y Widow como antinaturales. Por lo tanto, las tesis transhumanistas serían antinaturales porque responden a un deseo privado innecesario que supone una alteración para la naturaleza.

6.4.4. Peter Sloterdijk: la propuesta antropotécnica

La figura de Sloterdijk ya fue mencionada anteriormente en el primer capítulo, donde se reflejó la aportación que realiza desde la filosofía para un humanismo tecnológico. Los postulados de este filósofo alemán pueden ser abrazados desde las tesis transhumanistas por considerar a la técnica como uno de los pilares fundamentales para la proyección y perfeccionamiento del ser humano en el futuro. Recordemos, que Sloterdijk diagnostica un agotamiento de la tradición literaria para promover un proyecto humanista. En ese sentido, señala a las tecnologías, y concretamente a la ingeniería genética, como la punta de lanza para impulsar un nuevo humanismo contextualizado en la era tecnológica.

El universo de posibilidades que brinda la tecnología, y concretamente la biotecnología, representa una vía pragmática desde la que desarrollar nuevas antropotécnicas. No obstante, Sloterdijk reconoce una necesidad en este nuevo universo de posibilidades, a saber, un código de las antropotécnicas donde el ser humano participe consciente y activamente. Así pues, el postulado de este pensador puede servir como una fundamentación razonable, en términos filosóficos, para sustentar el discurso y el proyecto transhumanista.

6.4.5. Raymond Kurzweil: la integración con la tecnología

Raymond Kurzweil es otro de los íconos de ese transhumanismo de corte tecnocientífico inspirado en el campo de la IA y la ingeniería de computación en general. En su obra *La singularidad está cerca* promueve el concepto de singularidad que ha servido para impulsar un proyecto universitario muy prometedor como la *Singularity University* y que en su página web se presenta de la siguiente manera: «Preparación de líderes mundiales y organizaciones para el futuro: explore las oportunidades e implicaciones de las tecnologías exponenciales y conéctese a un ecosistema global que está configurando el futuro y resolviendo los problemas más urgentes del mundo». Kurzweil utiliza el concepto de singularidad para referirse al momento en el que acontece una explosión de superinteligencia capaz de perfeccionarse a sí misma y de fabricar otros sistemas inteligentes al margen de la intervención humana, y así sucesivamente por medio de un crecimiento exponencial que derivará en una entidad global que se caracterizará por poseer una IA que será superior a la de los seres humanos.

Hay un motivo muy importante para entender mejor el pensamiento transhumanista de Kurzweil. Es un autor agnóstico en lo que respecta a las creencias metafísicas porque considera que la argumentación teísta sobre un más allá representado por un ser divino no es descartable, y de la misma manera tampoco sería descartable el ateísmo. Su posición se debe a la formación religiosa que recibió durante su juventud, centrada principalmente en el conocimiento de varias religiones. En ese sentido, ese hecho formativo en su etapa de juventud lo marcará para toda la vida y sentará las bases en su pensamiento para cultivar la idea de convergencia. Kurzweil ha sabido aprovechar muy bien la idea de convergencia para entender que existe la posibilidad de imaginar más allá de la naturaleza humana entendida biológicamente y trascender hacia un espacio tecnológico, desafiando así las fronteras entre lo biológico y lo tecnológico.

El término «singularidad» es empleado en el ámbito de la teoría general de la relatividad para designar un sistema o estado en el que no se da lugar la aplicación de leyes físicas conocidas. Por lo tanto, este no es un término propio del campo transhumanista, sino que más bien se hace un uso del mismo como una mera analogía para referirse a lo discontinuo y lo impredecible, algo que tiene que ver con la entrada en un escenario desconocido y poshumano.

Anticipándose a posibles críticas, Kurzweil sostiene que la trascendencia desde lo biológico a lo tecnológico no implica una desaparición de la humanidad a expensas de las máquinas con superinteligencia. Por ello, su propuesta se centra en la integración con la máquina, en un proceso de síntesis que consiste en la transferencia de la mente en una de ellas. Dicho esto, la singularidad podría ser reconocida desde dos perspectivas: por un lado, como la creación de sistemas con superinteligencia; y, por otro lado, como el fortalecimiento de la inteligencia humana para alcanzar niveles poshumanos gracias a la integración con la tecnología.

La argumentación que Kurzweil utiliza para sostener sus postulados se fundamenta en la Ley de Rendimientos Acelerados, también conocida como «Ley de Moore» (Kurzweil, 2016b: 64-73). Esta ley hace referencia al poder computacional que se duplica en periodos que van de dieciocho meses a dos años, o lo que es lo mismo, tiene un crecimiento exponencial. No obstante, la Ley de Moore no brinda un sustento lo suficientemente sólido al postulado de Kurzweil, ya que es expresada con regularidad y no hay una base científica que haga posible una regularidad más potente, lo que ha sido objeto de sendas críticas. La apuesta más arriesgada del planteamiento sobre la singularidad se encuentra en la idea de inmortalidad, que es, según Javier Monserrat (2015), una desacertada predicción, tal y como está planteada. Monserrat no pone en duda los sorprendentes avances que depara la tecnología en diversos campos, sin embargo, afirma que está siempre sometida a los intereses humanos en función de criterios éticos y morales (2015: 1439-1440). Y, en ese sentido, es importante mencionar que Kurzweil ignora el carácter político y filosófico-histórico de la tecnología, motivo por el cual su planteamiento es susceptible de una razonable crítica.

Podría considerarse que el pensamiento de Kurzweil se encuentra cerca de las pretensiones teológicas, ya que el crecimiento exponencial al que hace referencia le aporta una visión muy peculiar de la evolución. Así pues, la singularidad es presentada como una nueva etapa en el proceso evolutivo que se verá motivada por la necesidad de trascender:

> La Singularidad denota un evento que tendrá lugar en el mundo material, el inevitable siguiente paso en el proceso evolutivo que empezó con la evolución biológica y que se ha extendido mediante la evolución tecnológica dirigida por los humanos. Sin embargo, es precisamente en el mundo de la materia y de la energía donde nos enfrentamos a la trascendencia, una de las connotaciones principales de lo que la gente llama espiritualidad. Consideremos la naturaleza de la espiritualidad en el mundo físico (Kurzweil, 2016b: 443-444).

A pesar de las críticas que los defensores de la singularidad han recibido, se siguen situando en la senda ideológica de un cambio tecnológico radical que supondrá un verdadero cambio de paradigma para la vida humana en varias de sus dimensiones. Las tesis de Kurzweil han levantado ampollas en los laboratorios de IA, aunque también cuenta con importantes reconocimientos.

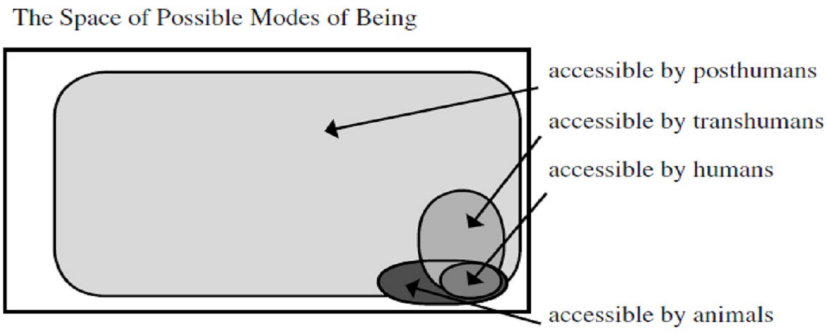
6.4.6. Nick Bostrom: la apertura de posibilidades

En su texto *In Defense of Posthuman Dignity*, Bostrom formula una crítica a los posicionamientos bioconservadores de figuras conocidas como las de Leon Kass, Francis Fukuyama, George Annas, Wesley Smith, Jeremy Rifkin y Bill McKibben. El fundador del *Future of Humanity Institute* afirma que la naturaleza brinda en ocasiones regalos envenenados como el cáncer, y que existe una obligación moral de hacer un uso de la tecnología para el beneficio de la especie. Además, una de las típicas críticas que los bioconservadores dirigen contra el transhumanismo se centra en la posibilidad de que se genere un escenario en el que los poshumanos atenten contra la vida de los humanos. En ese sentido, Bostrom reconoce el importante poder que se encuentra implícito en las biotecnologías, aunque afirma que la defensa de las biotecnologías para mejorar las

condiciones de vida de los seres humanos no es incompatible con estar a favor de una política reguladora de estas actividades (Bostrom, 2007: 206).

El postulado transhumanista de Bostrom no se caracteriza por un optimismo tecnológico acrítico, ya que reconoce que el enorme potencial tecnológico que depara el futuro también puede ser empleado de mala manera y por lo tanto provocar un daño enorme, que va desde la extinción de la vida inteligente hasta el incremento de las desigualdades sociales. Los riesgos que implica la tecnología en el ámbito del mejoramiento humano deben ser tomados muy en cuenta para que los proyectos que sean impulsados se caractericen por la precaución y tengan en cuenta la diversidad de factores de impacto (Bostrom, 2005: 4).

El ser humano es un ser que se caracteriza por su limitación natural, pues existen una serie de pensamientos, sentimientos, experiencias, etc., frente a las que existe la posibilidad de acceder para superar las limitaciones que impone la naturaleza biológica. Es asombroso pensar que existe un universo de posibilidades más allá de lo que se conoce. Es posible disfrutar ese universo de posibilidades por medio de una nueva realidad poshumana que trascienda lo humano. Aquí Bostrom establece una clara diferenciación entre lo transhumano y lo poshumano, considerando al primero como un momento de transición caracterizado por mejoras moderadas; mientras que el segundo va más allá del estado de transición y se sitúa en una nueva esfera. En lo que respecta a la accesibilidad de la realidad posible, Bostrom utiliza el siguiente cuadro para ilustrar gráficamente las parcelas de dominio de los animales, los humanos, los transhumanos y los poshumanos.



Fuente: Bostrom, 2005: 5.

Bostrom orienta las prioridades transhumanistas a partir de tres ejes (2008: 107):

- Salud: capacidad de mantenerse completamente sano, activo y productivo, tanto mental como físicamente.

- Cognición: capacidades intelectuales generales, como la memoria, el razonamiento deductivo y analógico, y la atención, así como facultades especiales tales como la capacidad de comprender y apreciar la música, el humor, el erotismo, la narración, la espiritualidad, las matemáticas, etc.

- Emotividad: capacidad de disfrutar la vida y responder con un efecto apropiado a las situaciones de la vida y a otras personas.

Debido a las limitaciones biológicas que los seres humanos presentan es muy probable que exista dificultad para reconocer los valores que subyacen tras los postulados del transhumanismo. Esto se debe principalmente a que el reconocimiento y el deseo de un valor requiere como condiciones de posibilidad la familiaridad con el mismo valor y que este esté sometido a un proceso de deliberación. Ante la falta de reconocimiento de los valores transhumanistas, Bostrom afirma que para llevar a cabo una aproximación a los mismos no es necesario renunciar a los valores actuales. Es importante también aclarar que el transhumanismo no promueve un beneficio para los poshumanos en detrimento de los humanos, sino más bien favorecer que los seres humanos sean capaces de ver más allá de sus limitaciones y realizar sus ideales de la mejor manera posible (2005: 8-9).

Para promover los valores transhumanistas y las actividades de mejoramiento de la especie humana es fundamental asumir tres condiciones básicas (Bostrom, 2005: 9-11):

- La seguridad global. La cuestión de la seguridad es presentada como un requisito ineludible en el desarrollo del proyecto transhumanista. Tanto en el capítulo 3, dedicado a Hans Jonas, como en el capítulo 5, dedicado a la importancia de la responsabilidad, se insistió en los riesgos que también representa la actividad tecnocientífica.

- La necesidad de desarrollo tecnológico. El desarrollo tecnológico es una condición de posibilidad para la realización de los ideales transhumanistas. Este desarrollo necesariamente debe ir acompañado de un crecimiento económico que permita invertir en el campo de la investigación tecnológica y científica.

- Acceso amplio a los beneficios. Los beneficios que traen consigo las técnicas de mejoramiento no deben ser exclusivos de una élite, pues se generaría un nuevo escenario de desigualdad. En ese sentido, es fundamental cultivar la igualdad, la solidaridad y el respeto a los demás seres humanos para que el proyecto transhumanista amplíe su campo de beneficios más allá de las élites económicas.

6.5. El factor de pertenencia a un grupo social y la identidad

El fenómeno transhumanista tiene importantes implicaciones políticas, pues es una actividad humana desarrollada necesariamente dentro de una comunidad concreta. Tiene razón Christine Overall al señalar que la ética dedicada a la reflexión en el campo del mejoramiento no debería centrarse únicamente en temas relacionados con los riegos, los beneficios, los costes o la autenticidad de la especie, sino que también debería incorporar a la discusión aspectos de índole social y política (Overall, 2009: 327). Las tecnologías de mejoramiento se encuentran estrechamente vinculadas con cuestiones políticas de poder, por ejemplo en temas de propiedad de los medios tecnológicos, oportunidades de acceso a los mecanismos de mejora que se encuentran dentro del ámbito de pertenencia a una clase social, y nuevas formas de opresión que serán discutidas más adelante.

Los mecanismos orientados a la gestión de las tecnologías del mejoramiento pueden ocasionar diversos impactos sobre la vida de los colectivos que conforman una sociedad. Estos impactos pueden ser fruto de una deliberación previa, pero también lo pueden ser de la lógica propia de dichas tecnologías. Por lo tanto, como antecedente para la discusión de este tema tendría que ser discutida la posibilidad que existe de que las tecnologías de mejoramiento tengan un impacto, ya sea positivo o negativo, sobre la sociedad. Se mencionan los impactos porque las tecnologías de mejoramiento son entendidas a partir de

una clara influencia sobre la identidad de los individuos de una sociedad. Mediante sus mecanismos de acción pueden generar problemas de opresión entre grupos con diferente identidad y también provocar claras relaciones de desigualdad. Parece que la utilización de la tecnología para mejorar la vida de los individuos tendría que ser algo positivo para una sociedad. Sin embargo, esas mejoras podrían ser exclusivas de determinados grupos y generar dinámicas de poder que dificultaran las oportunidades de acceso a otros grupos, provocando de ese modo estrategias de exclusividad. Estas estrategias de exclusividad obstaculizarían el establecimiento de relaciones equitativas e inclusivas entre los diversos grupos que conforman una sociedad y por lo tanto provocarían un problema de carácter político.

Las variaciones genéticas entre unos y otros seres humanos surgen a raíz de la denominada «lotería genética» como señala Overall (2009: 330). Sin embargo, no es esa la única lotería, ya que en el nacimiento nadie elige en qué colectivo ni en el seno de qué familia nacer. La posición en el mundo es diferente en función del grupo social de pertenencia. Por ejemplo, no es lo mismo nacer en el barrio de Brooklyn que en el barrio de Upper East Side, ambos pertenecientes al distrito de New York. La pertenencia a un determinado grupo social tiene un impacto político sobre la vida, como puede ser en el acceso a determinados servicios vinculados a la educación o la sanidad, y por lo tanto puede representar ventajas o desventajas.

Los grupos de pertenencia social que configuran la identidad pueden ser de diversa índole: sexo, género, etnia, clase económica, raza, etc. En ese sentido, la identidad influye sobre la posición dentro de una sociedad y por lo tanto en el disfrute de las oportunidades que las instituciones ofrecen. El acceso a las tecnologías se encuentra fuertemente marcado por la pertenencia a un grupo social, lo que implica que las tecnologías de mejoramiento humano también puedan representar un factor de desigualdad en materia de oportunidades. Un gobierno puede legislar en materia de mejoramiento, liberalizar ese sector al extremo, facilitando mecanismos políticos y jurídicos para hacer uso de la libertad, aunque eso no garantizaría bajo ningún concepto que todos los individuos de una sociedad tuvieran el mismo acceso a las oportunidades de estas tecnologías. Esto se debe esencialmente a algo

que ya ha sido indicado, a saber, el grupo social de pertenencia. El derecho del libre uso y acceso a los sistemas de mejoramiento no garantiza que todos los individuos de una sociedad puedan ejercer ese derecho, ya que se estaría hablando de cuestiones que requieren de cierta posición socioeconómica. Por lo tanto, existe un importante condicionamiento a partir de la pertenencia a un grupo social, aunque este fenómeno de condicionamiento social no es nada nuevo. La pertinencia de analizar el impacto del condicionamiento social surge a raíz del nuevo escenario tecnológico. Bostrom también es consciente de la problemática que suscita el mejoramiento humano en términos de desigualdad y sostiene lo siguiente:

> Los genéticamente privilegiados pueden convertirse en genios eternos, sanos, súper genios de una belleza física impecable, dotados de un ingenio deslumbrante y un sentido del humor despectivamente autodestructivo, irradiando calidez, encanto empático y confianza relajada. Los no privilegiados se mantendrían como las personas hoy en día, pero tal vez tendrían menos respeto por sí mismos y sufrirían episodios ocasionales de envidia. La movilidad entre las clases baja y alta podría desparecer y un niño nacido de padres pobres, sin mejores genéticas, podía encontrar que es imposible competir con éxito contra los súper hijos de los ricos (Bostrom, 2004: 502).

Razonamientos como el de Bostrom y Overall son los que conducen al planteamiento de la necesidad de discutir el fenómeno del transhumanismo en términos políticos incorporando criterios de responsabilidad y justicia. La tecnología en sí misma no tiene un carácter negativo, sino que depende de los humanos el hacer un buen uso de ella para gestionar la vida de las sociedades aspirando a mayores cuotas de felicidad. Y la estrategia a adoptar frente a la desigualdad no debería consistir en la prohibición de los mecanismos de mejoramiento, sino en diseñarlos de tal manera que adopten formas inclusivas. En ese sentido, promover laboratorios abiertos sobre ciencia cívica permitiría desarrollar la deliberación en torno a aquellos mecanismos de mejoramiento que generan controversia, incorporando a diversos representantes de los grupos de interés.

Puede considerarse una situación hipotética tomando como punto de partida los dos barrios del distrito de New York mencionados con anterioridad. Imaginen que Robert es un niño negro del barrio de Brooklyn y que Mathew es un niño blanco del barrio de Upper East Side. Robert es hijo de una familia de obreros, su padre trabaja en una fábrica de neumáticos y su madre es limpiadora en las oficinas de una empresa farmacéutica, la situación laboral de los progenitores es muy precaria y sus salarios apenas les alcanzan para pagar la hipoteca de la vivienda al banco y las necesidades que demandan la atención de un hogar. Mathew es hijo de un exitoso economista de un banco de inversión y de una agente de *marketing* de una famosa productora de cine. Robert está matriculado en un colegio público del barrio, donde todos los niños y niñas son de su misma clase social y con situaciones familiares similares a la suya, mientras que Mathew es alumno de uno de los colegios privados con mayor prestigio de New York y se relaciona con hijos de familias con grandes fortunas económicas. Teniendo en cuenta el contexto de los EE. UU., es muy probable que Mathew reciba una mejor formación académica debido a que la planta docente de su colegio posee estudios en mejores universidades que los de Robert. Sin embargo, esas diferencias educativas podrían superarse mediante las políticas subsidiarias que diversas instituciones brindan en los EE. UU.

En el ejemplo anterior existe una desigualdad educativa que podría ser corregida por políticas públicas de becas que garantizaran una igualdad de oportunidades. No obstante, en el campo del mejoramiento humano todo parece ser diferente. Mathew fue sometido desde que tenía un año hasta sus 15 a diversas terapias de mejoramiento cognitivo y muscular, pero no solo Mathew, sino la mayoría de los niños y niñas de su colegio. En cambio, tanto Robert como sus compañeros y compañeras no han dispuesto de dinero suficiente como para someterse a terapias de mejoramiento. Estas diferencias han provocado que existan campeonatos de béisbol para jóvenes mejorados y para jóvenes no mejorados, generando de ese modo una seria frustración en los jóvenes del colegio de Robert al ver que los otros jóvenes, los del colegio de Mathew, son físicamente superiores. Además, tanto Robert como Mathew van a comenzar su etapa universitaria y ambos optan por cursar sus estudios en la Universidad de Columbia. Robert ha conseguido una beca por cumplir los criterios

socioeconómicos que exigía la institución, pero no ha conseguido superar la prueba escrita de acceso al no tener la misma habilidad cognitiva que Mathew. La frustración en Robert es aún mayor que durante su adolescencia, pues ya no está siendo desplazado por su condición física, sino también por su condición cognitiva.

Este ejemplo, aunque hipotético, sirve para mostrar una situación que podría ocurrir en el futuro a raíz del avance de las tecnologías de mejoramiento y como consecuencia de la inacción de los poderes políticos para incorporar criterios de responsabilidad y justicia ante este fenómeno. Además, y siguiendo a Overall, debe mencionarse que las tecnologías de mejoramiento no solo podrían conducir a un escenario de desigualdad social, sino también de discriminación racial y étnica.

También Habermas (2010) invita a reflexionar sobre las implicaciones que podría tener en el futuro la posible tecnofilia acrítica ante el desarrollo del diagnóstico genético preimplantacional (DPI), que consiste en un método de diagnóstico prenatal para seleccionar aquellos embriones que cumplen con determinadas características y/o eliminar los que portan algún defecto. El DPI se encontraría dentro de lo que se denomina «eugenesia liberal», una práctica que responde al deseo de los progenitores o usuarios en lo que respecta a la manipulación del genoma de los embriones. Como señala Javier Aguirre Román:

> Los defensores de la eugenesia liberal suelen partir de la imposibilidad real de distinguir entre intervenciones terapéuticas, por una parte, y, por otra, intervenciones destinadas a «mejorar» la especie humana. Por esta razón, lo mejor y más sensato, es dejar la decisión de las finalidades de las intervenciones genéticas a las preferencias individuales de los «participantes en el mercado» (Aguirre Román, 2016: 8).

La eugenesia liberal despierta ciertas sospechas por parte de Habermas, ya que entiende que determinadas técnicas podrían poner en peligro la esencia natural del ser humano. No obstante, el argumento de Habermas puede ser sometido a posibles críticas como las que le dirige Ferry (2017: 98-101). En primer lugar, la distinción que Habermas lleva a cabo entre naturaleza y sociedad no es muy convincente desde el punto de vista ético. Además, otro

aspecto cuestionable del postulado del filósofo alemán en torno a la eugenesia liberal es que no ofrece los argumentos suficientes para creer que el azar sería preferible a la elección en materia de responsabilidad sobre los hijos.

Es pertinente pensar que las oportunidades de manipulación genética que brindan las biotecnologías podrían provocar un aumento de la discriminación racial en aquellas sociedades donde existe un caldo de cultivo favorable para dicha discriminación, si antes no se llevan a cabo proyectos formativos dirigidos a la ciudadanía y si la legislación competente brilla por su ausencia. Podrían imaginarse sociedades con serios problemas de racismo como la estadounidense donde el colectivo negro padece las fuertes consecuencias de la discriminación por parte de la población blanca y ciertas instituciones gubernamentales (Maestro Bäcksbacka, 2008; Navarro, 2015; Deroeux 2015). En ese sentido, es fundamental someter los ejercicios tecnológicos de mejoramiento a una evaluación contextual en el seno de los laboratorios abiertos sobre ciencia cívica que han sido propuestos para valorar los beneficios y costes que pueden tener para la humanidad.

En el ámbito del mejoramiento tecnológico, hay un asunto que está generando polémicas en las últimas décadas, como es la biomejora moral que ha sido defendida por expertos en la materia como Savulescu. En lo referente a esta biomejora moral, Cortina destaca la debilidad del argumento de Savulescu (2012: 229-243). El bioeticista australiano apela a la necesidad de la biomejora moral frente a los daños y desigualdades que podrían provocarse a raíz de una mejora cognitiva de la que solo pudiera beneficiarse un porcentaje muy reducido de la sociedad, ocasionando de ese modo marcadas diferencias en el seno de la ciudadanía. Para Cortina la debilidad del argumento reside verdaderamente en que la mejora cognitiva no representaría en términos reales un riesgo novedoso en lo referente a un incremento del poder:

> Hace décadas que la posibilidad de que un grupo de personas pueda destruir la tierra haciendo uso del poder científico-técnico es una realidad. Pero el riesgo no vendría tanto de los científicos como de gentes con poder político o económico suficiente como para tener en sus manos ese tipo de instrumentos, tales como la energía atómica o las armas de destrucción

masiva. Este peligro es una realidad. La posible mejora cognitiva puede incrementar un poder que ya existe, pero no supondría un riesgo nuevo (2017: 112).

Frente a esta situación se abre un espacio de discusión para aquellas disposiciones morales que tienen una base biológica que podrían motivar un mejoramiento. Cortina presenta en su obra *Aporofobia* la base biológica que tienen las disposiciones morales de los seres humanos. Esta autora sostiene que tantos siglos de educación no parecen haber sido suficientes para resolver muchos de los grandes males de la humanidad (2017: 116-117). En ese sentido, afirma que debe someterse a una profunda discusión la propuesta de la biomejora moral, pues si las disposiciones morales tienen una base biológica y esa base hoy puede mejorarse gracias a las biotecnologías, existe la obligación de poner encima de la mesa esas posibilidades para plantear un diagnóstico que debe incorporar las reflexiones que son detalladas a continuación (Cortina, 2017: 118-123):

1. Discutir si queremos potenciar u orientar en otra dirección los códigos inscritos en el cerebro en torno al desinterés hacia los otros lejanos.

2. La neuroética ha constatado que nuestras bases biológicas responden a un entorno social y físico periclitado.

3. El progreso moral en el nivel cultural parece no haberse dado de la misma manera que la evolución de nuestras disposiciones biológicas, las cuales han permanecido intactas durante siglos.

4. La pregunta sobre qué disposiciones morales es preciso reforzar, conduce directamente a otra pregunta: ¿qué se entiende por moral? Si, como señalan algunos autores como Jonathan Haidt (2012), la moral consiste en un conjunto de engranajes que persiguen la maximización del bien teniendo en cuenta que pertenecemos a un grupo, las tecnologías de mejoramiento contribuyen a un refuerzo moral. No obstante, como señala Cortina, las disposiciones morales no deben responder exclusivamente a necesidades grupales, sino también a las necesidades que presentan cada uno de los seres humanos.

5. Que la educación es un buen medio, pero no es suficiente, y puede complementarse con las tecnologías de mejoramiento.

6. En las últimas décadas se han realizado importantes descubrimientos en el campo de la biología en lo referente a una base biológica de nuestro comportamiento. No obstante, estas no son razones suficientes para defender sin una necesaria crítica los límites de la biomejora moral.

Por ello, en primer lugar es fundamental seguir profundizando las investigaciones en este terreno, para analizar caso por caso y valorar qué medios son los empleados (Cortina, 2017: 120). En segundo lugar, es necesario el consentimiento informado de la persona que es sometida a una intervención de este tipo y que además existan argumentos razonados (Cortina, 2017: 121). A pesar de estas propuestas que son discutidas por Cortina sobre los «melioristas morales», la autora sigue confiando en la educación formal e informal como medio para construir instituciones adecuadas que sepan hacer frente a las exigencias del futuro. Recogiendo el testigo de Cortina, es importante hacer hincapié en la necesidad de generar espacios de debate y discusión sobre las tecnologías de mejoramiento.

6.6. Mejoramiento, responsabilidad e igualdad democrática

Si el propósito de este trabajo no es otro que invitar a la reflexión sobre la necesidad de incorporar criterios de responsabilidad en el ámbito de la IA, es necesario contextualizar dicho propósito en el marco de las tecnologías de mejoramiento. A lo largo de este capítulo se han expuesto los rasgos fundamentales del transhumanismo y las perspectivas que de dicho movimiento surgen. Además, se ha insistido en el aspecto político de las tecnologías de mejoramiento para poner de relieve la necesidad de reflexionar sobre el impacto de estas tecnologías en lo que respecta a la igualdad de oportunidades y a la posibilidad de que se generen desigualdades de diversa índole. En ese sentido, es necesario plantear una discusión que gire en torno a la contribución que puede hacer la IAR en el contexto de las tecnologías de mejoramiento Esta contribución podría comenzar promoviendo el respeto y la preocupación por la igualdad de oportunidades, uno de los pilares fundamentales de las

democracias actuales, y a la vez el objetivo número 10 de los ODS, orientado a reducir la desigualdad en y entre los países. Un marco teórico útil para proporcionar un fundamento filosófico en el terreno de la cohesión democrática puede ser el principio de justicia esbozado por el filósofo estadounidense John Rawls (1997). El liberalismo rawlsiano aporta al MIAR el interés por la persecución de una igualdad democrática que no podría entenderse sin hablar antes de justicia distributiva. La obra de Rawls se caracteriza por una fuerte preocupación a la hora de encontrar un concepto de justicia que sea racional. Para ello establece una confrontación entre dos concepciones de lo justo: aquella que se encuentra representada por el principio de utilidad, y aquella que lo está por su versión contractualista (Martínez Navarro, 1999: 26). Pero para Rawls el principio de utilidad presenta algunos problemas que deben ser abordados y superados por la doctrina del contrato social. El contractualismo proporciona una sólida base desde la que concebir un concepto de justicia que surja del acuerdo de individuos racionales que forman parte de la sociedad. Esta visión rawlsiana de la justicia que toma como punto de partida el contrato social entiende que los principios rectores que determinan lo que es justo o injusto tienen como premisa fundamental el acuerdo de los ciudadanos en una posición inicial hipotética en la que nadie pretendería imponer sus propios intereses sobre los de los otros.

Las reglas que ordenan las sociedades deben ser fruto de un contrato social en el que diferentes personas y grupos sociales se pongan de acuerdo y persigan un arreglo justo. El acuerdo no es fácil cuando se trata de política, pues las personas responden a la tendencia de apostar en función de sus intereses y creencias, además de que pueden situarse en una posición social diferente. Ante esta situación marcada por la diferencia, Ralws propone tomar como punto de partida un experimento mental que él denomina «posición original». Esta posición original es reconocida como el comienzo apropiado desde el que garantizar acuerdos que sean imparciales y que sirvan para derivar posteriormente en la justicia como imparcialidad.

En la posición original se procede según el velo de la ignorancia que promueve la suspensión temporal de la identidad y los intereses, es decir, de la clase social, el género, raza, étnica, credo, etc., que no deben influir en la toma de decisiones para lograr el contrato social. Esa suspensión permite, según Rawls, partir desde una posición originaria de igualdad que facilitaría los acuerdos.

Parece razonable suponer que en la posición original los grupos son iguales, esto es, todos tienen los mismos derechos en el procedimiento para escoger principios; cada uno puede hacer propuestas, someter razones para su aceptación, etc. Obviamente el propósito de estas condiciones es representar la igualdad entre los seres humanos en tanto que personas morales, y como criaturas que tienen una concepción de lo que es bueno para ellas y que son capaces de tener un sentido de la justicia. Como base de la igualdad se toma la semejanza en estos dos aspectos. Los sistemas de fines u objetivos no están jerarquizados en cuanto a su valor, y se supone que cada quien tiene la capacidad necesaria para comprender y actuar conforme a cualquier principio adoptado. Estas condiciones, junto con el velo de la ignorancia, definen los principios de justicia como aquellos que aceptarían como seres iguales y personas racionales preocupadas por promover sus intereses, siempre y cuando supieran que ninguno de ellos se encuentra en una posición de ventaja o desventaja por virtud de contingencias sociales y naturales (Rawls, 1997: 31).

Rawls mantiene la creencia que del contrato hipotético que plantea surgirán dos principios de justicia. El primer principio promoverá iguales libertades básicas a todos los ciudadanos, siendo prioritario sobre otras consideraciones que tengan utilidad social y bienestar general. El segundo principio hace referencia a la igualdad social y económica, no requiriendo una igual distribución de las rentas y del patrimonio, sino solo permitiendo desigualdades sociales y económicas que sirvan para mejorar la situación de aquellos miembros menos prósperos de la sociedad. La desigualdad debe ser respetada en el sentido en que pueden elevar al máximo las expectativas de aquellos que sufren la posición más baja (Martínez Navarro, 1999: 27).

Es importante detenerse ante la cuestión de la igualdad, pues hablar de igualdad de oportunidades sin más resulta insuficiente y una mayor claridad es requerida. Las oportunidades no son en un principio objetos materiales que sean fácil ni inmediatamente mensurables. En ese sentido, las oportunidades podrían designar aquel conjunto de recursos de los que cada individuo se beneficia, la totalidad de las capacidades, las posibilidades de realización, o también la libertad real. En su obra *Teoría de la justicia*, Rawls concede una primacía al principio de igualdad de oportunidades sobre el principio de la diferencia. Este principio de igualdad brinda oportunidades de acceso a posiciones sociales y forma parte de esa libertad real de cada ser humano. Las desigualdades sociales resultantes de la raza, el sexo o la religión, pueden ser corregidas mediante una igualdad en la distribución. En ese sentido, el principio de igualdad de oportunidades rawlsiano representa una igualdad dentro del marco político del liberalismo, que tiene como principal función la oposición y compensación de los efectos de las diversas discriminaciones y desigualdades.

Así pues, en el contexto de la IAR los postulados de Rawls sirven para fundamentar una nueva dinámica de producción de conocimiento que tenga una función cívica y por lo tanto un claro beneficio para la sociedad en términos de igualdad. Las tecnologías de mejoramiento pueden jugar un importante papel en el proceso de fortalecimiento democrático y prestar un servicio para luchar contra la desigualdad a través de mecanismos de reconocimiento del principio de igualdad de oportunidades. Para alcanzar este objetivo resulta fundamental potenciar el debate y la discusión entre los diferentes grupos de interés en el marco de foros abiertos sobre ciencia cívica. Así pues, unas tecnologías avanzadas e impulsadas desde la IA que servirían como enriquecimiento de la democracia y los ODS por medio de una incorporación del criterio de responsabilidad y la deliberación inclusiva.

6.7. Tecnologías de mejoramiento con compromiso

El concepto de IAR esbozado en el capítulo 5 se caracteriza por tener un fuerte compromiso con el cumplimiento y fortalecimiento de los derechos humanos, los ODS y los límites planetarios. En ese sentido, es necesario articular un equilibrio entre la IA y los derechos humanos. Como se ha señalado, la propuesta de una IAR surge de una

preocupación cívica y democrática que fundamenta su despliegue en el marco de los derechos humanos. Es fundamental observar cómo los gobiernos de los países del mundo están respondiendo al desafío tecnológico desde parámetros de responsabilidad. En ese sentido, el Secretario General de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), señaló lo siguiente:

> Necesitamos analizar la digitalización de la economía y la sociedad desde el punto de vista de todo gobierno. Tenemos que ir más allá de los márgenes de nuestras estrategias y llegar y superarlos para observar mejor cómo la digitalización está cambiando nuestras vidas, cómo podemos abusar de ella, y cómo podemos ayudar a aquellos en peligro de quedarse atrás (Pariser, 2011: 4).

40 gobiernos de todo el mundo están dando forma a programas que se enfocan en el interés por la IA y las tecnologías de mejoramiento que incorporan aprendizaje automático, entre los que se encuentran los que se indican a continuación.

En EE. UU. el Gobierno ha impulsado el Instituto Nacional de Ciencia y el Subcomité del Consejo de Tecnología centrada en el aprendizaje automático. Estos organismos están orientados a la valoración sobre los avances concretos en el campo de la IA y el aprendizaje automático, tanto en el sector privado, como en los gobiernos federales y en el escenario internacional. Entre las primeras acciones de este instituto se encuentran la elaboración de dos documentos: *Preparing for the Future of Artificial Intelligence* y *National Artificial Intelligence Research and Development Strategic Plan*. El Gobierno de Canadá, través del Consejo de Tecnologías de la Información, llevó a cabo la invitación para un diálogo nacional en el que se propusieran metodologías que reflexionen sobre el desarrollo de la IA. Entre los resultados de este diálogo se encuentra la Declaración de Montreal, presentada en el capítulo 5. El Ejecutivo de Japón ha promovido en el G7 la formación de una visión compartida en materia de IA que se encuentre centrada en el ser humano a partir de parámetros de cuidado y confianza, mediante el fortalecimiento de la protección de los datos. De un modo parecido, en Singapur, el Gobierno impulsó en 2014 el programa «país inteligente» para promover un empleo de la IA en beneficio de la ciudadanía, centrado principalmente en el manejo de datos, el Internet de las cosas, etc. Además, en Corea del

Sur, la institución gubernamental ha constituido un grupo de expertos para aumentar la I+D+I por medio de la IA (Chakraborty, 2018: 25-26).

Los organismos responsables de los gobiernos del mundo deberían profundizar su interés en el enorme potencial que las tecnologías más avanzadas tienen para su integración en la condición humana y generar espacios de discusión abierta y participativa sobre ciencia cívica. Para ello es fundamental diseñar una serie de lineamientos que alimenten las estrategias en materia de investigación en el terreno de la IA. Estos lineamientos se sitúarían en el contexto de la Declaración Universal de los Derechos Humanos (DUDH) y de ese modo los resultados de la innovación tecnológica responderían a un compromiso con el bienestar de la humanidad.

Este compromiso con los derechos humanos lleva necesariamente al terreno de las regulaciones. La ética se encuentra en la base de todas aquellas propuestas y soluciones que planteen una transición transhumanista. El establecimiento de directrices y regulaciones legales, así como aquellas técnicas que sirvan para limitar los usos poco éticos, principalmente contrarios a los derechos humanos, deben ser una exigencia ineludible en la hoja de ruta de toda institución que promueva el diseño, producción o empleo de la IA para mejorar cualquier actividad o condición de los seres humanos. En materia regulatoria existen dos propuestas que han sido presentadas en los últimos años:

1. En Estados Unidos: el Plan Nacional Estratégico de Investigación y Desarrollo en Inteligencia Artificial, de octubre de 2016, del Consejo Nacional de Ciencia y Tecnología.

2. En la Unión Europea: la Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (Salgado, 2017)

Es importante mencionar que la propuesta estadounidense fue presentada durante la presidencia de Barack Obama y que en la actualidad, bajo el mandato del presidente Donald Trump, no se le ha dado la misma importancia. En cambio, la resolución europea ha

servido para impulsar proyectos de legislación en el sector de la robótica. El 20 de septiembre de 2018 la Comisión Europea reunió a un grupo de expertos para abordar temas relacionados con la ética de las personalidades artificiales. Esta iniciativa de la Comisión Europea gira en torno a los robots y la IA, buscando de ese modo legislar en este terreno. Se han planteado algunos elementos como fundamentales para la elaboración de este reglamento ético, entre los que se encuentran: la adaptación del comportamiento, la adquisición de autonomía, el autoaprendizaje, el soporte físico y la inexistencia de vida (Salgado, 2017). La Comisión Europea se ha propuesto crear un registro de robots inteligentes para poder llevar a cabo una identificación de los mismos y de ese modo posteriormente regular en el terreno de los derechos y deberes, tanto de los usuarios, como de los creadores.

El actual escenario transhumanista exige el compromiso con un nuevo enfoque jurídico y dogmático para el tratamiento de los derechos fundamentales, que debería partir del debate y la discusión abierta entre los grupos de interés, fortaleciendo así estrategias de abajo arriba en el proceso de toma de decisiones. Las diversas formas que adopta el empleo de la IA pueden producir un daño individual aparentemente imperceptible para esos derechos, y a pesar de esa falta de percepción, se correría el riesgo de un daño colectivo a los mismos, lo que supondría un impacto aún mayor. Para reconocer la necesidad de un nuevo enfoque jurídico y dogmático es fundamental asumir la potencialidad de la IA y la vez la posibilidad de dar una respuesta jurídica con mecanismos garantes de derechos fundamentales concretos (Cotino Hueso, 2017: 137). El Parlamento Europeo, conocedor de la necesidad de este compromiso, impulsó la elaboración del documento sobre robótica antes mencionado, y también una iniciativa bajo el título *Implicaciones de big data para los derechos fundamentales: respecto a la privacidad, protección de datos, no discriminación, seguridad y cumplimiento de la ley* (2016).

Por lo tanto, la IA promovida por el espíritu transhumanista no debe situarse al margen de los compromisos reglamentarios y jurídicos que infunden los derechos humanos. En ese sentido, la actividad de la IA debe perseguir un equilibrio y armonía con las exigencias humanistas de carácter universal que se encuentran implícitas en diversas legislaciones

actuales. Además, esta exigencia de compromiso por el equilibrio y la armonía puede ser asumida como una oportunidad desde la que plantear una orientación de carácter cívico y democrático en los postulados transhumanistas.

6.8. Tecnología, bienestar humano y medioambiental

El desarrollo de la IA puede aportar importantes beneficios a la lucha contra la degradación del medio ambiente y contribuir así a una mejora de la calidad de vida de los seres humanos y la biosfera. Una IAR que contribuya de esa manera a la mejora de la calidad de vida de los seres vivientes se fundamentaría en un humanismo tecnológico que centra su interés en la búsqueda de los mecanismos necesarios para un pleno florecimiento de la vida y sus facultades. De este modo las tesis transhumanistas podrían enriquecerse considerablemente, ya que no solo se centrarían en aspectos relacionados con las biotecnologías orientadas a la mejora directa de la condición física de los seres humanos o a su lucha contra el envejecimiento. Por medio de un MIAR se generan conocimientos innovadores que pueden participar del propósito transhumanista, sin embargo, trasladándolo a un plano ambiental, es decir, contribuyendo a la reproducción de las condiciones ideales para que los organismos vivos pudieran desplegar su modo de vida de la mejor forma posible en armonía con la biosfera.

Si el propósito principal del movimiento transhumanista se centra en la mejora de la especie y además se toman en cuenta los ODS números 7 (energía asequible y no contaminante), 13 (acción por el clima), 14 (vida submarina) y 15 (vida de ecosistemas terrestres), podría establecerse un vínculo entre estos ODS y este propósito fundamental. La justificación de la posibilidad de construir este vínculo radicaría en la urgencia de cuidado medioambiental, ya que las condiciones de vida de la especie humana también pueden perfeccionarse si las tecnologías avanzadas contribuyen a la mejora del medio ambiente. Así pues, no se trataría exclusivamente de promover las tesis transhumanistas mediante una incidencia directa de la tecnología en el cuerpo humano, sino también en su medio ambiente.

Es importante destacar que, cuando se establece un vínculo entre la IAR y los postulados transhumanistas en términos medioambientales, no se está situando este discurso en la estela del ecomodernismo. Este distanciamiento respecto del ecomodernismo se debe principalmente a que este movimiento defiende el marco liberal-capitalista como la base del proyecto político desde el que pretende impulsarse la eficiencia tecnológica como mecanismo de racionalización y adaptación al cambio climático (Cortina, 2017: 172-173). En ese sentido, en la base de la IAR se encuentra un fuerte compromiso de cuestionamiento crítico al capitalismo, pues defiende la redefinición de la intencionalidad proyectada en el diseño de los sistemas artificiales. El diseño tecnológico proyectado desde la IAR no debe responder exclusivamente a aspectos económicos, sino que se sitúa del lado de los grandes desafíos que enfrenta la humanidad y que se encuentran recogidos en los ODS.

A lo largo de este trabajo se han expuesto varios ejemplos sobre algunos de los ámbitos en los que la IA está comenzando a ejercer un papel importante, siendo el medioambiental uno de los principales para las investigaciones en IA. En ese sentido, pueden desarrollarse sistemas artificiales que obtengan y procesen datos a partir de estudios de campo que permitan conocer la realidad de un modo más aproximado. Microsoft ha desarrollado *Conservation Metrics*, un proyecto basado en IA para la protección de los ecosistemas que se está utilizando en países como Ecuador, para la detección de tala ilegal y otras actividades humanas, en colaboración con *Rainforest Connection*; o en el Congo, en la aplicación de modelos de redes neuronales para detectar y cuantificar vocalizaciones de elefantes en conjuntos de datos acústicos masivos, en colaboración con el Proyecto de Escucha de Elefantes de la Universidad de Cornell. *AI for Earth* es otro *software* de *Microsoft* desarrollado con IA que tiene como objetivo el impulso de proyectos en las siguientes cuatro áreas: clima, agricultura, biodiversidad y agua. Este *software* está siendo empleando en lugares como el Instituto Patagónico para el Estudio de los Ecosistemas, el Centro Intercultural de Estudios de Desiertos y Océanos (México), o el Centro Humboldt (Nicaragua) (Microsoft, 2018).

Existen por lo tanto claros ejemplos de IA que incorporan criterios de responsabilidad en la protección medioambiental, lo que significaría una evidente estrategia de IAR. El MIAR permite un acercamiento a la realidad de cada una de las esferas de la quíntuple hélice y contribuye a fomentar el debate y la discusión para desarrollar una ciencia cívica. Esto significa que cada esfera conoce de cerca su realidad y por lo tanto puede proporcionar datos lo suficientemente cercanos como para generar conocimientos que den posibles soluciones a las problemáticas existentes. Una vez que los datos han sido proporcionados pueden impulsarse propuestas de IAR en aquellos espacios en los que se considere la pertinencia de una mejora medioambiental que favorezca la condición biológica del ser humano. Las tecnologías avanzadas que integran IA en sus sistemas también pueden abrirse paso en una nueva senda de investigaciones que favorezcan esta condición más allá de una incidencia directa sobre sus organismos vitales. Así pues, la IAR proporcionaría un abanico de nuevas posibilidades al proyecto transhumanista mediante la ampliación de su espectro de acción hacia espacios medioambientales que a la vez permitirían combatir la degradación de la biosfera y cultivar las habilidades cívicas que favorecieran la sensibilización.

6.9. Narratividad y evaluación ética

Ante el poder que subyace en la técnica, anunciado por filósofos como Jonas (1995) o Ellul (2003), es importante dirigir una mirada reflexiva desde la que evaluar los postulados transhumanistas para considerar sus impactos y las problemáticas que de ellos se derivan, tal como se ha propuesto en esta tesis al considerar la importancia de fortalecer espacios de debate y discusión e impulsar una ciencia cívica.

Para la evaluación ética de los casos de mejoramiento que promueven los postulados transhumanistas, es importante rescatar el modelo narrativo empleado en el ámbito de la bioética que encuentra su origen en la dimensión fronética del pensamiento aristotélico y en la hermenéutica. Hay expertos en el campo transhumanista como Antonio Diéguez (2017), que señalan la necesidad de un ejercicio hermenéutico sobre las tecnologías de

mejoramiento. De la misma manera, Alfredo Marcos (2018: 112) apoya una de las tesis de Diéguez, que se detalla a continuación:

> Aquí, como en muchas otras ocasiones históricas en que los cambios fueron demasiado rápidos como para tener una visión clara de la totalidad de los acontecimientos y de su significado, habrá que atender a las situaciones concretas y a los matices diversos con la diligencia que se pueda. Es primordial, por tanto, evitar el error común de realizar juicios generales y definitivos, de lanzar condenas o alabanzas globales, puesto que, además de no ser de demasiada utilidad, apenas convencerían más que a los del coro (Diéguez, 2016: 194).

La bioética actual ha centrado una de sus líneas de investigación en la dimensión narrativa de la vida humana, donde es tomado en cuenta el contexto que condiciona la realidad y que debe ser sometido a un ejercicio hermenéutico. La perspectiva narrativa de la vida señala el carácter lingüístico de la dimensión experiencial del pensamiento y toma distancia de aquellos postulados que tratan de promover un racionalismo universal y abstracto. Así pues, debido a esta visión narrativa, la bioética ha experimentado un giro hermenéutico. Tomás Domingo Moratalla y Lydia Feito Grande destacan las siguientes características:

1. En la perspectiva narrativa se enfatiza la idea de lo particular, de la experiencia, del sentido único de la vivencia para los implicados, y de la necesidad de evaluar lo más específico del caso para poder tomar decisiones. La ética narrativa rechaza el modelo de los principios, especialmente cuando este se convierte en un mandato abstracto y alejado de la vida de las personas.

2. La ética narrativa intenta recuperar dimensiones de la moral que han sido relegadas u olvidadas, como la experiencia vital, el sentido personal que se otorga a los acontecimientos, o la dimensión de responsabilidad y compromiso con los otros seres humanos. Aporta, además, una reflexión sobre la educación de actitudes morales, subrayando que la enseñanza de contenidos y procedimientos racionales está incompleta si no trabaja también la dimensión actitudinal.

3. En relación con lo anterior, la ética narrativa se inscribe en un conjunto de aproximaciones que están insistiendo reiteradamente en la necesidad de completar el modelo moderno de la ética racionalista, decisionista y principialista, con una perspectiva desde la relación, el contexto, la atención a lo particular y los elementos emocionales o afectivos que influyen en la toma de decisiones y en las actitudes. Este enfoque desde luego no es novedoso, ya que hunde sus raíces en la ética aristotélica y, en general, en las éticas de la virtud. Sin embargo, su vigencia actual es enorme y aporta como novedad el intento de aplicación (2017: 38-39).

La perspectiva narrativa de la bioética centra su atención en el contexto concreto, histórico, en las circunstancias y en lo experiencial, para extraer aquellas notas características que pueden contribuir de alguna manera con algún grado de universalidad (Domingo Moratalla y Feito, 2017: 39). Además, la tarea hermenéutica parte de un reconocimiento de la experiencia concreta que no se acerca a la neutralidad debido a ese carácter narrativo. Hay aspectos que demandan ser puestos de relieve a través de un enfoque expresivo mucho más abierto y profundo. En ese sentido, la bioética narrativa, como señala Jonsen (2016), enriquece el trabajo casuístico que basaba su razonamiento en tres pasos:

- Determinación de temas.
- Interpretación de máximas y principios.
- Argumentación por analogía.

Esta perspectiva ética de carácter narrativo se sitúa en la estela de la casuística, pues el interés es orientado hacia el análisis de problemas éticos que subyacen en los casos particulares. El análisis de lo particular permite llevar a cabo una ilustración de ciertos problemas que sirven a modo de «puente narrativo» (Domingo Moratalla y Feito, 2017: 42) con reglas generales. Frente al enfoque principialista, que promueve una aplicación de principios universales que en ocasiones son difíciles de contextualizar en casos concretos, la identificación de problemáticas concretas destaca aspectos más relevantes para el análisis y la generación de nuevos conocimientos que sean producto de la experiencia. Cuando

Diéguez (2017), y posteriormente Marcos (2018), afirman la necesidad de evaluar caso por caso las estrategias transhumanistas y tomar distancia de posiciones generalizadoras, se están situando precisamente en esta línea narrativa alejada del principialismo.

El énfasis en la experiencia sirve para fundamentar una dimensión narrativa del enfoque ético propuesto en el ámbito transhumanista. Para la actuación con responsabilidad, y por lo tanto para contextualizar la IAR, resulta un requisito ineludible un conocimiento aproximado de las problemáticas concretas de los casos particulares, que podrían abordarse en el marco de una discusión abierta, con los grupos de interés, para construir una ciencia cívica. Un acercamiento aproximado a la experiencia permite adquirir un mejor conocimiento para la evaluación ética y la toma de decisiones a partir de los sentidos que los seres humanos plasman en la construcción de sus relatos de vida. Pues las vidas se desarrollan en el curso de un relato configurado por circunstancias contextuales que exigen ser reconocidas lo más cerca posible si lo que verdaderamente se pretende es valorar y lograr un mejor conocimiento para la evaluación de la experiencia moral implícita en los proyectos transhumanistas. Esta concepción narrativa de la experiencia lleva a cabo un acercamiento al modo que adopta la construcción de los sentidos que configuran la acción humana que impulsan los deseos transhumanistas.

En el capítulo 5 se mencionó la deliberación como uno de los pilares fundamentales del concepto de IAR. En ese sentido, la deliberación representa una herramienta esencial desde la que construir reflexiones éticas y proponer ponderaciones sobre los hechos concretos en los que se encuentren implicados aspectos morales que requieren una serie de acciones. La deliberación promueve la toma de decisiones frente a situaciones susceptibles de una problematización fronética que exige responsabilidad. En ese horizonte de comprensión que brinda la narración, la deliberación cobra su sentido y adquiere su valor como un método de descubrimiento de aspectos valorativos y normativos que deben someterse a un profundo ejercicio reflexivo y ético. Frente a los casos concretos que sugieren una problematización en el contexto de los postulados transhumanistas, la deliberación encarna una propuesta narrativa fundamental para impulsar una evaluación ética y un cultivo de habilidades cívicas y democráticas.

CAPÍTULO 7

EL DESAFÍO EN EL ÁMBITO DE LAS PROFESIONES

> Frente a lo desconocido, es difícil saber qué sentir o qué hacer. Resulta tentador mostrarse temoroso. Pero, ante esta inmensa e imponente fuerza transformadora, no deberíamos sentir temor, sino generosidad. Deberíamos ser tan generosos como esté en nuestra mano.
>
> (Avent, 2017: 335)

Este capítulo versa sobre el impacto de los sistemas artificiales en el ámbito de las profesiones y cuáles son los principales efectos que podrán generarse. Se explorarán los inminentes cambios y desafíos que el mundo del trabajo enfrentará en las próximas décadas como resultado directo del desarrollo de las nuevas tecnologías que integran IA.

La IA se encuentra inmersa en un proceso de desarrollo y adaptación en el ámbito de las profesiones. Es muy probable que se den cambios radicales para los que la sociedad no está preparada. Es importante reconocer la dificultad que existe para predecir el futuro con suficiente certeza, aunque pueden hacerse estudios estimativos y generar profundas reflexiones para medir los impactos y estudiar posibles alternativas con anticipación. *Sitra*, el Fondo de Innovación de Finlandia, presentó el informe *Megatrends 2017* en el que se reflexiona sobre la naturaleza del trabajo en el futuro y se destacan dos posibles resultados:

1. Las nuevas posiciones quedarán vacantes debido a que las personas se jubilarán y el hecho de que el aprendizaje de nuevas habilidades y una mejor combinación de las personas en el mercado laboral con las vacantes ayudará a que éste crezca. La digitalización y la inteligencia artificial se aprovecharán para ayudarnos y al trabajar todavía podremos llevar una vida buena y decente. Todavía habrá suficiente trabajo para un gran número de personas. La cantidad de trabajo y las formas de trabajo serán diferentes. Habrá un mayor énfasis en la interacción con los demás, o esto se llevará a

cabo virtualmente en equipos globales. En tal escenario el trabajo seguirá siendo una parte central de nuestras vidas.

2. Otro escenario ampliamente considerado es el del aumento de la desigualdad, donde una pequeña elite trabajará de manera extremadamente productiva y una elite aún más pequeña acumulará el capital, por ejemplo, al poseer la tecnología y las plataformas. Esto dará lugar a una enorme riqueza, pero no en el trabajo. La digitalización y la inteligencia artificial reemplazarán en gran medida el trabajo realizado por los humanos. En este caso, la distribución de la riqueza creada será el núcleo de la organización social (Sitra, 2017).

Estos dos posibles escenarios que presenta Sitra invitan a pensar en la necesidad de que el fenómeno de la automatización del ámbito de las profesiones sea sometido a un profundo ejercicio reflexivo para valorar sus impactos. Las reacciones frente a este fenómeno podrían encontrarse en dos grupos claramente diferenciados: tecnooptimistas y tecnopesimistas. En el lado del tecnooptimismo se sitúan figuras como Matt Ridley, Ronald Arkin y James Bessen, quienes observan la introducción de los intelectos sintéticos desde un punto de vista fundamentalmente optimista y en ocasiones carente de reflexión crítica y excesivamente confiado. Mientras que en el lado de los tecnopesimistas se encuentran Nick Bostrom, Tyler Cowen y Martin Ford. A pesar de la existencia de diferencias entre los postulados pesimistas, un elemento común es la alerta y desconfianza frente a la automatización. Más allá de estos dos postulados, excesivamente confiados y ciertamente desalentadores, puede existir una tercera vía fundamentada en un humanismo tecnológico y en un ejercicio de responsabilidad ética que permita valorar estos fenómenos desde un optimismo crítico consciente de los peligros y que extraiga de la IA el máximo provecho para el beneficio de la humanidad.

El sentido que se proyecta sobre la tecnología puede situarse en una estela de compromiso y preocupación por los derechos humanos y los ODS, así como por los límites planetarios, con el fin de que la irrupción de los intelectos sintéticos en el ámbito laboral no provoque una serie de consecuencias inesperadas que impliquen un impacto negativo para las sociedades, como podría ocurrir en materia de justicia y derechos. La IAR representa una propuesta reflexiva para someter el fenómeno de la automatización a una valoración

contextualizada en el marco de unas exigencias democráticas y de respeto al ser humano y a la biosfera.

7.1. Impactos de la tecnología en el ámbito profesional

Es conocido que la IA ha experimentado un importante progreso en las últimas décadas en diversos ámbitos. Esa amplitud de campos en los que la IA ha experimentado un importante progreso no ha permitido que el ámbito laboral se encuentre al margen, pues la mayoría de los avances en materia de IA tienen que ver con la automatización de determinadas profesiones.

La tecnología está transformando considerablemente el ámbito de las profesiones y es el principal detonante de dicha transformación. Cada vez hay más empresas que dedican sus investigaciones a la creación de una variedad de sistemas, máquinas, herramientas, etc., cuya finalidad es la recopilación y almacenamiento de gran cantidad de conocimiento y habilidades que tradicionalmente han sido propias de los seres humanos. Tal como señalan los ingleses Richard y Daniel Susskind, esta transformación impulsada por la tecnología ha tenido principalmente dos impactos: la automatización y la innovación (2016: 109).

Con el término «automatización» estos autores se refieren a aquel fenómeno que tiene como objetivo lograr eficacia y ahorrar costes, lo que tradicionalmente se ha conocido como optimización. Podría considerarse que la automatización ha estado vinculada con la mejora de aquellos sistemas habitualmente manuales, algo que no es incierto. Sin embargo, cuando en la actualidad se habla de automatización se está insistiendo más en la incorporación de la tecnología en la profesión que en la mejora de los aspectos manuales. La tecnología tiene un gran alcance en el ámbito de las profesiones en lo relativo al apoyo y la complementariedad con el trabajo de los profesionales. Aquellos trabajos caracterizados por la monotonía y la rutina suelen encontrarse en el foco de la automatización, aunque también otros que no son tan monótonos y que requieren de ciertas habilidades persuasivas, como algunas estrategias de *marketing* elaboradas por empresas importantes a partir de cuestionarios en la web. Un ejemplo de todo esto puede verse en la plataforma *Skype* de

Microsoft que utilizan algunas empresas sanitarias para prestar sus servicios médicos, como es el caso de *iGlobalMed*.

Considérese ahora el otro impacto que ha tenido la tecnología, a saber, la innovación. Tiene que ver con aquellos servicios novedosos que han sido prestados en ciertos espacios y que no necesariamente tienen como objetivo la sustitución de una tarea rutinaria que tradicionalmente desempeñaba un humano. La innovación tecnológica brinda una serie de conocimientos prácticos que permite ampliar el espectro de actuación. Esta apertura ha suscitado la ampliación del horizonte de muchas posibilidades que hasta hace pocos años nadie imaginaba. Existen aplicaciones para los teléfonos inteligentes que ofrecen un menú en función de las preferencias de gusto y disponibilidad económica sin necesidad de desplazarse del escritorio de la oficina, como es el caso de Adomicilioya, que contó en el año 2016 con 130.000 descargas, según el diario *El Universo* de Ecuador, o *Uber Eats*. Estudiosos del impacto de la tecnología sobre las profesiones como los Susskind muestran optimismo ante el fenómeno de la innovación, pues aseguran que facilitará la producción de nuevos conocimientos (Susskind y Susskind, 2016: 112).

Es evidente que la tecnología ha servido para contribuir a la innovación en numerosos sectores y que eso ha permitido acrecentar considerablemente los conocimientos y también ha ampliado el acceso a nueva y mejor información. Un claro ejemplo es la informatización de los servicios de gestión de rentas internas de los países que ha facilitado una mejor organización de los datos económicos de los contribuyentes, colaborando además con la persecución del fraude fiscal de una forma más rápida y eficiente. En ese sentido, el impacto de la tecnología sobre las profesiones ha sido de tal magnitud que al comienzo de la irrupción de los ordenadores en ciertos trabajos ni siquiera existía conciencia social suficiente sobre el horizonte de posibilidades que se vislumbraba. En la actualidad, las profesiones no pueden entenderse sin el uso de las tecnologías, lo que implica que la adopción de una actitud de resistencia frente a las mismas no representa una estrategia inteligente, ya que la cantidad de cursos de formación continua y capacitaciones que brindan numerosas instituciones educativas hoy en día son un claro ejemplo de la demanda de una actualización constante de la relación con la tecnología en el ámbito laboral.

Richard Sennett ha estudiado los efectos provocados por los cambios que están dándose en los últimos tiempos en el ámbito económico y profesional, sobre todo en el sentido que afectan a la vida personal de los profesionales a través de su participación en actividades institucionales y en las relaciones sociales. Sennett (2000) afirma que los cambios en los espacios mencionados están ejerciendo una influencia considerable sobre la elaboración de las narrativas personales y el sostenimiento del carácter:

> Poner el acento en la flexibilidad cambia el significado mismo del trabajo, y con ellos las palabras que usamos para hablar del trabajo. […] El capitalismo flexible ha bloqueado el camino recto de la carrera, desviando a los empleados, repentinamente, de un tipo de trabajo a otros. […] Es totalmente natural que la flexibilidad cree ansiedad; la gente no sabe qué le reportarán los riesgos asumidos ni qué caminos seguir. En el pasado, quitarle la connotación maldita a la expresión «sistema capitalista» dio lugar a muchas circunlocuciones como sistema de «libre empresa» o de «empresa privada». En la actualidad, el término flexibilidad se usa para suavizar la opresión que ejerce el capitalismo. Al atacar la burocracia rígida y hacer hincapié en el riesgo se afirma que la flexibilidad da a la gente más libertad para moldear su vida. De hecho, más que abolir las reglas del pasado, el nuevo orden implanta nuevos controles, pero éstos tampoco son fáciles de comprender. El nuevo capitalismo es, con frecuencia, un régimen de poder ilegible (Sennett, 2000: 9-10).

La orientación cortoplacista y el principio de flexibilidad predominantes en el mundo del trabajo durante las últimas décadas están creando nuevos escenarios de competitividad profesional que demandan respuestas innovadoras por parte del sector empresarial. Defectos como rigidez, rutinización, ineficiencia, lentitud y la limitación de la autonomía de los profesionales, caracterizan el actual modelo económico de muchas empresas. Un sector empresarial de vanguardia tendría que promover un modelo de organización novedoso que sea más flexible para superar así el estancamiento burocrático y dar respuestas con capacidad de adaptabilidad en un mundo cada vez más cambiante fruto de los avances tecnológicos.

7.2. La inteligencia artificial como sustento de la innovación y la automatización

Detrás de todo el diseño tecnológico para diversos campos profesionales se encuentra la IA. La familiaridad con los sistemas de IA ha ido en aumento conforme pasan los años y eso ha provocado que en ocasiones se convierta en una difícil tarea la identificación de estos sistemas y que por lo tanto se pasen por alto. La IA ha penetrado en muchos aspectos de la vida como las computadoras, la banca electrónica, las compras por Internet, etc. El crecimiento exponencial de variados sistemas de IA ha permitido ampliar el espacio de incidencia de la tecnología en las profesiones.

Las máquinas en general, y los intelectos sintéticos en particular, aumentan su poder sobre la vida hasta tal extremo que condicionan el futuro y la estabilidad de muchas profesiones. El aumento exponencial de diversos sistemas es más que evidente, y así lo demuestran los macrodatos (*big data*) que se encargan de la manipulación de gran cantidad de información recopilada mediante herramientas de minería (*data mining*). Las máquinas manipulan gran cantidad de datos sobre la vida, datos que se comunican de unas a otras, lo que representa un considerable aumento de poder. Si la IA que se encuentra detrás del progreso de la tecnología está estimulando una incidencia cada vez mayor de nuevas fórmulas tecnológicas en el área de las profesiones, es porque la investigación es creciente, lo que supone una recuperación del clásico cuestionamiento de Turing (1950) acerca de si es posible, o no, que las máquinas piensen. Como indica Andrés Ortega:

> Estamos ante una combinación de una cierta mecánica y de algoritmos que impulsa el progreso de las máquinas y entre estas, de los robots, con cambios en la manera cómo perciben, razonan, controlan y coordinan, según recuerda la experta en computación Daniela Rus. Nos encontramos ante inteligencias en el sentido de que las máquinas son capaces de procesar información y reaccionar ante su entorno, lo que es muy relevante. Se trata de «robótica desarrollista» o experimental (Ortega, 2016: 82).

En los últimos años existe un fuerte interés en el desarrollo de la IA que va en dos direcciones: la primera tiene que ver con una teoría de la información que sea más sólida para el aprendizaje artificial; la segunda se centra en el desarrollo del aspecto práctico y comercial de varios sistemas de resolución de problemas concretos y de ámbitos específicos. No existe una postura única en torno al futuro de la IA. Bostrom señala a este respecto: «Las opiniones de los expertos sobre el futuro de la IA varían enormemente. No hay acuerdo sobre la sucesión temporal de los acontecimientos ni sobre qué formas podría llegar a adoptar la IA. Las predicciones sobre el futuro desarrollo de la inteligencia artificial, señaló un estudio reciente, «son tan firmes como diversas» (2016: 19).

El desarrollo en los diferentes campos que conforman el grueso de la IA no se da de la misma manera, pues hay ámbitos que dependen a su vez de otros. Por ejemplo, el campo de la robótica no evoluciona al mismo ritmo que el del aprendizaje maquinal. *Boston Dynamics* es una empresa de ingeniería y robótica fundada en 1992 por Marc Raibert, exprofesor del MIT, que ha experimentado numerosos avances en los últimos 25 años. Así pues, el tiempo comprendido para desarrollar ciertos proyectos en este ámbito es en ocasiones superior que el del *machine learning* para sus proyectos.

Los investigadores de IA reconocen el papel tan relevante que juega el aprendizaje dentro de la inteligencia humana y se preguntan si es posible emular esa forma de aprendizaje en las computadoras. El aprendizaje maquinal tiene como objetivo la creación de programas que permitan la generalización de comportamientos a partir de ejemplos que le son suministrados, y que por lo tanto generan un patrón de comportamiento. Como señala Jerry Kaplan: «Como descripción general, los programas informáticos que aprenden extraen patrones de los datos» (2017: 32).

La IA promueve una automatización de las profesiones mucho mayor. Kaplan menciona que existen trabajos que requieren la utilización de la visión para la identificación de determinados objetos y para la ubicación de esos objetos en un lugar concreto, por ejemplo en la estantería de un almacén (2017: 123). En los aeropuertos los servicios policiales suelen encargarse del reconocimiento facial de los viajeros una vez revisado su

pasaporte, pero en la actualidad ya han sido incorporados sistemas de reconocimiento facial en aeropuertos como el de Ámsterdam-Schiphol, además de EE. UU. o Australia, para realizar la labor que normalmente había desempeñado la policía. Por lo tanto, la IA se encuentra inmersa en diversos campos profesionales de la actualidad, como en el reconocimiento óptico, también utilizado para la clasificación de correos o la digitalización de antiguos documentos con el objetivo preservar el patrimonio.

Los empleos más amenazados por la automatización son aquellos más rutinarios, pero la IA ha extendido su ámbito de actuación de la automatización hasta campos profesionales que antes no se habían visto afectados por ese fenómeno. Todo esto sugiere la necesidad de una profunda reflexión sobre el nuevo alcance de la automatización potenciada por la IA.

Para seguir el hilo conductor de este apartado, es necesario completar lo dicho hasta ahora con unas reflexiones sobre la robótica, con el fin de mostrar su alcance en el ámbito profesional.

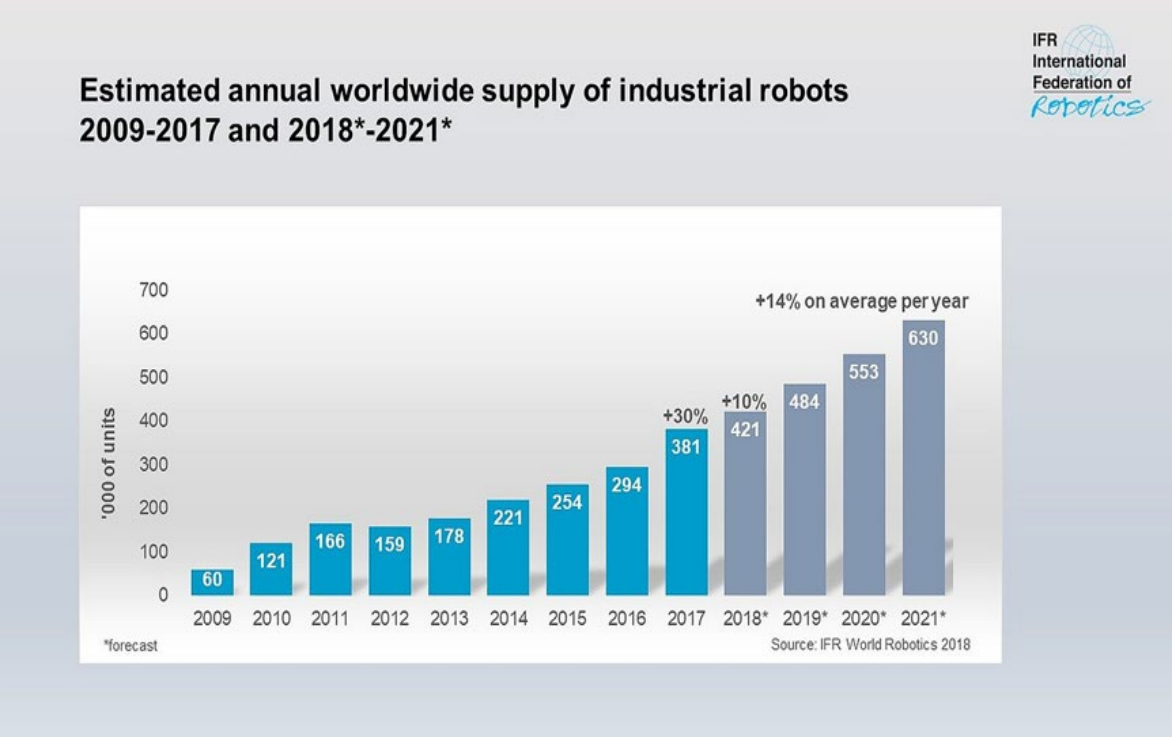
7.3. Los robots se abren camino

La IA se encuentra estrechamente vinculada con la robótica. La robótica tal y como se la conoce hoy en día no podría entenderse sin la IA. Esto no quiere decir que todos los robots sean en sí mismos IA, aunque en muchos casos ya posean IA.

La competencia entre los trabajadores y los «falsos trabajadores», en términos de Kaplan (2016), está servida. Los robots ya funcionan a mucha más velocidad que los seres humanos en muchas tareas profesionales. Un claro ejemplo es el robot construido por *Industrial Perception* que logra mover una caja por segundo, mientras que un humano tarda 6 segundos. Además, el robot no presenta cansancio ni sufre lesiones y, al menos por el momento, no es merecedor de derechos laborales, motivos que suponen un atractivo para numerosas empresas. El robot de *Industrial Perception* supone una clara amenaza para la estabilidad de muchos empleos rutinarios. *Baxter*, creado por *Rethink*, una famosa empresa utilizada por el ejército de los EE. UU. para desactivar bombas en Afganistán e Iraq, es otro ejemplo de robot humanoide, pero más ligero que el robot de *Industrial Perception* y

mucho más ágil a la hora de trabajar. Es importante destacar que la presencia de los robots en la industria no supone completamente ninguna novedad, aunque el nivel de automatización al que se está llegando sí es un fenómeno novedoso.

Un dato muy importante es el ofrecido por la International Federation of Robotics (IFR) (2018) que informa sobre del crecimiento de ventas anual de robots industriales. En el siguiente gráfico puede observarse el aumento de un 114 % en los últimos cinco años (2013-2017), lo que evidencia un importante crecimiento.



Fuente: International Federation of Robotics.

De la misma manera que Bostrom habla de una «explosión de superinteligencia», Martin Ford, en su obra *El auge de los robots*, hace referencia a lo que él denomina «la inminente explosión de la robótica« (2016: 23). Los ejemplos de robots mencionados anteriormente funcionan con el mismo *software*, ROS (en inglés, *Robot Operating System*), un sistema operativo para robots de código abierto. Si a este sistema operativo estándar para la robótica se le unen unos medios de programación que tengan un bajo coste, es muy

probable que se produzca una explosión de robótica, pues se estarían sentando las bases para la estandarización de *hardware* y *software* dedicado a los robots, lo que podría significar una mayor expansión de la producción y por lo tanto una disminución considerable de los costes, algo que supondría un atractivo para muchas empresas.

Uno de los sectores donde mayor impacto ha tenido tradicionalmente la robótica es la industria de manufactura, donde se ha visto incrementado el desempleo de trabajadores en los últimos años, sobre todo en EE. UU., aunque es importante matizar que actualmente en este país el empleo manufacturero representa el 10 % del total. China ha sido el lugar donde más empleos de la manufactura se han destruido, y parece que ese fenómeno irá en aumento, pues en el gigante asiático se diseñan programas de producción muy agresivos y en los que hay que desplegar la producción de forma repentina. Este fenómeno supone un crecimiento de la presión a la que están expuestos los trabajadores chinos, y que puede desembocar en una ola de suicidios como ocurrió en 2010 con *Foxconn*. A diferencia de los seres humanos, los robots no sufren estrés y son susceptibles para el adiestramiento en muchas labores, representando una clara competencia incluso frente a los bajos salarios de los trabajadores chinos. Sin embargo, la tendencia creciente hacia la automatización no solo se está produciendo en China, sino también en otros países asiáticos como Japón, Vietnam e Indonesia, y del mismo modo podrían destruirse miles de puestos de trabajo y dejar sin alternativa a muchas familias que dependen de esos trabajos. Las empresas ubicadas tradicionalmente en países asiáticos lo han hecho por los bajos salarios de los trabajadores, motivo por el que si la automatización va en aumento es muy probable que decidan regresar a sus países de origen, ya que no existiría una razón de peso para la deslocalización al contar con automatización suficiente para la producción. Esto podría provocar un aumento considerable de la pobreza en los países que han sido un foco para la ubicación de muchas empresas beneficiadas por el proceso de deslocalización.

7.4. Perspectivas comparadas: tecnooptismistas y tecnopesimistas

Las reacciones en lo que respecta al futuro de la robótica en el campo profesional son muy variadas. Por un lado están quienes se sitúan en el terreno del optimismo y plantean un argumentario de beneficios gracias al aumento de la automatización; por otro lado tenemos a los tecnopesimistas, entre los que podemos encontrar diferencias, aunque se caracterizan por un común denominador, a saber, la reivindicación de la necesidad de reflexionar filosóficamente sobre las implicaciones del aumento de la automatización en el futuro. Esta diversidad de posiciones se escenificó ostensiblemente en la Cumbre de las Ideas que se celebró en la ciudad de Puebla (México) en el año 2016 y que contó con la presencia de Matt Ridley, Ronald Arkin y James Bessen, del lado de los tecnooptimistas, y con Nick Bostrom, Tyler Cowen y Martin Ford, del lado de los tecnopesimistas.

7.4.1. El tecnooptimismo como una respuesta de confianza desmedida

Existen varias posiciones muy optimistas respecto al crecimiento de la automatización, como indica un informe de abril de 2017 de la Federación Internacional de Robótica sobre el impacto de la robótica en la productividad y el empleo. En ese informe se reflejan los efectos negativos del aumento de la automatización, sobre todo en aquellos empleos que no requieren de una alta cualificación, aunque existen economistas como Richard B. Freeman (2015) que sostiene que tras el aumento de la desigualdad no se encuentra la robótica. En cambio, otros economistas, como James Bessen (2015), afirman que la automatización tendrá un efecto positivo sobre todo para la relocalización de muchos trabajadores. Para Bessen la pérdida de puestos de trabajos es muy probable, pero no de forma masiva, y lo que tendrá lugar es un cambio de modelo de trabajo, ya que se demandarán nuevas habilidades. El problema para él no es la presencia de IA en el mundo del trabajo, sino más bien el desarrollo de las habilidades, pues argumenta que la IA ya ha estado presente en las últimas tres décadas y eso no ha ocasionado un desempleo masivo. Las habilidades podrían suponer un problema para aquellas personas escasamente formadas, pues el reto no es el supuesto desempleo ocasionado por la IA al que muchos temen, sino más bien el diseño de

estrategias para desarrollar habilidades de aprovechamiento de las nuevas tecnologías en aquellos trabajadores que carezcan de ellas. En definitiva, los efectos de la IA en el campo de las profesiones son positivos en términos generales, pues se reducen costes y aumenta la demanda de productos, además de incentivar la nueva formación de habilidades de aprovechamiento de la tecnología.

Otro argumentario similar al de Bessen, y que se encuentra en la línea de los que podrían denominarse «tecnooptimistas», es el que presentó el Banco *Barclays* en el año 2015. En un informe titulado *Robots at the gate: Humans and technology at work* el banco inglés considera que un aumento de la inversión en robótica para potenciar la automatización lograría convertirse en una garantía de futuro para el crecimiento de la economía inglesa y la generación de nuevos puestos de trabajo.

Matt Ridley es un escritor y científico británico que también puede situarse entre los tecnooptimistas. Ridley publicó en el diario *The Times* del Reino Unido un texto titulado *Artificial intelligence is not going to cause mass unemployment*, donde critica a quienes defienden la idea de que la IA puede suponer una amenaza para la estabilidad del empleo. El británico argumenta que, aunque los robots ocupen puestos de trabajo y desplacen a muchos trabajadores, siempre crearán otros puestos de trabajo como consecuencia de la aparición de nuevas exigencias fruto de la innovación tecnológica. El inglés sostiene que la tecnología ha modificado una y otra vez el campo del trabajo a lo largo de la historia fruto de la adaptación, motivo por el que aquellas posturas pesimistas no serían justificadas. Además, afirma que «en 1700, casi todos tuvimos que excavar el suelo desde el amanecer hasta el anochecer o todos murieron de hambre. La tecnología, nos liberó de ese mundo precario y terrible» (2016). En ese sentido, Ridley es un claro defensor del tecnooptimismo, incluso él mismo se hace llamar «optimista racional», y piensa que la tecnología seguirá mejorando las condiciones de vida de los humanos. Además, para él no existen motivos de preocupación en el futuro, pues aunque los robots desplacen a los seres humanos de sus puestos de trabajo, aparecerán nuevos nichos de empleo gracias a la tecnología y en ese momento la ciudadanía y las instituciones deben estar preparadas para plantear soluciones.

7.4.2. Tecnopesimismo como visión trágica

Frente a los tecnooptimistas se sitúan los tecnopesimistas, para quienes el crecimiento de la automatización tendrá graves consecuencias en el empleo de millones de personas, lo que implicaría la ruina para las sociedades. Hay un hecho histórico importante para entender mejor este pesimismo tecnológico. Martin Ford comienza el capítulo 2 de su obra *El auge de los robots* mencionando el discurso que Martin Luther King pronunció la mañana del domingo 31 de marzo de 1968 en el púlpito de la catedral de Washington, donde se menciona una triple revolución: la tecnológica, la armamentística y la que lucha por la defensa de los derechos civiles. Esta «triple revolución» que menciona Luther King hace referencia a un informe que elaboraron y entregaron en 1964 al presidente Johnson un grupo de tecnólogos, académicos y periodistas. Ese informe sirvió como detonante para despertar la preocupación de la ciudadanía por la oleada de automatización surgida tras la Segunda Guerra Mundial. Incluso Norbert Wiener (1985; 1989), un erudito matemático que trabajaba para el MIT, sostuvo que la revolución tecnológica tendría efectos muy negativos para la industria, hasta tal punto que ya no sería necesario contratar a personas.

Ford basa su análisis en el contexto de los EE. UU. y hace un repaso histórico de la economía y el empleo. Sostiene que el desempleo de larga duración es un grave problema que debería generar preocupación, ya que desemboca en un aumento de la desigualdad, además de que la precarización del trabajo va en aumento. Se está refiriendo a la »polarización del mercado laboral», que él define de la siguiente manera:

> La tendencia de la economía a eliminar trabajos de mediana cualificación propios de la clase media y reemplazarlos con una combinación de trabajos en el sector de servicios con sueldos bajos, y trabajos para profesionales muy cualificados que no suelen estar al alcance de la mayoría de los trabajadores (Ford, 2016: 59).

Para Ford la principal causa de la polarización se origina en la automatización de aquel trabajo que es rutinario, aunque también, pero con menor importancia, en la internacionalización de los mercados laborales. En principio, quienes se sitúan en la línea del tecnooptimismo sostienen que la automatización incidirá principalmente sobre aquellos trabajos que son rutinarios, mientras que Ford, por el contrario, sostiene que la automatización también afectará a aquellos trabajos que aparentemente no son rutinarios por el momento. Ford se opone a la estrategia argumentativa de autores como Matt Ridley, que basa sus afirmaciones en la historia económica. En última instancia, la respuesta a la cuestión de si las máquinas inteligentes eclipsarán algún día la capacidad de la gente para llevar a cabo gran parte del trabajo que impone la economía surgirá de la tecnología del futuro y no de los datos que ofrece la historia económica (Ford, 2016: 68).

La tecnología de la información irrumpe sin precedentes, con fuerza y transformando el medio laboral sin vuelta atrás. Piénsese por ejemplo en la incorporación de Internet a los trabajos. Aquellos trabajos llamados «de cuello blanco se encuentran en riesgo según Ford, y eso se debe, entre otras cosas, al *machine learning*. Ese riesgo se basa en el adiestramiento de las máquinas para el desarrollo de tareas que suelen hacer los humanos, analizar documentos en materia de derecho, o también enviar correos. Los macrodatos (*big data*) también tienen que ver con ese cuestionamiento de la estabilidad de los empleos de cuello blanco. Muchos trabajos se están automatizando por medio del desarrollo de *software* que manipula macrodatos y algoritmos inteligentes que condicionan el funcionamiento de las empresas. El sector financiero es uno de los principales espacios donde la IA ha irrumpido con fuerza, por ejemplo, ya están siendo utilizadas aplicaciones para la gestión de préstamos y la detección del fraude como *ZestFinance, Lending Club, LendUp*, entre otros, como señala el Equipo Fintech (2017).

Del lado de los tecnooptimistas se afirma con convencimiento que la automatización destruirá empleo, principalmente en aquellos trabajos que requieren una baja cualificación y que una de las soluciones está en la formación de nuevas habilidades. Sin embargo, como señala Ford, el problema radica en que las máquinas también se están haciendo cargo de aquellas tareas que requieren una mayor cualificación.

La educación superior será susceptible de una transformación como consecuencia del progreso de la IA. Los algoritmos de evaluación son una herramienta que está siendo impulsada para la valoración de ensayos por parte de máquinas, pues tras ellos se esconde una importante tecnología con IA bastante avanzada. No obstante, el sector educativo todavía parece haberse mostrado inmune a este mecanismo de evaluación. Pero las muestras del creciente impacto de la tecnología en ese sector están viniendo de las plataformas llamadas MOOC, o cursos abiertos en línea, como Coursera, EdX, MiríadaX, etc. Son muchos los usuarios inscritos en los cursos que ofrecen esas plataformas. Más allá de ese dato, lo destacable aquí es que los cursos en línea ya están compitiendo con la modalidad presencial, lo que podría provocar una reducción considerable de los puestos de trabajo que giran en torno al campo presencial. Por lo tanto, si los cursos en línea siguen aumentando su demanda, la educación será otro espacio profesional afectado, y no precisamente porque en él se desempeñen labores rutinarias y de baja cualificación.

Otro ámbito en el que también está incidiendo la IA es el médico. La cantidad de información a la que se podría tener acceso es asombrosa para el desarrollo de nuevos diagnósticos y tratamientos. Ya existe experiencia de trabajo con IA en el campo médico desde hace unos años. Un claro ejemplo es *Watson*, un equipo de IBM que trabaja en el Centro Médico de Anderson para el Cáncer de la Universidad de Texas. El sistema *Watson* es un intelecto sintético que no solo se utiliza en el campo médico, sino también en el de las finanzas. Normalmente se conoce por haber participado en el programa de televisión estadounidense *Jeopardy!* Los expertos consideran que una de las principales ventajas más importantes de la IA en el campo médico es la gran probabilidad que existe para prevenir y paliar ciertos errores que en ocasiones pueden llegar a ser fatales. La IA puede desempeñar una función de asesoramiento, pues la cantidad de legislación en materia de medicina en ocasiones provoca que el personal desempeñe sus funciones a la defensiva y una segunda opinión no estaría de más. En las próximas dos décadas se necesitarán médicos entrenados para utilizar IA (Ford, 2016: 146-147).

Pero la IA en el ámbito sanitario no ha irrumpido únicamente en el terreno médico, sino también en el farmacéutico. Empresas como *GlaxoSmithKline* (GSK), *Merck & Co*, *Johnson & Johnson* y *Sanofi*, representan algunos de los casos empresariales que han decidido acelerar la investigación de nuevos fármacos y tratamientos con IA. *Exscientia* es la empresa que se dedica a suministrar IA al sector farmacéutico. Además, otro sector en el que la IA podría tener un fuerte impacto es en el de los cuidados de personas mayores. La población de los países más desarrollados está envejeciendo a pasos agigantados. Una noticia del diario *Expansión*, con fuente en el Instituto Nacional de Estadística (INE), da a conocer que en España hay 118 mayores por cada 100 menores de 16 años. Además, el INE señala que para 2050 el porcentaje de personas mayores superará el 30 % del total de la población. El creciente número de población mayor está dando lugar a que empresas dedicadas a la robótica encuentren un buen nicho de mercado en el sector de los cuidados para desarrollar nuevas tecnologías. La empresa francesa *Alderaban Robotics* está dedicándose al diseño de robots para el cuidado de personas mayores. Un ejemplo es *Pepper*, un sistema utilizado desde el año 2016 en geriátricos.

Ford retrata un panorama bastante pesimista, pues considera que la pirámide laboral, en lo relativo a la capacitación de los trabajadores, irá siendo ocupada progresivamente por la automatización en la mayor parte, y por la IA en aquellas áreas que requieran una mayor cualificación. Ante este escenario una solución viable no pasa por seguir otorgando títulos superiores, pues eso no implica un aumento de la empleabilidad. Frente a tal situación, Ford propone una renta básica garantizada para hacer frente a un futuro sin empleo.

Vivek Wadhwa, tecnólogo de origen indio residente en los EE. UU., considera que el aumento del uso de la tecnología es exponencial, y que por tanto la IA estará cada vez más presente en la vida. Para Wadhwa la IA supone una clara amenaza para el empleo, y el caso de los EE. UU. es una evidente muestra de ello, como señala en un artículo publicado en el *Washington Post* donde critica la inacción del presidente Donald Trump en el sector de los transportistas, que es uno de los más vulnerables. Según este tecnólogo, además de las propuestas de renta básica y del aumento de formación en habilidades de programación, es necesario ir más allá para que el progreso de la tecnología no se convierta en una distopía.

Hay otros autores como Tyler Cowen (2013) que también se sitúan en la senda del tecnopesimismo. En su obra *Se acabó la clase media* reflexiona sobre los efectos negativos de la IA en el futuro, que originarán, según él, una fuerte desigualdad social a partir de dos estratos bien diferenciados, una élite altamente cualificada y una clase empobrecida al ver sus salarios reducirse considerablemente y su puesto de trabajo ocupado por un intelecto sintético. No todos los tecnopesimistas se sitúan en la crítica del aspecto profesional, sino también del vital en todas sus dimensiones.

Según Nicholas Carr (2014), existe una gran dependencia de las máquinas desde hace un tiempo, lo que provoca que el carácter ontológico adquiera su forma en función de esa dependencia. Esta dependencia de las máquinas se viene forjando desde hace tiempo y para entenderla es importante considerar algunos temas. La utilización de los intelectos artificiales es creciente, pues cada vez son más las profesiones sometidas a la informatización y a la automatización. Por ejemplo, el sentido de la orientación se está mediatizando por medio del GPS, lo que implica que en la actualidad exista una dependencia cada vez mayor del intelecto artificial que dirige el GPS. Lo mismo ocurre con las relaciones sociales, pues numerosos estudios en psicología y sociología están analizando el fenómeno de la virtualidad de las relaciones interpersonales a través de plataformas como *WhatsApp, Facebook o Twitter*. Así pues, numerosos sistemas de IA se están convirtiendo en una herramienta diaria de dependencia.

Lo que está claro es que más allá de las dos posiciones estrictamente diferenciadas, a saber, tecnooptimismo y tecnopesismismo, la introducción de los sistemas artificiales en el campo de las profesiones es un hecho evidente y así lo demuestra un artículo publicado por el turco Daron Acemoglu y el colombiano Pascual Restrepo (2017) en *Vox Eu* para los EE. UU. Su trabajo analiza la evidencia con la que los robots industriales redujeron el empleo y los salarios entre 1990 y 2007. Las estimaciones sugieren que un robot extra por cada 1.000 trabajadores reduce la relación empleo en la población en 0.18-0.34 puntos porcentuales y los salarios en 0.25-0.5 %. Además, un estudio del Oxford Economics (2019) señala que, solo en Europa, la automatización en el sector de la manufactura ha ocasionado la destrucción de 400.000 puestos de trabajo entre los años 2000 y 2016, una situación que,

alertan, aumentará en las próximas décadas, al igual que en EE. UU., China o Corea del Sur. Sin embargo, la postura que debería adoptarse no es recomendable que sea la del alarmismo catastrofista, sino más bien la de una crítica responsable desde la que enfrentar este desafío para sacar el máximo provecho destinado a la sociedad en su conjunto. Parece ser que no son tantos los puestos de trabajo que se destruirán en el futuro, a diferencia de lo que sostiene Tyler Cowen, aunque es interesante tomar como referencia aquellas posturas que son más pesimistas, ya que pueden servir como un punto de referencia para la reflexión. La clave está en situarse en el terreno de lo crítico, más allá del optimismo y el pesimismo exacerbados, asumiendo el desafío que depara el futuro desde un marco de responsabilidad tecnológica.

7.5. Los impactos políticos y sociales

El fin del trabajo provocado por el aumento exponencial de la automatización representará un importante desafío en términos políticos y económicos para muchas sociedades del planeta. Los sistemas políticos y económicos parecen no haber comenzado a pensar en alternativas frente a este fenómeno en masa que se dará en las próximas décadas. Este fenómeno ya fue alertado por pensadores como John Maynard Keynes (2010) en 1930 o Wassily Leontief (1986), quienes destacaron las graves consecuencias que tendría el desempleo tecnológico en las sociedades del siglo XX. No obstante, sus temores vinculados con el desempleo tecnológico parecen no haberse materializado. Más allá de los pronósticos de estos dos economistas clásicos, existen importantes estudios que alertan sobre las importantes implicaciones sociales que tendrá la automatización. Entre esos informes destaca el elaborado por Carl Benedikt Frey y Michael A. Osborne (2013) que examina el grado de susceptibilidad que ciertos empleos tienen para verse afectados por la disrupción tecnológica. En este informe se calcula que la estabilidad de alrededor del 47 % del empleo de los EE. UU. se encuentra en riesgo por la automatización y que los salarios, así como el sector educativo, se verán negativamente afectados. A diferencia del informe de Frey y Osborne, existen otros trabajos que apuntan a la baja en el impacto de la automatización. Por ejemplo, el informe titulado *The Risk of Automation for Jobs in OECD*

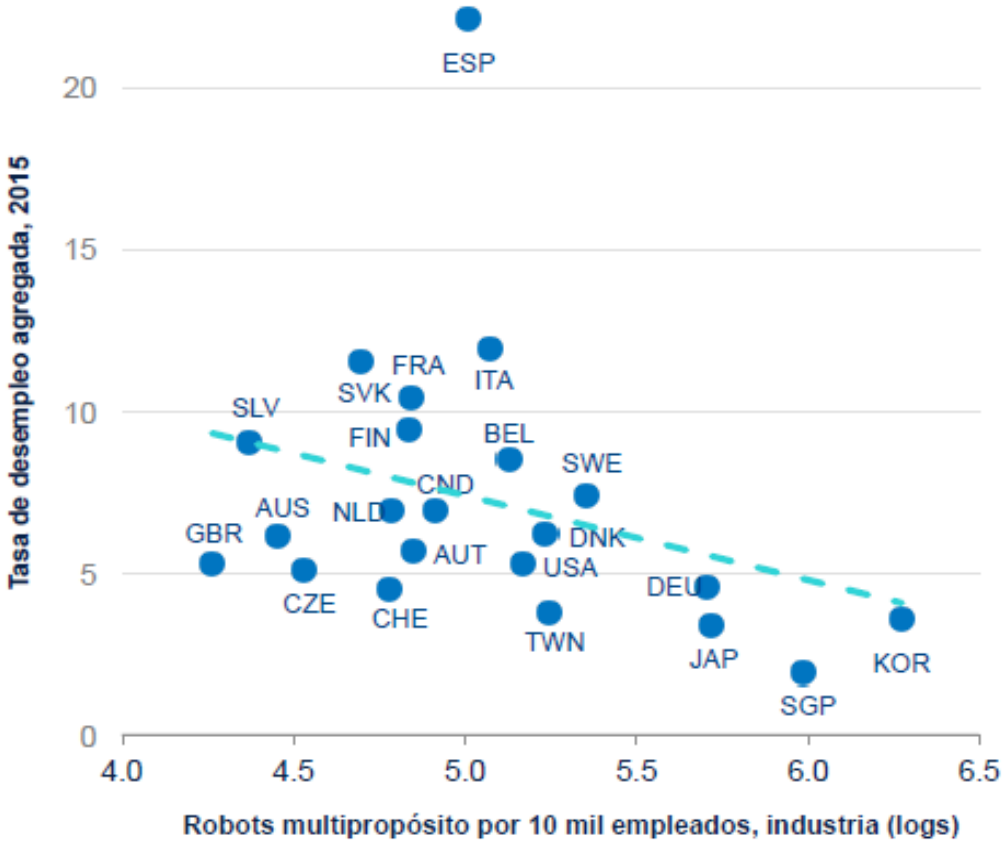
Countries: A Comparative Analysis, que toma como muestra 21 Estados miembros de la OCDE, refleja que la media de afectación de la automatización rondaría el 9 %. Además, a pesar de las estimaciones que se hagan del impacto de la automatización en el campo profesional, deben tenerse en cuenta dos importantes factores que ponen en evidencia que el aumento del desempleo no se debe necesariamente a la tecnología. En primer lugar, la incorporación de herramientas tecnológicas no es un proceso directo y carente de obstáculos, ya que existen barreras de diversa índole, económicas, sociales o jurídicas, que pueden dificultar su integración. En segundo lugar, las tecnologías no solo pueden sustituir los trabajos existentes, sino que también pueden ayudar a promover la creación de nuevos focos de empleo, que pueden impulsarse desde el debate y la discusión entre los grupos de interés en el seno de laboratorios abiertos sobre ciencia cívica. En ese sentido, son variadas las reacciones que se presentan ante la automatización y se pone de relieve una cuestión importante, a saber, que tendrán un fuerte impacto en las próximas décadas, afectando así a las instituciones que dan forma a las sociedades contemporáneas.

En lo que respecta a la creación de nuevos empleos fruto de la integración de la tecnología en el ámbito profesional, se suceden opiniones dispares. Acemoglu y Restrepo (2016), aseguran que el aumento adicional en determinadas profesiones con nuevos títulos de trabajo ha sido posible gracias al desempeño de tareas que han presentado una novedad frente a las más tradicionales. En cambio, están también quienes afirman que la tecnología no ha tenido un impacto tan positivo en la actualidad como en el pasado, refiriéndose al ferrocarril o el automóvil, entre quienes se encuentran Thor Berger y Carl Benedikt Frey (2015) y Jeffrey Lin (2011). Existen por lo tanto dos fenómenos que tienen un profundo impacto: por un lado, la automatización de los empleos existentes y, por otro lado, la creación de nuevos empleos fruto de la aparición de nuevas tareas tras el desarrollo tecnológico. En ese sentido, el resultado sobre el mercado laboral y la estabilidad de los empleos dependerá de que la tecnología cree más puestos de trabajo de los que destruya la automatización. En las últimas décadas ha existido un equilibro entre estos dos fenómenos, ya que en realidad hay una dependencia entre ambos. Acemoglu y Restrepo (2016) sostienen que la búsqueda de un equilibrio entre la automatización y la innovación en

tecnologías que promuevan nuevas profesiones es algo fundamental, pues en caso de que los esfuerzos únicamente se concentren en la automatización y no reconozcan la importancia de la innovación surgiría un grave problema económico, político y social.

Luciano Floridi (2014; 2017) ha reflexionado sobre las serias implicaciones que tendrá la entrada de IA a gran escala en los ámbitos profesional y económico. El italiano muestra la necesidad de que las instituciones políticas asuman responsabilidad y diseñen estrategias desde las que poder afrontar el fenómeno de la automatización y para ello pone como ejemplo la reciente aprobación por parte del Parlamento Europeo de una directriz que camina en la dirección de crear un marco ético y legal que refleje este nuevo escenario de robots autónomos. La propuesta de responsabilidad de Floridi adopta una forma estrictamente política, ya que se encuentra dirigida al establecimiento de regulaciones que enfrenten de una forma más realista y anticipada los adelantos en el campo de la robótica. Además, la asunción de responsabilidad también tendría que verse materializada en políticas fiscales por medio de tasas, y aunque el filósofo italiano no lo especifica claramente, también se establecerían impuestos a la actividad robótica. Pero eso no es todo, Floridi también destaca el problema en la gestión del tiempo libre que provocará el desempleo tecnológico e invita a repensar cuál será el papel ser humano en su tiempo de ocio. Como puede comprobarse no solo centra su preocupación en la cuestión económica, sino también en la existencial.

La Fundación BBVA publicó un documento con los resultados de una investigación bajo el título *El impacto del cambio tecnológico y el futuro del empleo,* donde, a través de una tabla, presenta el impacto de la robótica en los 22 países más especializados:



Fuente: Doménech, García, Montañez y Neut, 2017: 21.

En su obra *La Cuarta Revolución Industrial*, Klaus Schwab indica cómo este acontecimiento histórico se caracteriza por la convergencia de tecnologías digitales, físicas y biológicas que transformarán el mundo tal y como se lo conoce. La idea de la convergencia es fundamental para comprender la magnitud de la Cuarta Revolución Industrial, ya que en ella se da el encuentro de diversos desarrollos tecnológicos que promueven la transición hacia nuevos sistemas. Uno de los focos de interés de esta revolución es el campo productivo de las profesiones. Todo parece indicar, como señala David Ritter (2016) en un artículo publicado en *The Guardian*, que el futuro del empleo dependerá de trabajos que no existen, de nuevas industrias que usarán tecnologías fruto de la innovación. No obstante, para no parecer exageradamente optimista, es importante que la innovación tecnológica que sirva como detonante de la aparición de nuevos empleos vaya

acompañada de medidas políticas y económicas que sepan amortiguar los efectos de la automatización y pensar en dinámicas económicas y educativas que sean fruto de reformas. Estas reformas deberían perseguir la adaptación responsable a los contextos concretos para caracterizarse por una verdadera utilidad cívica para las sociedades.

7.6. La inteligencia artificial responsable y su incorporación en el campo profesional

El futuro presenta grandes incertidumbres y eso puede conducir a la adopción de diversas estrategias. La incertidumbre suscita la generación de inseguridad, miedo, o también excesiva confianza frente a un fenómeno que puede provocar un alejamiento de cualquier actitud reflexiva. Es por ello que se torna necesario enfrentar el desafío de la irrupción de la IA en el campo profesional con responsabilidad. El marco teórico de una IAR nos proporciona las bases desde las que estudiar la irrupción de la IA en el ámbito del trabajo y su impacto en la empleabilidad.

Las pretensiones de un progreso irreflexivo y ciego, despojado de valores, pueden conducir hacia catastróficas consecuencias. La fe ciega en el progreso de la IA sobre el campo profesional provocaría consecuencias muy importantes en el ámbito político y económico de las sociedades. En ese sentido, es fundamental proporcionar unos criterios abiertos desde la ética que entren en diálogo y se construyan en base a un MIAR que promueva la participación de los sectores implicados según el modelo de la quíntuple hélice para afrontar la innovación y los desafíos sociales y cívicos, y considerar el lugar del ser humano en el campo profesional, para definir sus límites y complejidades, y también para ver los diferentes efectos que puede tener la irrupción de la IA a gran escala. Es importante reflexionar sobre la articulación entre el ser humano y sus producciones tecnológicas para no despojar a esa actividad de una reflexividad moral. La lógica del progreso tecnológico provoca un constante dinamismo orientado hacia el futuro bajo la idea de un progreso que debe ser cuestionado desde la óptica de la responsabilidad anticipada.

La búsqueda de la armonía entre progreso de la IA y bienestar humano debería ser el punto de partida de cualquier investigación. El desarrollo de la IA en el ámbito profesional no solo debe encontrar su motivación en aspectos propios de la lógica desarrollista de la tecnología ni tampoco exclusivamente por el ahorro en el capítulo de personal, sino también por la búsqueda de una compatibilidad y equilibrio entre innovación y bienestar humano. Los tecnooptimistas heredan su actitud de la Ilustración y le dan forma a la luz del progreso de la IA. Ese optimismo tecnológico exacerbado propio de la Ilustración no ha tenido todas las consecuencias positivas que se esperaban para la humanidad, como señala Andrew Feenberg (1991) en su crítica a la preeminencia de la administración tecnocrática.

En este punto es crucial someter el fenómeno del impacto de la IA en el campo profesional a un debate público, con el fin de que los sectores implicados discutan este asunto desde la perspectiva de una ciencia cívica. La ciudadanía se verá afectada por este fenómeno y por lo tanto deben valorarse los efectos que pueden existir para la misma. El espacio de discusión que nos ofrecen los laboratorios abiertos representa una clara oportunidad para deliberar sobre estos asuntos en profundidad, pues las valoraciones no solo deben surgir en el seno de las academias o de las empresas de turno. La incertidumbre del futuro requiere agudizar el sentido de responsabilidad, al igual que ciertos animales agudizan sus sentidos cuando sienten acechar un peligro. Así pues, la asunción de responsabilidad en el seno de laboratorios abiertos sobre ciencia cívica podría contribuir positivamente para que no ocurra como a Rose en *La corrosión del carácter* de Sennett, para quien el riesgo representó de entrada un fracaso y provocó en ella una desmovilización, no hacer nada parece una actitud pasiva y no prudente» (Sennett, 2000: 94).

La magnitud de las investigaciones en el campo de la IA propiciará un cambio de paradigma en la concepción del trabajo, lo que representa la necesidad de plantear una nueva idea de responsabilidad ética de carácter tecnológico para promover el respeto por los derechos humanos y el interés por los ODS. La estabilidad del empleo para millones de personas es muy frágil, su situación económica podría caracterizarse por la extrema vulnerabilidad si no se toman importantes medidas con anticipación. Así pues, ante tal

panorama es importante cultivar el sentimiento de responsabilidad en aquellos espacios e instituciones decisivas. En ese sentido, la ética orientada al futuro de la que hablaba Jonas (1995: 63-70), invita a desarrollar una ética del cuidado y la preocupación ante lo que aún está por venir.

En lo referente a la formación, los responsables de la actividad tecnológica, tecnólogos de los sectores público y privado, políticos, representantes laborales y representantes de la sociedad civil, deberían formarse para afrontar los diversos impactos que subyacen en la incorporación de los intelectos sintéticos al ámbito profesional. Esa diversidad de impactos provoca el surgimiento de una serie de problemáticas jurídicas, sociales, económicas, políticas, etc. Este proceso formativo, que debería profundizar en el desarrollo de habilidades comunicativas y democratizadoras, facilita las herramientas prácticas y teóricas desde las que estudiar los posibles impactos. Además, la perspectiva ética es un aspecto ineludible dentro del proceso formativo, ya que permite introducir elementos de reflexividad para reconocer posibles impactos y otras variables que no son consideradas debido a la priorización de los criterios estrictamente técnicos. Así pues, educar en una ciencia cívica es fundamental para la construcción de una IAR (Dignum, 2017b: 4-5).

Por otro lado, y como ya se ha indicado, Dignum destaca la importancia de generar espacios políticos de discusión de problemáticas mediante el encuentro de la pluralidad de perspectivas que caracterizan a los grupos de interés (*stakeholders*) implicados en esta actividad. En ese sentido, el desarrollo de laboratorios cívicos sobre ciencia cívica brindarían la oportunidad de poder someter a discusión cuestiones que por motivos económicos en ocasiones no se abordan.

El modelo de innovación tradicional, caracterizado por el cierre y el aislamiento, no consigue afrontar el tratamiento de problemáticas desde la pluralidad de perspectivas que ofrecen los grupos de interés (González Esteban, 2007), ya que asume una posición dogmática e incuestionable. En cambio, los laboratorios abiertos representan un espacio en el que se tienen en cuenta un conjunto de factores que determinan las acciones tecnológicas inevitablemente y que no puede ser evitados o relegados a un segundo plano. El escenario

de los impactos de los sistemas artificiales en el campo profesional es muy complejo y por ello deben promoverse modelos de innovación abierta en las diversas esferas de acción política para elaborar las agendas que posteriormente se implementarán. La fórmula de la quíntuple hélice implica un nexo de unión entre los diferentes grupos de interés y genera una dinámica de trabajo colaborativo con retos comunes que afrontar, donde el tratamiento de las problemáticas complejas es emprendido desde una asunción de responsabilidad. Este modelo es capaz de reconocer que su despliegue afronta la complejidad desde otra complejidad, la de una quíntuple hélice que con un origen en la pluralidad entiende que la innovación solo es posible contando con la participación de todos los grupos de interés, y en especial la ciudadanía, en la introducción de IA en el ámbito de las profesiones.

El MIAR se construye sobre una base pragmática que establece un fuerte vínculo entre lo teórico y lo práctico. Este modelo proporciona un espacio de diálogo entre los sectores implicados. De ese modo se parte de un terreno deliberativo común desde el que asumir las problemáticas que surgen al introducir los sistemas artificiales en el campo de las profesiones. Además, este trabajo deliberativo, que supone la incorporación de criterios de responsabilidad dentro del marco de la IA, permite observar los fenómenos desde una perspectiva cívica y buscar siempre un beneficio social desde diversas ópticas que puede ser sociales, políticas, económicas, jurídicas, educativas, culturales, etc. Es posible que las problemáticas que se deriven de la automatización y de la aparición de nuevos empleos demanden nuevos mecanismos de corte político o económico que proporcionen alternativas viables a la sociedad. En ese sentido, el MIAR busca un tratamiento participativo de los problemas. La producción de nuevos conocimientos que ayuden a enfrentar dichos problemas tendrá un carácter pragmático y justificado en la realidad que se origina en la quíntuple hélice. El trabajo colectivo y participativo de los grupos de interés está introduciendo una lógica novedosa para el tratamiento de problemas al reconocer que la responsabilidad debe ir de la mano de la innovación y que ésta no puede entenderse sin contar con la premisa fundamental de la cooperación.

La incorporación de criterios de responsabilidad, fruto del reconocimiento de la posibilidad de plantear una IAR, implica considerar el conocimiento científico que promueve la investigación tecnológica como un recurso público. Este reconocimiento permite replantear la función social de la tecnología y orientarla hacia el cultivo de las habilidades cívicas y las prácticas democráticas. La integración de la IAR en el desarrollo de sistemas artificiales en el contexto de las profesiones debe servir para la búsqueda de un aprovechamiento beneficioso de la tecnología para la ciudadanía. Frente a un tecnopesimismo exacerbado al estilo de los luditas de comienzos del siglo XIX, es importante sentar las bases por medio del MIAR para la producción de nuevos conocimientos que promuevan un beneficio cívico. Existen otras respuestas frente a los avances tecnológicos que no necesariamente debe ser la de una renta básica universal que mercantilice la vida de la ciudadanía mediante una exposición radical al libre mercado (Hayek, 2014).

7.6.1. Inteligencia artificial responsable en la práctica

Ante la pérdida de confianza de la ciudadanía en las instituciones políticas y en la democracia es importante resaltar la necesidad de potenciar aquellos mecanismos políticos que sirvan de algún modo para el fortalecimiento de los sistemas políticos democráticos y las habilidades cívicas. En el contexto de las profesiones la incorporación de una IAR debe tomar en cuenta precisamente esa necesidad, a saber, la de servirse del fortalecimiento de los valores democráticos para poder emprender sus proyectos con un impacto positivo y en beneficio para la sociedad en su conjunto. Las exigencias que imponen las metas establecidas por los ODS impulsan a la IAR para llevar a cabo una adaptación de todos sus proyectos mediante el reconocimiento de la legitimidad de instituciones como la ONU y la asunción de un compromiso con la humanidad. No sería razonable que el desarrollo tecnológico se situase al margen de las exigencias de una institución de alto reconocimiento internacional como es la ONU y que camine por un horizonte alejado del compromiso que impone este tiempo con el ser humano. Así pues, a continuación se van a detallar algunos

asuntos sobre los que podría resultar muy beneficiosa la deliberación en el marco de un laboratorio abierto sobre IAR en el contexto de las profesiones:

- Mejora de las profesiones médicas gracias a la IA

El empleo de las nuevas tecnologías que incorporen IA puede favorecer el estudio de determinadas enfermedades y contribuir a un perfeccionamiento del ejercicio profesional sanitario. En ese sentido, los sistemas artificiales permitirían dibujar un nuevo horizonte en las investigaciones científicas para la búsqueda del bienestar humano y la salud.

- Generación de nuevos espacios educativos que favorezcan el intercambio de conocimiento de la actividad docente.

Las tecnologías aplicadas a la educación y la información (TIC) favorecen la gestación de nuevos espacios para el proceso de enseñanza-aprendizaje donde impulsar nuevas metodologías. Además, el diálogo entre academia, empresas, organizaciones de la sociedad civil y administración pública contribuye a un intercambio y democratización del conocimiento gracias a su amplia accesibilidad.

- Nuevos nichos de empleo destinados al estudio del cambio climático y la protección de los ecosistemas.

Anteriormente se mencionó la necesidad de orientar el MIAR hacia la generación de conocimientos innovadores que asumieran el compromiso con los desafíos que impone el cuidado del medio ambiente. En ese sentido, la IA potenciará un mejor y mayor conocimiento sobre los entornos ambientales y los ecosistemas mediante el acceso a una gran cantidad de datos que en otro tiempo no eran accesibles.

- Nuevos empleos que mejoren la convivencia en las ciudades mediante los proyectos de ciudades inteligentes.

La decreciente legitimidad en las instituciones y las necesidades de nuevas formas de organización social y planificación de las políticas públicas demandan nuevos mecanismos para la actividad política. Tomando en cuenta esa demanda, la IA brinda interesantes

herramientas desde las que generar innovadoras dinámicas sociales y políticas que fortalezcan las habilidades cívicas y democráticas de las sociedades actuales.

- Mejorar aquellas profesiones que tengan una excesiva carga burocrática y gubernamental.

El crecimiento del poder y las competencias de los Estados ha supuesto un aumento de los procesos burocráticos y administrativos. En ocasiones, eso dificulta una agilización en los procesos y por lo tanto un empobrecimiento de las relaciones entre la ciudadanía y las instituciones, que como ya se ha señalado, cada vez cuentan con menor legitimidad. Así pues, los sistemas artificiales conducen a un camino de optimización de esos procesos y por lo tanto en un cultivo de las habilidades cívicas y democráticas en el seno de las sociedades.

Estos son algunos de los ejemplos en los que podría aplicarse la IAR en el espacio profesional por medio de un compromiso con los siguientes ODS:

- Salud y bienestar (3).
- Educación de calidad (4).
- Energía asequible y no contaminante (7).
- Trabajo decente y crecimiento económico (7).
- Industria, innovación e infraestructura (9).
- Ciudades y comunidades sostenibles (11).
- Acción por el clima (13).
- Vida submarina (14).
- Vida de ecosistemas terrestres (15).
- Paz, justicia e instituciones sólidas (16).

Es importante destacar que bajo ningún concepto la incorporación de IA en el contexto de las profesiones debe significar un desplazamiento o eliminación del componente humano. El componente humano es sometido a un proceso de cultivo y fortalecimiento mediante la introducción de criterios de responsabilidad en el ámbito de los intelectos sintéticos. En ese sentido, fortalecer una IAR supone una aportación enriquecedora y fructífera para las profesiones en términos prácticos.

7.6.2. Sociedades inclusivas, innovadoras y reflexivas

El título de este apartado hace referencia al reto social del Horizonte 2020 (H2020) (Comisión Europea, 2014), el Programa Marco de Investigación e Innovación de la Unión Europea. La Comisión Europea ha adquirido un compromiso para fortalecer sus sociedades a través del cultivo de la inclusión, la innovación y la reflexión en un contexto de constantes transformaciones en el marco de la globalización. Europa, al igual que otros territorios del planeta, se enfrenta a importantes retos económicos y sociales que afectan considerablemente al futuro de sus sociedades. Estos retos tienen que ver con la pobreza, la desigualdad, la migración, la brecha digital o el empobrecimiento de la democracia, que repercute en una creciente desconfianza de la ciudadanía en las instituciones, etc. Estos retos de gran impacto requieren de un importante compromiso por parte del conocimiento científico compartido que ofrecen, no solo las ciencias naturales y exactas, sino también las sociales y humanas. En ese sentido, se trata de cultivar y fortalecer un humanismo tecnológico para que medidas innovadoras no tengan efectos debilitadores para la cohesión social, como ha sido el caso de la brecha digital. La búsqueda de mecanismos para la compatibilidad y el equilibrio entre la innovación y el bienestar social, económico y político, significa un aspecto clave en el marco del H2020.

Debido a que la ciencia cívica en el contexto de la IAR parte de la consideración del conocimiento científico como un recurso público al servicio del cultivo y fortalecimiento de las habilidades cívicas y democráticas, el compromiso con la inclusión, la innovación, la reflexividad y la seguridad, se torna un aspecto fundamental para su ejercicio. La IAR se sitúa en el contexto de la compatibilidad y el equilibrio que persigue el H2020, y por lo

tanto, aquellas acciones emprendidas para la introducción de intelectos sintéticos en el campo de las profesiones tendrían que adquirir un compromiso equilibrado entre la inclusión, la innovación, la reflexividad y la seguridad. En ese sentido, la IAR puede contribuir a un crecimiento económico por medio de un aumento de la productividad y la eficiencia a una velocidad nunca antes vista. Lasse Rouhiainen menciona en su obra *Inteligencia Artificial. 101 cosas que debes saber hoy sobre nuestro futuro*, la ley de las capacidades exponenciales de Thomas Frey (2017), para quien un incremento de la automatización provocaría a su vez una disminución exponencial del esfuerzo, lo que a su vez generaría un aumento considerable del número de actividades que podrán ejecutarse. Es decir, que cuanto más esfuerzo ahorre la automatización, mayores serán las nuevas actividades que pueden desarrollarse. Según la ley de Frey, la automatización conduce a un nuevo escenario de posibilidades y esto podría ser aprovechado para cultivar un humanismo tecnológico que persiguiera el beneficio de la ciudadanía en su conjunto.

Con la automatización también existe cierto peligro, pues podría crearse una enorme brecha de conocimiento que debería abordarse con profundidad. Precisamente una IAR impulsa proyectos que reconocen la posibilidad de esa brecha y que de ese modo se promueva la búsqueda de un equilibrio a la hora de incorporar los intelectos sintéticos en las profesiones. La IA aumentará la productividad entre un 30 % y un 60 % en muchos países como Finlandia, Suecia o EE. UU. (Rouhiainen, 2018: 162). La automatización de las actividades repetitivas permitirá la creación de innovadoras oportunidades para que los trabajadores humanos puedan dedicar su tiempo a otras diligencias. *Forrester Research*, una empresa independiente de investigación de mercado que centra su interés en el impacto de las tecnologías en la sociedad, estima que en los próximos diez años se crearán en los EE. UU. 15 millones nuevos empleos como resultado directo de la introducción de la IA en el ámbito profesional (Passy, 2017).

Este nuevo escenario, en el que la IAR puede jugar un papel muy importante, debería estar marcado por un imperativo vinculado al cultivo de habilidades que representen nuevos focos de empleo y que a la vez mejoren las condiciones de vida de la ciudadanía. Recordemos que para Bessen (2015) la IA también provocará la demanda de nuevas

habilidades. Un claro ejemplo de las nuevas oportunidades que se han creado en los últimos años y que han demandado nuevas habilidades, han sido aplicaciones como *Glovo* y *Uber Eats* que han permitido crear puestos de trabajo como fruto de recientes exigencias tecnológicas.

En el contexto de las nuevas habilidades requeridas, Rouhiainen (2016) presenta el siguiente listado:

Habilidades personales para el futuro:

1. Conciencia de uno mismo y autoevaluación.
2. Inteligencia emocional.
3. Inteligencia social.
4. Inteligencia interpersonal.
5. Empatía y escucha activa.
6. Flexibilidad cultural.
7. Perseverancia y entusiasmo.
8. Enfoque en el bien común.
9. *Mindfulness* y meditación.
10. Entrenamiento físico.
11. Contar historias

Habilidades empresariales para el futuro:

12. Resolución de problemas.
13. Creatividad.
14. Adaptabilidad a las nuevas tecnologías.
15. Mentalidad empresarial.

16. Ventas y *marketing*.

17. Análisis de datos.

18. Habilidades para hace presentaciones.

19. Inteligencia ambiental.

20. Pensamiento a gran escala.

21. Contabilidad y gestión del dinero.

22. Habilidad de desconectarse.

23. Detectar tendencias.

24. Pensamiento y mentalidad de diseño.

Además, Rouhiainen extrae otras habilidades de la lectura de *The Next Era of Human/Machine Partnerships*:

- Habilidades técnicas relacionadas con la IA y la *blockhain*.
- Habilidad de inteligencia social.
- Mentalidad creativa.
- Aprendizaje a lo largo de la vida.

Una IAR comprometida con el equilibrio entre la inclusión, la innovación, la reflexividad y la seguridad debe orientar su preocupación al cultivo formativo de esas habilidades a través de los sistemas educativos. ¿Se están formando personas en las habilidades requeridas para el futuro tecnológico? La educación ejerce un papel fundamental en el despliegue de la IAR en los próximos años, pues es importante recordar que la incorporación de responsabilidad al ámbito tecnológico que proponía Jonas (1995) se caracterizaba por el ejercicio anticipativo y la preocupación por el futuro. En democracia la educación asume una función fundamental en la formación cívica mediante el cultivo de las habilidades necesarias para poder vivir en paz y en comunidad, educando a una ciudadanía

libre, crítica y responsable. Por esa razón la IAR reconoce la importancia de las habilidades en este nuevo contexto y por ello impulsa una potenciación de la enseñanza de habilidades desde las etapas iniciales del crecimiento para que los futuros ciudadanos y ciudadanas estén en plena capacidad de hacer una buena gestión de la tecnología para el cumplimiento de los propósitos de bienestar de la humanidad. La propuesta formulada en este trabajo, destinada a la asunción de responsabilidad frente al desafío del campo de las profesiones, consiste en plantear la necesidad de enriquecer la educación cívica mediante el cultivo de nuevas habilidades para que se vaya adquiriendo un compromiso cada vez mayor con los derechos humanos, los ODS y los límites planetarios.

A medida que la IA aumenta su poder de influencia sobre las profesiones y la vida humana en general, es fundamental el diseño de proyectos con celeridad y contundencia en materia educativa, para formar a personas en aquellas habilidades orientadas a enfrentar el futuro desde la responsabilidad. Este nuevo rumbo educativo supone una revolución, ya que no se centraría en la formación de personas en trabajos que posteriormente realizarán intelectos sintéticos, pues de nada serviría esa inversión de tiempo y esfuerzo.

En materia educativa la tarea de los gobiernos en las próximas décadas debe centrarse en la generación de conocimientos innovadores que sepan enfrentar los retos surgidos a raíz de la automatización de las profesiones. En ese sentido, el cultivo de nuevas habilidades es una exigencia ineludible para la innovación. Las instituciones políticas tienen la responsabilidad de impulsar la incorporación y ampliación de los avances tecnológicos en la órbita educativa (Parra y Arenas-Dolz, 2015: 336). La integración de la IAR en las esferas de la vida humana se corresponde con el reconocimiento del valor de una educación novedosa en el contexto de las habilidades para la generación de dinámicas que den respuesta a las problemáticas que enfrenta la humanidad. Así pues, la propuesta de una IAR pasa por el fortalecimiento de los sistemas educativos para saber hacer frente a los retos tecnológicos que depara el futuro. En esta línea, y aplicados al contexto de la educación superior, aunque también podrían ser extrapolables a la educación primaria y secundaria, es oportuno subrayar algunas de las contribuciones de la tecnología desde la perspectiva de un humanismo tecnológico:

1. Cambian su mentalidad prevalente, la que promueve el mantenimiento del statu quo, por la del cambio, a fin de que la universidad asuma formas experimentales de renovación para continuar cumpliendo con su misión en el nuevo marco científico y social.

2. Reemplazan el ecosistema del proceso de enseñanza-aprendizaje, al igual que las metodologías educativas tradicionales, por ejemplo, el aula cerrada y la enseñanza vertical, pues no se trata de seguir utilizando las mismas a través de los nuevos medios.

3. El estudiante deja de ser pasivo y tiene ahora un papel activo y es parte dinámica del proceso de enseñanza-aprendizaje junto con el profesor y sus compañeros, con quienes trabaja en red de forma colaborativa.

4. El conocimiento ya no es propiedad exclusiva del docente, sino que está abierto al acceso libre de todos.

5. Crean en sus estudiantes una atmósfera favorable a la investigación, la innovación y la creatividad.

6. Sustituyen el sistema de enseñanza conductista por el sistema constructivista o conectivista en el que el alumno tiene que idear, imaginar soluciones a las situaciones planteadas, con lo que se estimula la creatividad, la experimentación, la innovación y la conectividad.

7. Transforman los ambientes físicos de aprendizaje, facilitando la utilización de tecnologías y medios virtuales, hasta la creación misma de aulas virtuales y de una educación abierta y ubicua o mixta en la que lo virtual y lo presencial se complementan (Parra y Arenas-Dolz, 2015: 337-338).

Por lo tanto, una IAR que apueste por sociedades inclusivas, innovadoras y reflexivas, debe centrar su interés en el fortalecimiento de un sistema educativo que adquiera compromiso con el bienestar social y el cultivo de aquellas habilidades que sirvan para enriquecer los sistemas democráticos y los desafíos que depara la automatización del mundo del trabajo.

CAPÍTULO 8

EL DESAFÍO DE LOS COCHES AUTÓNOMOS

> Las disciplinas se diferencian en parte por razones históricas y por razones de conveniencia administrativa, y en parte porque las teorías que construimos para resolver nuestros problemas tienen una tendencia a construir sistemas unificados. Pero todas estas clasificaciones y distinciones son relativamente poco importantes y superficiales. No estudiamos temas, sino problemas; y los problemas pueden atravesar los límites de cualquier objeto de estudio o disciplina.
>
> (Popper, 1994: 95)

Uno de los avances más significativos en el campo de la IA se encuentra orientado a la investigación de los automóviles autónomos. Este avance tecnológico que incorpora intelectos sintéticos en su sistema cambiará de forma radical la vida cotidiana de la ciudadanía en términos de movilidad, economía y otros aspectos también importantes para la vida. Es importante destacar que, a pesar de que en este ámbito se suele hablar en términos generales de vehículos autónomos, este concepto abarca un campo muy amplio que va desde los automóviles, a los barcos, aviones y drones. Así pues, este capítulo estará dedicado exclusivamente a los automóviles sin conductor.

Los automóviles autónomos han sido creados para poder ir desde un punto geográfico a otro sin necesidad de intervención humana. La mayoría de los automóviles que han sido fabricados en los últimos años tienen asistentes tecnológicos que permiten dar los primeros pasos para la autoconducción, aunque todavía la conducción del volante es confiada a los seres humanos. En el futuro los automóviles autónomos no solo no tendrán conductor, sino que sus ocupantes se convertirán en pasajeros y podrán disfrutar de una serie de recursos tecnológicos interactivos en su interior que van desde ofertas de lugares de interés hasta formación cívica.

Este capítulo no versará sobre las controversias éticas que giran en torno a la introducción de los vehículos autónomos en las vías en materia de seguridad y que se abordan desde la ética empírica (Bonnefon *et al*., 2016), ni tampoco sobre las implicaciones morales de la conducción autónoma en el momento del reconocimiento óptico de personas u objetos (Sandberg *et al*., 2015). Más bien se profundizará en la línea que se ha desarrollado en esta tesis, es decir, cómo la IAR puede influir en el diseño y en el impacto cívico de los automóviles autónomos. La IAR integrada en las tecnologías autónomas podría implicar una asunción de compromiso ético y político más claro en el contexto en el que estas tecnologías despliegan su actividad para cultivar habilidades cívicas y fortalecer la democracia.

8.1. Desarrollo histórico del automóvil sin conductor

El sector de los coches sin conductor ha experimentado asombrosos cambios en función de los avances tecnológicos que han ido sucediéndose con el paso del tiempo. Así pues, un conocimiento introductorio de dichos avances a lo largo de la historia puede brindar unas interesantes coordenadas desde las que acceder a una mejor orientación en este campo que presenta tantas novedades y desafíos en la actualidad y en las próximas décadas.

8.1.1. Un temprano comienzo

A principios de agosto de 1925 un extraño espectáculo dejaba atónitos a un conjunto de espectadores en Broadway, Nueva York. Un automóvil vacío iba de un lado para otro por la calle, detrás lo seguía de cerca otro automóvil repleto de aparatos de radio y varios hombres desempeñando cada uno una función tras un conjunto de controles. Del primer automóvil sale una antena desde la parte de atrás, donde debería estar el asiento trasero, pues en su lugar hay un enjambre de cables, baterías y tubos. Este es el histórico caso del automóvil de Francis Houdina, la primera experiencia de radiocontrol impulsada por la *Houdina Radio Control Company* (Lee Stayton, 2011).

Aunque las investigaciones más recientes e importantes en el campo de los automóviles automatizados se remontan un par de décadas atrás, impulsadas principalmente por *Tesla*, *Google* o *Uber*, cualquiera podría pensar que los automóviles sin conductor son relativamente una novedad. Esta idea se sostiene principalmente en los avances que se han dado en los últimos años gracias al desarrollo de la IA. No obstante, la historia demuestra que ya en 1925 la compañía *Houdina Radio Control* había introducido el primer prototipo de automóvil autónomo en las calles de Broadway. Pensar que este acontecimiento histórico fue el inicio de un desarrollo constante y progresivo hasta el automóvil autónomo actual es un error colosal. El desarrollo de este campo ha sido el producto de la convergencia de varias fuentes de conocimiento que se han derivado de diversos paradigmas dados en el curso histórico. El espíritu que se encuentra tras los coches automáticos es el mismo que ha servido de impulso para otros inventos, pues pertenece a una cultura de electrificación. La síntesis máquina-herramienta, la interconexión en red, etc., son producto de las denominadas «acciones modo-máquina» (Collins, 1990: 42).

Cuando la compañía *Houdina Radio Control* comenzó a explorar el campo de los automóviles autónomos, la radio era una tecnología que estaba en boga. Un año más tarde, en 1926, la compañía *Aachen Motors's* exhibió un automóvil a control remoto bajo el nombre de *Phantom Auto* (Lee Stayton, 2011: 13). Aunque en la historia de los automóviles autónomos estos acontecimientos representan el comienzo de un espíritu revolucionario para la movilidad, durante la época significaron únicamente una forma de entretenimiento para las personas, en lugar de verdaderamente un avance tecnológico. Esto se debió en gran parte a que el vehículo todavía presentaba una gran dependencia de control humano. Un claro ejemplo de esta utilidad de la tecnología orientada hacia el ocio y el entretenimiento fue la Feria Mundial de Nueva York de 1939, donde acudieron 44 millones de visitantes para disfrutar de las muestras que allí se exhibían sobre los últimos adelantos científicos y tecnológicos. El gran triunfador de esta ideología tecnoutópica durante esta edición de la feria fue Futurama, una exhibición diseñada por Norman Bel Geddes para representar el mundo del futuro por medio de la escenificación de autopistas automatizadas,

gigantescos barrios, y otras obras de ingeniería sorprendentes durante esa época (Wetmore, 2003: 3-4).

8.1.2. La década de los 60

Durante la década de 1960 el interés por los vehículos sin conductor siguió presente en el imaginario colectivo de muchos ingenieros automotrices estadounidenses como Andrew Kucher, vicepresidente del departamento de ingeniería e investigación de Ford. Kucher fue mencionado en un artículo que se publicó en el *Chicago Daily Tribune* con el título: *In 50 Years: Cars Flying Like Missiles!*. Unos años antes, en 1953, las compañías *General Motors* y *Radio Corporation of América* exploraron la posibilidad de construir vehículos automáticos mediante experimentos caracterizados por el desarrollo de sistema de dirección y control a distancia (Wetmore, 2003: 6). Como fruto de las investigaciones realizadas por la compañía *General Motors* pueden encontrarse dos vehículos conceptuales, *Firebird y Firebird II*, aunque no tenían capacidades automatizadas en absoluto. El espíritu de Futurama siguió presente durante la Feria Mundial de Nueva York de 1964, donde *General Motors* presentó un sistema de carreteras automatizado muy parecido al que una década antes había presentado Vladimir Zworykin de *Radio Corporation of America* (Wetmore, 2003: 9)

8.1.3. La investigación desde los albores del siglo XX hasta la actualidad

Es importante mencionar que no todas las investigaciones de vehículos autónomos estuvieron dirigidas a los automóviles y que tampoco se dieron necesariamente en EE. UU. Los esfuerzos se centraron principalmente en otras tecnologías como los peajes electrónicos e información para el conductor facilitada a través de sistemas ubicados fuera del vehículo. A diferencia de la radio, se abrieron paso los sistemas eléctricos que implicaban cambios en las infraestructuras. La investigación de automóviles sin conductor comenzó con unos enfoques novedosos orientados a la visión computarizada. Estas técnicas de visión

computarizada se iniciaron en 1969, aunque no fue hasta 1980 cuando despegó en Alemania.

Este nuevo camino de investigación se centró en la visión computarizada que se dio al otro lado del Atlántico, concretamente en Alemania. En este país el ingeniero Ernst Dickmanns se esforzó en la creación de vehículos autónomos a mediados de la década de 1980. Por aquel entonces creó un vehículo 100 % autónomo usando como soporte material una furgoneta de la compañía Mercedes-Benz. Este profesor experto en IA de la Universidad Bundeswehr de Múnich utilizó visión sacádica, que consiste en una serie de movimientos rápidos de los ojos u otras partes de animales o dispositivos, además de cálculos probabilísticos y computación paralela, que permite resolver muchos problemas de forma simultánea (Dickmanns *et al*., 1994).

El proyecto alemán fue emulado en EE. UU. con un automóvil desarrollado por la Universidad Carnegie Mellon basado en el soporte físico proporcionado por un Chevrolet, probado con éxito en carreteras sin tráfico, donde alcanzó hasta 60 millas por hora (Bogost, 2014). Posteriormente, EUREKA, una organización intergubernamental para la financiación y coordinación europea en investigación y desarrollo, impulsó el proyecto PROMETEO que fue el mayor proyecto de I+D nunca antes conocido en el ámbito de los coches sin conductor, recibiendo 749 millones de euros. PROMETEO contó con la participación de numerosas instituciones universitarias y fabricantes de automóviles. Este proyecto se desarrolló entre 1987 y 1995 e involucró el vehículo VaMP14 creado por el laboratorio de investigación de Dickmann junto con su Daimler-Benz, VITA-II. Este modelo utilizaba video analógico digital con señales digitalizadas para detectar carriles y otros vehículos, además de sensores adicionales de presión para detectar el frenado, la temperatura, el ángulo de giro, la aceleración, entre otros (Ulmer, 1994: 2). Finalmente este proyecto europeo acabó con 1000 kilómetros de operación recorridos por parte de automóviles autónomos en condiciones normales de tráfico en las autopistas de París, así como también un viaje entre Múnich y Copenhague (Dickmann *et al*., 2014) 2014).

En EE. UU. Se emprendió también el camino de la investigación basada en la visión. En 1991 se aprobó la Ley de Eficiencia en el Transporte Intermodal de Superficie, que permitió que 650 millones de dólares se invirtieran para fondos de investigación en vehículos sin conductor durante los siguientes seis años (Novak, 2013). Esta ley fomentó la iniciativa para crear un automóvil automatizado basado en computación orientada a la seguridad y el impacto medioambiental. Finalmente todas estas investigaciones sirvieron para que en 2004 y 2005 se impulsara el gran desafío de la Agencia de Proyectos de Investigación Avanzados de Defensa (DARPA), que despertaría un gran interés y reuniría a importantes expertos para ampliar las capacidades de los automóviles autónomos.

En la actualidad, *Google* basa sus investigaciones en el escáner LIDAR, una red de cámaras y un complejo sistema de mapeo y GPS de gran precisión. En cambio, Mercedes-Benz se centra más en un complejo de cámaras y radares, y no en LIDAR. Por su lado, Tesla está trabajando en un sistema de enfoque interactivo para la autonomía del coche, enfocándose principalmente en actualizaciones de *software* que agregan nuevas funciones de automatización. La compañía Volvo probó sus automóviles hace unos años en las carreteras de Suecia y tiene especial interés en la investigación de sistemas avanzados de asistencia al conductor y en los sistemas de control de vehículos avanzados. Como puede comprobarse existe un amplio espectro en el desarrollo de la autonomía de los coches que va desde cambios más drásticos a otros más prudentes.

8.2. Inteligencia artificial y autonomía

Después de uno de los accidentes en los que estuvo involucrado el prototipo de *Waymo* en la ciudad de Chandler, Arizona, la empresa *Waymo* aprovechó la conferencia *Google I/O* para explicar con detalles la tecnología que se encuentra tras los automóviles sin conductor y explicar el papel de la IA en la función de la conducción. Como señala un informe de *Waymo* (Peña, 2018), el trabajo que han desarrollado los investigadores en este campo no podría haber sido posible sin su familiaridad con la IA, una herramienta que consideran fundamental, junto con el aprendizaje automático, que está permitiendo promover el carácter autónomo de los coches en los últimos años.

Los avances de la IA han sido decisivos en muchos ámbitos como el reconocimiento óptico o de voz. Los expertos en IA de *Google* colaboraron con *Waymo* para fortalecer la autonomía de sus coches. Los ingenieros de *Waymo* trabajaron codo con codo con el equipo *Google Brain* gracias a los últimos avances en *deep learning* de las máquinas y redes, concretamente en el sistema de detención de peatones. La tasa de error en este sistema de detención consiguió reducirse hasta cien veces gracias a la incorporación de la IA en estos proyectos, contribuyendo de ese modo a que estos sistemas sean más eficaces y seguros para las vías.

Los avances experimentados en *Waymo* en los últimos tiempos gracias a la IA han permitido franquear la delgada línea roja que existe entre la realidad y la ciencia ficción en materia de vehículos autónomos. Hoy en día *Waymo* es la única compañía en el mundo que tiene una flota de automóviles verdaderamente autónomos recorriendo las vías públicas. La IA juega un papel crucial en el engranaje de todas las partes que configuran el sistema de conducción autónoma. Aunque la percepción y el reconocimiento óptico son las áreas que han experimentado una mayor madurez en el aprendizaje profundo de las máquinas, también se utiliza la IA para fortalecer otras esferas que van desde la predicción, hasta la *planificación*, el mapeo y la simulación. A través de *machine learning* pueden emprenderse navegaciones por situaciones con diversos escenarios y variadas exigencias y complejidades. Hasta la fecha *Waymo* ha recorrido más de 6 millones de millas en vías públicas y se han recopilado cientos de millones de datos a través de la observación de interacciones entre otros automóviles, peatones, ciclistas, etc.

La infraestructura algorítmica de *Waymo* se basa en el ecosistema *TensorFlow* y los centros de datos que les facilita *Google*, incluida la unidad de procesamiento tensorial o TPU (del inglés *tensor processing unit*), un circuito integrado y desarrollado también por *Google* que se encuentra orientado para el aprendizaje automático. El TPU le brinda a *Wingo* un entrenamiento más eficiente de sus redes, lo que permite fortalecer las pruebas que se realizan, mejorando así sus modelos y la implementación a una mayor velocidad de las últimas redes en sus automóviles autónomos.

El reto de los automóviles autónomos es llevar la tecnología de conducción autónoma a todas partes y ante cualquier adversidad, ya sea una intensa lluvia o nevada. En ese sentido, la IA es un importante fundamento tecnológico para que la autonomía pueda fortalecerse y se alcancen nuevos logros. La IA permitirá a este nuevo sector impulsar sus investigaciones y hacer más fácil alcanzar la meta de un transporte más seguro, fácil y accesible para toda la ciudadanía.

8.3. Los coches autónomos se abren camino

La investigación en el sector de los automóviles sin conductor aumenta a un ritmo considerable como consecuencia de otros avances tecnológicos. El complejo que configura el ámbito de estos artefactos se encuentra impulsado desde varias aristas, lo que hace que su diseño y fabricación se demore tanto tiempo. Ese tiempo de espera es un punto a favor para las sociedades, las instituciones y los organismos que experimentarán su impacto, pues pueden hacer uso de él para valorar la incorporación de esta tecnología en las formas culturales de movilidad. Existen una serie de detonantes que sirven como impulso para este campo de investigaciones. La consultora *McKinsey & Company* ha identificado 10 elementos que representan las condiciones de posibilidad para la creación de estos automóviles:

Actuación	**Nube**	**Percepción y análisis de objetos**	**Control de conducción**	**Toma de decisiones**
Dirección, frenado y aceleración	Aprendizaje y actualización de mapas de alta definición, incluidos datos de tráfico, así como algoritmos para la detección de objetos, clasificación y toma de decisiones	Detección, clasificación y seguimiento de objetos y obstáculos	Conversión de salidas de algoritmos en señales de accionamiento para actuadores	Planificación del recorrido del vehículo, trayectoria y maniobras

Localización y mapeo	**Análisis**	**Sistema operativo**	***Hardware* del ordenador**	**Sensores**
Fusión de datos para cartografía de entornos y localización de vehículos	Plataforma para monitorear el funcionamiento del sistema autónomo, detectar fallas y generar recomendaciones	*Middleware* y sistema operativo en tiempo real para ejecutar algoritmos	Sistema de alto rendimiento y bajo consumo de energía en un chip (SOC) con alta confiabilidad	Múltiples sensores, incluyendo lidar, sonar, radar y cámaras

Fuente: McKinsey & Company, 2017.

En este momento la creación de automóviles autónomos se concentra en el desarrollo de tecnologías de asistencia en la conducción. SAE Internacional, la Sociedad de Ingenieros de Automoción, identifica seis niveles de autonomía en los vehículos en general:

- 0-sin automatización: el conductor humano tiene todo el control y realiza todas las tareas.
- 1-asistencia al conductor: algunas funciones simples y mínimas, como la dirección, pueden ser realizadas por el vehículo, aunque todas las funciones restantes recaen bajo estricto control humano.
- 2-automatización parcial: el sistema de asistencia del vehículo comienza a ayudar en algunos elementos de la conducción como la aceleración, se utiliza información que se obtiene del entorno de conducción (sensor de lluvia). En este nivel es necesario todavía un conductor preparado que se haga responsable de las tareas más primarias.
- 3-automatización condiciona: el automóvil comienza conducir en la mayoría de situaciones, como el cambio de carril, aunque el conductor todavía puede intervenir en el manejo de estas situaciones.
- 4-alta automatización: el automóvil tiene la habilidad de asumir la conducción frente a cualquier situación sin la necesidad de intervención humana. No obstante,

en situaciones climáticas severas se puede desactivar el piloto automático. El desarrollo de los coches de Google se encuentra en este nivel.

- 5-automatización completa: este nivel se caracteriza principalmente por la completa autonomía frente a todas las situaciones por muy complejas que sean sin necesidad de intervención humana (SAE International, 2019)

Es importante destacar que existe una especie de «carrera de la automatización» entre las empresas de este sector en pugna por alcanzar los niveles 4 y 5. Esto se debe principalmente a que aquellas empresas que alcancen esos niveles podrán hacerse con la exclusividad comercial. Entre las empresas más destacadas de este sector se encuentran: *Google, Tesla, Uber, Apple, BMW, Toyota, Audi, Mercedes Benz, Volvo* y *Volkswagen*, entre otras.

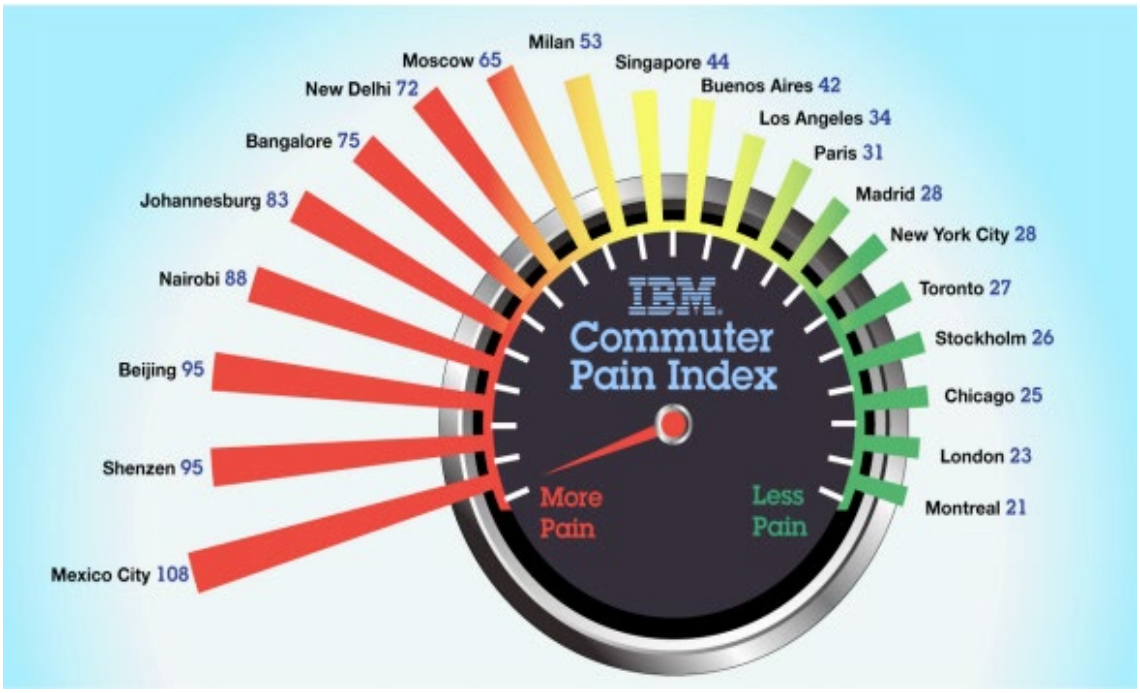
8.4. Cambio cultural de la movilidad

Más allá de los desafíos tecnológicos que enfrentan los automóviles autónomos, el cambio en la cultura de movilidad que se impone a la sociedad es uno de los desafíos más difíciles. No es de extrañar que debido a los accidentes que han ocurrido en el año 2018 en EE. UU. la ciudadanía muestre su desconfianza ante esta tecnología. Incluso hay quienes se han opuesto a que se realicen pruebas con este tipo de transportes autónomos en áreas urbanas donde hay muchos transeúntes (Berboucha, 2018).

Los coches autónomos forzarán un cambio en la cultura de la movilidad de la ciudadanía pues plantearán nuevas formas y posibilidades desde las que entenderla. Personas que en la actualidad han abandonado el uso del coche, como los ancianos o los invidentes, podrán hacer uso nuevamente de los vehículos sin conductor, asumiendo que se les brinda una nueva posibilidad para ampliar su horizonte móvil. Además, dos estudios, uno de la OCDE (Organización para la Cooperación y el Desarrollo Económicos) y otro de *Barclays*, ambos de 2015, revelan una reducción significativa de las flotas de coches, debido a que los coches sin conductor estarían en condiciones de sustituir los desplazamiento realizados tanto por vehículos particulares como por taxis u otros medios

de transporte público. Los denominados SAV (un automóvil con un servicio múltiple que depende de la demanda de cada momento, similar a los taxis pero sin conductor) y los PSAV (un vehículo similar a los SAV, pero de uso compartido simultáneo) promoverán otra movilidad diferente a la conocida hasta ahora. Además, se liberarán espacios viales gracias a la menor necesidad de plazas de estacionamiento, lo que favorecería la descongestión de la circulación y también el aprovechamiento de esos espacios para otros fines.

Además, un menor tiempo dedicado a la conducción permitirá un desplazamiento de los horarios diarios de domicilio-trabajo y los replanteará de otro modo, lo que posiblemente influirá de manera positiva en la salud y la productividad de las personas en el trabajo y en una mejor la conciliación con la vida familiar. Una encuesta realizada por IBM pone de relieve que los conductores de los países emergentes presentan más situaciones de estrés frente al volante:



Fuente: IBM, 2011: 7.

Este gráfico se elaboró teniendo en cuenta nueve factores:

- La media de tiempo en el desplazamiento para llegar al trabajo.
- El tiempo perdido en los atascos.
- Los precios de los carburantes.
- La percepción de la evolución sobre las condiciones del tráfico.
- La sensibilidad de los conductores a las frecuentes paradas durante un atasco.
- El estrés al frente del volante.
- El nerviosismo generado tras el volante.
- El impacto que el tráfico tiene sobre la productividad en el trabajo.

Así pues, los coches autónomos podrían contribuir al bienestar de la ciudadanía que dedica tiempo a la conducción porque representarían una menor preocupación frente a dicha actividad.

Gracias a las redes de conexión que presentan los coches sin conductor, los servicios de geolocalización permitirán acceder a un mayor conocimiento sobre el conductor en materia de costumbres y preferencias, y las tecnologías integradas en estos vehículos, como sensores y cámaras, brindarían aplicaciones fundadas en el estudio de comportamiento. Esto facilitará que los conductores puedan localizar a lo largo de su trayecto o en el destino aquellos servicios que puedan interesarle en función de los patrones surgidos del análisis de su comportamiento. Entre esta información pueden encontrarse: restaurantes, hoteles, miradores, teatros, cines, lugares de interés, sitios turísticos, aeropuertos, etc. Además, se podrán elegir trayectos alternativos si así se desea. A pesar de lo interesante de esta propuesta, es importante destacar que el acceso a los datos personales puede plantear problemas de confidencialidad.

Por lo tanto, el coche 100 % autónomo posibilitará una liberalización de tiempo en el trayecto de cada desplazamiento, permitiendo la dedicación para otras ocupaciones según el deseo del conductor o su estilo de vida. La posibilidad de no conducir podría provocar una reducción de las situaciones de estrés y repercutirá positivamente en el bienestar social y personal.

La presencia de automóviles autónomos en las vías no repercutirá solamente en la cultura de la movilidad, sino también en la seguridad vial y en cómo se la percibe. Debido a la falta de necesidad de control humano, no habrá conductores en estado de embriaguez, ni con muestras de somnolencia al volante. Sería una solución a muchos de los problemas de mortalidad en las carreteras, ya que reduciría los errores humanos. No obstante, debido a la coexistencia durante un tiempo de conducción autónoma y humana, la transformación de esta cultura de la conducción en materia de seguridad vial se demorará un tiempo.

Por último, también tendrá un impacto positivo en la relación con el medio ambiente, pues contribuirá de manera positiva al reconocimiento del valor que se le confiere a la naturaleza, ya que los automóviles autónomos que están siendo desarrollados se basan en fuentes de energía renovables o en la electricidad.

8.5. Las ventajas de la autonomía

Las investigaciones en el sector de los automóviles autónomos se ven motivadas por los beneficios que ofrecerá la incorporación de esta tecnología en el futuro. Debido a la distancia temporal es complejo imaginar muchas de estas ventajas. Lasse Rouhiainen destaca las siguientes:

- Más seguridad en las calles: según datos de la Organización Mundial de la Salud (OMS), al año se producen 1,35 millones de muertes en todo el mundo por accidentes de tráfico. Entre las principales causas se encuentran las distracciones al volante y el consumo de alcohol. Con la autonomía de los vehículos esto ya no será un problema y las vías serán más seguras, tanto para conductores como para peatones.

- Reducción de los gastos de hospital: estrechamente relacionado con la seguridad se encuentran los gastos sanitarios. Esto quiere decir que si hay menos siniestros de tráfico, los gastos sanitarios se reducirán considerablemente.

- Incremento de la productividad: debido a la conducción automática, tanto el conductor como sus acompañantes liberarán tiempo que podrán dedicar a otras tareas u obligaciones.

- Distribución más rápida de los negocios: los modernos sistemas de navegación permitirán ganar tiempo y optimizar nuestras tareas de forma más eficiente, lo que tendrá un efecto positivo en el capítulo económico.

- Mejora en la eficiencia del tráfico: los coches autónomos no repetirán los patrones culturales del manejo inadecuado de determinados episodios de tráfico y por lo tanto la movilidad será más fluida y segura.

- Menos problemas de aparcamiento: los vehículos autónomos no tendrán siempre la necesidad de aparcar y por lo tanto de ocupar un espacio. El espacio libre podrá ser entonces utilizado con otras finalidades y los aparcamientos no serán un problema para la planificación urbana de las instituciones políticas.

- Opciones de movilidad más económicas: debido al ahorro en el capítulo de personal, los coches sin conductor implicarán una reducción en los costos de movilidad.

- Menor impacto medioambiental: las investigaciones en los automóviles autónomos se centran en la búsqueda de fuentes de energías renovables o en la electricidad, por lo que se producirán menos emisiones de CO2 a la atmósfera y por lo tanto se contribuirá con la calidad de vida de la ciudadanía (Rouhiainen, 2018: 196-198).

En definitiva, existen importantes beneficios en el ámbito de los coches sin conductor, aunque eso no quiere decir que la cuestión de la crítica y la responsabilidad ética deban ser descuidadas. Con más motivos, la incorporación de la responsabilidad en este contexto es

fundamental para orientar todos los diseños y proyectos hacia un fortalecimiento de los pilares democráticos y cívicos que sostienen las sociedades actuales a través de un compromiso con los ODS y los límites planetarios. Esta tecnología puede contribuir sin duda a la garantía de cuidado y compromiso con los retos actuales, ya que representa una valiosa herramienta para la consecución de los mismos desde sus concretas posibilidades.

8.6. Desafíos, vulnerabilidades y amenazas

Como en cualquier otro despliegue de la actividad tecnológica más avanzada, el contexto de los automóviles sin conductor también presenta importantes desafíos en torno a las vulnerabilidades y amenazas. Esta tecnología sostenida sobre el espíritu de la autonomía conduce a un horizonte de nuevas posibilidades de diversa índole que pueden interpretarse y considerarse a la luz de la voluntad del ser humano. Así pues, no solo pueden ser aprovechadas para el fortalecimiento de los desafíos que enfrenta la humanidad en el ámbito de los ODS, por ejemplo, sino también con fines maliciosos y perjudiciales para la ciudadanía. Esta exposición hacia fines que se alejan de un compromiso con el bienestar y la democracia encuentra su origen en las tecnologías de doble uso. Así pues, a continuación se detallarán algunos de estos desafíos relacionados con las vulnerabilidades y las amenazas que deparan los automóviles sin conductor:

- Brechas de seguridad:

  Los vehículos que comenzaron a fabricarse a partir de 2009, y concretamente a partir de 2015, poseen cada vez más sistemas de control electrónico que permiten testear el sistema interno o estar conectados a una centralita de control de la marca de nuestro automóvil. En ese sentido, con el internet de las cosas y la comunicación entre máquinas, los automóviles se encuentran más expuestos a ciberataques, por ejemplo el clonado de mecanismos de acceso a los aparatos como las llaves. El sistema integrado de control de los automóviles se sitúa en al área de una red como *Flexuras, Local InterConnect Network* (LIN), *Ethernet*, etc. Fernando Ruiz Domínguez (2017) destaca que los ciberataques a un automóvil pueden darse de

forma directa o inalámbrica: mediante el acceso a los puertos USB, que permite la conexión de dispositivos electrónicos; mediante una computadora de diagnóstico, que facilita el acceso al ordenador de a bordo, o también conocido como cerebro o centralita; o también a través del CD-ROM, un acceso cada vez más inusual. Además, los ciberataques también pueden producirse por control remoto a través de las llaves del vehículo; de la monitorización de la presión de los neumáticos; de los sistemas de asistencia al conductor; y por último a través de Bluetooh o wifi.

De los dos tipos de ataque, el que se realiza por medios inalámbricos es más peligroso porque permite atacar múltiples objetivos al mismo tiempo; en cambio, el caso contrario requiere de un contacto directo con el ordenador de a bordo, lo que limita el riesgo (Ruiz Rodríguez, 2017: 4-5). Así pues, algunos estudios demuestran el riesgo al que se encuentran expuestos los automóviles debido al entramado de redes que los configuración y que se encuentran integradas en los mismos (Checkoway *et al*. 2011).

- Acceso y tratamiento de los datos:

  Los automóviles sin conductor dependen en mayor medida de las bases de datos que le son suministradas para optimizar su funcionamiento y configurarlo. Esta cantidad de datos al que tendrán acceso plantea la necesidad de llevar a cabo un ejercicio de preocupación por el tratamiento de los mismos y una incorporación de criterios de responsabilidad bajo parámetros de seguridad. Rouhiainen (2018: 199-200) menciona un informe elaborado por *Intel* y *Strategy Analytics* bajo el título *Accelerating the Future: The Economic Impact of the Emerging Passenger Economy*, para poner de relieve la cantidad de control de datos a los que están expuestos los consumidores.

  Rouhiainen advierte de que los gobiernos de distintos niveles deben comenzar a considerar esta exposición para que los datos que son utilizados en los automóviles sin conductor respeten la identidad del consumidor y procedan con mecanismos de transparencia. Esta advertencia se debe no solo a la información facilitada de forma

voluntaria por los usuarios, sino también a aquellos datos de conducción que son recopilados por sistemas electrónicos de control de los vehículos almacenados en los ordenadores de a bordo y transferidos a algún servidor de almacenamiento de información.

- Controversias éticas:

  Entre los desafíos más importantes están las implicaciones éticas que la llegada de los coches autónomos y robotizados tendrá sobre la industria y las instituciones públicas. Son los desafíos éticos los que verdaderamente están provocando importantes rompecabezas en los principales laboratorios de IA del mundo. El MIT impulsó el proyecto *Moral Machine*, que como se indica en su página web es «una plataforma para recopilar una perspectiva humana sobre las decisiones morales tomadas por las máquinas inteligentes, como los coches autónomos» (Massachusetts Institute of Technology, 2016).

  La responsabilidad adoptará una nueva dimensión tras la irrupción a gran escala de los coches sin conductor. Tanto es así que el mundo de los seguros y del derecho ya ha comenzado a discutir la implantación de nuevas medidas ante este tipo de novedades tecnológicas. En torno al principio ético de la responsabilidad están girando importantes controversias que ponen a los automóviles autónomos en el centro del debate.

- Uso malicioso:

  Una vez que han sido mencionados algunos de los riegos ante los que se encuentran los automóviles autónomos, es más fácil afirmar que estas máquinas pueden ser un medio empleado para la generación de daños personales y materiales. Los coches autónomos podrían ser utilizados para el crimen organizado o para ataques terroristas, dado que son un elemento fácilmente accesible y manipulable mediante las técnicas de hackeo orientado a la ciberguerra (Ethical Hacking News Tutorials, 2016).

Puede hacerse un doble uso de estas tecnologías avanzadas para que promuevan fines para los que no fueron concebidas. Así pues, como otras creaciones tecnológicas avanzadas, el automóvil autónomo implica importantes riesgos dada la vulnerabilidad de sus sistemas informáticos integrados. De la misma manera que algunos ataques terroristas de los últimos tiempos se han perpetrado con coches o camiones, puede pensarse que los automóviles inteligentes son un objeto muy atractivo para grupos criminales especializados en el robo en sociedades a nivel global.

- Impacto político y económico:

Igualmente hay que darle importancia a las políticas públicas que reglamentarán los ciberataques y las manipulaciones de los coches sin conductor, así como a las causas y consecuencias de los accidentes, y las cargas de responsabilidad que de las mismas se derivan. Los mercados deben estudiar las posibles consecuencias de las campañas que pueden generarse en contra de este producto tecnológico cuando se produzcan algunos siniestros.

Al igual que la industria robótica está teniendo un impacto positivo en la economía industrial, los coches autónomos pueden tener un efecto similar. No obstante, un informe de *Barclays* en 2015 estima que el número de automóviles por hogar se reducirá drásticamente: en EE. UU. el número autos caerá de los 2.1 actuales a 1.24, mientras que en Reino Unido pasará de 1.2 a 0.7. Ese fenómeno también tendrá un impacto en la economía de los hogares, pues el modelo común de movilidad de un coche-una persona, se verá desplazado hacia nuevas formas de movilidad, como ya se ha señalado al mencionar los SAV y los PSAV. Además, el mercado de los automóviles tradicionales (con conductor) se reducirá considerablemente en las próximas dos décadas.

- Reconocimiento público:

  Uno de los desafíos más importantes que enfrenta la introducción de los coches sin conductor consiste en la aceptación pública que debe ir precedida de un convencimiento de su necesidad. Los accidentes ocurridos durante el año 2018 no representan un hecho a favor de este reconocimiento público, aunque eso no puede ser un motivo suficiente como para mostrar una firme oposición a estas tecnologías. Así pues, este tipo de vehículos deben necesariamente aceptarse por parte de la ciudadanía, aunque antes es importante que ofrezcan las suficientes muestras de seguridad.

8.7. Bienestar y compromiso cívico

Antes se señaló el impulso en materia de IA que se encuentra tras los coches autónomos. Dado el propósito de este trabajo, que consiste en promover una IAR comprometida con la humanidad y su bienestar, es necesario contextualizar las exigencias de responsabilidad en el marco de la Agenda 2030 y los límites planetarios. Así pues, a continuación se detallan algunas de las prioridades de responsabilidad que podrían asumir estas tecnologías inteligentes del ámbito de la movilidad:

- Educación con medios interactivos para promover los ODS:

  Debido al cambio cultural al que conduce el carácter autónomo de los automóviles y el tiempo liberado como resultado de la ausencia de control humano en la conducción, ese espacio temporal podría ser destinado a la educación. Un tiempo dedicado a la educación, mediante las herramientas interactivas que estos automóviles ofrecen, podría favorecer el cultivo de habilidades cívicas y democráticas y por lo tanto fortalecer el compromiso con las exigencias del mundo actual.

  Esta apertura de posibilidades debido a la liberación temporal, sugiere una nueva gestión del tiempo y por lo tanto un cambio cultural respecto a él. María Ángeles Durán Heras (2010; 2012) es una investigadora española que durante décadas se ha

especializado en el estudio del trabajo no renumerado y su relación con estructuras sociales y económicas. Como señala Durán Heras, el tiempo es un recurso escaso y cada persona hace un uso diferente, aunque sobre lo que esta investigadora trata de reflexionar es sobre si la gestión del tiempo es voluntaria u obligada y si en realidad existen posibilidades de cambios en el futuro (2010: 15) A propósito del uso cívico del tiempo que liberan los coches autónomos, es importante realizar una diferenciación conceptual entre trabajo y empleo:

> La delimitación de la frontera entre trabajo y empleo no es una cuestión lingüística, es, sobre todo, una cuestión política, porque el estatuto del trabajador va asociado con algunos de los más importantes derechos y obligaciones sociales y económicas. En España, el derecho laboral, tal como afirma el Estatuto de los Trabajadores, solo se aplica a una pequeña parte de lo que puede considerarse trabajo. Quedan excluidos, en el sentido de que no se rigen por estas normas, numerosos trabajadores que desarrollan otros tipos de trabajo:
>
> a) El trabajo no retribuido.
>
> b) El trabajo forzoso.
>
> c) El trabajo de los familiares que conviven con el empresario y no son asalariados.
>
> d) El trabajo por cuenta propia.
>
> e) El trabajo independiente.
>
> f) Algunos tipos especiales de trabajo que se regulan por normas propias (funcionarios de la Administración Pública y otros) (Durán Heras, 2010: 21-22).

Las investigaciones de Durán Heras sobre el impacto directo que tiene el tiempo en las estructuras económicas y sociales resultan muy útiles para esclarecer que las tecnologías más avanzadas liberan un espacio temporal en el que pueden cultivarse conocimientos que contribuyan a poner solución a las problemáticas de esta época. En ese sentido, el tiempo estaría siendo destinado para un trabajo no remunerado pero que representa una acción directa sobre el mundo.

La educación ciudadana es fundamental en los contextos democráticos y una liberación de tiempo, acompañada de un aprovechamiento de las herramientas tecnológicas de estos automóviles, favorecería considerablemente la formación cívica. Por ello, habría que orientar el diseño de los asistentes tecnológicos de los coches sin conductor hacia un nuevo núcleo de posibilidades que posibilitaran una educación para la ciudadanía. Una ciudadanía formada es una potente herramienta para el compromiso que están demandando los desafíos del planeta en la actualidad. En ese sentido, la IAR podría impulsar el diseño de los automóviles sin conductor y asumir un compromiso con el ODS 4, educación de calidad, que se encuentra estrechamente vinculado con los demás exigencias de la Agenda 2030, pues sin una formación de calidad resulta muy difícil la toma de conciencia cívica.

- Compromiso con el medio ambiente y cero emisiones:

El sistema capitalista ha ocasionado la destrucción del medio ambiente como consecuencia de los altos niveles de contaminación que generan sus industrias y modelos de consumo. La nube de polución que acecha a las ciudades y pueblos es más que evidente y son escasas las políticas públicas que los gobiernos están promoviendo para solucionar esta grave crisis y su repercusión en la salud pública y el equilibrio de la biosfera. Por ello es urgente plantear un modelo de movilidad alternativo al actual que surja desde un cambio cultural en la movilidad ciudadana, basada mayoritariamente en el uso del automóvil privado e impulsado por un motor de combustión de energías fósiles, poco eficiente y sostenible como medio de transporte, y que como sostiene un informe de Ecologistas en Acción (2007) es la principal causa de la contaminación urbana.

En este panorama de vulnerabilidad medioambiental está surgiendo una disrupción tecnológica que podría provocar una transformación cultural en el ámbito de la movilidad, pues la humanidad se encuentra a las puertas de la incorporación en las vías de los automóviles sin conductor. Estas tecnologías autónomas generarán una revolución sin precedentes en la movilidad urbana y en la configuración de las

ciudades, teniendo una influencia directa sobre las condiciones medioambientales. Al promover una nueva cultura de la movilidad, el vehículo autónomo dejará de ser visto como un objeto de consumo como tal para convertirse en un medio tecnológico que brindará un servicio que podrá utilizarse en función de las necesidades específicas de movilidad. Además, un estudio desarrollado por el *Rocky Mountain Institute* ha puesto de relieve que los coches sin conductor tendrán un impacto positivo en el medio ambiente, por ejemplo en la reducción de los recursos utilizados para la fabricación de partes destinadas a la protección y la seguridad, ya que se espera que estos automóviles reduzcan la frecuencia de los accidentes (Walker, 2014).

Por último, dado que las fuentes que proporcionarán el suministro energético para los coches autónomos proceden de recursos renovables y no fósiles, no solo no se emitirán gases contaminantes, sino que también se producirá una incidencia directa sobre la disminución de CO2. Así pues, los coches autónomos como medio tecnológico contribuirían a desarrollar una IAR comprometida con los siguientes ODS: Salud y bienestar, energía asequible y no contaminante, industria, innovación e infraestructura, ciudades y comunidades sostenibles, producción y consumo responsable, acción por el clima y vida de ecosistemas terrestres.

- Contribución para que las ciudades sean más seguras y habitables al haber menos siniestralidad:

Los accidentes ocasionados por los coches sin conductor durante el año 2018 en EE. UU., concretamente en Arizona, con un prototipo de *Uber* que tuvo un desenlace fatal tras la muerte de una mujer (Limón, 2018), y en California, con un prototipo de *Tesla* que no tuvo víctimas mortales (Álvarez, 2018), sirvieron para que las principales empresas del sector que están investigando en esta materia fortalecieran sus esfuerzos para exigir mejoras en materia de seguridad. A pesar de estos accidentes, los datos que arroja la OMS cifran los accidentes de automóviles con conductor en 1,35 millones al año a nivel mundial. Estos datos ponen de relieve que supondría un error el magnificar de forma impertinente los accidentes que han tenido los coches sin conductor, ya que

según la *National Higway Traffic Safety Administration* (NHTSA) de EE. UU., el 94 % de los accidentes en ese país se deben a errores humanos (Sánchez, 2018).

Al contrario de lo que promueven algunas portadas de diarios sensacionalistas, los automóviles sin conductor podrían favorecer la seguridad vial, ya que se reducirían progresivamente, en función de la incorporación de estos automóviles, ciertas conductas viales que provocan que el tránsito sea más inseguro. Además, una conducción más segura y prudente por parte de estos automóviles serviría de ejemplo para que los seres humanos fueran adquiriendo ciertos parámetros de comportamiento orientados a fortalecer las habilidades cívicas. En definitiva, la conducción autónoma podría tener un impacto positivo, tanto cualitativo, al servir de ejemplo formativo y mejorar la calidad de la conducción y la habitabilidad de los entornos urbanos, como cuantitativo, al reducir los costos destinados a material de seguridad y también las víctimas por accidentes de tráfico. Esta sería una IAR comprometida con el bienestar de la ciudadanía, la habitabilidad de las ciudades y las vías más seguras. En este caso, estarían involucrados ODS como los siguientes: salud y bienestar, industria, innovación e infraestructura y ciudades y comunidades sostenibles.

- Liberación de espacios anteriormente dedicados a aparcamientos para zonas verdes u otras finalidades cívicas:

Uno de los efectos derivados del cambio cultural en materia de movilidad es una menor utilización de los automóviles privados y personales y una consecuente despreocupación por la búsqueda de estacionamiento. Si los vehículos autónomos utilizados para la movilidad en el ámbito urbano asumen una función de transporte compartido, como los SAV y los PSAV, existirá una menor flota debido al correspondiente uso colectivo.

La movilidad compartida permitiría una reducción considerable de la flota de automóviles y por lo tanto una liberación del espacio dedicado al estacionamiento. Esos espacios destinados al estacionamiento de automóviles podrían adquirir otra función que los grupos de interés acordarían mediante un ejercicio deliberativo. Las

nuevas funciones de estos espacios deberían asumir la necesidad de reorientar su finalidad para potenciar un uso público y cívico de los espacios y de este modo contribuir con los ODS, entre los que podrían encontrarse los siguientes: salud y bienestar, agua limpia y saneamiento, energía asequible y no contaminante, industria y ciudades y comunidades sostenibles. Así pues, la IAR incorporada en los coches autónomos tendría una consecuencia directa sobre el hábitat, ya que contribuirían a una resignificación del uso y las funciones cívicas y democráticas de los espacios que tradicionalmente han estado destinados a estacionamientos.

En conclusión, es importante destacar el gran poder transformador que tiene la IAR y cómo puede contribuir a la consecución de los logros de la Agenda 2030. Otra forma de concebir la IA es posible y eso se ha demostrado en este capítulo a través de los coches autónomos, que plantean notables ventajas frente a los coches con conductor, uno de los principales causantes del deterioro medioambiental y del caos en los espacios urbanos contemporáneos. La revolución en materia de movilidad es inminente y los actores implicados tienen la responsabilidad de promover un aprovechamiento de las tecnologías más avanzadas en este ámbito para el fortalecimiento de las exigencias cívicas y la democracia en términos de una mejora de las condiciones de vida de la ciudadanía.

CAPÍTULO 9

LOS DESAFÍOS MILITARES Y LA CIBERSEGURIDAD

> La ciencia y la tecnología no son neutrales, sino que pueden implicar desde el comienzo hasta el final de un proceso diversas intenciones o posibilidades, y pueden configurarse de distintas maneras. Nadie pretende volver a la época de las cavernas, pero sí es indispensable aminorar la marcha para mirar la realidad de otra manera, recoger los avances positivos y sostenibles y a la vez recuperar los valores y los grandes fines arrasados por un desenfreno megalómano.
>
> (Francisco, 2015: 94)

En este capítulo se estudian los retos de la IA aplicada al ámbito de la defensa y la ciberseguridad. En el contexto actual la robótica está experimentando avances significativos en ámbitos que se centran en la ciberseguridad y el manejo de datos masivos, y además se puede constatar una proliferación en la investigación y desarrollo de capacidades militares mediante la IA. En este capítulo se atenderá en particular a los desafíos éticos en el uso militar de la IA y la ciberseguridad y se analizarán los nuevos retos que plantean estos dos campos de despliegue tecnológico.

El uso de los drones se está normalizando en los ejércitos desde la última década del siglo XX. Existen varias definiciones de drones, aunque la más acertada y amplia es la propuesta por Grégoire Chamayou, que lo define como «un vehículo de tierra, mar o aire, controlado a distancia o de forma automática» (Chamayou, 2016: 11). Este uso ha generado importantes controversias éticas relacionadas con la imputación de responsabilidad en el ámbito de estas tecnologías. En ese sentido, han proliferado las perspectivas éticas que han tratado de arrojar luz sobre este nuevo fenómeno militar y sus controversias.

Una propuesta de IAR aplicada al ámbito de los drones debería tener en cuenta a los diversos afectados por esta actividad y fomentar su participación en el diseño y producción de estos artefactos autónomos. Este mecanismo de participación permitiría considerar a la

tecnología del campo militar como un conocimiento-recurso público que es susceptible de tratamiento cívico y, por lo tanto, objeto de fortalecimiento democrático. Además, el ejercicio ético aplicado a estas tecnologías militares también permitiría una resignificación de sus propósitos, entre los que se encuentra una mayor preocupación por la cultura humanitaria.

Por último, en este capítulo se abordará la aparición de la ciberguerra como un nuevo espacio de despliegue de los conflictos bélicos que presenta novedosos desafíos para las instituciones que conforman las sociedades actuales. Debido al avance tecnológico, las formas de hacer la guerra están experimentado una transformación de la que surgen nuevos modos de enfrentamiento y expresiones de la violencia. En ese sentido, una IAR contribuye a la generación de mecanismos de seguridad que se encuentran motivados por la protección de los valores cívicos y democráticos.

9.1. Tecnología y campo militar

A medida que la IA y la robótica continúan desarrollándose, el interés que despiertan en el ámbito militar es cada vez mayor, ya que consideran que estas tecnologías avanzadas se pueden convertir en importantes herramientas para la obtención de beneficio en las estrategias de los ejércitos. La IA permite la creación de nuevas capacidades militares que sin duda tendrán un impacto considerable en las estrategias de las instituciones de defensa de los Estados. Entre la gama de efectos disruptivos que pueden generar los sistemas artificiales en el ámbito militar pueden encontrarse los que van desde el espionaje y la vigilancia a los nuevos mecanismos ofensivos y defensivos.

Los profesionales que conforman el grupo de la actividad militar están empleando cada vez más intelectos sintéticos o maquinas teledirigidas. Peter Warren Singer (2009a) ha denominado a este fenómeno «deshumanización de la guerra», pues la tecnología permite tomar distancia respecto al campo de batalla. No obstante, el fenómeno de la toma de distancia en el campo de batalla viene dándose a lo largo de la historia con arcos y fechas, catapultas, cañones, etc. En cambio, la tecnología más avanzada ha permitido ampliar de

forma disruptiva esta distancia con máquinas como los drones, situándose en una era posheroica, que incluso algunos investigadores como Christopher Coker (2015) identifican como la ausencia de heroicidad de las máquinas. Más allá de asuntos relativos a la heroicidad, es importante esclarecer los orígenes de los robots militares, como hacen Javier Jordán y Josep Baqués (2014), así como el desarrollo histórico que han experimentado hasta llegar a la actualidad, pues representan el antecedente de la influencia que la IA está teniendo en la actualidad sobre las estrategias de las instituciones de defensa.

9.1.1. Los orígenes de los robots militares

Fue durante la Gran Guerra cuando EE. UU. desarrolló el torpedo aéreo *Liberty Eagle*, un avión no tripulado con carga explosiva. En 1942, en plena Segunda Guerra Mundial, la marina estadounidense compró mil unidades del TDR-1, un avión bimotor que era capaz de transportar material detonador. Pero no fue EE. UU. la única potencia que mostró interés por las nuevas tecnologías aplicadas al ámbito militar, sino también Alemania, país que desarrolló la bomba volante V-1, un primer intento de dron o vehículo aéreo no tripulado (UAV, siglas en inglés de *unmanned aerial vehicle*). Además, la Unión Soviética dotó al Ejército Rojo durante la Segunda Guerra Mundial de un pequeño número de carros dirigidos por un sistema de control remoto. Sin embargo, no fue hasta la guerra de Vietnam cuando se produjo un salto cualitativo de los drones, mediante el encargo de una función novedosa centrada en el reconocimiento aéreo para contribuir a la táctica militar. Dadas las limitaciones de la época se cancelaron los programas de drones de reconocimiento táctico en la década de 1960. Posteriormente vendrían otros modelos como el *Firebee*, el *Ryan Model 147A Fire Fly* o el *Lightning Bug* que sentarían las bases de los futuros drones. Además del reconocimiento óptico, también se centró la investigación en el terreno del espionaje con el d-21 que no tuvo los éxitos esperados por parte del ejército de EE. UU.

Posteriormente, tras las lecciones aprendidas en la guerra del Vietnam, el impulso que experimentó el campo de los drones fue creciente. En la Guerra de los Seis Días librada en el valle de la Becá, el ejército israelí empleó un gran número de drones que permitieron el ahorro de aviones de combate y la disminución de bajas militares. Los drones *Ryan 147I*,

Scout y *Mastiff* permitieron localizar emplazamientos antiaéreos sirios. Así pues, la tecnología militar empleada permitió al ejército de Israel imponerse a los sistemas integrados de defensa del área de fabricación soviética, algo que asestó un importante golpe psicológico a los mandos militares del Pacto de Varsovia. La campaña del Líbano despertó un gran interés en EE.UU. para adquirir un millar de *ADM-141ª Tactical Air Launched Decoys* (TALD), y posteriormente el *RQ-2ª Pioneer*, que sería también utilizado en la Guerra de Iraq, hasta su sustitución total por el *RQ-7 Shadow* y el *ScanEagle* en 2007. Otros países con importantes ejércitos, como Francia, Alemania, Canadá, Reino Unido e Italia, también emplearon durante la década de 1990 drones como el *CL-289.*

9.1.2. El presente de la robótica militar

Los robots se han abierto camino con el paso del tiempo en el campo militar gracias a los avances tecnológicos, como ha podido comprobarse en el crecimiento de las unidades y la diversidad empleada. El comienzo puede situarse en los drones destinados a misiones de inteligencia centradas en el espionaje (Jordán y Baqués, 2014: 35-38).

El dron *RQ-4A Global Hawk* es el más sofisticado y de mayor tamaño de la actualidad. Está en servicio desde el año 2003 y puede mantenerse en vuelo durante más de veintiocho horas. Desde el año 2014 se impulsó la versión *Block 40* que permitía señalar objetivos terrestres en movimiento. Además de este dron, existen otros ejemplos de actualidad como el *MQ-4C Triton*, empleado por la Armada de los Estados Unidos. Para aquellas misiones más discretas en materia de espionaje, EE. UU. utiliza el dron *RQ-170 Sentinel* en la vigilancia de objetivos como el complejo donde vivía Osama Bin Laden, aunque también ha sido utilizado para obtener datos de inteligencia sobre Corea del Norte (Sweetman, 2010).

En este terreno de la inteligencia se están impulsando en los últimos años microdrones que simulan la forma de un ave o un insecto. Su principal ventaja es el tamaño y la similitud que presentan con ciertos animales o insectos, facilitando considerablemente que puedan pasar inadvertidos (Ackerman, 2011).

Más allá del tamaño y el objeto de los drones, ya sean de espionaje o para cualquier otro uso, el interés que despiertan por parte de las instituciones es su poder de autonomía. Es cierto que los robots militares autónomos no son todavía una realidad, aunque el desarrollo de la IA está permitiendo traspasar fronteras nunca antes conocidas en el terreno militar, permitiendo incluso que los misiles seleccionen sus objetivos (Markoff, 2015). Hay varios ejemplos de esta creciente autonomía: el sistema israelí *Drome*, el norteamericano *Phalanx CIWS*, el inglés *Taranis*, etc. Una de las principales prioridades del Pentágono, dentro de la llamada Alianza Tecnológica Colaborativa Robótica (RCTA, siglas en inglés de *Robotics Collaborative Technology Alliance*), se centra en la inversión de las investigaciones en el desarrollo de diversas capacidades entre las que se encuentran las siguientes: pensar (adaptar el razonamiento táctico), mirar (centrarse en la percepción de la situación), moverse (de forma segura y adaptativa), hablar (comunicación eficiente) y trabajar (interacción con el mundo físico) (Ortega, 2016: 198).

9.1.3. Nuevos horizontes para la robótica militar

Las posibilidades que se están abriendo gracias a la tecnología más avanzada, así como el interés con el que los ejércitos observan el desarrollo de los sistemas autónomos, han permitido abrir la puerta a nuevas misiones, como señalan Jordán y Baqués (2014: 45-46), entre las que se destacan las siguientes:

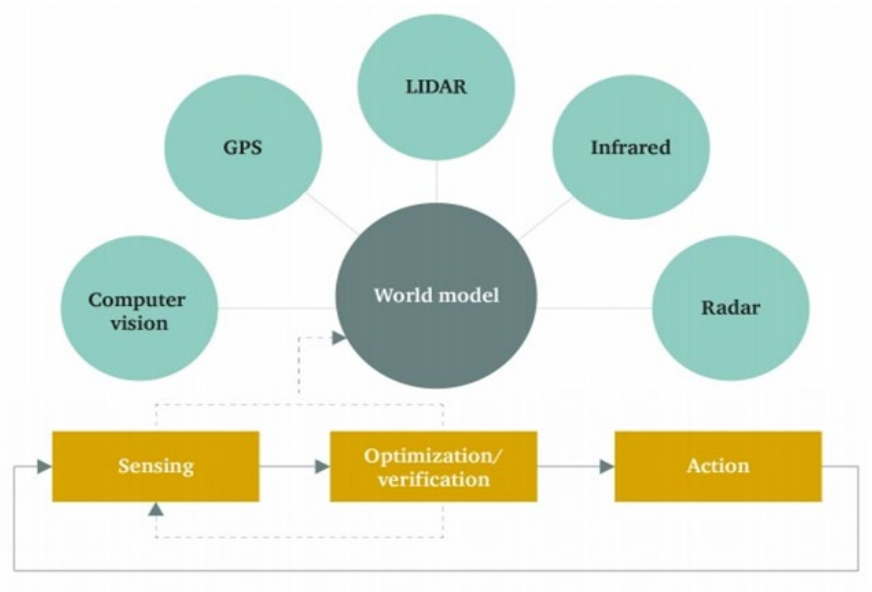
1. Asistencia médica en combate.
2. Transporte.
3. Lucha contra los incendios.
4. Vigilancia terrestre de fronteras y perímetros.
5. Desenvolverse con soltura en entornos humanos.

Durante el siglo XX el campo militar ha sabido adaptarse muy bien a los avances tecnológicos y en la actualidad la robótica y la IA también han contribuido en el desarrollo e investigación. Es importante reconocer que muchos de los proyectos se encuentran todavía en una etapa inicial, aunque cabe mencionar que, conforme se van dando nuevos avances en el campo de la IA, se abrirán nuevos espacios para el despliegue de las estrategias militares. El futuro de incertidumbre en el terreno militar suscita diversos interrogantes: ¿participarán los robots en misiones donde tengan que determinar si un sujeto es un civil o un terrorista? ¿Serán los seres humanos desplazados de aquellas situaciones que requieren una alta capacidad deliberativa? ¿Destinarán los ejércitos a los robots únicamente para fines bélicos? ¿Están las esferas de la quíntuple hélice del MIAR discutiendo y valorando los desafíos éticos que giran en torno a la introducción de los intelectos sintéticos en el campo militar?

9.2. Situación actual de la investigación

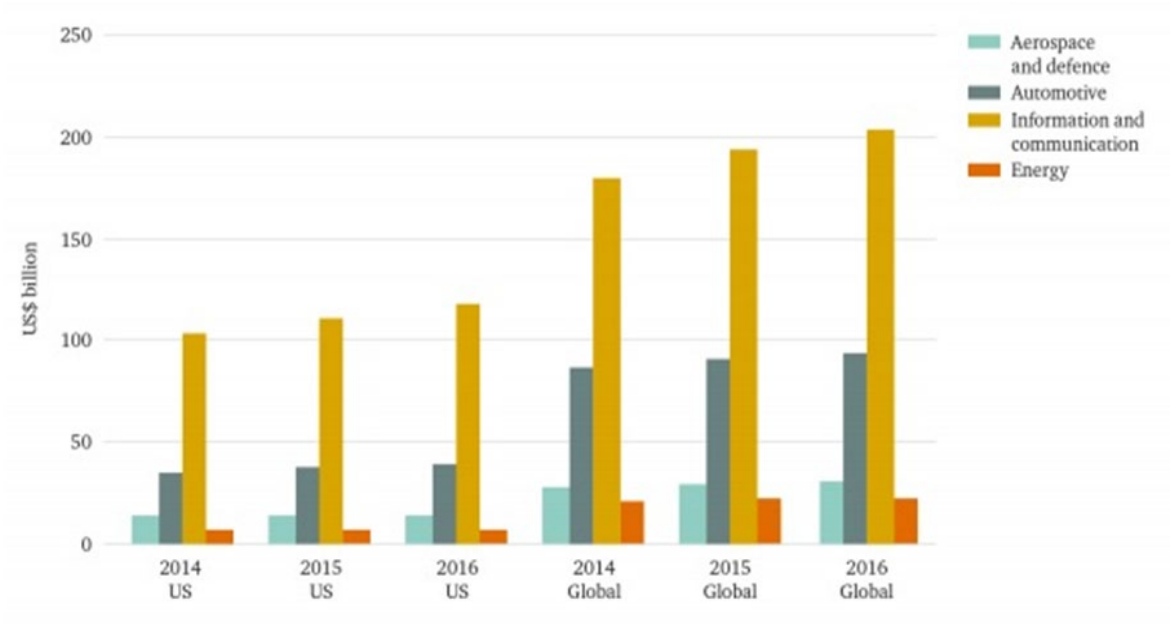
El futuro de la IA en el ámbito militar se encuentra estrechamente ligado a la capacidad de los laboratorios de IA para diseñar sistemas autónomos, es decir, sistemas que puedan desarrollar una capacidad independiente para realizar un razonamiento basado en conocimientos y en la experticia. De momento no hay sistemas autónomos que operen de esta manera y esto se debe principalmente a que los humanos todavía se encuentran detrás de esos robots controlándolos desde una cierta distancia. Esto también ocurre con los drones militares, aunque presentan una ligera sofisticación respecto a los robots militares, ya que poseen cierta autonomía, sin la necesidad de intervención humana, aunque aún requieren de control humano para la ejecución de las misiones que les son encargadas.

La inteligencia humana sigue normalmente la frecuencia percepción-cognición-acción para el procesamiento de información obtenida a través de la percepción de los entornos que le rodean. La IA está desarrollándose de forma similar, tratando de emular estas secuencias de los circuitos cerebrales y trasladándolas a una computadora a través de algoritmos de optimización y verificación. Así pues, el siguiente cuadro ilustra cómo un sistema autónomo con IA procesa y emprende sus acciones (Cummings, 2017: 3):



Fuente: Cummings, 2017: 3.

En los últimos años están invirtiéndose una gran cantidad de recursos para el desarrollo de sistemas autónomos. Hay países como EE. UU., Rusia, China, Japón o Corea del Sur que están experimentado importantes avances en esta materia, aunque, como se ha indicado, no cuentan con una autonomía completa, ya que este fenómeno está directamente relacionado y es dependiente del desarrollo de la IA. Este desarrollo militar de la autonomía se encuentra principalmente con el obstáculo de los altos costes en la producción e investigación de estos sistemas, así como con problemas técnicos imprevistos (Thompson, 2013). Otro aspecto a tener en cuenta es la introducción de los UAV en los ejércitos, lo que supone un cambio cultural en el campo de batalla, pues de momento en el ejército solo se aceptan estas tecnologías si asumen un rol de apoyo y no de un total desplazamiento de aquellas acciones bélicas que desarrollan los humanos, ya que cuentan con un alto prestigio (Hendrix, 2015). Además, los esfuerzos destinados al desarrollo de los sistemas autónomos en el campo militar no ha sido del mismo nivel que el de los drones comerciales o el de los coches sin conductor. El siguiente gráfico del *Industrial Research Institute* (2016) arroja datos de I+D del sector aeroespacial y de defensa, del automovilístico, de la energía y el de la información y comunicación.



Fuente: Cummings, 2017: 9.

El sector aeroespacial y de la defensa es el encargado del desarrollo e investigaciones de los sistemas autónomos para las Fuerzas Armadas, aunque, como puede comprobarse en el anterior gráfico, el gasto en I+D es muy inferior que en los otros sectores. Los ejércitos tienen especial interés en el desarrollo de estos sistemas autónomos, aunque, debido a los elevados costes, priorizan las inversiones y la investigaciones en los medios tradicionales de combate como los aviones. En definitiva, solo un pequeño porcentaje de los presupuestos de defensa se invierte en el desarrollo de sistemas militares autónomos.

En materia de competitividad e investigación en IA, el campo militar es uno de los que menos interés provoca en los especialistas de robótica. Un claro ejemplo puede observarse en el caso de *Boston Dynamics*, una empresa estadounidense líder en robótica militar que fue comprada por *Google* para el desarrollo de robots domésticos y que recientemente ha despertado el interés en Toyota para su compra y posterior investigación en vehículos sin conductor (Lunden, 2017). Con un mercado altamente competitivo, los especialistas en robótica prefieren dedicar sus investigaciones a otros sectores que cuentan con una mayor financiación, por lo que el sector aeroespacial y de defensa ha pasado a convertirse en una segunda o tercera opción en el interés de estos especialistas. Como señala Cummings: «es

más probable que el estadounidense promedio tengan un vehículo sin conductor antes que los soldados en el campo de batalla, y que los terroristas puedan ser potencialmente capaces de comprar drones por internet con tanta o mayor capacidad que aquellos disponibles para los militares» (Cummings, 2017: 11).

Debido a las limitaciones y dificultades de los laboratorios para el desarrollo de intelectos sintéticos capaces de desenvolverse como los humanos en entornos complejos, la llegada de los robots autónomos en el ámbito militar se demorará un tiempo. Además, fruto de las complejidades del ámbito militar, es razonable, y fácilmente entendible, que no se quieran incorporar a las prácticas militares armas autónomas. Aún queda mucho camino por recorrer, aunque, partiendo de la heurística del temor de Jonas (1995), se torna necesario reflexionar con anticipación desde el reconocimiento y la incorporación del principio de responsabilidad en la práctica tecnológica en el marco de la cooperación cívico-militar.

9.3. Un contexto de enjambres

Para iniciar una reflexión ética sobre el campo de la robótica militar que incorpora IA es fundamental partir del reconocimiento de una serie de premisas. Mark Coeckelbergh (2011) identifica una serie de aspectos esenciales para dicho reconocimiento:

- La tecnología contiene una serie de ideas que son proyectadas en su diseño.
- La discusión sobre la robótica, considerada como un medio, también significa de algún modo la consideración de los fines de las acciones militares.
- Las tecnologías son condiciones de posibilidad para cambiar las cosas.
- Los robots militares no deben ser necesariamente maliciosos.
- El despliegue de la robótica militar se da lugar en un enjambre de relaciones y responsabilidades.

La actividad tecnológica se encuentra inmersa en contextos políticos que proyectan sobre los diseños determinadas ideas que luego tienen un efecto perceptible cuando se despliegan. Por ello, los robots, considerados como medios para la consecución de un fin, empujan a situar a los fines también en el centro de la discusión. La discusión sobre los medios es también una discusión sobre los fines, pues los medios tienen de algún modo el poder de influir tales fines. Coeckelbergh ejemplifica con claridad cómo los medios pueden influir sobre los fines, que en este caso serían militares. La pólvora y la tecnología representan dos medios que se encontraron estrechamente vinculados con fines políticos y militares y sirvieron para dar forma al concepto de guerra del siglo XX y a la política internacional. La pólvora permitió la conquista de tierras lejanas y las bombas atómicas crearon símbolos de poder a nivel internacional durante la Guerra Fría (Coeckelbergh, 2011: 272). Así pues, estas tecnologías no representaron simples medios para un fin, sino que motivaron política y militarmente para condicionar los fines de sus estrategias.

Las tecnologías no son axiológicamente neutrales, pues encarnan una serie de ideas que se proyectan en su diseño y además se encuentran inmersas en determinados contextos sociales. En ese sentido, pensadores como P. W. Singer (2009a; 2009b) y Ronald C. Arkin (2008) consideran que los robots militares cambiarán la forma de concebir la guerra para siempre, pues sitúan a las tecnologías como las condiciones de posibilidad para cambiar las cosas.

A pesar de la creencia en la imagen de la robótica militar como intelectos sintéticos que actúan al margen del control humano directo, cabe mencionar que por el momento no hay robots militares autónomos que estén siendo empleados por parte de los ejércitos. Para Coeckelbergh esa imagen ha servido para alimentar la literatura de figuras como Robert Sparrow (2007; 2009) y Arkin (2008), creando así una imagen maliciosa de los robots. Esto se debe a que parten del reconocimiento de los mismos como entidades autónomas, cuando en realidad no han experimentado un desarrollo lo suficientemente avanzado debido a las limitaciones que presentan las investigaciones en IA.

Este hecho, que implica cierta lentitud en el desarrollo de la robótica militar autónoma, conduce a una reorientación de la reflexión ética sobre la robótica en términos contextuales y no tanto individuales. Es decir, el despliegue de la actividad robótica se da en términos de una red de relaciones, o como Coeckelbergh señala, en medio de un «enjambre» (2011: 273). Considerar la actividad robótica en el centro de un enjambre de relaciones conduce a replantear la metodología que trata de arrojar luz reflexiva, en términos éticos, sobre dicha actividad. El filósofo belga afirma que «un enjambre consiste en partes «independientes» (no hay un controlador central), pero al igual que un enjambre de abejas o pájaros, todos los nodos están vinculados a todos los demás, y de esta manera, el conjunto puede actuar al unísono, autoorganizándose» (Coeckelbergh, 2011: 273). En los enjambres se produce una descentralización del poder, ya que la agencia es distribuida, colectiva y emergentemente. No obstante, la descentralización no tiene siempre por qué fijar una determinada organización, ya que no existe una conexión directa entre enjambre y organización social.

La descentralización supone un reconocimiento de la responsabilidad desde una dimensión más amplia y holística, ya que la agencia se disuelve en el entramado de relaciones morales. El enjambre de relaciones difícilmente puede reducirse a la decisión de una única entidad, por lo que «tendríamos que ir más allá del análisis de la distribución de responsabilidades entre un solo actor humano y un único artefacto (por ejemplo un robot)» (Coeckelbergh, 2011: 274). La amplitud de la red y sus circunstancias van más allá de la simpleza de identificar la exigencia de responsabilidad en un humano o en un artefacto en concreto. La actividad tecnológica se distribuye entre una diversidad de actores, tanto humanos como no humanos, y por lo tanto la actividad no puede controlarse centralmente ni monopolizada. Esto último podría contextualizarse desde el MIAR y establecer un paralelismo entre las esferas de la cuádruple hélice y la diversidad de actores que participan en la actividad tecnológica.

Como vemos, el MIAR es relevante para impulsar la IAR en este contexto, pues la reflexión sobre el fenómeno de los robots militares debe situarse en un marco caracterizado por la amplitud y el reparto de responsabilidad que puede compararse con el proceso de hibridación entre las esferas. Por ello la discusión ética en el ámbito de la robótica militar

no puede limitarse a la autonomía de los sistemas militares, sino que más bien tiene que reconocer que el despliegue de la acción robótica no se restringe a una sola entidad. Como ya se ha indicado, esto se debe a que forma parte de un complejo de redes o un enjambre, como lo llama Coeckelbergh, donde la responsabilidad se distribuye entre varias esferas, tanto humanas como no humanas.

9.4. Controversias éticas en el despliegue de la tecnología militar

La tecnología de la IA ha llegado para quedarse en el campo militar y existen numerosos ejemplos en la actualidad, pues hay varias posibilidades de aplicación de la IA. De este asunto ha sido muy consciente el exsecretario de Estado de Defensa estadounidense James Mattis, que, en una visita a *Amazon*, *Google* y otras compañías de Silicon Valley, se comprometió a seguir desarrollando la IA en el campo militar (Metz, 2018). La IA puede utilizarse en sistemas de entrenamiento proporcionando enemigos impredecibles, el aumento de tropas, como por ejemplo con el *Big Dog*, la automatización del combate, la optimización en la identificación de objetivos, etc.

Estos casos ponen de relieve la existencia de un importante impulso militar tras la IA, orientado hacia la autonomía de los robots, como el caso del CIWS (siglas en inglés de *Close-in Weapon System*, traducible como «sistema de armamento de proximidad»), que podría utilizarse con otros mecanismos y otros fines, o los drones, surgiendo así cuestiones éticas que tendrían que abordarse. En lo referente a los cuestionamientos éticos, es importante destacar un informe elaborado por Patrick Lin, George Bekey y Keith Abney (2008) para el Departamento de la Marina de EE. UU., donde se plantean las siguientes preguntas:

- ¿Podrán los robots autónomos seguir las pautas establecidas por las leyes de guerra y las reglas de enfrentamiento, tal y como se especifican en las convenciones de Ginebra?

- ¿Sabrán los robots diferenciar entre el personal civil y el militar?

- ¿Reconocerán a un soldado herido y se abstendrán de disparar?

Ronald Arkin, profesor del Instituto de Tecnología de Georgia, aborda algunas de estas preguntas a partir de estudios técnicos en el informe *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative / Reactive Robot Architecture* (2007).

Se está produciendo un importante cambio de perspectiva en la manera de concebir el ámbito militar y de seguridad y en la necesidad de reflexionar desde la ética, teniendo siempre en cuenta el principio de responsabilidad con la humanidad y con el futuro. Si verdaderamente existe una preocupación por la garantía de los derechos y la dignidad de los seres humanos, como dicta el derecho internacional humanitario, es importante la incorporación de responsabilidad en las políticas de defensa, pensando alternativas políticas y jurídicas desde el marco de una ciencia cívica que promueva una IAR. La Comisión Mundial de Ética del Conocimiento Científico y la Tecnología (COMEST por sus siglas en inglés) ha reafirmado su compromiso con el mandato de la UNESCO para promover y construir la paz, pues el objetivo principal es partir de análisis críticos y éticos sobre el uso de las tecnologías robóticas.

La COMEST reconoce que el uso de drones presenta un nuevo abanico de desafíos éticos, ya que se sitúa en un terreno de completa desconexión física entre el operador y el dron, lo que supondría que en ocasiones su actividad se confundiera con la de un videojuego (UNESCO, 2017: 22). La ONU ha mostrado su preocupación por el nuevo escenario al que conduce la tecnología en el campo militar, pues ha surgido la inquietud por la efectividad que puede tener el derecho internacional humanitario debido al progreso de la IA y al uso de tecnologías nunca antes conocidas. Una primera forma de asumir responsabilidad para estar a la altura de los nuevos tiempos consiste en la revisión del derecho internacional humanitario a la luz del fenómeno de los intelectos sintéticos empleados en el campo militar.

Surgen importantes debates frente a este fenómeno, por ejemplo, en lo relativo al reconocimiento de objetivos. Ante la pregunta planteada acerca de cómo sería posible que un dron cumpliera con el mandato de la Convención de Ginebra de 1949 e hiciera una

distinción entre un militar y un civil o personal sanitario, se plantean dudas. El reconocimiento de un objetivo no solo se reduce a factores de reconocimiento visual. El importante progreso de reconocimiento visual que ha experimentado la IA en la última década es asombroso. Sin embargo, en la toma de decisiones militares, no solo se tienen en cuenta factores de reconocimiento visual, sino que existen una serie de condicionantes mucho más amplios que, al menos por el momento, los drones no están preparados para manejar.

La COMEST afirma que la responsabilidad en el uso de cualquier sistema robótico recae sobre el operador, es decir, sobre el ser humano que controla ese sistema. En ese sentido, es importante no eludir la responsabilidad, pues es el ser humano quien decide utilizar este tipo de artefactos en el campo de batalla y asume todas las posibles consecuencias que ello implica. Los ejércitos están obligados a asumir su responsabilidad a este respecto y a buscar las estrategias necesarias para que el uso de la robótica no viole lo establecido en el terreno legal y tampoco en el ético.

9.5. Dronética: una mirada ética sobre el despliegue de los drones

El uso de los drones por parte de diversas potencias está cada vez más extendido, por ello la ética aplicada a este campo justifica su pertinencia y demanda. Además, el uso de los UAV despierta cada vez un mayor interés en la ciudadanía y pueden establecerse diferencias entre los argumentos presentados en torno al empleo de este tipo de vehículos. Los drones han generado una crisis en el *ethos* militar, en palabras de Chamayou (2013), pues la invisibilidad de los ataques le confiere a esta tecnología una forma de invulnerabilidad y por lo tanto la posibilidad de una garantía de impunidad. Existe por lo tanto un riesgo moral debido al distanciamiento del campo de batalla entre el propietario del dron y las víctimas. Chamayou se refiere a crisis del *ethos* militar en los siguientes términos:

> El problema acaso esté en que, visto con el prisma de los valores tradicionales, matar con el dron, aplastar al enemigo sin arriesgar el cuero, siempre apareció como el *summum* de la cobardía y el deshonor. La discordancia entre la realidad técnica de la conducción de la guerra y su resto ideológico conjuga una poderosa contradicción, que incluye al personal de las fuerzas armadas. Y suscita la colisión, en el seno del personal, entre las nuevas armas y los antiguos marcos, quizás vetustos pero aún potentes; se trata de la *crisis en el ethos militar* (2016: 95).

A propósito de esa crisis, no se han hecho esperar algunas reacciones en contra de los UAV, ya que han surgido diversas posturas sostenidas sobre fundamentos éticos y políticos. Existen una variedad de principios y valores éticos que sirven para sustentar las visiones sobre el uso de los drones en el mundo. Entre esa variedad de opiniones pueden encontrarse razones relativas a la erosión de las normas, al impacto provocado sobre las víctimas, aspectos morales de la toma de decisiones individuales en el uso de estas tecnologías, en el cambio cultural de la relaciones entre adversarios, etc. Mary Manjikian establece una clasificación de los cinco argumentos que sostienen las posturas en contra del uso de los drones de la siguiente manera:

Argumento	Teoría ética	A favor	Preguntas clave
Las especificaciones tecnológicas	Utilitarismo y consecuencialismo	Los pacifistas: • *Society of Friends Mennonite Church USA United Brethren* • *Human Rights Watch* • *Harvard Law School's International Human Rights Clinic (IHRC)*	¿Los nuevos y únicos aspectos de estas armas –es decir, una mayor precisión, capacidad de discriminación; velocidad, etc.– hacen más o menos morales/éticos a las generaciones anteriores a estas armas? ¿Estas armas pueden salvar vidas? Y si es así ¿de quién?
Los argumentos de identidad	Ética de la virtud	Campaña para detener robots asesinos	¿Sería honorable un guerrero –o nación– que utiliza estas armas? ¿Qué clase de personas o nación lo hace?

Relación con el adversario	Ética levinasiana y ontología	• *UN Convention on Certain Conventional Weapons* • *Amnesty International*	¿Es el uso de esta arma una forma apropiada de tratar a mis enemigos?
Efectos sobre la comunidad internacional	La legitimidad de las normas	• Parlamento europeo • *Independent and Peaceful Australia Network*	¿Qué clase de ejemplo se da en la comunidad internacional? ¿Estamos cambiando las normas/leyes de la guerra?
Relación con doctrinas específicas, estrategias, tácticas	Guerra justa y ética situacional	• Stephen Hawking, Elon Musk y Steve Wozniak, cofundador de *Apple* • Electronic Frontier Foundation • Proyecto secreto del gobierno sobre la asociación de científicos americanos	¿De qué manera el uso de estas armas afecta a nuestra capacidad para compartir una guerra justa? ¿La posesión de estos complementos armamentísticos nos hace más propensos a adoptar ciertas políticas y doctrinas? ¿Cuáles son los efectos de los UAV de vigilancia sobre los derechos de la ciudadanía?

Fuente: Manjikian, 2017: 4-5.

Como pone de relieve Manjikian, no existe una postura única en contra del empleo de los drones, pues las diversas teorías éticas han sabido facilitar un sustento argumentativo para proponer diferentes miradas. En ese sentido, es fundamental promover una ética aplicada que reconozca las tradiciones éticas como valiosas herramientas desde las que llevar a cabo un diálogo abierto y participativo entre los sectores implicados. El impulso de una IAR en este ámbito implica el reconocimiento de las diversas tradiciones éticas y la riqueza que el diálogo abierto tiene para la resolución de problemáticas con altura de miras, desde la perspectiva de la ciencia cívica. La tipología de argumentos que ofrece Manjikian sirve precisamente para poner en práctica ese ejercicio deliberativo que promueve la IAR. De esa forma se estarían sentado las bases para una dronética surgida de un trabajo democrático en el que pueda considerarse al conocimiento generado en esta área como un

recurso público que presenta una profunda complejidad y que por lo tanto no puede ser objeto exclusivo de ninguna de las esferas de la triple hélice (Estado, mercado y academia), sino que debe incorporar una mirada más amplia y enriquecedora para poder deliberar acerca de las diversas tradiciones éticas que arrojan luz sobre el fenómeno de los drones.

Debido al espíritu que sustenta los sistemas democráticos es difícil concebir el despliegue de los UAV sin previamente contextualizarlos en los valores de una sociedad. Teniendo en cuenta la tradición democrática que con el paso de las décadas se ha plasmado en una gran diversidad de manifestaciones humanas, es importante destacar que este sentido democrático también se encuentra proyectado en la doctrina militar. En última instancia, las acciones emprendidas en el ámbito militar son una cuestión de valores. En ese sentido, las reglas militares promueven también la responsabilidad en la toma de decisiones y resolución de conflictos. Por ello, las normas y principios que fundamentan la ética militar exigen en la actualidad un ejercicio de contextualización en función de las capacidades tecnológicas de la IA. En ocasiones, las Fuerzas Armadas han primado la eficiencia a corto plazo frente al sentido democrático, un hecho que no debería darse, pues la ética debe incorporarse en el diseño de toda forma tecnológica avanzada que comprometa la vida de la ciudadanía (Swarte, Boufous y Escalle, 2019: 4).

A pesar de que la problemática de la imputación de los actos de responsabilidad en el terreno de las tecnologías autónomas, y con mayor razón en el de los UAV destinados a la guerra, es un asunto de gran importancia para discutir en el seno de la ética, es necesario también considerar otra dimensión de la ética y su vínculo con las relaciones políticas que subyacen en este campo.

9.6. Diseño responsable de los UAV

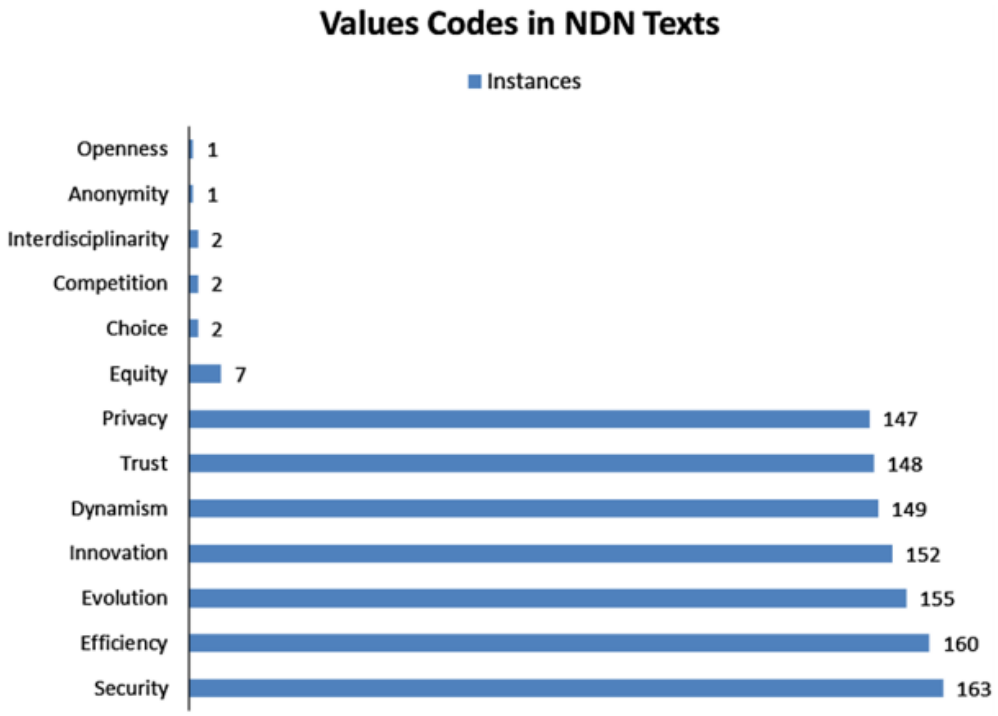
El despliegue de la actividad de los UAV pone de relieve importantes problemáticas en materia de ética y seguridad debido a los impactos que tienen para la seguridad pública. Daniel Weld y Oren Etzioni (2009) analizan algunos de los desafíos que se derivan de los sistemas autónomos y la garantía de seguridad que presentan para que puedan aceptarse por

la sociedad. Por ello se torna fundamental contar con la información suficiente para deliberar acerca de las dimensiones éticas implícitas en el diseño de estos sistemas. A pesar del carácter autónomo de esta tecnología aérea no tripulada, hay que reconocer que son los seres humanos quienes verdaderamente tienen el poder de controlar el diseño y los valores proyectados. Por lo tanto, es fundamental la incorporación de criterios de responsabilidad ética en el ejercicio profesional vinculado al diseño de los UAV, con el objetivo de alcanzar un desempeño de los algoritmos respetuoso con las normas de la sociedad y sus códigos éticos en el contexto de laboratorios abiertos sobre ciencia cívica.

Normalmente los tecnólogos no abordan cuestiones relativas a los valores proyectados, sino que se sitúan en la senda de los aspectos estrictamente técnicos (Shilton, 2014). Frente a esta situación la eticista del ámbito informático Deborah G. Johnson (2007) aboga por una intervención directa en el diseño centrada en una intervención anticipada en el desarrollo de la tecnología para influir en la construcción del diseño proyectado por los tecnólogos. Además, Johnson reconoce la dificultad de realizar este ejercicio de anticipación-reflexión (2011). Sin embargo, esta dificultad podría superarse con una visión deliberada e inclusiva del diseño, donde la participación vaya más allá de los expertos de las esferas del Estado, el mercado o la academia, y se encamine hacia el reconocimiento del conocimiento como un recurso de interés público. La propuesta de Johnson pasa necesariamente por asumir que la actividad tecnológica no es axiológicamente neutra y se encuentra conectada con valores y creencias sociales. Para entender con mayor claridad la necesidad de una ética de la anticipación en el diseño de los UAV es importante rescatar un análisis que Katie Shilton (2015) utiliza sobre la influencia de los valores en las *Named Data Networking* (NDN). A pesar de que el diseño de los UAV se diferencia de las redes de datos de Internet, es necesario establecer un paralelismo con el análisis de Shilton. Pues existen algunas cuestiones de profundidad en lo relativo a los valores y su influencia en el diseño a través de la selección y el posterior impacto social, debido al predominio de consideraciones explícitas en parte de la práctica del diseño tecnológico (Friedman, Kahn y Borning, 2006; Knobel y Bowker, 2011). Dada esta complejidad, la propuesta de los laboratorios abiertos permite someter a un ejercicio deliberativo el diseño de los UAV,

priorizando aquellos aspectos que sirvan para el cultivo de habilidades cívicas y el fortalecimiento de la democracia.

El siguiente cuadro refleja cómo los valores contribuyen en el diseño de la tecnología para dar forma a su uso e influencia sobre los contextos sociales en los que despliegan su actividad, aunque no solo tienen un impacto social, sino también individual, pues inciden en los patrones a partir de los cuales los individuos evalúan sus comportamientos, formulan sus juicios y establecen sus relaciones sociales (Queraltó, 2003; A. Domingo Moratalla, 2013):



Fuente: Shilton, 2015: 8.

El resultado del programa *Future Internet Architecture* (FIA) financiado por el *National Science Foundation* (NSF) destaca la importancia de los valores en el diseño tecnológico. Pueden agruparse los valores más destacables de la siguiente manera: los que responden a las presiones técnicas y oportunidades; los que se centran en las libertades

personales; y las influencias por un interés en las preocupaciones colectivas de una información compartida (Shilton, 2015: 8).

Jörg P. Müller (1996) señala que en la investigación del diseño de los intelectos sintéticos convergen la teoría de control, la psicología cognitiva y la teoría clásica de la planificación de la IA. Debido a esta convergencia se estrechan lazos entre los postulados informáticos, filosóficos y religiosos para poder esclarecer y definir qué es lo moralmente correcto en el diseño. Este ejercicio de convergencia y su posterior vínculo interdisciplinario contribuye al interés creciente por el estudio de la ética en el ámbito tecnológico, pues el factor ético es esencial y representa un marco de referencia para aclarar el grado de autonomía de los UAV. No obstante, más allá del diseño interior de los algoritmos y de los aspectos estrictamente técnicos que configuran el despliegue de la actividad de los vehículos aéreos autónomos, es también importante considerar las relaciones políticas implícitas en el esquema de este tipo de máquinas.

A diferencia de lo que sostiene Seth D. Baum (2017), para quien hay que rechazar la inclusión de la sociedad por carecer de una visión ética compartida, es fundamental realizar un ejercicio inclusivo e interactivo entre las diversas esferas que configuran la quíntuple hélice. Además, los cuatro pilares fundamentales sobre los que se sostiene el modelo de RRI, a saber, anticipación, reflexión, inclusión y sensibilidad, juegan un rol decisivo en la promoción del diseño ético de los UAV. Del mismo modo, una evaluación de carácter constructivo, que introduzca dinámicas de acción-participación en el diseño, puede enriquecer el trabajo de los tecnólogos al contar con mayores conocimientos como producto del encuentro de perspectivas.

Además de la inclusión de los grupos de interés en el diseño, la ética de la anticipación también se entiende como una práctica ética emergente que pretende dar a conocer aquellas problemáticas éticas que podrían derivarse del objeto de la discusión. Una combinación de anticipación y reflexión puede contribuir de forma positiva a poner de relieve cuáles son los valores que se proyectan sobre el diseño y que posteriormente configuran un sentido.

El diseño de la tecnología es muy importante, por lo que se torna necesario someter a una profunda reflexión colectiva aquellos valores que se proyectan y posteriormente dan forma a los sentidos que se insertan en los contextos sociales. Este ejercicio de inclusión y anticipación en el diseño de vehículos aéreos no tripulados, introduciendo criterios éticos, podría realizarse en el marco de laboratorios abiertos sobre ciencia cívica. En definitiva, para avanzar en el diseño ético de los drones debe establecerse una convergencia con criterios éticos de responsabilidad que respondan a las necesidades de los individuos y las comunidades en un sentido cívico y democrático.

9.7. Desarrollo sostenible y tecnología militar

El desarrollo sostenible tiene una relación directa con las políticas de defensa de los Estados. Las Fuerzas Armadas representan una figura clave en el esfuerzo que emprende la sociedad para garantizar un futuro de sostenibilidad. Para impulsar el compromiso social e institucional con los ODS es fundamental contar con el apoyo de las Fuerzas Armadas. Los ODS surgen como una respuesta de la comunidad internacional ante los graves problemas y desafíos que debe enfrentar la humanidad en las próximas décadas, incorporando en las prácticas el compromiso de todas las instituciones, ya sean de carácter privado o público, así como a la sociedad en general.

El espíritu del compromiso sustentable surgió con la idea de sostenibilidad impulsada en 1972 en la Conferencia del Medio Humano de Estocolmo, posteriormente asumida por la ONU en 1983 a través de la Comisión Mundial sobre el Medio Ambiente y el Desarrollo que difundiría años después en 1987 el texto *Nuestro futuro común* y que finalmente se impondría en 1992 en la Cumbre de la Tierra de Río de Janeiro. Este espíritu debe estar presente en todas las actividades militares y en las instituciones de defensa de los países como una política oficial que señale el camino a seguir de las Fuerzas Armadas frente a los desafíos de este siglo XXI. Ante esa exigencia, la IAR puede contribuir de diversas formas para que los organismos militares orienten sus proyectos hacía un compromiso con los ODS, haciendo de este mundo un lugar más habitable.

En relación con esta exigencia entra en escena el concepto de desarrollo sostenible militar, entendido según Luis B. Olivares Dysli como «un cambio fundamental en la forma como se implemente el desarrollo de las Fuerzas Armadas modernas con capacidades y potencial eficiente para enfrentar los desafíos del presentes y del futuro inmediato« (2014: 4). Podría identificarse que junto a este concepto de sostenibilidad se encuentra presente el compromiso con los desafíos del presente y del futuro de la humanidad y la biosfera. Olivares Dysli entiende que todas aquellas operaciones militares y cualquier actividad que implique el uso del potencial bélico, incluyendo personal humano y tecnología, en acciones, ya sean de paz o conflicto, deben incorporar recursos sostenibles y renovables para tratar de disminuir el volumen de consumo, mejorar la seguridad, y asumir una preocupación por los ecosistemas en los que despliegan su actividad, asumiendo de ese modo un grado de responsabilidad con la humanidad y sus instituciones (Olivares Dysli, 2014: 4-5).

Para que las instituciones militares puedan realizar sus actividades asumiendo un compromiso con los ODS es fundamental impulsar una serie de mecanismos que posibiliten tal realidad. Entre esos mecanismos pueden encontrarse la reducción de costes en gasto militar a través de la optimización de los recursos, el uso de recursos renovables, un mejor conocimiento de los escenarios para medir el impacto de factores externos, nuevas estrategias operativas, etc. En esa necesidad de mecanismos que faciliten el desarrollo de un conjunto de actividades comprometidas con los ODS, la IA supone una oportunidad tecnológica desde la que abrir un nuevo horizonte de posibilidades orientadas a la innovación. No obstante, la IA aplicada al campo militar debe incorporar el principio de responsabilidad en su ejercicio para que la generación de conocimientos innovadores adquiera ese compromiso sostenible que tanto reivindica Olivares Dysli. Ninguna organización de defensa debería impulsar el desarrollo de sus fuerzas sin antes contemplar las exigencias de los contextos nacionales e internacionales en materia de ODS.

Las Fuerzas Armadas son también un instrumento del Estado para ocuparse de aquellos conflictos que existen con otros países en materia de seguridad externa, aunque también lo pueden ser asuntos que tienen que ver con aspectos de orden y control democrático. Todo

lo anterior pone de relieve las nuevas situaciones a las que tienen que dar respuesta las Fuerzas Armadas y el papel que podrían tener los avances tecnológicos. El pasado represivo y violento que caracterizó la función de los ejércitos de muchos países durante el siglo XX está dejándose a un lado gracias a los procesos democráticos experimentados en las últimas décadas. La importancia del papel de las Fuerzas Armadas debe someterse a un profundo debate en el contexto de la sociedad del conocimiento. El despliegue de sus actividades no puede concebirse al margen de los retos tecnológicos y concretamente de la IA. Los intelectos sintéticos pueden hacer importantes aportaciones al trabajo de los ejércitos para el cultivo de habilidades cívicas y el fortalecimiento de la democracia frente a las nuevas amenazas emergentes y los nuevos escenarios.

En el contexto tecnológico las actividades militares pueden aparecer como respuestas efectivas ante los desafíos que plantean los ODS. Las funciones de las Fuerzas Armadas se sitúan en la estela de un enfoque interagencial (Proaño Cortez, 2008) donde su papel puede ser muy relevante para responder a los desafíos que la humanidad se ha fijado en la Agenda 2030. Esta premisa parte del reconocimiento de una consideración fundamental, a saber, la incidencia positiva que las instituciones de defensa pueden tener en la construcción de una política democrática. Las Fuerzas Armadas juegan un papel activo en el planteamiento de políticas que giran en torno a la cuestión de la seguridad, que en términos generales tiene una incidencia directa sobre las agendas públicas que configuran los sistemas democráticos.

El fortalecimiento de la democracia puede venir por una disminución del riesgo del autoritarismo, la falta de transparencia, las misiones destinadas a la defensa de los intereses económicos de bloques imperialistas, etc. En ese sentido, el desarrollo de los medios tecnológicos puede representar un incentivo para la formación de los profesionales de las instituciones y cuerpos de defensa, así como para la formación en dinámicas participativas, y contribuir considerablemente a un desarrollo los valores democráticos en las filas de los ejércitos. Por lo tanto, la IA contribuye de manera positiva a esta tarea formativa y democratizadora, ya que facilita herramientas tecnológicas que permiten un mejor conocimiento sobre el mundo y los desafíos que enfrenta la humanidad, para que las Fuerzas Armadas asuman así un compromiso activo en la defensa de la democracia.

La incorporación de responsabilidad en el despliegue de los sistemas artificiales del contexto militar y de seguridad representa un mecanismo de asunción de compromiso con los derechos humanos y los ODS que puede promoverse a través de la tecnología más avanzada. En este escenario de compromiso responsable potenciar una IAR es fundamental como medio para iniciar un ciclo que permita reorientar la actividad de defensa hacia nuevos horizontes. La IAR podría tener un impacto significativo en áreas como las siguientes:

- Lucha contra la destrucción del medio ambiente.

Las Fuerzas Armadas no pueden situarse al margen del escenario mundial, el cual enfrenta importantes retos como consecuencia de las diversas formas que adopta la degradación del medio ambiente: cambio climático, deforestación, contaminación del suelo, el agua y el aire, desertificación, falta de protección de los espacios naturales y protegidos, residuos contaminantes y peligrosos, etc. En ese sentido, es fundamental que las instituciones de defensa se sumen a la exigente tarea de combatir la degradación del medio ambiente mediante la adopción de estrategias innovadoras: eficiencia energética, mantenimiento innovador de los edificios y campos de ejercicio, control de los residuos y su reutilización, planificación ecológica, uso de energías renovables no convencionales, etc. La IA puede promover estrategias y fomentar el sentido de responsabilidad con la biosfera.

- Mejora de la comunicación y la calidad de los datos.

El volumen de datos creciente a los que tienen acceso las instituciones de defensa es cada vez mayor, por lo que el uso de herramientas como la IA puede facilitar el ejercicio de las actividades profesionales de este sector. Es importante facilitar el soporte para un tratamiento óptimo de los datos mediante una filtración y procesamiento de los mismos. La selección de las fuentes de datos debe ser fiable, y ante esa exigencia la IA también puede ser una útil herramienta. Además, estos datos pueden agilizar las telecomunicaciones entre diversas instancias y contribuir a generar conocimientos innovadores.

- Mecanismos de transparencia para fortalecer el acceso de la ciudadanía a los datos.

Katherine Dixon, directora del Programa de Seguridad y Defensa de Transparencia Internacional, alerta de la necesidad de valorar los altos niveles de corrupción en el sector militar, con el objetivo de reducirlos drásticamente. La corrupción es un grave obstáculo para la consecución de los logros que plantea la Agenda 2030. En el ámbito militar es necesaria la introducción de mecanismos de transparencia para fortalecer un control más estricto de los activos. No es posible asumir un compromiso con los ODS si antes no se depuran las responsabilidades en instituciones tan opacas como las de defensa, así como si no facilitan los cauces para la transparencia y el control de las cuentas. En ese sentido, una IAR impulsada desde el MIAR puede brindar un escenario favorable para formular interesantes iniciativas orientadas al control de los activos por medio de una intervención de la cuádruple hélice y también mediante la generación de plataformas digitales de control y transparencia (Belda, 2018).

- Mejor conocimiento para el desempeño del ejercicio profesional.

Un mejor y mayor acceso a los datos y el cultivo de nuevas habilidades como resultado de la apertura de posibilidades que ofrecen los sistemas artificiales permitirán fortalecer el ejercicio profesional en las próximas décadas. El manejo de datos de una mayor calidad puede favorecer las técnicas empleadas en el ejercicio de las Fuerzas Armadas y esto a su vez asegurar mejor sus objetivos. Esto permitirá un mejor conocimiento de las problemáticas a las que se enfrentan los ejércitos para contribuir a la paz y a la seguridad a los territorios.

9.8. Tecnologías humanitarias y responsabilidad

Existen ayudas humanitarias que no necesariamente deben realizarse por humanos. Es posible promover un uso cívico de los drones para que su finalidad no se destine exclusivamente a fines bélicos, sino que también sirvan como medio para promover los ODS referidos a la paz y el bienestar de la ciudadanía. El poder de las tecnologías puede orientarse de manera segura y efectiva en términos cívicos que respondan a criterios

humanitarios. *Sense Fl*y, la empresa de drones del grupo Parrot, destaca un conjunto aplicaciones humanitarias de los UAV:

- Respuestas de emergencias ante los desastres.
- Planificación urbana y gestión del suelo.
- Distribución y planificación de la ayuda.
- Identificación de poblaciones en riesgo.
- Monitoreo del territorio.
- Mapeo comunitario.
- Creación de capacidades en las comunidades locales.
- Administración del agua.
- Gestión de la propiedad terrestre.
- Desminado.

Estas aplicaciones representan casos evidentes de compromiso con los ODS que se pueden desarrollar mediante sistemas de IA integrados en drones. Entre los ODS con los que se asume un compromiso pueden encontrarse el fin de la pobreza, hambre cero, salud y bienestar, agua limpia y saneamiento, trabajo decente y crecimiento económico, industria, innovación e infraestructura, reducción de las desigualdades, ciudades y comunidades sostenibles, acción por el clima, vida de ecosistemas terrestres y paz, justicia e instituciones sólidas.

Swissnex Boston, una organización sin ánimo de lucro suizo-estadounidense, organizó junto a *WeRobotics*, una plataforma en red de carácter multidisciplinar, un evento donde se discutieron asuntos sobre el uso humanitario de los drones. En ese evento se escenificó un encuentro caracterizado por la pluralidad, pues contó con la presencia de empresas tecnológicas del sector privado, instituciones gubernamentales locales, centros académicos como Harvard y el MIT y el Comité Internacional de la Cruz Roja (Luterbacher, 2018). Este evento pone de relieve la importancia de promover una IAR fundamentada en un

modelo de innovación deliberativo, abierto y participativo, en el que se asuma un compromiso con los derechos humanos, los ODS y los límites planetarios. El trabajo participativo realizado en esta actividad, impulsada por *Swissnex Boston* y *WeRobotic*, representa un claro ejemplo acerca de cómo las instituciones de las diversas esferas representadas en la quíntuple hélice pueden fortalecer en el ámbito humanitario, mediante el uso de los drones, una cultura cívica y democrática.

En el terreno humanitario la respuesta más común de los drones se encuentra vinculada al mapeo y monitoreo, a la entrega de suministros y medicamentos, al reconocimiento de áreas de difícil acceso, así como a las tareas de búsqueda y rescate. A pesar del buen uso de estas tecnologías para fines humanitarios, también están surgiendo importantes desafíos éticos y reglamentarios en torno a la gestión del tráfico aéreo y la privacidad de los datos.

Para resolver los conflictos éticos y reglamentarios derivados de los drones humanitarios, *UAViators,* una organización con más de 3300 miembros en más de 120 países, cuya misión es promover el uso seguro, coordinado y efectivo de los drones en contextos humanitarios y de desarrollo, impulsó en el año 2014 el *UAV Code of Conduct* (Humanitarian UAV Network, 2014). Este código tiene como finalidad informar acerca del uso seguro, responsable y efectivo de los drones civiles en contextos humanitarios. En su elaboración participaron más de 60 organizaciones de diversos países en un proceso que duró dos años. En su página web se detallan los siguientes puntos del código de conducta:

1. Priorice la seguridad por encima de todas las demás preocupaciones: los beneficios humanitarios deben superar claramente los riesgos para las personas o propiedades.
2. Identifique la solución más adecuada: solo opere vehículos aéreos no tripulados cuando no haya medios efectivos disponibles y cuando los propósitos humanitarios sean claros, como la evaluación de las necesidades y la respuesta a las mismos. Las misiones de UAV deben ser informadas por profesionales humanitarios y expertos en operaciones de UAV con conocimiento directo del contexto local.
3. Respete los principios humanitarios de humanidad, neutralidad, imparcialidad e independencia: priorice las misiones de UAV en función de las necesidades y vulnerabilidades, asegúrese de que las acciones no sean, y no se perciban, influenciadas

política o económicamente; no discrimine ni haga distinciones por motivos de nacionalidad, raza, género, creencias religiosas, clase u opiniones políticas.

4. No haga daño: evalúe y mitigue las posibles consecuencias no deseadas que las operaciones de UAV puedan tener sobre las comunidades afectadas y la acción humanitaria.

5. Operar con los permisos pertinentes: las operaciones de vehículos aéreos no tripulados deben cumplir con las leyes internacionales y nacionales pertinentes, y los marcos regulatorios aplicables, incluidos aduanas, aviación, responsabilidad y seguros, telecomunicaciones, protección de datos y otros. Cuando no existan leyes nacionales, los operadores deberán adherirse a la Circular 328-AN / 190 RPAS de la OACI con la aprobación de las autoridades nacionales.
6. Comprometerse con las comunidades: el compromiso de la comunidad es importante y obligatorio. El desarrollo de la confianza y la participación de las comunidades locales fomenta la asociación activa, desarrolla las capacidades locales y el liderazgo y aumenta el impacto de su misión. Debe proporcionarse continuamente información a las comunidades sobre la intención y el uso de los UAV. Consulte las pautas de participación comunitaria humanitaria en UAV.
7. Sea responsable: los planes de contingencia siempre deben estar en su lugar para las consecuencias no deseadas. Los equipos de UAV deben asumir la responsabilidad y resolver cualquier problema que implique daños a personas y bienes.
8. Coordine para aumentar la efectividad: busque y establezca contactos con actores y autoridades locales e internacionales relevantes. Los equipos de UAV no deben interferir y siempre deben buscar la complementación entre mecanismos y operaciones formales de coordinación humanitaria.
9. Considere las implicaciones ambientales: los UAV operativos no deben representar un riesgo indebido para el medio ambiente natural y la vida silvestre. Los operadores de vehículos aéreos no tripulados deben asumir la responsabilidad de cualquier impacto ambiental negativo que su misión cause.

10. Sea sensible al conflicto: todas las intervenciones en zonas de conflicto se convierten en parte de la dinámica del conflicto y pueden tener consecuencias muy graves no deseadas, incluida la pérdida de vidas. Debe tenerse precaución extraordinaria al desplegar vehículos aéreos no tripulados en zonas de conflicto. Consulte las pautas de zonas de conflicto de UAV humanitarias.
11. Recopile, use, administre y almacene datos de manera responsable: recopile, almacene, comparta y descarte datos de manera ética utilizando un enfoque basado en las necesidades, aplicando el consentimiento informado cuando sea posible y empleando medidas de mitigación donde no lo sea. Debe evaluarse la posibilidad de que la información ponga en riesgo a individuos o comunidades si se comparte o se pierde, y se deben tomar medidas para mitigar ese riesgo (por ejemplo, limitar o dejar de recopilar o compartir). Consulte las Pautas de ética de datos humanitarios de UAV.
12. Desarrolle asociaciones efectivas en la preparación y para las crisis y en respuesta a ellas: trabaje con grupos que ofrezcan conjuntos de habilidades complementarias (acción humanitaria, operaciones de UAV, contexto local, análisis de datos, comunicaciones) durante, y preferiblemente antes de las crisis. Consulte las pautas de asociaciones efectivas del UAV humanitario.
13. Sea transparente: comparta las actividades de vuelo lo más ampliamente posible, idealmente públicamente, según corresponda al contexto. Transmita lecciones o problemas a las comunidades, autoridades relevantes y organismos de coordinación lo antes posible.
14. Contribuya al aprendizaje: lleve a cabo y comparta cualquier evaluación y revisión posterior a la acción para informar el mejoramiento del uso de UAV para la acción humanitaria.
15. Sea abierto y colaborativo: la coordinación es un proceso de múltiples partes interesadas. Esto significa que las lecciones aprendidas y las mejores prácticas sobre el uso y la coordinación de vehículos aéreos no tripulados en entornos humanitarios deben permanecer abiertas y transparentes junto con los talleres, capacitaciones y simulaciones relacionadas (Humanitarian UAV Network, 2014).

En la actualidad existen diversas experiencias en torno a la aplicación humanitaria de los drones. Benjamin Meiches (2019) destaca una misión de ayuda en la se ha realizado un uso humanitario de los drones. Para Meiches los drones producen modelos de ética y

generosidad que son completamente nuevos e inimaginables, como fruto de las posibilidades que brinda la tecnología (2019: 9). La primera experiencia humanitaria en la que fueron desplegados drones se localizó al este de la República Democrática del Congo. Esta tecnología proporcionó imágenes visuales del conflicto civil de Kinshasa en tiempo real, contribuyendo así a las reformas políticas y la presión diplomática. Dicha experiencia sirvió para implementar los drones en Darfur y Sudán del Sur, pues como argumentan los funcionarios de la ONU, los drones poseen ventajas estrictamente informativas, ya que son usados para la obtención de información relevante en las misiones de vigilancia, la resolución de conflictos y el establecimiento de la paz.

Este tipo de relatos nos lleva a reconsiderar el valor de la relación entre el dron y la ayuda humanitaria a la luz del concepto de «inmunidad humanitaria»(Smirl, 2015). Este tipo de inmunidad permite cambiar las prácticas de la ayuda humanitaria: en primer lugar porque brinda a las organizaciones la capacidad de prestar ayuda sin necesidad de interactuar directamente, aumentando de ese modo la distancia social entre los grupos involucrados; en segundo lugar, según Meiches, la introducción de los drones en situaciones conflictivas rompe con la »unidad fenomenológica», permitiendo que el acto de matar psicológicamente sea difícil, y generando la formación de simpatías, sentimientos y afectos; y en tercer lugar, los drones en ocasiones provocan temor y ansiedad en aquellas poblaciones sometidas que se sienten observadas desde el plano aéreo, pudiendo provocar comportamientos impredecibles en los conflictos que pueden resultar ambiguos. En función de estos aspectos destacables de la inmunidad humanitaria que presentan los drones, Meiches (2019:10) sostiene que esta tecnología es considerada en este terreno como una simple herramienta, un producto de la racionalidad instrumental que ignora cómo pueden modificar las prácticas y generar otras situaciones inesperadas que van más allá de la simple asistencia. No obstante, la experiencia de la República Democrática del Congo no es la única, pues existen también otros casos que sirven para ilustrar cómo los drones pueden contribuir a la ayuda humanitaria: en las inundaciones de Dar es-Salam, en un terremoto en Ecuador, en las tareas de desminación en Bosnia y Herzegovina, en la entrega de material médico en Papúa Nueva Guinea, en la gestión del territorio de campamentos en Haití, etc.

El uso humanitario de los drones representa una propuesta de responsabilidad impulsada desde la IAR, entendiendo que la artificialidad cognitiva puede proporcionar un soporte tecnológico beneficioso para los valores cívicos y democráticos de una sociedad. Además, diversas organizaciones e instituciones han sido conscientes de la necesidad de trabajar de forma colaborativa, entendiendo que existe una necesidad imperiosa por innovar y dar respuestas pragmáticas a las problemáticas que enfrenta el mundo. Antes de finalizar este apartado, es importante mencionar el documento *Drones in Humanitarian Action. A guide to the use of airborne systems in humanitarian crises*, elaborado conjuntamente por *CartONG* y *Zoï Environment Network*, consultores independientes de redes humanitarias de drones, el *Peace Research Institute* de Oslo, la *Federation of Air Traffic Controllers' Association* y la *Fondation Suisse de Déminage* (FSD). Estas organizaciones han sido conscientes de la necesidad que existe a la hora de orientar el uso y diseño de estas tecnologías para proporcionar soluciones reales a la ciudadanía, cultivando así habilidades cívicas y democráticas. Este es un claro ejemplo de humanismo tecnológico pues se parte de una concepción beneficiosa dentro de unos parámetros cívicos para impulsar un tipo de tecnología responsable con la humanidad fruto de un trabajo participativo que ha sabido reconocer que sin una diversidad de agentes y perspectivas resulta verdaderamente difícil generar un conocimiento innovador que brinde respuestas a las necesidades de la ciudadanía.

9.9. El nuevo escenario de la ciberguerra

La ciberguerra aparece como una continuación de los conflictos digitales y representa un fenómeno nuevo de carácter sorpresivo (Floridi, 2011). En el pasado los conflictos presentaban caracteres motorizados, pues se establecía un vínculo entre la industria y sus cadenas de montaje con el campo de batalla. A diferencia del pasado, en la actualidad los *bytes* han sustituido a las balas y las computadoras a las armas de fuego. El campo de batalla ha experimentado una profunda variación. Mientras que antes los conflictos se libraban en entornos físicos, ahora el espacio cibernético representa el terreno donde se realizan los enfrentamientos, como señala Gema Sánchez Medero:

> La ciberguerra puede ser entendida como una agresión promovida por un Estado y dirigida a dañar gravemente las capacidades de otro para imponerle la aceptación de un objetivo propio o, simplemente, para sustraer información, cortar o destruir sus sistemas de comunicación, alterar sus bases de datos, es decir, lo que habitualmente hemos entendido como guerra, pero con la diferencia de que el medio empleado no sería la violencia física sino un ataque informático que va desde «la infiltración en los sistemas informáticos enemigos para obtener información hasta el control de proyectiles mediante computadores, pasando por la planificación de las operaciones, la gestión del abastecimiento, etc.» (Sánchez Medero, 2012: 125)

Este fenómeno de cambio de escenario no tiene necesariamente que suponer una disminución de los niveles de violencia, sino un cambio en la forma de concebirla. Esta modificación en la concepción de la violencia deriva de la alta dependencia que las sociedades tienen de los medios tecnológicos más avanzados. Alejado de este enfoque que reconoce un componente violento en la ciberguerra que no tiene que ser necesariamente físico, se encuentra Thomas Rid (2012). Este politólogo duda que, debido a la falta de empleo de la fuerza física para causar daños potencialmente letales, puedan categorizarse a los ataques cibernéticos como «guerra». Para entender por qué la ciberguerra implica violencia, y por lo tanto se sitúa dentro de la categoría «guerra», es importante destacar las tres características de violencia que ofrece Christopher J. Finlay:

> Agencia: implica infligir daño a través de la acción intencional.
>
> Daño corporal: las variantes de la definición tienden a tratar el daño físico-físico como paradigmático. Otras posibilidades se incluyen en la medida en que parecen similares o son en algún sentido derivados o parientes cercanos de daños corporales, por ej. daño psicológico a la persona o daños a la propiedad.
>
> Violencia en un sentido descriptivo: la mediación entre ellos es la expectativa que los «actos de violencia» (como dice John Harris) también serán «actos violentos«en algún sentido. Serán repentinos, quizás fuertes y contundentes (Finlay, 2018: 362).

Además, como señala Juan José Díaz del Río Durán, la ciberguerra es asimétrica, ya que el bajo coste de los equipos informáticos no implica la fabricación de armamento caro y sofisticado por aquellos grupos que suponen una amenaza (2011: 251). El entorno en el que la guerra despliega sus acciones está cambiado, motivo por el que es necesario pensar este novedoso fenómeno desde conocimientos innovadores. La ciberguerra está provocando profundas transformaciones sobre la concepción de los conflictos. Las extensas reflexiones que se han venido haciendo a lo largo de la historia sobre la guerra justa van a experimentar cambios radicales debido a la naturaleza de la ciberguerra, ya que los impactos de determinadas estrategias pueden tener un alcance mucho mayor que en el pasado. Sobre el carácter de novedad que presenta la ciberguerra, podrían destacarse unas esclarecedoras palabras de Julio Albert Ferrero:

> Se trata de un nuevo concepto de guerra, cuyas acciones han aparecido desde finales del pasado siglo, que ha dado lugar a la aparición de las llamadas armas cibernéticas, genéricamente los virus, que responden a un nuevo concepto de arma. No se trata de armas de destrucción masiva, puesto que no destruyen nada a pesar de que su efecto puede llegar a ser demoledor, sino que interrumpen la actividad cibernética, por lo que se les denomina de interrupción masiva (2013: 82).

El espacio virtual permite una deslocalización de los ataques, ampliando de ese modo el potencial de actuación. Esto representa una grave amenaza para la paz de las sociedades, pues un número reducido de expertos hackers pueden poner en tela de juicio la estabilidad a través de un ejercicio informático. Se trata de un nuevo escenario de conflictos frente al que las instituciones de la quíntuple hélice deben estar preparadas y contar con mecanismos suficientes para asegurar la paz. Entre los ataques que pueden realizarse en la ciberguerra se encuentran: virus, gusano, troyano, código dañino, bomba lógica y *botnet.* Además, estos ataques tienen diversas procedencias: patrocinados por los Estados, servicios de inteligencia y contrainteligencia, terrorismo y extremismo político e ideológico, ataques de delincuencia organizada y ataques de perfil bajo (Albert Ferrero, 2013: 83-84).

A pesar de que todavía no ha habido ningún ataque que permita hablar de ciberguerra como tal, se han sucedido en las últimas décadas algunos ataques e intrusiones que han causado daños políticos, económicos y psicológicos. Entre esos casos podrían destacarse *Stuxnet* y *Anonymous*. En la actualidad millones de ciudadanos y ciudadanas dependen de internet para desarrollar su trabajo y poder beneficiarse de servicios esenciales para la vida. En ese sentido, la economía y la seguridad dependen en gran medida de las tecnologías más avanzadas y de la infraestructura de las telecomunicaciones. Por lo tanto, es lógico que las instituciones que conforman la quíntuple hélice tomen conciencia de esta situación y generen un conocimiento innovador en el ámbito de la IA para hacer frente a los desafíos del futuro en materia de seguridad. Un posible mecanismo para promover responsabilidad y reforzar la seguridad de la tecnología de la información y comunicaciones, como se verá seguidamente, podría consistir en la construcción de infraestructuras críticas que permitan establecer redes cívicas entre los grupos y las instituciones de los distintos niveles que forman la realidad política.

9.10. Ciberseguridad e inteligencia artificial responsable

La actividad profesional de los expertos de las TIC no involucra únicamente aspectos relativos a la técnica y sus habilidades, sino que también plantea cuestiones éticas relativas al tratamiento de la información. Por ello es importante reflexionar sobre las problemáticas éticas que giran en torno al concepto de cultura informativa de las sociedades actuales. El sistema educativo, el estado de la infraestructuras de información, el nivel de democratización, así como la situación económica de un país, son los principales factores que afectan el nivel cultural de la información de una sociedad (Malyuk y Miloslavskaya, 2016: 206). El concepto de cultura informativa que una sociedad posee influye en la ciudadanía y sus instituciones en materia pedagógica y psicológica. Estos son algunos de los procesos en los que se puede observar el impacto sobre esta dimensión cultural:

1. Generación de significados personales maduros y formación de una imagen adecuada y dinámica del mundo.

2. Intercambio efectivo de información asegurado por la formación de un conjunto de habilidades de información que contiene: evaluación de la utilidad y validez de la información recibida; búsqueda y selección de información personal relevante, incluyendo métodos de procesamiento; habilidades de comunicación y lenguaje (percepción y transmisión); defensa psicológica de la información.

3. Desarrollo y mejora de la eficacia individual y formas de guardar y asimilar información.

4. Información psicohigiénica (ecología) como autorregulación de los procesos de información y su correlación con los actuales estados del cuerpo.

5. Ética de la información que regula el acceso a la información de otra persona, uso de información para autoservicio o presión de la persona restringiendo el acceso a información útil de otros (Malyuk y Miloslavskaya, 2016: 205-206)

Por lo tanto, la cultura de la información se encuentra estrechamente vinculada con las TIC y sus sistemas de gestión. Su desarrollo influye sobre la capacidad de la ciudadanía para utilizar los recursos de información disponible, así como los medios de comunicación en un contexto político y social. En ese sentido, los mecanismos de información influyen de forma directa sobre el comportamiento humano y el desarrollo social. Por ello es fundamental considerar aspectos de seguridad en torno al tratamiento de la información en un contexto democrático mediante el diseño, evaluación e implementación de medidas contra el daño (*Cyberharm*) y las amenazas.

En diciembre de 2002 la Asamblea General de las ONU adoptó la resolución 57/239, que establece los principios para la creación de una cultura de la ciberseguridad global. Todos los participantes de la sociedad de la información global –gobiernos, empresas, otras organizaciones y usuarios individuales– que desarrollan, poseen, proporcionan, gestionan, mantienen y utilizan los sistemas y redes de información, deben seguir los siguientes principios:

a) Conciencia. Los participantes deben ser conscientes de la necesidad de una seguridad para sistemas de información y redes y lo que pueden hacer para mejorar ésta.

b) Responsabilidad. Los participantes son responsables de la seguridad de sistemas de información y redes de manera adecuada a sus roles individuales. Deben revisar sus propias políticas, prácticas, medidas y procedimientos, regularmente, y deben evaluar si son apropiados para su entorno.

c) Respuesta. Los participantes deben actuar de manera oportuna y cooperativa para prevenir, detectar y responder a incidentes de seguridad. Deben compartir información sobre amenazas y vulnerabilidades, según corresponda, e implementar procedimientos para una rápida y una cooperación efectiva para prevenir, detectar y responder a incidentes de seguridad. Esta puede implicar el intercambio de información y la cooperación transfronteriza.

d) Ética. Dada la omnipresencia de los sistemas y redes de información en las sociedades modernas, los participantes deben respetar los intereses legítimos de los demás y reconocer que su acción o inacción puede dañar a otros.

e) Democracia. La seguridad debe implementarse de manera consistente con los valores reconocidos por las sociedades democráticas, incluida la libertad de intercambio, pensamientos e ideas, el libre flujo de información, la confidencialidad de la información y la comunicación, la protección adecuada de la información personal, la apertura y la transparencia.

f) Evaluación de riesgos. Todos los participantes deben realizar evaluaciones de riesgos de forma periódica para identificar amenazas y vulnerabilidades; son suficientemente amplios para abarcan factores internos y externos clave, como la tecnología, la física y factores humanos, políticas y servicios de terceros con implicaciones de seguridad; permite la determinación del nivel aceptable de riesgo; y asiste en la selección de controles apropiados para gestionar el riesgo de daños potenciales a los sistemas de información y redes a la luz de la naturaleza e importancia de la información a ser protegida.

g) Diseño e implementación de seguridad. Los participantes deben incorporar la seguridad como elemento esencial en la planificación y diseño, operación y uso de sistemas y redes de información.

h) Gestión de la seguridad. Los participantes deben adoptar un amplio enfoque de la gestión de la seguridad basado en una evaluación de riesgos que es dinámica, que abarca todos los niveles de las actividades de los participantes y todos los aspectos de sus operaciones.

i) Reevaluación. Los participantes deben revisar, reevaluar la seguridad de sistemas y redes de información y deberían hacer las modificaciones apropiadas a políticas de seguridad, prácticas, medidas y procedimientos que incluyen abordar nuevas vulnerabilidades y amenazas (ONU, 2003).

Más allá de la prescripción de estos principios, es importante reconocer la dificultad y complejidad de la seguridad en el ámbito tecnológico, como señala el conocido criptógrafo Bruce Schneier (IT Security, 2007).

> A medida que los ataques cibernéticos aumentan su número y complejidad, la IA puede contribuir en la detención de estos ataques para combatirlos. Los diversos mecanismo que proporciona la IA como el *automatic learning* o el procesamiento de lenguaje natural, permiten a los expertos en seguridad identificar las amenazas y brindar respuestas. La IA presenta un gran potencial a la hora de proporcionar mecanismo para la defensa de la ciberseguridad. La adaptación de los intelectos sintéticos a los entornos, la recopilación de datos y el automatic learning puede conducir a un concepción beneficiosa de la IA en materia de seguridad, y en ese sentido a un cultivo de la responsabilidad cívica (Ava, 2019).

Las soluciones basadas en IA contribuirán en la reducción de la monotonía y agregarán más valor a las organizaciones. Las soluciones proporcionadas por la IA requieren de gente experta en el comienzo y un gran esfuerzo, aunque posteriormente representará un beneficio. Entre las soluciones que la IA puede proporcionar a las organizaciones en materia de seguridad se encuentran:

1. Análisis de vulnerabilidad.
2. Análisis de *malware*.
3. Detección de amenazas.
4. Seguimiento y respuesta de seguridad.

5. Detección de anomalías del alojamiento web (Passi, 2018).

La IA también puede utilizarse para tareas maliciosas. Hasta ahora se han destacado las ventajas de la IA y cómo ésta puede mejorar la ciberseguridad. Pero esto no significa que la IA no pueda ser derrotada en una determinada situación. Por ello, la IA necesita de la intervención humana para ser entrenada y garantizar la eliminación de los falsos positivos. Dejar trabajar a la IA de forma autónoma puede conducir a respuestas incorrectas. Debe ser entrenada de manera que pueda asumir las tareas monótonas primero y luego avanzar hacia tareas complejas. Finalmente los sistemas artificiales podrían analizar su desarrollo e idear soluciones optimizadas.

Los sistemas artificiales ya han mostrado un potencial ilimitado en varias aplicaciones de diferentes industrias. Del mismo modo, la implementación de la IA para soluciones de ciberseguridad permitirá proteger a las organizaciones de aquellas amenazas cibernéticas que pongan en riesgo la gestión de la información y también contribuirá en la identificación de nuevos tipos de programas malignos (*malware*). Además, los sistemas de ciberseguridad impulsados por IA pueden garantizar estándares de seguridad efectivos y ayudar a crear mejores estrategias de prevención y recuperación de datos (Joshi, 2019).

Naveen Joshi (2019) destaca cuatro mecanismos a través de los cuales la IA puede favorecer la seguridad de las organizaciones:

1. Inicio de sesión biométrico.

   La implementación de IA para la ciberseguridad permite la introducción de técnicas de inicio de sesión de carácter biométrico para la seguridad. Los sistemas artificiales pueden escanear huellas dactilares, retina, entre otros, con precisión. Estos inicios de sesión biométricos pueden usarse en combinación con contraseñas que ya están en uso en dispositivos como teléfonos inteligentes.

2. Detección de amenazas y actividades maliciosas.

   Las empresas de ciberseguridad están capacitando a los sistemas con IA para la detección de programas malignos (*malware*) y virus con la ayuda de varios conjuntos de datos que incluyen algoritmos y códigos. Usando tales datos, la IA puede realizar un reconocimiento de patrones que ayuda a identificar el comportamiento malicioso en el *software.* Además, la IA puede impulsar análisis predictivos para alternativas basadas en respuestas automáticas que siempre serán más rápidas y más efectivas que un enfoque manual.

3. Aprendizaje con procesamiento de lenguaje natural.

   Los intelectos sintéticos pueden recopilar automáticamente datos mediante el escaneo de artículos, estudios y noticias sobre amenazas cibernéticas. Estos intelectos se sirven del procesamiento del lenguaje natural para seleccionar información útil de los datos escaneados. Dicha información puede proporcionar información sobre ataques cibernéticos, anomalías, mitigación y estrategias de prevención. Utilizando la información analizada, las empresas de ciberseguridad pueden identificar escalas de tiempo, calcular riesgos, recopilar datos y hacer predicciones.

4. Aseguramiento del acceso condicional.

   El uso de IA para la seguridad cibernética favorecerá la creación de un marco de autenticación dinámico, en tiempo real y global que modifique los privilegios de acceso a la información, según la ubicación o la red. Con este enfoque, el sistema artificial recopilará información del usuario para analizar su comportamiento, la aplicación, el dispositivo, la red, los datos y la ubicación. Usando dicha información, el sistema artificial podría cambiar automáticamente los privilegios de acceso de cualquier usuario a los datos para garantizar la seguridad en redes remotas.

Así pues, en el terreno de la ciberseguridad los intelectos sintéticos pueden realizar una importante aportación inspirada en el cultivo de la responsabilidad cívica y democrática. La salvaguarda y preocupación por la acción segura es una muestra de cómo la tecnología puede desempeñar un papel relevante en el fortalecimiento de las habilidades cívicas y democráticas de una sociedad a través de un compromiso con los derechos humanos y los ODS orientado a la paz y la justicia.

CONCLUSIÓN

Una vez recorrido este extenso camino es necesario esbozar una serie de conclusiones ante la celeridad de un tiempo marcado por los grandes y disruptivos avances de la IA. A lo largo de este trabajo se han expuesto algunos de los profundos impactos y transformaciones que subyacen tras el despliegue de los intelectos sintéticos y que plantean importantes desafíos, a la vez que exigen un profundo ejercicio reflexivo de carácter ético. Como *homo faber*, el ser humano es creador de nuevas realidades a través de los medios que el mundo le brinda. Esa condición creativa, constitutiva de la inteligencia humana, supone un irrenunciable recurso desde el que someter su ejercicio a una mirada crítica y responsable. La IA es una de las creaciones de mayor alcance que la humanidad ha impulsado en las últimas décadas y que James Barrat (2015) ha denominado «nuestra invención final». Por ello resulta fundamental recoger el testigo de Jonas, un anunciador del riesgo mayor, como diría Jorge Enrique Linares (2018), para pensar el poder de la tecnología más avanzada en este siglo y valorar su impacto desde criterios éticos de responsabilidad, situando el respeto y cuidado del ser humano como su máxima prioridad.

El interés por la relevancia política y social de la IA en las diversas esferas de la vida humana es un aspecto central de este trabajo, cuyo hilo conductor ha sido la necesidad de incorporar criterios de responsabilidad ética ante los desafíos de los sistemas artificiales. Estos desafíos han permitido la formulación de una propuesta ética que pretende enriquecer el ejercicio de la tecnología y plantear ideas que permitan asumir un compromiso con el cultivo de las habilidades cívicas y el fortalecimiento de la democracia, situando al ser humano en el centro de sus preocupaciones.

En el contexto de la sociedad del conocimiento resulta fundamental promover un reconocimiento del valor de lo humano por encima de las lógicas tecnocráticas. Ortega señaló la condición técnica como un medio antropológico para la autoproyección vital. Esta condición implica la necesidad de un ejercicio hermenéutico de carácter crítico de las actividades en un mundo marcado por la carrera industrial en materia tecnológica. La

empresa hermenéutica demanda el reconocimiento de los límites y la importancia de comprender el contexto en el que se configuran y despliegan los conocimientos científicos que dan forma a los productos tecnológicos. De este modo, la hermenéutica permite situar el valor de la democracia frente al peligro tecnocrático que marca la agenda diaria de la vida de millones de seres humanos. Para ello es esencial un cambio cultural en la concepción de los artefactos tecnológicos, capaz de responder con procedimientos pragmáticos y fronéticos que fortalezcan las bases que hacen posible la democracia. No obstante, este asunto no sería posible sin una integración de las humanidades con los saberes científicos y tecnológicos que haga posible mostrar en el legado humanista un basamento que enriquece a la IA. Precisamente el humanismo tecnológico representa el impulso de un tiempo que persigue un nuevo modo de obrar que oriente la teoría y la práctica que da forma y despliega los intelectos sintéticos.

El planteamiento de una IAR ha sido posible gracias al impulso humanista que asume la tecnología como una creación que precisa ser reconocida no solo desde el optimismo, para un fructífero aprovechamiento del florecimiento humano, sino también desde la crítica, para no tomar distancia de esa condición de profunda reflexividad que debe estar presente en toda actividad científica. El humanismo tecnológico ha servido para orientar la IA hacia el terreno de lo político, de modo que pueda proponerse una integración de las capacidades de la IA para el fortalecimiento del carácter ético y político.

Debido al reconocimiento de la responsabilidad ética como una premisa ineludible en el ejercicio proyectivo de cualquier tecnología, el lector ha podido observar que el postulado de Jonas resulta necesario para fundamentar un ejercicio ético que trate de iluminar el fenómeno tecnológico en cuestión. No obstante, es prioritario salvar las diferencias históricas entre la formulación jonasiana del principio de responsabilidad, que se remonta a 1979, y las exigencias éticas del presente en relación con el desarrollo de la IA. Ante los impactos de diversa índole que está generando la IA, es fundamental proponer la contextualización de una nueva ética desde una fundamentación ontológica y metafísica. La esperanza de un futuro sostenible exige el surgimiento de un deber con perspectiva de futuro que favorezca una heurística del temor para avistar los efectos y las implicaciones de

la tecnología. La reflexión de Jonas sobre el poder de la tecnología y su estímulo para una ética de la tecnología favorecen: la búsqueda de aquello que es vulnerable para invitar al compromiso; la respuesta a los llamamientos que deben atenderse; la actitud crítica y reflexiva ante la tecnología y su poder; el cuestionamiento del empleo de la técnica como un medio para el dominio; y el imperativo de una ética que supere un espíritu antropocéntrico.

El recorrido realizado entre las diversas perspectivas éticas esbozadas en el estado de la cuestión ha ofrecido una panorámica sobre la amplitud de los análisis éticos y filosóficos que pueden proyectarse sobre el campo de la IA. El conocimiento de este conjunto de perspectivas constata la relevancia filosófica de un objeto de estudio que, en ocasiones, aparenta ser estrictamente técnico. Sin embargo, sus implicaciones para la vida humana trascienden el plano técnico y se sitúan en la senda filosófica y, en particular, en el ámbito de la reflexión ética. Las contribuciones de Kurzweil, Bostrom, Vallor, Hibbard, Nath y Sahu alimentan el empeño de la ética para orientar las actividades humanas desde una hermenéutica crítica. Al mismo tiempo han servido como unas coordenadas referenciales desde las que articular la propuesta de una IAR e inspirar un discurso alternativo establecido desde la responsabilidad ética y política.

El propósito de fortalecer la dimensión ética y política de la IA ha permitido promover un modelo de innovación que asume la necesidad de cuestionar los enfoques que privilegian los monopolios y oligopolios a partir de la defensa de la dimensión cívica de la ciencia y la sensibilidad frente a la vulnerabilidad del medio ambiente. El concepto de ciencia cívica, así como un modelo de generación de conocimiento que afronta la innovación social desde una dinámica de encuentro y participación de diversas esferas en el contexto de una innovación abierta y responsable, sostenido en la quíntuple hélice, han servido para orientar este trabajo desde sus inicios. La confluencia heterogénea facilita un ejercicio reflexivo de carácter participativo y colaborativo, y contribuye así a valorar con sensatez la relación existente entre la aceptabilidad y la deseabilidad en el quehacer científico, promoviendo una ética basada en el diálogo.

Resulta esencial poner en valor el carácter político de la tecnología para promover habilidades cívicas y democráticas. El presente trabajo ha puesto de relieve que el poder de la tecnología debe ser objeto de deliberación para reflexionar sobre los sentidos que configuran las relaciones políticas subyacentes en su espacio y la posibilidad de una ética aplicada. Por ello es importante trasladar la discusión tecnológica al ámbito de lo público, como un espacio caracterizado por el diálogo, donde se negocia la distribución del poder desde el encuentro de perspectivas e intereses. En este sentido, la IAR ha sido fundamentada en un concepto de ciencia cívica, centrada en el cultivo de las habilidades comunicativas de los tecnólogos, que favorezca un encuentro desde el que hacer posible la confluencia y el entrelazamiento de saberes con el objetivo de poder afrontar unos desafíos tecnológicos caracterizados por la complejidad y la controversia. Por lo tanto, la IAR representa una respuesta a las exigencias de un tiempo que precisa una innovación científica familiarizada con la comunicación social y la transparencia, situando el bienestar humano como una ejemplar disposición.

El interés pragmático y fronético del MIAR, así como las dinámicas deliberativas y participativas impulsadas desde la sociedad civil, permiten asumir responsabilidad con los desafíos que enfrenta la humanidad y que los poderes gubernamentales a nivel internacional han recogido en la Agenda 2030. En este modelo adquiere un papel relevante la sociedad civil, esfera decisiva desde la que contribuir a la generación de un conocimiento social, innovador y crítico que fundamente su legitimidad para ofrecer respuestas a los problemas que la ciudadanía y sus instituciones enfrentan día tras día. Ese carácter político de la tecnología se refleja en el compromiso con los derechos humanos, los ODS y los límites planetarios. Este compromiso es asumido por la IAR, entendida como una tecnología avanzada de carácter político y con capacidad para beneficiar a la sociedad, sirviendo como mecanismo para ofrecer respuestas que no descuiden la responsabilidad social.

El primer desafío abordado para contextualizar la IAR ha sido el de las tecnologías de mejoramiento de la especie integradas en el proyecto transhumanista. Son varios los caminos hacia el poshumanismo que ofrecen este tipo de tecnologías y también las posiciones filosóficas que provocan sus efectos: bioprogresistas y bioconservadores.

Reconociendo el valor de las aportaciones que los pensadores expuestos esbozan sobre la variedad de manifestaciones que subyacen tras el transhumanismo, es necesario plantear una vía reflexiva que permita problematizarlo desde una perspectiva ética y política. En ese sentido se han planteado unas líneas discursivas que permiten enriquecer el debate: la importancia del reconocimiento de la pertenencia a un grupo social y la identidad para el acceso a estas tecnologías; la necesidad de establecer criterios de igualdad de oportunidades en el beneficio tecnológico; el respeto de los proyectos transhumanistas a los derechos humanos; la prioridad de situar el bienestar humano y medioambiental como un aspecto ineludible; el establecimiento de procesos deliberativos en los entornos tecnológicos enfocados en el mejoramiento, y la importancia de la evaluación ética. Estas líneas favorecen la exploración de nuevos caminos reflexivos en un novedoso espacio de autogénesis y proyección que exige una consideración ética y política.

Otro desafío que está despertando un gran interés en numerosos ámbitos es la automatización del campo de las profesiones y el impacto de la IA sobre el trabajo. Se están invirtiendo grandes sumas de dinero para la investigación en robótica aplicada a las profesiones, lo que a grandes rasgos demuestra que los robots se abren camino. Los efectos de la automatización han propiciado la aparición de dos perspectivas: tecnooptimistas y tecnopesimistas. Ambas posiciones discuten sobre los impactos políticos y sociales de la irrupción de la IA en el mundo del trabajo. Más allá de las posiciones exageradamente optimistas y de las desmedidas catastrofistas, la IAR permite abrir la senda para una mirada crítica y reconocedora de las posibilidades. Por ello es importante que, como una forma de asumir responsabilidad e implicación, el desarrollo tecnológico orientado a las profesiones se comprometa con trabajar hacia una sociedad inclusiva, innovadora y reflexiva.

Los ingenieros del ámbito de la movilidad autónoma afirman que son pocos los años que quedan para ver circular por las vías de nuestras ciudades coches sin conductor. El sueño de Francis Houdina parece hacerse realidad tras décadas de investigación. No obstante, a pesar del interés que puede despertar este producto tecnológico, al igual que otros tantos, es importante observar su gestación y despliegue desde una óptica reflexiva. Es cierto que los coches sin conductor presentan una serie de ventajas, aunque eso no

implica que igualmente puedan ocasionar nuevas amenazas para la convivencia. Por lo tanto, el desarrollo de esta tecnología debe implicar un cambio cultural que vaya de la mano de la IAR para contemplar la posibilidad de un fundamento cívico en su beneficio. Reconociendo las posibilidades que ofrece la automatización, desde la óptica de la responsabilidad ética en el contexto de la IA, es esencial perseguir el bienestar para el ser humano y fomentar el compromiso cívico.

Por último, el campo militar y de la seguridad no podía permanecer al margen de las investigaciones y aplicaciones de la IA, ni tampoco de la reflexión realizada en las páginas anteriores. El empleo de los robots por parte de los ejércitos viene desarrollándose desde hace unas cuantas décadas, aunque ha sido en la actualidad cuando ha despertado un mayor interés. Este interés creciente se debe principalmente a los drones y al enjambre que se encuentra tras su aparataje. Los drones han generado diversas controversias éticas en su uso militar. En este sentido, la importancia de una ética aplicada a los drones se ha desarrollado principalmente de manera orientada al diseño y a la redefinición humanitaria de su utilidad. Además, en el contexto de las instituciones de defensa, una IAR debe fomentar el desarrollo sostenible, con el fin de generar una actividad responsable. El progreso de la tecnología ha propiciado la aparición del ciberespacio como un nuevo escenario que demanda protección de la infraestructura computacional y de todo lo relacionado con esta. Así pues, las potencialidades de la IA pueden invertirse en la ciberseguridad como una respuesta de responsabilidad para fortalecer y defender la democracia.

Expertos en el sustrato político de la IA como Félix Arteaga y Andrés Ortega (2019) han destacado la necesidad de generar las condiciones de posibilidad para construir un ecosistema de IA en España. La propuesta de este ecosistema señala que deben abordarse asuntos relativos al aumento de las inversiones públicas y privadas en el ámbito de los intelectos sintéticos, promover una preparación para los cambios socioeconómicos y garantizar un marco ético y legal adecuado para tal escenario. En este sentido, el concepto de una IAR va precisamente en esa dirección, pues reconoce el conocimiento científico desde una perspectiva pública, susceptible de tratamiento en la esfera pública, entendida como ese espacio de encuentro y convergencia de esferas. Este tratamiento público surge

ante la necesidad de configurar medidas de innovación que permitan ofrecer una solución a aquellas problemáticas y demandas surgidas de las transformaciones derivadas del desarrollo de la IA.

Existe una evidente posibilidad para impulsar ese ecosistema desde un cambio cultural en la consideración del conocimiento científico como un recurso de interés público que rompa con la visión monopolista y oligopolista de los modelos de generación de conocimiento. La exposición desarrollada en la tercera parte de este trabajo, donde han sido analizados los diversos desafíos y se realizado una contextualización de la IAR en cada uno de ellos, manifiesta precisamente esa posibilidad.

Un cambio cultural es esencial para poder observar la IA desde una mirada ética y política, pues de lo contrario se seguirá en la senda del consumo irreflexivo y carente de crítica, propio de la tecnocracia, en lo que respecta a los medios tecnológicos más avanzados. Sin un cambio cultural en el diseño, producción y despliegue tecnológico, será muy difícil promover el talante político y ético que permita asumir compromisos específicos ante los grandes retos. Las universidades, las empresas, las administraciones públicas y las diversas organizaciones de la sociedad civil tienen la responsabilidad de colaborar para proyectar un diseño innovador y fomentar el compromiso cívico y democrático en el ámbito de la IA. No se persigue eliminar por completo los propósitos con los que los ingenieros proyectan el diseño de los intelectos sintéticos, pues eso sería absurdo, sino contribuir a que el compromiso ético y político que exige la Agenda 2030 esté presente en estos procesos. Una orientación alternativa en el uso de los intelectos sintéticos es posible desde una perspectiva pragmática y fronética.

No solo es importante fomentar en el contexto de la IA una cultura ética y comunicativa centrada en el diálogo y el encuentro en los entornos generadores de conocimiento científico, sino también incorporar en unos casos y fortalecer en otros una materia de ética aplicada en los programas de estudio tecnológico, tanto de formación profesional, como universitaria. Este asunto surge del rechazo de la visión de la actividad tecnológica como un quehacer humano caracterizado por la neutralidad axiológica. La formación del *ethos* de

los estudios técnicos abre la puerta a la posibilidad de promover habilidades cívicas y democráticas en un entorno que habitualmente se encuentra habitado por la tecnocracia.

En el despliegue de los intelectos sintéticos y en el quehacer de los agentes involucrados en su diseño, producción e implementación, se encuentran implicadas una serie de problemáticas que pueden ir surgiendo con el paso del tiempo. En ese sentido, el progreso moral de las sociedades debe hacer frente a los diversos desafíos subyacentes en esta actividad con herramientas de orientación ética caracterizadas por el análisis y la reflexión en torno a la condición moral del ser humano. Es importante que el desarrollo del conocimiento de los agentes tecnológicos incluya un interés creciente por la importancia de la responsabilidad ética y social. Precisamente el humanismo tecnológico reconoce esta situación y sugiere una perspectiva enriquecedora de la actividad tecnológica desde un marco mucho más amplio e integral, caracterizado principalmente por un conocimiento más consciente de los productos humanos y su impacto en la realidad.

Estrechamente vinculada a la importancia de una ética aplicada a los estudios tecnológicos en el ámbito de la IA, está presente la necesidad de impulsar desde las instituciones, tanto públicas como privadas, investigaciones donde confluyan saberes entrelazados. Debido a la complejidad de los desafíos que plantea el despliegue de los intelectos sintéticos, la investigación científica tendría que comprometerse con la superación del espíritu de las dos culturas, pues la exacerbada especialización ha demostrado tener ciertas limitaciones para hacer frente a algunas problemáticas. Los reduccionismos científicos no siempre contribuyen a la formulación de respuestas pragmáticas, sino que, como ha sido señalado en varias ocasiones, es esencial también una concepción fronética del conocimiento para combatir el empobrecimiento de la experiencia humana y sus complejidades morales.

El fomento de espacios de encuentro como los laboratorios abiertos contribuyen favorablemente a cultivar el carácter público del conocimiento científico y permiten la participación de diversos grupos de interés. No obstante, tanto estos laboratorios, como las investigaciones surgidas de la confluencia de saberes entrelazados, exigen un cambio

cultural al momento de concebir la ciencia. En ese sentido, es fundamental plantear una noción cívica del conocimiento e impulsar la reflexión teórica y práctica de la ciencia cívica en aquellas instituciones dedicadas a la investigación y gestión de los saberes científicos.

Pero el cambio cultural en los entornos científicos también debe ir acompañado de reflexiones filosóficas sobre la tecnología y sus nuevos escenarios. Haciendo un repaso a los planes de estudio de los grados universitarios en Filosofía de algunas universidades españolas, puede observarse la ausencia en el listado de asignaturas troncales de una Filosofía de la Tecnología, a excepción de algunos casos donde sí se encuentra presente como materia optativa. Resulta difícil entender cómo los estudios académicos en el ámbito filosófico se resisten a incorporar en sus planes de estudio, como asignatura obligatoria, la reflexión sobre un fenómeno esencial de la condición humana.

BIBLIOGRAFÍA

Access Now (2018). *Human Rights In The Age Of Artificial Intelligence*. Access Now. Recuperado de https://medium.com/@eraser/human-rights-in-the-age-of-artificial-intelligence-941946be1c44?fbclid=IwAR3zzN1B1I1vT8qd7mhE--Qvt6yPynC0fMs61FWfoQRzmhdesmWvg2U1eC8

Acemoglu, D. y Restrepo, P. (2016). The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares and Employment. *NBER Working Papers*, *22252*.

Acemoglu, D. y Restrepo, P. (2017). *Robots and Jobs: Evidence from the US*. Voxeu . Recuperado de https://voxeu.org/article/robots-and-jobs-evidence-us

Ackerman, S. (2011). Air Force Keeps 'Micro-Aviary' Of Tiny, Bird-like 'Bots. *Danger Room.*

Adams, T. (2 de junio de 2016). Artificial intelligence: 'We're like children playing with a bomb'. *The Guardian*. Recuperado de https://www.theguardian.com/technology/2016/jun/12/nick-bostrom-artificial-intelligence-machine

Aguirre Román, J. (2016). Habermas y el rol de la religión en la esfera pública: el caso de la eugenesia liberal. *Franciscanum*, *166*(58), 49-85.

Aguirre Romero, J. M.ª (2002). Ciencia, Humanismo, Humanidades y Tecnología. *Espéculo. Revista de estudios literarios*. Recueprado de http://webs.ucm.es/info/especulo/numero19/humanism.html

Albert Ferrero, J. (2013). La ciberguerra. Génesis y evolución. *Revista general de marina*, *264*(1), 81-97.

Allen, C., Smith, I. y Wallach, W. (2005). Artificial morality: Top-down, bottom-up, and hybrid approaches. *Ethics and Information Technology*, *7*, 149-155.

Álvarez, R. (28 de marzo de 2018). El accidente de un Tesla Model X provoca la muerte de su conductor y una nueva investigación en torno a la conducción autónoma. *Xataka* Recuperado de https://www.xataka.com/automovil/el-accidente-de-un-tesla-model-x-provoca-la-muerte-de-su-conductor-y-una-nueva-investigacion-en-torno-a-la-conduccion-autonoma

Álvarez, J. J. (2013). *Aproximación crítica a la inteligencia artificial. Claves filosóficas y prospectivas de futuro*. Madrid, España: Universidad Francisco de Vitoria.

Amnistía Internacional y Access Now (2018). *Toronto's Declaration on Machine Learning*. Recueprado de https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf.

Annas, George J. (2017). Health and Human Rights: Of Bridges and Matrixes. *Am J Bioeth*, *17*(9), 13-15.

Arana Cañedo-Argüelles, J. (2017). Ante los desafíos del posthumanismo y transhumanismo. *Nueva revista de política, cultura y arte*, *162*, 171-199.

Arendt, H. (2012). *La condición humana*. Barcelona, España: Paidós.

Aristóteles (2014). *Ética a Nicómaco*. Madrid, España: Alianza.

Arkin, R. C. (2008). *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*. United States: U.S. Army Research Office. Recuperado de https://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf

Arkin, R.C. y Ulam, P. (2009). An Ethical Adaptor: Behavioral Modification Derived from Moral Emotions. *IEEE International Symposium on Computational Intelligence in Robotics and Automation* (CIRA-09).

Arntz, M.; Gregory, T.; Zierahn, U. (2016). The Risk of Automation for Jobs in OECD Countries: A Comparative analysis. *OECD Social, Employment and Migration Working Papers*, *189*. Recuperado de http://www.ifuturo.org/sites/default/files/docs/automation.pdf

Arteaga, F. y Ortega, A. (2019). Hacia un ecosistema español de Inteligencia Artificial, *Elcano Policy Paper*, *225*.

Asimov, I. (1978). *Yo, robot*. Barcelona, España: Edhasa.

Ava, I. (2019). AI and cybersecurity: Is it a complication? *Toolbox*. Recuperado de https://it.toolbox.com/guest-article/ai-and-cybersecurity-is-it-a-complication.

Avent, R. (2017). *La riqueza de los humanos. El trabajo en el siglo XXI*. Barcelona, España: Ariel.

Barben, D., Fisher, E., Selin, C. y Guston, D. (2008) Anticipatory governance of nano-technology: foresight, engagement, and integration. En E. Hackett, M Lynch y J. Wajcman (Eds.), *The Handbook of Science and Technology Studies* (979-1000) Cambridge, MA, United States: MIT Press.

Barber, B. (2004). *Democracia fuerte*. Córdoba, España: Almuzara.

Barclays (2015). *Disruptive Mobility: A Scenario for 2040*. London, England: Barclays Bank PLC. Recuperado de https://www.investmentbank.barclays.com/content/dam/barclaysmicrosites/ibpublic/documents/investment-bank/global-insights/barclays-disruptive-mobility-pdf-120115-459kb.pdf

Barclays (2016). *Driverless vehicles: A new engine for economic transformation?* London, England: Barclays Bank PLC. Recuperado de https://www.barclayscorporate.com/content/dam/barclayscorporate-com/documents/insights/innovation/barclays-corporate-driverless-vehicles-oct-2016.pdf

Barclays (2018). *Robots at the gate: Humans and technology at work*. London, England: Barclays Bank PLC Recuperado de https://www.investmentbank.barclays.com/our-insights/robots-at-the-gate.html?icid=FoW_110418_PR

Baker, E. J. (2018). *Artificial intelligence and national security law: A dangerous nonchalance*. Massachusetts Institute of Technology, EU: Star Forum. Recuperado de https://cis.mit.edu/sites/default/files/documents/StarrForumReport_18-01.pdf

Barrat, J. (2015). *Our final version. Artificial Intelligence and the End of the Human Era*. New York, EU: Thomas Dunne Books.

Baum, S. D. (2017). Social choice ethics in artificial intelligence. *AI and Society*, 1-12.

Bauman, Z. (2017). *Modernidad líquida*. Madrid, España: Fondo de Cultura Económica de España.

Bechi, P. (2008). El itinerario filosófico de Hans Jonas. Etapas de un recorrido. *Isegoría*, *39*, 101-128.

Beck, U. (1992). *The Risk Society. Towards a New Modernity.* London, United Kingdom: Sage.

Belda, C. (2018). Para cumplir las metas de desarrollo, hay que mirar con lupa el gasto militar. *El País*. Recuperado de https://elpais.com/elpais/2018/01/22/planeta_futuro/1516635931_952181.html

Bello Reguera, E. (2007). El humanismo según Sartre y Heidegger: los ecos de una polémica. En C. Ballestín Cucala y J. L. Rodríguez García. (Coord.), *Estudios sobre J. P. Sartre* (pp. 33-56). Madrid, España: Mira.

Benjamin, M. (2014). *Las guerras de los drones. Matar por control remoto*. Barcelona, España: Anagrama.

Berboucha, M. (2018). Uber Self-Driving Car Crash: What Really Happened. *Forbes* Recuperado de: https://www.forbes.com/sites/meriameberboucha/2018/05/28/uber-self-driving-car-crash-what-really-happened/#708feea4dc41

Bernstein, R. J. (1994). Rethinking Responsibility. *Social Research*, *61*(4), 833-852.

Bessen, J. (2015). How computer automation affects occupations: Technology, jobs, and skills. *Law & Economics Working Paper*, *13*, 15-49.

Bimber, B.A. (1996). *The politics of expertise in congress: The rise and fall of the office of technology assessment*. Albany: State University of New York Press.

Bloch, E. (2013). *El principio de esperanza I*. Madrid, España: Trotta.

Boden, M. A. (2017). *Inteligencia Artificial*. Madrid, España: Turner Noema.

Bogost, I. (1 de noviembre de 2014). The Secret History of the Robot Car. The Atlantic. Recuperado de https://www.theatlantic.com/magazine/archive/2014/11/the-secret-history-of-the-robot-car/380791/

Bonnefon, J.-F., Sharrif, A., y Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573-1576.

Bostrom, N. (2003a). The Transhumanist FAQ: A General Introduction. *World Transhumanist Association*.

Bostrom, N. (2003b). Transhumanist values. En F. Adams. (Ed.), *Ethical Issues for the 21st Century* (3-14). Virginia, EU: Philosophical Documentation Center Press.

Bostrom, N. (2004). Human Genetic Enhancements: A Transhumanist Perspective, *The Journal of Value Inquiry*, *37*, 493-506.

Bostrom, N. (2007). In Difense of Posthuman Dignity. *Bioethics*, *19*(3), 202-214.

Bostrom, N. (2008). Why I Want to be a Posthuman When I Grow Up. En B. Gordijn y R. Chadwick (Eds.), *Medical Enhancement and Posthumanity* (107-137). London, United Kingdom: Springer.

Bostrom, N y Yudkowsky, E. (2011). The Ethics of Artificial Intelligence. En Ramsey W. y Frankish (Eds.), *Cambridge Hankbook of Artificial Intelligence* (316-334). Cambridge, England: Cambridge University Press.

Bostrom, N. (2016). *Superinteligencia, caminos, peligros, estrategias*. Madrid, España: Teell.

Bostrom, N. (2017). *Asilomar AI Principles*. Recuperado de https://futureoflife.org/ai-principles/?cn-reloaded=1

Bostrom, N. y Sandberg, A. (2017). El mejoramiento como desafío práctica. En N. Bostrom y J. Savulescu (Eds.), *Mejoramiento humano* (391-435). Madrid, España: Teell.

Botti, V. J. y Julián, V. (2000). Agentes Inteligentes: el siguiente paso en la Inteligencia Artificial. *Novática: Revista de la Asociación de Técnicos de Informática*, *145*, 95-99.

Brockman, J. (1996). *La tercera cultura: más allá de la revolución científica*. Barcelona: Tusquets.

Brooks, R. A. (1991). Intelligence Without Reason. *Massachusetts Institute of Technology, A.I. Memo*, 1293. Recuperado de https://people.csail.mit.edu/brooks/papers/AIM-1293.pdf

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., hÉigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evans, O., Page, M., Bryson, J., Yampolskiy, R., y Amodei, D. (2018). *The Malicious Use of Artificial Intelligence; Forecasting, Prevention, and Mitigation*. Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Center for a New American Security, Electronic Frontier Foundation y OpenAI. Recuperado de https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

Buchanan, A. E. (2011). *Beyond Humanity? The Ethics of Biomedical Enhancement*. Oxford, England: Oxford University Press.

Buchanan, A., Brock, D. W., Daniels, N. y Wikler, D. (2012). *From Chance to Choice: Genetics & Justice*. Cambridge, England: Cambridge University Press.

Bureau of European Policy Advisers (2011). *Empowering people, driving change: Social Innovation in the European Union*. Luxemburgo: Oficina de Publicaciones de la Unión Europea.

Camps, V. (2002). ¿Qué hay de malo en la eugenesia? *Isegoría*, *27*, 55-71.

Carayannis, E. G. y Campbell D. F. J. (2010). Triple Helix, Quadruple Helix and Quintuple Helix and how do knowledge, innovation and the environment relate to each other? A proposed framework for a trans-disciplinary analysis of sustainable development and social ecology. *International Journal of Social Ecology and Sustainable Development*, *1*(1), 41-69. Recuperado de https://www.igi-global.com/article/triple-helix-quadruple-helix-quintuple/41959

Carayannis, E. G., Barth, T. D. y Campbell D. F. J. (2012). Dimensions of Environmentally Sustainable Innovation: the Structure of Eco-Innovation Concepts. *Journal of Innovation and Entrepreneurship.* Recuperado de https://innovation-entrepreneurship.springeropen.com/articles/10.1186/2192-5372-1-2

Carayannis, E. G., y Campbell, D. F. J. (2014). Developed democracies versus emerging autocracies: arts, democracy, and innovation in Quadruple Helix innovation systems. *Journal of Innovation and Entrepreneurship*. Recuperado de https://innovation-entrepreneurship.springeropen.com/articles/10.1186/s13731-014-0012-2

Carr, N. (2014). *Atrapados por las máquinas. Cómo las máquinas se apoderan de nuestras vidas*. Madrid, España: Alfaguara.

Cepeda Mayorga, I. (2017). La necesidad de reconocer y comprender la formación en una ciudadanía activa. En F. Díaz Estrada y M. A. Martínez Martínez (Eds.), *Realidades, teóricas y prácticas. Fractales de una ciudadanía en tránsito* (185-210). México: Castellanos.

Chamayou, G. (2015). *A Theory of the Drone*. New York, United States: The New Press.

Chamayou, G. (2016). *Teoría del dron. Nuevos paradigmas de los conflictos del siglo XXI*. Madrid, España: Ned.

Chakraborty, S. y Bhojwani, R. (2018). Artificial intelligence and human rights: are they convergent or parallel to each other? *Novum Jus*, *12*(2), 13-38.

Chavarría, J. (2003). El principio de responsabilidad: Ensayo de una axiología para la tecnociencia. *Isegoría*, *29*, 125-137.

Checkoway, S., McCoy D., Kantor, B., Anderson, D., Shacham, H. y Savage, S. (2011). *Comprehensive Experimental Analyses of Automotive Attack Surfaces*. Recuperado de http://www.autosec.org/pubs/cars-usenixsec2011.pdf

Chilvers, J. (2010). *Sustainable Participation? Mapping Out and Reflecting on the Field of Public Dialogue in Science and Technology*. Harwell, England: Sciencewise-ERC.

Coeckelbergh, M. (2011). From Killer Machines to Doctrines and Swarms, or Why Ethics of Military Robotics Is not (Necessarily) About Robots. *Philosophy & Technology*, *24*, 269-278.

Coeckelbergh, M. (2014). The Moral Standing of Machines: Towards a Relational and Non-Cartesian Moral Hermeneutics. *Philosophy & Technology*, *27*(1), 61-77.

Coello Coello, C. A. (2003). *Breve historia de la computación y sus piones*. Ciudad de México, México: Fondo de Cultura Económica.

Cohen, J.L. y Arato, A. (2000). *Sociedad civil y teoría política*. Ciudad de México, México: Fondo de Cultura Económica.

Coker, C. (2015). *Future War*. Cambridge, England: Polity.

Collin, H. M. (1990). *Artificial Experts*. Cambridge MA, United States: MIT Press.

Collingridge, D. (1980). *The Social Control of Technology*. Milton Keynes: Open University Press.

Comisión Europea (1995). *Green paper on innovation*. Recuperado de http://europa.eu/documents/comm/green_papers/pdf/com95_688_en.pdf

Comisión Europea (2007). *Taking European Knowledge Seriously. Report of the Expert Group on Science and Governance to the Science, Economy and Society Directorate.* Recuperado de https://ec.europa.eu/research/science-society/document_library/pdf_06/european-knowledge-society_en.pdf

Comisión Europea (2009). *The World in 2025: Rising Asia and Socio-ecological Transition. The World in 2025*. Recuperado de https://ec.europa.eu/research/social-sciences/pdf/policy_reviews/the-world-in-2025-report_en.pdf

Comisión Europea (2010). *Europa 2020 Una Estrategia para un crecimiento inteligente, sostenible e integrador*. Recuperado de https://www.aragon.es/estaticos/GobiernoAragon/Departamentos/PresidenciaJusticia/Areas/PJ_04_Informacion_de_la%20Union_europea/01_Europe_Direct_Aragon/Publicaciones%20de%20la%20Uni%C3%B3n%20Europea/Europa%202020%20-%20la%20estrategia%20europea%20de%20crecimientoA.pdf

Comisión Europea (2013). *Guía de la innovación social*. Recuperado de http://movil.asturias.es/Asturias/descargas/PDF_TEMAS/Asuntos%20Sociales/guia_innovacion_social.pdf

Comisión Europea (2014). *Horizon 2020. El Programa Marco de Investigación e Innovación de la Unión Europea*. Luxemburgo: Oficina de Publicaciones de la Unión Europea. Recuperado de https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/H2020_ES_KI0213413ESN.pdf.

Comisión Europea (2015). *White Paper on Citizen Science in Europe*. Recuperado de https://ec.europa.eu/futurium/en/system/files/ged/socientize_white_paper_on_citizen_science.pdf

Comisión Europea (2017). *Citizen Science*. Shaping Europe's digital future. Recuperado de https://ec.europa.eu/digital-single-market/en/citizen-science

Comisión Europea (2019). *Directrices éticas para una IA fiable*. Recuperado de https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Conill Sancho, J. (2010). *Ética hermenéutica. Crítica desde la facticidad*. Madrid, España: Tecnos.

Conill Sancho, J. (2013). *Horizontes de economía ética. Aristóteles, Adam Smith, Amarty Sen*. Madrid, España: Tecnos.

Conill, J. (2016). De la razón pura a la razón vital orteguiana a través de Nietzsche. *Revista de Hispanismo Filosófico*, *21*, 71-96.

Copeland, J. (1996). *Inteligencia artificial*. Madrid, España: Alianza.

Cordeiro, J. L. y Wood, D. (2018). *La muerte de la muerte. La posibilidad científica de la inmortalidad física y su defensa moral*. Barcelona, España: Deusto.

Cortina, A. (1996). El estatuto de la ética aplicada. Hermenéutica crítica de las actividades humanas. *Isegoría*, *13*, 119-127.

Cortina, A. (1998). *Ciudadanos del mundo*. Madrid, España: Alianza.

Cortina, A. (2013).*¿Para qué sirve realmente la ética?* Barcelona, España: Paidós.

Cortina, A. (2017). *Aporofobia, el rechazo al pobre. Un desafío para la democracia*. Barcelona, España: Paidós.

Cortina, A. [Albert] (2017). *Humanismo avanzado. Para una sociedad biotecnológica*. Madrid, España: Teconté.

Cotino Hueso, L. (2017). Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales. *Dilemata*, *24*, 131-150.

Cowen, T. (2013). *Se acabó la clase media: cómo prospera en un mundo digital*. Barcelona, España: Antoni Bosch.

Cummings, M. L. (2017). *Artificial Intelligence and the Future of Warfare.* International Security Department and US and the Americas Programme. Recuperado de https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings-final.pdf

Dascal, M. y Dror, I. E. (2013). The impact of cognitive techonologies: Towards a pragmatic approach. *Pragmatic & Cognition*, *13* (3).

De Grey, A. y Rae, M. (2013). *El fin del envejecimiento: los avances que podrían revertir el envejecimiento humano durante nuestra vida*. Berlín, Alemania: Lola Books.

Della Mirandola, P. (1984). *De la dignidad del hombre*. Madrid, España: Nacional.

Deroeux, I. (2015). Ta-Nehisi Coates: un grito de rabia afroamericano. *Nueva Sociedad*, *259*, 27-32.

Dewey, J. (1948). *La experiencia y la naturaleza*. México: Fondo de Cultura Económica.

Dewey, J. (1984). Lo que yo creo. En J. A. Boydston. (Ed.), *John Dewey: The later Works, 1925-1953* (276-278). Carbondale, United States: Southern Illinois Press.

Diamond, L. (1999) *Developing democracy. Towards consolidation*. Baltimore, United States: The John Hopkins University Press.

Díaz del Río Durán, J. (2011). La ciberseguridad en el ámbito militar. *Cuadernos de Estrategia*, *149*, 215-256.

Dickmann, J., Appenrodt, N. y Brenk, C. (2014). Making Bertha: Radar is the key to Mercedes-Benz's robotic car. *IEEE Spectrum*, 44-49.

Diéguez, A. (1993). Tecnología y responsabilidad. *Revista de Filosofía*, *9*, 189-200.

Diéguez, A. (2017). *Transhumanismo. La búsqueda tecnológica del mejoramiento humano.* Barcelona, España: Herder.

Diéguez, A. (13 de septiembre de 2018). Transhumanismo y filosofía. *El País*. Recuperado de https://elpais.com/elpais/2018/09/12/opinion/1536752872_112358.html

Dignum, V. (2017a). Responsible Autonomy. International Joint Conference on Artificial Intelligence.

Dignum, V. (2017b). Responsible artificial intelligence: designing AI for human values. ITU Journal: ICT Discoveries, 1.

Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issuse. *Ethics and Information Technology*, *20*, 1-3.

Doménech, R., García, J. R., Montañez, M. y Neut, A. (2017). *El impacto del cambio tecnológico y el futuro del empleo*. VI Encuentro de Derecho Financiero y Tributario-Instituto de Estudios Fiscales: BBVA Research. Recuperado de https://www.bbvaresearch.com/wp-content/uploads/2018/02/El_empleo_del_Futuro-IEF-26feb2018.pdf

Domingo Moratalla, A. (1991). *Ecología y solidaridad. De la ebriedad tecnológica a la sobriedad ecológica*. Madrid, España: Fe y Secularidad.

Domingo Moratalla, A. (1995). *Responsabilidad bajo palabra. Desafíos éticos para una democracia joven*. Valencia, España: Arzobispado de Valencia.

Domingo Moratalla, A. (2007). *Hábitos de ciudadanía activa. De la democracia escrita a la democracia vivida*. Madrid, España: Fundación Emmanuel Mounier.

Domingo Moratalla, A. (2013). *Educación y redes sociales. La autoridad de educar en la era digital*. Madrid, España: Encuentro.

Domingo Moratalla, T. (2007). La ética antropológica de Hans Jonas en el horizonte de la fenomenología hermenéutica. *Thémata: Revista de Filosofía*, *39*, 373-380.

Domingo Moratalla, T. y Feito Grande, L. (2017). Bioética narrativa: clave de la deliberación. En F. J. León Correa y P. Sorokin. (Coord.), *Bioética y salud pública en y para América Latina* (37-59). Santiago de Chile, Chile: Federación Latinoamericana y del Caribe de Instituciones de Bioética.

Duque Pajuelo, F. (1996). Martín Heidegger: En los confines de la Metafísica. *Anales del Seminario de Historia de la Filosofía*, *13,* 19-38.

Duque Pajuelo, F. (2002). *En torno al humanismo. Heidegger, Gadamer, Sloterdijk*. Madrid, España: Tecnos.

Durán Heras, M. A. (2010). *Tiempo de vida y tiempo de trabajo*. Madrid, España: Fundación BBVA. Recuperado de http://digital.csic.es/bitstream/10261/101047/3/Duran_Tiempo_vida_trabajo.pdf

Durán Heras, M. A. (2012). *El trabajo no remunerado en la economía global*. Madrid, España: Fundación BBVA. Recuperado de http://digital.csic.es/bitstream/10261/76517/3/Duran_Trabajo_No_Remunerado.pdf

Durbin, P. T. (1992). *Social Responsibility in Sciencie, Technology and Medicine.* Pennsylvania, United States: Lehigh University Press.

Ecologistas en Acción (2007). *Los problemas del coche en la ciudad*. Madrid, España: Ecologistas en Acción. Recuperado de https://spip.ecologistasenaccion.org/IMG/pdf_Cuaderno_1_Problemas_Coche.pdf

Eggers, W. D. y Macmillan, P. (2014). *La revolución de las soluciones*. Madrid, España: LID.

Einsiedel, E. F. (2005). Understanding publics in the public understanding of science. En M. Dierkes y C. von Grote. (Ed.), *Between Understanding and Trust: The Public, Science and Technology*. Routledge Taylor & Francis Group eLibrary.

Ellul, J. (2003). *La edad de la técnica*. Barcelona, España: Octaedro

Ellul, J. (2004). El orden tecnológico. En C. Mitcham y R. Mackey (Eds.), *Filosofía y tecnología*. Madrid, España: Ediciones Encuentro.

Esquirol, J. (2006). *El respeto o la mirada atenta. Una ética para la era de la ciencia y la tecnología.* Barcelona, España: Gedista.

Esquirol, J. (2011). *Los filósofos contemporáneos y la técnica. De Ortega a Sloterdijk*. Barcelona, España: Gedisa.

Ethical Hacking News Tutorials (2016, agosto 6). Hacking a Car with an Ex NSA Hacker CYBERWAR Clip YouTube [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=OlP8An2t15w

Equipo Fintech (2017). *Aplicaciones de la inteligencia artificial en el sector financiero.* Recuperado de https://www.fin-tech.es/wp-content/uploads/2017/09/Aplicaciones-de-inteligencia-artificial-en-el-sector-financiero.pdf

Etzkowitz, H. (2003). Innovation in Innovation: The Triple Helix of University-Industry-Government Relations. *Social Science Information*, *42*(3), 293-337.

Feenberg, A. (1991). *Critical Theory of Technology*. New York, United States: Oxford University Press.

Feito Grande, L. (2013). El debate ético sobre la mejora humana. *Diálogo Filosófico*, *8*, 45-51.

Fernández-Beltrán,F., García-Marzá, D., Sanahuja, R., Andrés Martínez, A. y Barberá Forcadell, S. (2017). La gestión de la comunicación para el impulso de la Investigación e Innovación Responsables: propuesta de protocolo desde la ética dialógica. *Revista Latina de Comunicación Social*, *72*, 1.040 a 1.062. Recuperado de http://www.revistalatinacs.org/072paper/1207/57es.html

Ferry, L. (2017). *La revolución transhumanista. Cómo la tecnomedicina y la uberización del mundo van a transformar nuestras vidas*. Madrid, España: Alianza.

Finlay, C. J. (2018). Just War, Cyber War, and the Concept of Violence. *Philosophy and Technology*, *31*, 357-377.

Flyvbjerg, B. (2001). *Making Social Science Matter: Why Social Inquiry Fails and How It Can Succeed Again*. Cambridge, England: Cambridge University Press.

Flyvbjerg, B. (2004). Cinco malentendidos acerca de la investigación mediante los estudios de caso. *Revista Española de Investigaciones Sociológicas*, *106*(4), 33-62.

Flyvbjerg, B (2006a). Social Science That Matters. *Foresight Europe*, *2*, 38-42.

Flyvbjerg, B. (2006b). Making Organization Research Matter: Power, Values, and Phronesis. En S. R. Clegg, C. Hardy, T. B. Lawrence y W. R. Nord (Eds.), *The Sage Handbook of Organization Studies* (370-387). California, United States: Thousand Oaks.

Floridi, L. (2002a). Information ethics: An environmental approach to the digital divide. *Philosophy in the Contemporary World*, *9*, 39-45.

Floridi, L. (2002b). On the intrinsic value of information objects and the infosphere. *Ethics and Information Technology*, *4*(4), 287-304.

Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics Information Technology*, *7*(4), 185-200.

Floridi, L. (2006a). Four challengues for a theory of informational privacy. *Ethics and Information Technology, 8*(3), 109-119.

Floridi, L. (2006b). Information technologies and the tragedy of the good will. *Ethics and Information Technology*, *8*(4), 253-262.

Floridi, L. (2007a). Global information ethics: The importance of being environmentally Earnest. *International Journal of Technology and Human Interaction*, *3*(3), 1-11.

Floridi, L. (2007b). Understanding information ethics. *American Psychology Association Newsletters*, *8*(2), 3-12.

Floridi, L. (2011). A Children of the Fourth Revolution. *Philosophy and Technology*, 24, 227-232.

Floridi, L. (2014). Techonological unemployment, leisure occupation, and the human project. *Philosophy and Technology*, *27*, 143-150.

Floridi, L. (2017). Robots, jobs, taxes, and responsabilities. *Philosophy and Technology*, *30*, 1-4.

Floridi, L. (2001). *The philosophy of information*. Oxford, England: Oxford University Press.

Ford, M. (2016). *El auge de los robots. La tecnología y la amenaza de un futuro sin empleo*. Barcelona, España: Paidós.

Forester, J. (1993). *Critical Theory, Public Policy, and Planning Practice: Toward a Critical Pragmatism*. Albany, United States: State University of New York Press.

Forester, J. (1999). *The Deliberative Practitioner*. Cambridge, MA, United States: MIT Press.

Forester, J. (2009). *Dealing with Differences: Dramas of Mediating Public Disputes*. Oxford, England: Oxford University Press.

Forge, J. (2010). A note on the definition of "dual use". *Science and Engineering Ethics*, *16*(1), 111-118.

Francisco, S. P. (2015). *Laudato si'*. Recuperado de: http://www.vatican.va/content/francesco/es/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html

Freeman, R. B. (2015). Knowledge, Knowledge… Knowledge for My Economy. *KDI Journal of Economic Policy*, *37*, 1-21.

Frey, C. B., y Osborne, M. A. (2013). The Future of Employment: How Susceptible are Jobs to Computerisation. *Oxford Martin School Working Paper*, *7.*

Frey, C. B. y Berger, T. (2015). *Technology, Globalisation and the Future of Work in Europe*. Oxford, England: Oxford University Press.

Frey, T. (2017). *Epiphany Z: Eight Radical Visions for Transforming Your Future*. New York, United States: Morgan James.

Friedman, B., Kahn, P. H. y Borning, A. (2006). Value sensitive design and information systems. En D. Galletta & P. Zhang (Eds.), *Human-computer interaction and management information systems: Applications*. New York, United States: M.E. Sharpe

Fromm, E. (1970). *La revolución de la esperanza*. México: Fondo de Cultura Económica.

Fukuyama, F. (2002). *El fin del hombre: consecuencias de la revolución biotecnológica.* Barcelona, España: Ediciones B.

Garcés, M. (2017). *Nueva ilustración radical*. Barcelona, España: Anagrama.

Garcés, M. (2019). *Humanidades en acción. Utilicemos el Aula oberta para saber, hacer y comprender*. Barcelona, España: Rayo Verde.

García Gibert, J. (2010). *Sobre el viejo humanismo. Exposición y defensa de una tradición*. Madrid, España: Marcial Pons.

García, M. (2018). Los laboratorios ciudadanos en los sistemas de experimentación e innovación. En Oliván, R., Pascale, P., Savazoni, R., Satana, B., Serrano, F., Mendonça, C., Güemes, C., Resina, J., Bantes, I., Ruiz, L., García, M., Peña-López, I., Coral, P., Apolaro, A.,

Ceballos, D., Fernández, J. y Romo, C. (Comp.), *Abrir instituciones desde dentro*. Zaragoza, España: Laboratorio de Aragón Gobierno Abierto. Recuperado de http://www.laaab.es/hackinginside/

García Inda, A. (2003). Ciudadanía y cultura de los derechos: el ciudadano consumidor. En M. J. Bernuz Benítez y R. Susín Betránn. *Ciudadanía: dinámicas de pertenencia y exclusión*. Logroño, España: Universidad de la Rioja.

García-Marzá, D. (2011). *Ética empresarial. Del diálogo a la confianza*. Madrid, España: Trotta.

Garlick, J. y Levine, P. (2016). Where civics meets science: Building science for the public good through Civic Science. *Oral Diseases*, *23* (6).

Geddes, P. (2009). *Ciudades en evolución*. Oviedo, España: KRK.

Genus, A. (2006). Rethinking constructive technology assessment as democratic, reflective, discourse. *Technological Forecasting and Social Chang*e, *73*, 13-26.

Giner, S. (2008). El destino de la sociedad civil. *Revista Española del Tercer Sector*, *10*, 17-49.

Gobierno de Santa Fe (2016-2019). SantaLab: Laboratorio de Innovación Pública. Provincia de Santa Fe, Argentina: Modernización del Estado. Recuperado de https://www.santafe.gob.ar/index.php/web/content/view/full/203591/(subtema)/93686

Goering, S. y Yuste R. (2016). On the Necessity of Ethical Guidelines for Novel Neurotechnologies, *Cell*, *167*(3), 882-885.

González Esteban, E. (2007). La teoría de los stakeholders. Un puente para el desarrollo práctico de la ética empresarial y de la responsabilidad social corporativa. *Veritas*, *17*(2), 205-224.

González Melado, F. J. (2010). *Transhumanismo (humanity +). La ideología que nos viene*. Recuperado de https://www.academia.edu/3621186/Transhumanismo_humanity_La_ideolog%C3%ADa_que_nos_viene

Gracia, D. (2016). Problemas con la deliberación. *Folia Humanistica. Revista de Salud, Ciencias Sociales y Humanidades*, *3*.

Griffin, A. (31 de julio de 2017). Facebook's Artificial Intelligence Robots shut down after they start talking to each other in their own language. *The Independent*. Recuperado de https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html

Grunwald, A. (2014). Technology Assessment for Responsible Innovation. En J. Van den Hoven, N. Doorn, T. Swierstra, B.-J. Koops y H. Jomijn (Eds.), *Responsible Innovation 1: Innovate Solutions for Global Issue*s (15-31). New York, United States: Springer.

Guisan, E. (2013). El utilitarismo. En V. Camps, O. Guariglia y F. Salmerón (Eds.), *Concepciones de la ética* (269-296). Madrid, España: Trotta.

Habermas, J. (2000). *Aclaraciones a la ética del discurso*. Madrid, España: Trotta.

Habermas, J. (2010a). *El futuro de la naturaleza humana*. Barcelona, España: Paidós.

Habermas, J. (2010b). *Teoría de la acción comunicativa*. Madrid, España: Trotta.

Hajer, M. (2009). *Authoritative Governance: Policy Making in the Age of Mediatizatio*n. Oxford, England: Oxford University Press.

Hamilton, K., Karahalios, K. y Langbort, C. (2016). When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software. *International Journal of Communication*, *10*, 4972-4990.

Harris, J. (2017). Los mejoramientos son una obligación moral. En N. Bostrom, y J. Savulescu (Eds.), *Mejoramiento humano* (137-161). Madrid, España: Teell.

Hayek, F. (2014). *Derecho, legislación y libertad*. Madrid: Unión.

Heidegger, M. (1958). *La época de la imagen del mundo*. Santiago de Chiler, Chile: Anales de la Universidad de Chile.

Heidegger, M. (1994). La pregunta por la técnica. En *Conferencias y artículos*. Barcelona, España: Odós.

Heidegger, M. (2002). *Serenidad*. Barcelona, España: Serbal.

Heidegger, M. (2003a). *Acerca del evento. Aportes a la filosofía*. Buenos Aires, Argentina: Biblioteca Internacional Martín Heidegger.

Heidegger, M. (2003b). *Introducción a la metafísica*. Barcelona, España: Gedisa.

Heidegger, M. (2004). *Carta sobre el humanismo*. Madrid, España: Alianza editorial.

Heidegger, M. (2017). *¿Qué es la filosofía*? Barcelona, España: Herder.

Henwood, K., Pidgeon, N. (2013). *Whatis the Relationship between Identity and Tech- nological, Economic, Demographic, Environmental and Political Change Viewed through a Risk Lens?* London: Government Office for Science Recuperado de https://www.gov.uk/government/publications/identity-and-risk

Hellström, T. (2007). Dimensions of Environmentally Sustainable Innovation: the Structure of Eco-Innovation Concepts. *Sustainable Development*, *15*, 148-159.

Hendrix, J. (2015). *Retreat from Range: The Rise and Fall of Carrier Aviation*. Washington DC, United States: Center for a New American Security.

Henwood, K. y Pidgeon, N. (2013) *What is the Relationship between Identity and Technological, Economic, Demographic, Environmental and Political Change Viewed through a Risk Lens?* London, England: Government Office for Science. Recuperado de https://www.gov.uk/government/publications/identity-and-risk

Hibbard, B. (2002). *Super-Intelligent Machines*. New York: Springer.

Hibbard, B. (2012). Avoiding Unintended AI Behaviors. En J. Bach, B. Goertzel y M. Iklé, M. (Eds.), *Artificial General Intelligence: 5th International Conference* (107-116). New York, United States: Springer, 107-116.

Hibbard, B. (2015). *Ethical Artificial Intelligence*. Recuperado de https://arxiv.org/ftp/arxiv/papers/1411/1411.1373.pdf

Hickman, L. A. (1992). *John Dewey's Pragmatic Technology.* New York, United States: John Wiley & Sons.

Horkheimer, M. (2010). *Crítica de la razón instrumental*. Madrid, España: Trotta.

Horkheimer, M. y Adorno, Th. W. (2016). *Dialéctica de la Ilustración. Fragmentos filosóficos*. Madrid, España: Trotta.

Humanitarian UAV Network. (2014). *UAV Code of Conduct: Humanitarian UAV Code of Conduct*. Recuperado de https://uavcode.org/

Humanity + (2018). *Transhumanist Declaration*. Recuperado de https://humanityplus.org/philosophy/transhumanist-declaration/

Huxley, A. (2017). *Literatura y ciencia. El humanismo frente al progreso científico y tecnológico.* Barcelona, España: Página Indómita.

IBM. (2014). *IBM Global Commuter Pain Survey: Traffic Congestion Down, Pain Way Up*. Recuperado de https://newsroom.ibm.com/2011-09-08-IBM-Global-Commuter-Pain-Survey-Traffic-Congestion-Down-Pain-Way-Up,1

Ida, R. (2017). ¿Deberíamos mejorar la naturaleza humana? Un interrogante planteado desde una perspectiva asiática. En N. Bostrom y J. Savulescu (Eds.), *Mejoramiento humano* (63-74). Madrid, España: Teell.

Institute for the Future y Dell Technologies (2018). *The Next Era of Human□Machine Partnerships.* Recuperado de http://www.iftf.org/fileadmin/user_upload/downloads/th/SR1940_IFTFforDellTechnologies_Human-Machine_070717_readerhigh-res.pdf

Irwin, A. (2006). The politics of talk: coming to terms with the 'new' scientific governance. *Social Studies of Science*, *36*, 299-330.

Istvan, Z. (2013). *The Transhumanist Wager*. New York, United States: Futurity Imagine Media

Izuzquiza, I. (2000). *Caleidoscopios. La filosofía occidental en la segunda mitad del siglo XX*. Madrid, España: Alianza.

Jasanoff, S. (2003) Technologies of humility: citizen participation in governing science. *Minerva, 41*, 223-244.

Jensen, S. J. y Widow, J. L. (2018). Unnatural Enhancements. *SAGE Journals*, *83*(4), 347-363.

Jonas, H. (1995). *El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica*. Barcelona, España: Herder.

Jonas, H. (1997). *Técnica, medicina y ética. Sobre la práctica del principio de responsabilidad.* Barcelona, España: Paidós.

Jonas, H. (2001). *Más cerca del perverso fin y otros diálogos y ensayos*. Madrid, España: Los libros de la Catarata.

Jonas, H. (2005). *Memorias*. Oviedo, España: Losada.

Jordán, J. y Baqués, J. (2014). *Guerra de drones. Política, tecnología y cambio social en los nuevos conflictos*. Madrid, España: Biblioteca Nueva.

Jonsen, A. R. (2016). Razonamiento casuístico en la ética médica. *Dilemata*, *20*, 1-14.

Johnson, D. G. (2007). Ethics and technology ‘‘in the making’’: An essay on the challenge of nanoethics. *NanoEthics*, *1*(1), 21-30.

Johnson, D. G. (2011). Software agents, anticipatory ethics, and accountability. En G. E. Marchant, B. R. Allenby y J. R. Herkert (Eds.), *The growing gap between emerging technologies and legalethical oversight* (61-76). Netherlands: Springer.

Joshi, N. (2019). Can AI Become Our New Cybersecurity Sheriff? *Forbes*. Recuperado de https://www.forbes.com/sites/cognitiveworld/2019/02/04/can-ai-become-our-new-cybersecurity-sheriff/#22d7c0a936a8

Kant, I. (1999). *Fundamentación metafísica de las costumbres*. Barcelona, España: Ariel.

Kaplan, J. (2016). *Abstenerse humanos. Guía para la riqueza y el trabajo en la era de la inteligencia artificial*. Madrid, España: Teell.

Kaplan, J. (2017). *Inteligencia Artificial. Lo que todo el mundo debe saber*. Madrid, España: Teell.

Karinen, R. y Guston, D. (2010) Toward anticipatory governance: the experience with nanotechnology. Governing future technologies. *Sociology of the Sciences Year book, 27*, 217-232.

Kartha, S.; Siebert, C.K; Mathur, R.; Nakicenovic, N.; Ramanathan, V.; Rockström, J.; Schellnhuber, H.J; Srivastava, L.; Watt, R. (2009). *A Copenhagen Prognosis: Towards a safe climate future*. Report by the Potsdam Institute for Climate Impact Research, Stockholm Environment Institute, and The Energy and Resources Institute. Recuperado de https://mediamanager.sei.org/documents/Publications/Climate-mitigation-adaptation/a20copenhagen20prognosis.pdf

Kasparov, G. (2017). *Deep Thinking. Where Machine Intelligence Ends*. Croydon, England: John Murray.

Kelly, K. (2017). *Lo inevitable. Entender las 12 fuerzas tecnológicas que configurarán nuestro futuro*. Madrid, España: Teell.

Keynes, J. M. (1930). *Economic Possibilities for Our Grandchildren. In Essays in Persuasion*. New York, United States: Norton & Co.

Keynes, J. M. (2010). *Tratado sobre el dinero*. Madrid, España: Síntesis.

Knobel, C. P. y Bowker, G. C. (2011). Values in design. *Communications of the ACM*, *54*(7), 26-28.

Kravetz, K. (2013). Civic Studies: Bringing Theory to Practice. *The Good Society*, *22*(2), 162-171.

Krutzinna, J. (2016). Can a welfarist approach be usted to justify a moral duty to cognitively enhance children? *Bioethic*, *30*(7), 528-535.

Kurzweil, R. (2016a). *Cómo crear una mente. El secreto del pensamiento humano.* Berlín, Alemania: Lola Books.

Kurzweil, R. (2016b). *La singularidad está cerca*. *Cuando los seres humanos trascendamos la biología.* Berlín, Alemania: Lola Books.

Hawking, S., Russell, S., Tegmark, M. y Wilczek F. (1 de mayo de 2014). Stephen Hawking: "Transcendence looks at the implications of artificial intelligence – but are we taking AI seriously enough?". *The Independent*. Recuperado de https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-9313474.html

Lasalle, J. M. (2019). *Ciberleviatán. El colapso de la democracia liberal frente a la revolución digital*. Barcelona, España: Arpa Editores.

Latorre, J. I. (2019). *Ética para máquinas*. Barcelona, España: Ariel.

Lee Stayton, E. (2011). *Driverless Dreams: Technological Narratives and the Shape of the Automated Car*. Massachusetts, United States: MIT Press.

Leontief, W. W. (1982). The Distribution of Work and Income. *Scientific American*, *247*(3), 188-204

Leontief, W. (1986). *Ensayos sobre economía*. Barcelona, España: Planeta-Agostini.

Levine, P. (2014). Civic Studies. *Philosophy and Public Policy Quarterly*, *31*(1), 29-33.

Lévy-Leblond, J. M. (1975). *La ideología de/en la física contemporánea y ostros ensayos*. Barcelona, España: Anagrama.

Leydesdorff, L. y Etzkowitz, H. (2000). Le "Mode 2" et la globalisation des systèmes d'innovation "nationaux": le modèle à Triple hélice des relations entre université, industrie et gouvernement. *Sociologie et sociétés*, *32*(1), 135-156. Recuperado de http://goo.gl/t1dMn

Limón, R. (12 de junio de 2018). El coche del accidente mortal de Uber tenía inhabilitada la frenada de emergencia. *El país.* Recuperado de https://elpais.com/tecnologia/2018/06/12/actualidad/1528798311_923424.html

Lin, J. (2011). Technological Adaptation, Cities, and New Work. *Review of Economics and Statistics*, *93*(2), 554-574.

Lin, P., Bekey, G. y Abney K. (2012). *Robot Ethics: The Ethical and Social Implications of Robotics*. Massachusetts, United States: MIT Press.

Locke, J. (2006). *Segundo tratado sobre el gobierno civil*. Madrid, España: Tecnos.

Lozano Aguilar, J. F. (2004). *Códigos éticos para el mundo empresarial*. Madrid, España: Trotta.

Lozano Aguilar, J. F. (2011). *Qué es la ética de la empresa*. Barcelona, España: Proteus.

Lunden, I. (2017). SoftBank is buying robotics firms Boston Dynamics and Schaft from Alphabet. *TechCrunch*. Recuperado de https://techcrunch.com/2017/06/08/softbank-is-buying-robotics-firm-boston-dynamics-and-schaft-from-alphabet/

Luterbacher, C. (5 de junio de 2018). How drones are transforming humanitarian aid. *Swissinfo.ch* [Mensaje en un blog]. Recuperado de https://www.swissinfo.ch/eng/sci-tech/-dronefrontier_how-drones-are-transforming-humanitarian-aid/44141254

Lynch, M. (2000) Against reflexivity as an academic virtue and source of privileged knowledge. Theory, *Culture & Society, 17*, 26-54.

McKinsey & Company (2017). Self-driving car technology: When Will the robots hit the road? Our Insights. Recuperado de https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/self-driving-car-technology-when-will-the-robots-hit-the-road

Maestro Bäcksbacka, F. J. (2008). «El dilemma norteamericano»: de la esclavitud a la institucionalización de la discriminación racial. *Studia histórica. Historia contemporánea*, *26*, 53-78.

Malyuk, A. y Miloslavskaya, N. (2016). Cybersecurity culture as an element of IT professional training. *Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC).* Recuperado en https://ieeexplore.ieee.org/document/7529390

Manjikian, M. (2017). *A Typology of Arguments about Drone Ethics*. Carlisle, England: Strategic Studies Institute.

Marcos, A. (2018). Bases filosóficas para una crítica al transhumanismo. *Revista Artefactos*, *7*(2), 107-125.

Markoff, J. (2015). *Machines of Loving Grace. The Quest for Common Ground Between Humans and Robots*. Nueva York, United States: HarperCollins.

Martínez Navarro, E. (1999). *Solidaridad liberal. La propuesta de John Rawls*. Granada, España: Comares.

Massachusetts Institute of Technology (2016). *Moral Machine*. Recuperado de http://moralmachine.mit.edu/hl/es.

Mateo, J. L. (2006). Sociedad del Conocimiento. *Arbor: Ciencia, pensamiento y cultura*, *718,* 145-151.

Max-Neef, M. A. (1993). *Desarrollo a escala humana. Conceptos, aplicaciones y algunas reflexiones*. Barcelona, España: Icaria Editorial.

Mayz Vallenilla, E. (1984). *El sueño del futuro*. Caracas, Venezuela: Ateneo de Caracas.

McCarthy, J., Minsky, M. L, Rochester, N. y Shannon C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*, *27*(4), 12-14.

McCorduck, P. (1991). *Máquinas que piensan*. Barcelona, España: Tecnos

Meiches, B. (2019). Non-human humanitarians. *Review of International Studies*, *45*(1), 1-19.

Mendieta, E. (2002). El debate sobre el futuro de la especie humana: Habermas critica la eugenesia liberal*. Isegoría*, *27*, 91-114.

Metz, C. (26 de agosto de 2018). Artificial Intelligence Is Now a Pentagon Priority. Will Silicon Valley Help? *El País*. Recuperado de https://www.nytimes.com/2018/08/26/technology/pentagon-artificial-intelligence.html

Microsoft (2018). *AI for Earth*. Recuperado de https://www.microsoft.com/en-us/ai/ai-for-earth

Midgley, M. (1994). *The Ethical Primate: Humans, Freedom and Morality*. London, England: Routledge.

Ministerio del Interior de España (2018). Anuario estadístico del Ministerio del Interior. Recuperado de https://estadisticasdecriminalidad.ses.mir.es/jaxiPx/Datos.htm?path=/Datos5//l0/&file=05001.px&type=pcaxis

Minsky, M. L. (1967). *Computation: Finite and Infinite Machines*. New York, United States: Prentice-Hall.

Minsky, M. L. (2010). *La máquina de las emociones. Sentido común, inteligencia artificial y el futuro de la mente*. Bogotá, Colombia: Debate.

Misselhorn, C. (2018). Artificial Morality. Concepts, Issues and Challenges. *Society*, *55*, 161-169.

Mitcham, C. (1989). *¿Qué es la filosofía de la tecnología?* Barcelona, España: Antrophos

Mitcham, C. y Mackey, R. (2004). *Filosofía y tecnología*. Madrid, España: Encuentro.

Monserrat, J. (2015). El transhumanismo de Ray Kurzweil. ¿Es la ontología biológica reductible a computación? *Pensamiento: Revista de investigación e información filosófica, 71*(209), 1417-1441.

Moliner González, J. A. (2018). Algunos problemas éticos de las tecnologías militares emergentes. Bie3: Boletín I.E.E.E., *9*, 522-541.

Molinuevo, J. L. (2004). *Humanismo y nuevas tecnologías*. Madrid, España: Alianza.

Moor J. H. (2006) The nature, importance and difficulty of machine ethics. IEEE Intell Syst, 21(4), 18-21.

Morales, A. C. (2008). Innovación social: un proceso emergente en las dinámicas de desarrollo. *Revista de Fomento Social*, *63*,411-444.

Morales, A. C. (2009a). Innovación social: un ámbito de interés para los servicios sociales. Zerbitzuan: *Revista de servicios sociales*, *45*, 151-175.

Morales, A. C. (2009b). Innovación «abierta» en el tercer sector: el modelo organizativo 2.0. *Revista Española del Tercer Sector*, *13*, 17-37.

Morales, A. C. (2012). Innovación social y cooperativas. Convergencias y sinergias. *Ekonomiaz: Revista Vasca de Economía*, *79*, 146-167.

Moravec, H. (1988). *El hombre mecánico. El futuro de la robótica y la inteligencia humana*. Madrid, España: Ediciones Temas de Hoy.

More, M. (1990). Transhumanism: Toward a Futurist Philosophy. *Extropy*, *6*, 6-12. Recuperado de http://fennetic.net/irc/extropy/ext6.pdf

Mumford, L. (2011). *El mito de la máquina*. Logroño, España: Pepitas de Calabaza.

Müller, J. P. (1996). *The Design of Intelligent Agents: A Layered Approach*. Berlin, Germany: Springer.

Müller, W. E. (1989). Zur Problematik der Verantwortungsbegriffes bei Hans Jonas. *Zetschrif für evangelische Ethik*, *3*3, 204-216.

Müller, V. C. y Bostrom, N. (2014). Future progress in artificial intelligence: A Survey of Expert Opinion. En V. C. Müller (Ed.), *Fundamental Issues of Artificial Intelligence* (555-572). Berlín, Germany: Springer.

Muñoz-Alonso López, G. (1997). La evaluación de tecnologías (ET): origen y desarrollo. *Revista General de Información y Documentación*, *7*(1), 15-30.

Nadella, S. (28 de junio de 2016). The Partnership of the Future [Mensaje de un blog]. Recuperado de https://slate.com/technology/2016/06/microsoft-ceo-satya-nadella-humans-and-a-i-can-work-together-to-solve-societys-challenges.html

Nath, R. y Sahu, V. (2017). The problem of machine ethics in artificial intelligence. *AI & Society*, *1*, 1-9.

Navarro, P. A. (2015). La brecha racial USA: los asesinatos de ciudadanos negros a manos de políticas blancos desatan la tensión. *El siglo de Europa*, *1109*, 32-38.

Nedon, V. (2014). *Open Innovation in R&D Departments*. Hamburg, Germany: Springer.

Newell, A. y Simon, H. A. (1974). Simulación del pensamiento humano. *Revista Teorema*, *4*(3), 335-378.

Nietzsche, F. (1985). *Así habló Zaratustra*. Madrid, España: Alianza.

Novak, M. (16 de mayo de 2013). The National Automated Highway System That Almost Was. *Smithsonian*. Recuperado de http://www.smithsonianmag.com/history/thenational-automated-highway-system-that-almost-was-63027245/

Nussbaum, M. (2010). *Sin fines de lucro. Por qué la democracia necesita de las humanidades.* Buenos Aires, Argentina: Katz.

Observatorio CETELEM (2016). *El coche autónomo. Los conductores, dispuestos a ceder la conducción a la tecnología*. Recuperado de https://elobservatoriocetelem.es/wp-content/uploads/2016/03/observatorio_cetelem_auto_2016.pdf

Oelmüller, W. (1988). Hans Jonas. Mythos-Gnosis-Prinzip Verantwortung. *Stimmen der Zeit*, *5*(206), 343-351.

Olivares Dysli, L. B. (2014). Desarrollo Sostenible, Defensa y Seguridad Nacional. *Bie3: Boletín I.E.E.E.* Recuperado de http://www.ieee.es/Galerias/fichero/docs_opinion/2014/DIEEEO119-2014_DesarrolloSostenible_Domingo_Olivaress.pdf

Open Data en Europa y Asia Central (2016). About Open Data en Europa y Asia Central. Recuperado de https://www.odecanet.org/about/.

Organización de Naciones Unidas (1948). *Declaración Universal de los Derechos Humanos*. Recuperado de http://www.un.org/es/universal-declaration-human-rights/#health

Organización de Naciones Unidades (2003). *Creación de una cultura global de ciberseguridad*. Recuperado de https://www.itu.int/ITU-D/cyb/cybersecurity/docs/UN_resolution_57_239.pdf

Organización de Naciones Unidas (2015). *Objetivos de Desarrollo Sostenible*. Recuperado de https://www.undp.org/content/undp/es/home/sustainable-development-goals.html

Organización de Naciones Unidas (2017). *Mandato del Relator Especial sobre ejecuciones extrajudiciales, sumarias o arbitrarias*. Recuperado de http://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/RES/35/15

Ortega, A. (2016). *La imparable marcha de los robots*. Madrid, España: Alianza.

Ortega y Gasset, J. (1965). El mito del hombre allende la técnica. En Ortega y Gasset, J., *Obras completas* (617-624), tomo IX. Madrid, España: Revista de Occidente.

Ortega y Gasset, J. (1965). *Meditación de la técnica*. Madrid, España: Espasa-Calpe.

Ortega y Gasset, J. (2008). *La rebelión de las masas*. Madrid, España: Tecnos.

Ortiz Llueca, E. (2013). Bioética personalista y bioética utilitarista. *Cuadernos de Bioética*, *24*(1), 57-65.

Overall, C. (2009). Life Enhancement Technologies: Significance of Social Category Membership. En N. Bostrom y J. Savulecu (Eds.), *Human Echancement* (327-340). Oxford, England: Oxford University Press.

Oxford Economics (2019). *How Robots Change the World. What Automation Really Means for Jobs and Productivity*. Recuperado de https://n9.cl/zljk

Palavicini Corona, G. y Cepeda Mayorga, I. (Coord.). (2019). *Innovación y Emprendimiento Social en Instituciones de Educación Superior: Students4Change*. Ciudad de México, México: Instituto Tecnológico y de Estudios Superiores de Monterrey. Recuperado de https://www.uestudents4change.org/repositorio

Palm, E. y Hansson, S. O. (2006). The case for ethical technology assessment (eTA). *Technological Forecasting & Social Change*, *73*, 543 - 558.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. London, England: Penguin Books.

Parlamento Europeo (2017). *Resolución de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)*. Recuperado de https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=fr&reference=2016/2225(INI)

Parra, N.-H. y Arenas-Dolz, F. (2015). *Revolución tecnológica y democracia del conocimiento. Por una universidad innovadora*. Valencia, España: Laboratorio de la Sociedad del Conocimiento.

Passi, H. (2018). How will Artificial Intelligence and Machine Learning redefine and transform CyberSecurity? *Greycampus*. Recuperado de https://www.greycampus.com/blog/information-security/how-will-artificial-intelligence-and-machine-learning-redefine-and-transform-cybersecurity

Pateman, C. (1970). *Participación and Democratic Theory*. Cambridge, England: Cambridge University Press.

Pateman, C. (1985). *The Problem of Political Obligation: A Critique of Liberal Theory*. Cambridge, England: Polity Press.

Persad, G. 201: Downward mobility and Rawlsian justice. *Philosophical Studies*, *175*(2), 277-300.

Passy, J. (9 de abril de 2017). This i show many U.S. Jobs robots will créate over the next 10 year. *Market Watch*. Recuperado de https://www.marketwatch.com/story/this-is-how-many-us-jobs-robots-and-automation-will-create-over-the-next-10-years-2017-04-04

Peña, M. (8 de mayo de 2018). Waymo explica el rol esencial de la inteligencia artificial en la tecnología autónoma. *Digital Trends*. Recuperado de https://es.digitaltrends.com/autos/waymo-inteligencia-artificial-autos-autonomos/

Phills, J.A; Deiglmeier, K. y Miller, D.T. (2008) Rediscovering Social Innovation. *Stanford Social Innovation Review*, *6*(4), 34-43.

Pigem, J. (2013). *La nueva realidad*. Barcelona, España: Kairós.

Popper, K. (1994). *Conjeturas y refutaciones. El desarrollo del conocimiento científico*. Barcelona, España: Paidós.

Proaño Cortez, M. F. (2008). *Construyendo roles. Democracia y Fuerzas Armadas*. Buenos Aires, Argentina: CELS.

Programa de las Naciones Unidas para el Desarrollo (2000). *Objetivos del Desarrollo del Milenio.* Recuperado de https://www.undp.org/content/undp/es/home/sdgoverview/mdg_goals.html

Queraltó, R. (2003). *Etica, tecnología y valores en la sociedad global. El caballo de Troya al revés*. Madrid, España: Tecnos.

Quintanilla, M. Á. (2002). La democracia tecnológica. *Arbor: Ciencia, pensamiento y cultura*, *683-684*, 637-652.

Quintanilla, M. Á. (2006). *Tecnología: un enfoque filosófico*. Ciudad de México, México: Fondo de Cultura Económica.

Quintanilla, M. Á., Parselis, M., Sandrone, D. y Lawler, D. (2017). *Tecnologías entrañables. ¿Es posible un modelo alternativo de desarrollo tecnológico?* Madrid, España: Los Libros de la Catarata.

Radnitzky, Gerard (1973). Hacia una teoría de la investigación que no es ni reconstrucción lógica ni psicología o sociología de la ciencia. *Teorema, 3*, 197-264.

Rapp, F. (1985). Humanism and Technology. The Two-Cultures Debate. *Technology in Society*, *7*, 423-435.

Rawls, J. (1997). *Teoría de la justicia*. Ciudad de México, México: Fondo de Cultura Económica.

Raworth, K., (2017). A Doughnut for the Anthropocene: humanity's compass in the 21st century. *The Lancet*, *1*(2), 48-49.

Regan, T. (1983). *The case for animal rights*. Berkeley, United States: The University of California Press.

Rid, T. (2012). Cyber war will not take place. *Journal of Strategic Studies*, *35*(1), 5-32.

Ridley, M. (2016). *Artificial intelligence is not going to cause mass unemployment*. Recuperado de http://www.rationaloptimist.com/blog/artificial-intelligence/

Rip, A., Misa, T. y Schot, J. (Eds.) (1995). *Managing Technology in Society: The Approach of Constructive Technology Assessment*. London, England: Thomson.

Ritter, D. (2016). It's up to organised people to ensure the new economy serves the greater good. *The Guardian* [Mensaje en un blog]. Recuperado de https://www.theguardian.com/sustainable-business/2016/oct/07/its-up-to-organised-people-to-ensure-the-new-economy-serves-the-greater-good

Rip, A. (2001a). Assessing the impact of innovation: new developments in technology assessment. En *OECD Proceedings: Social Sciences and Innovation* (197-213). Paris, Francia: OECD.

Rip, A. (2001b). Contributions from Social Studies of Science and Constructive Technology Assessment. En A. Stirling (Ed.,. *On Science and Precaution in the Management of Technological Risk, Vol. II: Case Studies, Institute for Prospective Technology Studies* (94-122). Sevilla, España: European Commission Joint Research Centre.

Rodríguez García, R. (1991). *Heidegger y la crisis de la época moderna*. Madrid, España: Editorial Cincel.

Rodríguez Gil, G. (7 de julio de 2017). España alcanza un nuevo récord de envejecimiento con 118 mayores por cada 100 menores de 16 años. *Expansión*. Recuperado de http://www.expansion.com/economia/politica/2017/07/07/595e2ca8468aeb3a398b45eb.html

Rosales Rodríguez, A. (2004). Naturaleza orgánica y responsabilidad ética: Hans Jonas y sus críticos. *Transformação*, *27*(2), 97-111.

Rouhiainen, L. (2018). *Inteligencia artificial. 101 cosas que debes saber hoy sobre nuestro futuro*. Barcelona, España: Planeta.

Rowe, G. y Frewer, L. J. (2000). Public Participation Methods: A Framework for Evaluation. *Science, Technology, and HumanValues*, *25*(1), 3-27.

Ruiz Domínguez, F. (2017). La implantación del automóvil inteligente: ¿un riesgo calculado para la seguridad global? *Instituto Español de Estudios Estratégicos*. Recuperado de http://www.ieee.es/Galerias/fichero/docs_opinion/2017/DIEEEO60-2017_Automovil_Inteligente_FRuizDominguez.pdf

Ruiz Marcos, L. (2018). *Experimentar en las instituciones culturales. El ejemplo de los laboratorios ciudadanos*. Zaragoza, España: Laboratorio de Aragón Gobierno Abierto. Recuperado de http://www.laaab.es/2018/07/experimentar-en-las-instituciones-culturales-el-ejemplo-de-los-laboratorios-ciudadanos/

Russel, S. J. y Norvig, P. (1996). *Inteligencia Artificial: un enfoque moderno*. México: Prentice-Hall.

SAE International (2019). *Levels of Driving*. Recuperado de https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic

Salgado, V. (2017). *Derechos humanos e inteligencia artificial. Leyes de la robótica en la UE*. Recuperado de https://pintos-salgado.com/2017/03/17/derechos-humanos-e-inteligencia-artificial-leyes-de-la-robotica-en-la-ue/

Sánchez, C. (12 de abril de 2018). "Parar de investigar el coche autónomo por los accidentes sería un gran error". *Eldiario.es*. Recuperado de https://www.eldiario.es/hojaderouter/movilidad/Parar-investigar-coche-autonomo-accidentes_0_760124707.html

Sánchez Medero, G. (2010). Los Estados y la ciberguerra. *Boletín de Información*, *317*, 63-75.

Sandberg, A., Bradshaw-Martin, H. y Gérardin-Laverge, M. (2015). La voiture autonome et ses implications morales. *Multitudes*, *1*(58), 62-68.

Sandel, M. (2016). *Contra la perfección. La ética en la era de la ingeniería genética.* Barcelona, España: Marbot.

Sandel, M. (2017). Contra la perfección: lo que pasa con los niños de diseño, los atletas biónicos y la ingeniería genética. En N. Bostrom y J. Savulescu (Eds.), *Mejoramiento humano* (75-94). Madrid, España: Teell.

Sanders, M. E. (2006). *A rationale for new approaches to STEM education and STEM education graduate programs.* Invited paper presented at the 93rd meeting of the Mississippi Valley Technology Education Conference, Nashville, TN.

Sanders, M. E. (2009). Integrative STEM: Primer [in some places titled STEM, STEM Education, STEMmania]. *The Technology Teacher*, *68*(4), 20-26.

Sanmartín, J. (1990). *Tecnología y futuro humano*. Barcelona: Anthropos.

Sarewitz, D. y Karas, T. H. (2007). Policy Implications of Technologies for Cognitive Enhancement, *Sandia National Laboratories*. Recuperado de https://prod-ng.sandia.gov/techlib-noauth/access-control.cgi/2006/067909.pdf.

Savulescu, J. (2012). *¿Decisiones peligrosas? Una bioética desafiante*. Madrid, España: Tecnos.

Schomberg, R. von (2011). Prospects for technology assessment in a framework of responsible research and innovation. En M. Dusseldorp and R. Beecroft (Eds), *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden* (39-61). Wiesbaden, Germany: Vs Verlag.

School of Arts and Sciences. Tufts University (2009). *Civic Studies*. Recuperado de https://as.tufts.edu/civicstudies/about

Schwab, K. (2016). *La Cuarta Revolución Industrial*. Barcelona, España: Debate.

Schutz, Alfred (1974). *El problema de la realidad social*. Buenos Aires, Argentina: Amorrortu.

Sennett, Richard (2000). *La corrosión del carácter. Las consecuencias personales del trabajo en el nuevo capitalismo*. Barcelona, España: Anagrama

Sepúlveda Férriz, J. L. y Domingo Moratalla, T. (2011). La transformación del obrar humano en la época de la civilización tecnológica y la exigencia de una nueva ética. *Principios: Revista de Filosofía*, *18*(30), 5-26.

Shilton, K. (2014). Anticipatory Ethics for a Future Internet: Analyzing Values During the Design of an Internet Infrastructure. *Sci Eng Ethics*, *21*, 1-18.

Simondon, G. (2009). *La individuación a la luz de las nociones de forma y de información*. Buenos Aires, Argentina: Cactus-La Cebra.

Singer, P. W. (1975). *Animal liberation*. New York, United States: Random House

Singer, P. W. (2009a). *Wired for War: The Robotics Revolution and 21st Century Conflict*. London, England: Penguin.

Singer, P. W. (2009b). Wired for war? Robots and military doctrine. *Joint Force Quarterly*, *52*, 105-110.

Singer, P. W. y Friedman, A. (2014). *Cybersecurity and Cyberwar: What Everyone Needs to Know*. Oxford, England: Oxford University Press.

Sitra (2017). *Megatrendit 2017*. Recuperado de https://www.sitra.fi/aiheet/megatrendit/#megatrendit-2017

Siurana, J. C. (2003). *Una brújula para la vida moral*. Granada, España: Comares.

Skeem, J. L. y Lowenkamp Ch. (2016). Risk, Race, and Recidivism. Predictive Bias and Disparate Impact. *Criminology*, *54*(4), 680-712.

Sloterdijk, P. (2011). *Sin salvación. Tras las huellas de Heidegger*. Madrid, España: Akal.

Smirl, L. (2015). *Spaces of Aid: How Cars, Compounds and Hotels Shape Humanitarianism.* London, England: Zed Books.

Smith, B. y Shum H. (2018). *The Future Computed. Artificial Intelligence and its role in society*. Washington, United States: Microsoft Corporation.

Snow, C. P. (2013). *The two cultures and the scientific revolution*. Eastford, United States: Martino Fine Books.

Sociedad de Ingenieros de Automoción (2019). *Levels of Driving Automation*. Recuperado de https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic

Sousa, D. A. y Pilecki, T. (2013). *From STEM to STEAM: Using Brain-Compatible Strategies to Integrate the Arts*. Thousand Oaks. United States: SAGE.

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., Biggs, R., Carpenter, S. T., de Vries, W., de Wit, C. A., Foke, C., Gerten, D., Heinke, J., Mace, G. M., Persson, L. M., Ramanathan, V., Reyers, B. y Sörlin, S. (2015). Planetary boundaries: Guiding human development on a changing planet. *Science*, *349*(6254), 1286-1287.

Stern, N. (2009). *The global deal: climate change and the creation of a new era of progress and prosperity*. New York, United States: Public Affairs.

Stilgoe, J., Owen, R. y Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*, 1568-1580.

Susskind, R. y Susskind, D. (2016). *El futuro de las profesiones. Cómo la tecnología transformará el trabajo de los expertos humanos*. Madrid, España: Teell.

Swarte, T., Boufous, O. y Escalle, P. (2019). Artificial intelligence, ethics and human values: the cases of military drones and companion robots. *Artificial Life and Robotics*, *1*, 1-6.

Sweetman, B. (2010). *Beast Sighted In Korea*. *Aviation Week*. Recuperado de https://web.archive.org/web/20110813065330/http://www.aviationweek.com/aw/blogs/defense/index.jsp?plckController=Blog&plckBlogPage=BlogViewPost&newspaperUserId=27ec4a53-dcc8-42d0-bd3a-01329aef79a7&plckPostId=Blog%3a27ec4a53-dcc8-42d0-bd3a-01329aef79a7Post%3a088e4448-9e53-492f-8a14-7671361e1743&plckScript=blogScript&plckElementId=blogDest

Swiss Foundation for Mine Action y European Union (2017). *Drones in Humanitarian Action. A guide to the use of airborne systems in humanitarian crises.* Recuperado de https://drones.fsd.ch/wp-content/uploads/2016/11/Drones-in-Humanitarian-Action.pdf

Tännsjö, T. (2017). El mejoramiento médico y los valores del deporte de élite. En N. Bostrom y J. Savulescu (Eds.). *Mejoramiento humano* (327-338). Madrid, España: Teell.

Taylor, J. B. (1970) Introducing Social Innovation. 1970. *The Journal of Applied Behavioral Science*, *6*(6), 69-77.

Tegmark, M. (2017). *Life 3.0: being human in the age of Artificial Intelligence*. London, England: Penguin Random House.

Terrones Rodríguez, A. L. (2018). Deliberación, responsabilidad y prudencia: fundamentos para construir una ética aplicada a la inteligencia artificial. *Revista Estudios*, *36*.

Thompson, M. (2013). Costly Flight Hours. *Time*. Recuperado de http://nation.time.com/2013/04/02/costly-flight-hours/

Towarnicki, F. y Palnier, J.-M. (1981). Conversación con Heidegger. *Revista Palos de la Crítica*, 4. Recuperado de http://www.bolivare.unam.mx/cursos/TextosCurso10-1/CONVERSACI%D3N%20CON%20HEIDEGGER.pdf

Trías, E. (2000). *Ética y condición humana*. Barcelona, España: Península.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *49*, 433-460.

UAViators (2014). *UAV Code of Conduct*. Recuperado de https://uavcode.org/code-of-conduct/

Uhlemann, E. (2018). Time for Autonomous Vehicles to Connect [Connected Vehicles]. *IEEE Vehicular Technology Magazine*, *13*(3), 10-13.

Ulmer, B. (1994). VITA II - Active Collision Avoidance in Traffic. *Proceedings of the Intelligent Vehicles '94 Symposium*, 1-6.

Unamuno, M. de (2013). *Del sentimiento trágico de la vida en los hombres y en los pueblos*. Madrid, España: Alianza.

Université de Montréal (2018). *Montreal Declaration for responsible AI development*. Recuperado de https://docs.wixstatic.com/ugd/ebc3a3_c5c1c196fc164756afb92466c081d7ae.pdf

Urra Canales, M. (2017). *Estado, mercado, academia… y comunidad. Una cuádruple hélice para el desarrollo integral y la innovación* (tesis doctoral). Universidad Pontificia Comillas, Madrid. Recuperado de https://repositorio.comillas.edu/xmlui/handle/11531/26826

Vaccari, A. (2010). Vida, técnica y naturaleza en el pensamiento de Gilbert Simondon. *Revista iberoamericana de ciencia, tecnología y sociedad*, *5*(14), 153-165.

Vallor, S. (2015). Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. *Philosophy and Technology*, *28*, 107-124.

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford, England: Oxford University Press

Vega Reñón, L. (2016). Variaciones sobre la deliberación. *Dilemata*, *22*, 203-230.

Vidarte, F. J. y Rampérez, J. F. (2005). *Filosofías del siglo XX*. Madrid, España: Síntesis.

Vincent, J. (2 de abril de 2018). Badly implemented AI could 'jeopardize democracy', says French president Emmanuel Macron. *The Verge*. Recuperado de https://www.theverge.com/2018/4/2/17187736/france-ai-strategy-emmanuel-macron-dangers-democracy

Wadhwa, V. (2017). *There is a jobs crisis brewing that the Trump administration should not ignore.* Recuperado de http://wadhwa.com/2017/04/01/jobs-crisis-brewing-trump-administration-not-ignore/

Walker, J. (2014). Robots Don't Drink and Drive. Recuperado de https://rmi.org/robots-dont-drink-drive/

Walker, W. R. y Herrmann, D. J. (Eds.) (2005). *Cognitive Technology: Essays On the Transformation of Thought and Society*. Jefferson, United States: McFarland & Company.

Walsh, T. (2017). *Android Dreams: The Past, Present and Future of Artificial Intelligence*. London, England: C. Hurst & Co.

Walton, D. N. (2004). Criteria of rationality for evaluating democratic public rhetoric. En B. Fontana, C. J. Nederman y G. Reimer (Eds.), *Talking democracy* (295-330). Pennsylvania, United States: University Park PA.

Walton, D. N. (2006). How to make and defend a proposal in deliberation dialogue. *Artificial Intelligence and Law*, *14*, 117-239.

Weld, D. y Etzioni, O. (2009). The first law of robotics (A Call to Arms). En M. Barley, H. Mouratidis, A. Unruh, D. Spears, P. Scerri y F. Massacci (Eds.), *Safety and Security in Multiagent Systems* (90-100). New York, United States: Springer.

Werner, M. H. (2003). Hans Jonas 'Prinzip Verantwortung'. En M. Düwell y K. Steigleder. *Bioethik. Eine Einführung* (41-56). Frankfurt , Germany: Suhrkamp.

Wetmore, J. (2003). Driving the Dream: The History and Motivations Behind 60 Years of Automated Highway Systems. *Automotive History Review*. Recuperado de https://cspo.org/library/driving-the-dream-the-history-and-motivations-behind-60-years-of-automated-highway-systems-in-america/

Wiener, N. (1949). *The Machine Age.* Cambridge, England: MIT Press.

Wiener, N. (1985). *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, England: MIT Press.

Wiener, N. (1989). *The Human Use Of Human Beings: Cybernetics And Society*. London, England: Free Association Books.

Wilson, S. W. (1991). The animat path to AI (15-21). En J. A. Meyer y S. W. Wilson (Eds.). *From animals to animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior*. Cambridge MA, United States.: MIT Press.

Wilsdon, J. y Willis, R. (2004) *See-through Science: Why Public Engagement Needs to Move Upstream*. London, England: Demos.

Winner, L. (2008). *La ballena y el reactor*. Barcelona, España: Gedisa.

Wittgenstein, L. (1988). *Investigaciones filosóficas*. Barcelona, España:Crítica.

Wittgenstein, L. (2003). *Tractatus logico-philosophicus*. Madrid, España: Alianza.

Witthøfft Nielsen, L. (2011). The Concept of Nature and the Enhancement Technologies Debate. En J. Savulescu *et al*. (eds.). *Enhancing Human Capacities*. Chichester, England: Wiley-Blackwell.

Wolin, R. (2003). *Los hijos de Heidegger. Hannah Arendt, Karl Löwith, Hans Jonas y Herbert Marcuse*. Madrid, España: Cátedra.

Wynne, B. (1992). Misunderstood misunderstandings: social identities and the public uptake of science. *Public Understanding of Science*, *1*, 281-304.

Wynne, B. (2002). Risk and environment as legitimatory discourses of science and technology: reflexivity inside-out? *Current Sociology*, *50*, 459-477.

Wynne, B. (2011). Lab work goes social, and vice-versa: strategising public engage- ment processes. *Science and Engineering Ethics*, *17*, 791-800.

Yakman, G. (2006). STEAM Pyramid History. Recuperado de https://steamedu.com/pyramidhistory/

Yakman, G. (2011a). *STEAM: A Framework for Teaching Across the Disciplines*. Recuperado de https://steamedu.com/

Yakman, G. (2011b). STEAM: Learning That is Representative of the Whole World. Recuperado de https://www.researchgate.net/publication/328030888_STEAM_Learning_That_is_Representative_of_the_Whole_World

Yakman, G. (2016). *Maker Education with TE / STEM & STEAM for Global Innovation*. Recuperado de https://www.researchgate..net/publication/327449077_Maker_Education_with_TE_STEM_STEAM_for_Global_Innovation

Yuste, R., Goering, S., Agüera y Arcas, B., Bi, G., Carmena, J. M., Carter, A., Fins, J. J:, Friesen, P., Gallant, J., Huggins, J. E., Illes, J., Kellmeyer, P., Klein, E., Marblestone, A., Mitchell, C., Parens, El, Pham, M., Rubel, A., Sadato, N., Specker Sullivan, L, Teicher, M., Wasserman, D., Wexler, A., Wittaker, M. y Wolpaw, J. (2017). Four ethical priorities for neurotechnologies and AI. *Nature*, 551 (7679), 159-163.