2 ¿Qué relación tiene la ética con la tecnología? Una visión práctica

La ética no es una disciplina alejada de lo técnico, sino que, por el contrario, ambos conceptos se encuentran vinculados; para poder alcanzar la perfección en lo técnico es necesario tener en cuenta la ética. En palabras de Ortega, "de puro llena de posibilidades, la técnica es mera forma hueca (...), es incapaz de determinar el contenido de la vida". Pero entonces ¿cómo lo hacemos?

Como afirman Bilbao et al (2006) hemos descubierto que somos capaces de cambiar la vida humana a través del salto cualitativo que hemos alcanzado debido a la tecnología. **Nunca hemos tenido en nuestras manos tanto poder para transformar el mundo en el que vivimos** y, por otro lado, este salto tiene unos riesgos cuya magnitud no podemos conocer por el momento. Vivimos en un contexto en el que la tecnología, la economía y la ética están totalmente vinculadas: ¿lo posible a nivel técnico o lo que supone una rentabilidad económica, es humanamente deseable?

Y es que, antes de comenzar con esta revolución tecnológica, el desarrollo tecnocientífico era la gran variable del progreso económico y, por tanto, se mantenía alejado de la ética. Sin embargo, en la actualidad, es difícil sostener estas bases ya que las consecuencias sociales que acarrea la tecnología no son irrelevantes (UNESCO, 2019).

En la actualidad estamos viviendo una revolución algorítmica con la que se está introduciendo la tecnología en todas las facetas de nuestra vida, y que, como dice Bodinggton (2017) está integrada de forma invisible en las herramientas que usamos cada día, de forma que son estos algoritmos los que de alguna manera se encargan de decidir y organizar incluso nuestra información de consumo, lo que afecta directamente a nuestra vida profesional y personal.

También es innegable que las posibilidades que ofrece la tecnología son una prioridad en la estrategia de los países y las empresas más avanzadas. Pero es importante no olvidar las consecuencias de sus aplicaciones y tener presente, además de los retos y las posibilidades, los riesgos que pueden suponer para poblaciones más vulnerables (Blázquez, 2022).

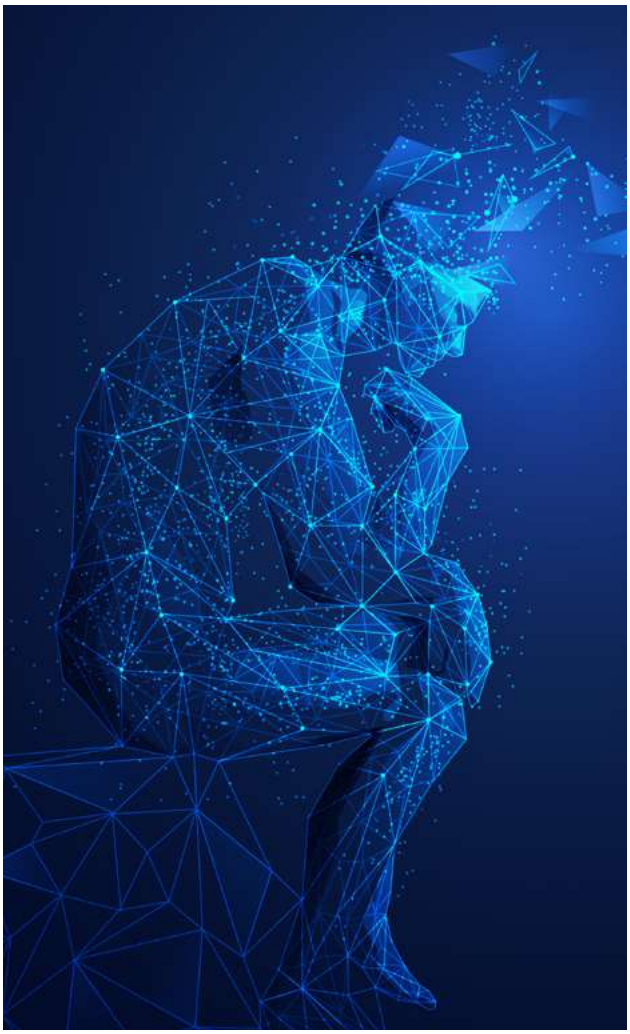
2.1 Un repaso por la ética y la filosofía de la tecnología

El progreso exponencial de las tecnociencias pone sobre la mesa preguntas éticas y filosóficas para las que es necesario encontrar respuestas. Y es que, como dice Quintanilla (2017) la tecnología trae consigo problemas de interés filosófico. Incluso sería posible afirmar que, en la actualidad, cualquier problema de cualquier índole se topa con factores relacionados con la tecnología en uno u otro momento. Tanto el problema como su posible solución y el cambio que trae consigo tiene implicaciones éticas que es necesario analizar.

Heidegger, ya en el siglo pasado en su texto "La pregunta por la técnica" hizo un alegato en contra de la visión antropocentrista e instrumentalista de la técnica con el objetivo de modificar la visión que hasta entonces se tenía de la misma. Consideraba que la técnica era una forma de entender el mundo que, de alguna manera, estaba enmarcada por los agentes de los que se rodeaba: usuarios, necesidades, ingenieros, máquinas, desarrollos, etc.

Apuntaba de una manera casi premonitoria que la existencia humana pertenecía ya a la técnica en tanto que era esta última la que ponía las reglas del juego, lo que de manera implícita viene a significar que son la economía, las ciencias y nuestro afán por modificar la naturaleza y el entorno (a través de la tecnología) lo que marca esas reglas. Esto, según él, conllevaba dos peligros: los efectos y daños que la tecnología desataba en la sociedad y en el entorno, y las consecuencias que podía acarrear el hecho de convertir el mundo en el máximo exponente de la eficacia, evitando que existiera un cuestionamiento en torno a las consecuencias de ese desarrollo.

Para Heidegger la técnica podía ser un medio para conseguir un fin o un "hacer" del hombre en el cual ambos (hombre y técnica) quedaban vinculados y debían cooperar. Afirmaba que el objetivo del hombre era el de dominar la técnica y manipularla como medio, pero en realidad, era la técnica la que estaba ejerciendo poder sobre nosotros, ya que mecanizaba y aceleraba los procesos (incluso el proceso mismo del pensar). Muy vinculado con el presente, Heidegger afirmaba que esta aceleración del pensar producía una evolución que estaba condicionada por la innovación tecnológica y no por las capacidades del ser humano.



De una manera u otra, Cortina (2019) haciendo referencia a Kant, viene a explicar una de las consecuencias de lo anterior, considerando que los avances tecnológicos preceden a la orientación moral que explicaría cómo hacer uso de ellos. **La tecnología va más rápido que nuestra capacidad para crear un marco que la delimite.**

Sí, los avances científicos y tecnológicos son tan rápidos que no es posible avanzar a la misma velocidad en el plano moral pese a que sería necesario para un equilibrio en la vida humana. La cuestión es que, a pesar de que Sócrates ya en el siglo V a. C. sugiriera la importancia de preocuparnos primero por reflexionar acerca de nosotros mismos y nuestra naturaleza ("Conócete a ti mismo"), el ser humano se ha dedicado en mayor medida a explorar hacia el exterior, el entorno y la naturaleza, lo que ha producido una brecha entre ambas cuestiones que ahora parece irresoluble.

> *"La ciencia nos da una sensación de progreso que la vida ética y política no puede proporcionarnos"*
>
> *John Gray en "Perros de paja"*

En 1939 Ortega y Gasset escribió "Meditación de la técnica" donde advertía ya las consecuencias que podría tener la tecnología en el ser humano: predecía que nos sumergiríamos en ella sin darnos cuenta. Y en efecto, de alguna manera así está sucediendo. Somos conscientes y no lo somos, al mismo tiempo. Quizás porque no conocemos (o no existe) la opción de salir de ella, quizás porque no sabemos cómo hacerlo. Además, ya entonces describía el mundo como un lugar tecnificado en constante desarrollo en el que no había límites, lo que consideraba como un arma de doble filo ya que dejaba en manos del hombre la capacidad de construir un nuevo mundo, pero también de destruirlo.

Ortega exponía la necesidad del hombre de transformar el medio para ser capaz de

satisfacer sus necesidades, unas necesidades que, en realidad, habían sido creadas por él mismo. En la misma línea, Langdon Winner en su libro "La ballena y el reactor: una búsqueda de los límites en la era de la alta tecnología" escrito en 1986 confirmaba que esta transformación del medio a través de la tecnología era realizada mediante acciones con las que el hombre se alejaba de su necesidad primitiva, con el objetivo de encontrar una solución que le permitiera preocuparse de otras cuestiones no de vital importancia.

Así, lo que se creaba (crea) finalmente era (es) un mundo sobrenatural que tenía (tiene) consecuencias en el ser humano y en su forma de vivir, ya que modificaba (modifica) el entorno con ella y, además, aseguraba (asegura) la satisfacción de sus necesidades con el mínimo esfuerzo, lo que hacía (hace) a los humanos capaces de poner el foco en crear otras necesidades nuevas. En palabras de Winner, "son fuerzas que actúan para remodelar dicha actividad y su significado".

Siguiendo con el libro de Winner, el autor utiliza el concepto de sonambulismo tecnológico. Citándole de manera literal: "caminamos sonámbulos de buen grado a través del proceso de reconstrucción de las condiciones de la existencia humana". Quizás debamos tener en cuenta este concepto y exprimirlo para entenderlo y tratar de evitarlo; quizás la ignorancia de la sociedad ante el proceso de cambio de la humanidad debido a la tecnología esté causada por la integración y la naturalización de la misma en nuestra vida diaria, en nuestro día a día.

Y es que, a partir de ahora la tecnología tiene la capacidad (la hemos dotado de la capacidad) de enunciar la verdad en el sentido en que la definía la filosofía griega antigua; **lo digital ahora puede erigir lo real de una forma más fiable que los humanos** (Sadin, 2020). Pero entonces… ¿Cómo podemos hacer encajar la ética en todo esto?

La ética es, como decía Aristóteles, el camino que nos lleva al objetivo común de la "buena vida" y tiene que ver con la manera en la que tomamos buenas decisiones. Tiene que ver con los procesos y, como explican Olmeda e Ibañez (2022) con los criterios y valores que se incorporan a él. Se trata de un concepto que se va transformando (o no) a lo largo del proceso, y que implica una acción. Para convertirnos en agentes éticos debemos poseer libertad, conciencia e intención; con estas tres características debemos involucrarnos en los procesos de decisión para, efectivamente, convertirlos en éticos de principio a fin.

Siguiendo con Olmeda e Ibáñez (2022) y vinculado a la idea de Heidegger de que la técnica está enmarcada por los agentes por los que se rodea, es importante mencionar que en la ética profesional también intervienen muchos agentes (la ética personal de cada persona unida a la de su gremio, la de la empresa en la que trabaja, la de los clientes, usuarios, partners con los que relaciona, etc.) y es necesario tenerlos en cuenta.

A pesar de ser la ética el arte de vivir bien y de comprender qué es bueno y qué no lo es a través del estudio moral de las reglas que nos llevarán a un comportamiento deseable, Peter Singer reflexiona acerca de la existencia de diferentes puntos de vista éticos en relación con la tecnología, distinguiendo dos enfoques filosóficos:

- El primero, en el que él mismo se encuentra, habla de maximizar los intereses de todos los seres humanos por igual con el objetivo de mejorar el bienestar y reducir el dolor. Esta visión utilitarista de la ética la seguían ya en el siglo XVIII autores como Jeremy Bentham y John Stuart Mill.
- El segundo enfoque, que se asocia con Kant, afirma que el límite de la ética debe ser la dignidad humana en tanto que hay ciertas cuestiones que son inviolables ya que harían perder la esencia del ser humano.

Para Singer, la "visión utilitarista" del primer enfoque, no olvida la dignidad humana, sino que sienta las bases de una sociedad en la que se promueve el bienestar global, en la que se fomenta el mejoramiento. Como comenta Olmeda e Ibáñez (2022) este es, probablemente, el enfoque más utilizado en el mundo de la empresa, siendo su objetivo el de encontrar soluciones que produzcan el mayor beneficio/utilidad para el mayor número de personas o agentes.

Esta forma de plantear los problemas nos obliga a conocer las consecuencias de la decisión, valorar quiénes pueden estar afectados por ella (cada uno de los agentes que hemos mencionado previamente) o medir la probabilidad y el impacto a corto, medio y largo plazo. Pero también tiene limitaciones como por ejemplo la dificultad misma de medir esas consecuencias, la necesidad de simplificar u omitir ciertas variables o la disquisición entre el punto de vista utilitarista y lo que está dejando al otro lado (fuera) de esa decisión (Olmeda e Ibáñez, 2022)

En cualquier caso y más allá de toda reflexión filosófica, lo que es evidente es que la tecnología aporta beneficios a la sociedad; las intenciones con las que se produce su desarrollo son normalmente positivas y tienen un impacto directo en la vida de las personas pero, efectivamente, y como explica Coeckelbergh (2021) en la actualidad, y ya vaticinaron otros pensadores anteriores, produce problemas éticos que son en realidad consecuencias no deseadas de la tecnología y que analizaremos a lo largo de este texto. Son estos los que debemos conocer para poder tenerlos en cuenta en nuestros procesos individuales y empresariales.

2.2 La inteligencia artificial y la ciencia de los datos

Aunque no existe una única definición de IA, UNESCO (2019) apunta que las máquinas que se basan en la IA o en la "computación cognitiva" son potencialmente capaces de imitar o incluso superar las capacidades cognitivas humanas.

Efectivamente, la idea fundamental para enmarcar la definición de Inteligencia Artificial es, como también dice Rodríguez (2018) la de crear programas o máquinas que sean capaces de llevar a cabo conductas consideradas inteligentes si fueran realizadas por un ser humano. Sin embargo, como cuenta Coeckelbergh (2021), algunos investigadores prefieren alejar la definición de la inteligencia humana y hacerla más neutral. Y es que la inteligencia no es algo exclusivo de los humanos.

La filosofía continental suele estar de acuerdo en que los seres humanos y sus mentes son distintas de las máquinas.

Otros filósofos, muchos de ellos seguidores de la tradición analítica de la filosofía, apoyan a los investigadores de la IA que creen que el cerebro y la mente son y funcionan como ordenadores, como por ejemplo Daniel Dennet, quien niega la existencia de cualquier cosa que exista fuera del cuerpo. Sin embargo, ambos grupos están de acuerdo en argumentar contra el dualismo cartesiano, el problema mente-cuerpo que Descartes resolvía afirmando que el cuerpo funciona según unos principios físicos y la mente no está subordinada a esos principios, pero ambos (cuerpo y mente) interactúan y se influyen, interaccionan entre sí.

En este aspecto, Crawford (2023) explica que "la vinculación de la IA con la inteligencia humana coloca a la primera en un lugar en el que se limita a ser entendida como una inteligencia incorpórea, liberada de cualquier relación con el mundo material". Se extrae de esta afirmación que Crawford entiende la IA como algo material, compuesto de un conjunto de estructuras sociales, políticas, técnicas, culturales, etc.

A lo largo de este texto tratamos de entender las consecuencias éticas y sociales de la IA como una tecnología práctica, la cual desempeña un papel importante en el análisis de los datos y en su interpretación (UNESCO, 2019), siendo una tecnología que está alineada con la función utilitarista de la ética que mencionamos anteriormente y que tiene un fin práctico.

Con independencia de la forma física que tenga, del comportamiento del sistema como un todo y de la influencia que tenga en su entorno (lo que es importante a nivel fenomenológico - basado en la experiencia - y ético), la base de la inteligencia de una IA es el software con el que cuenta: uno o varios algoritmos (Coeckelbergh, 2021)

Los algoritmos son secuencias de código que pueden interferir en el comportamiento de las personas a través de diferentes dispositivos. Es decir, no solo analizan los millones de datos que generamos en nuestro día a día, sino que, además, esos datos son empleados para tomar decisiones que influyen en nuestra vida (Kearns & Roth, 2020).

La RAE define un algoritmo como un "conjunto ordenado y finito de operaciones que permite hallar la solución de un problema". Se trata según Kearns & Roth (2020) de "una serie de instrucciones especificadas con precisión para llevar a cabo una tarea concreta" y en el diseño de algoritmos se pueden desarrollar varios, todos diferentes entre sí -con diferentes fortalezas y debilidades en función del problema que se quiera resolver.

Muchas cuestiones éticas en las que profundizaremos más adelante tienen que ver con tecnologías que están basadas en el aprendizaje automático y, por tanto, con la ciencia de los datos. Nos referimos a aprendizaje automático como el software que "puede aprender" (Coeckelbergh, 2021) - a pesar de que sabemos que el término es controvertido y que no se trata de una cognición real. Este aprendizaje se realiza a través de un proceso estadístico basado en datos cuya tarea principal es el reconocimiento de patrones o reglas que les permiten hacer predicciones no programadas previamente. Se trata de hacer observaciones particulares para alcanzar descripciones generales a través de correlaciones.

A nivel filosófico la cuestión es que este tipo de programas carecen de intencionalidad ya que el verdadero entendimiento no se puede generar en simbología computacional (Coeckelbergh, 2021).

Sadin (2020) afirma que estas tecnologías son capaces de aportar mejoras gracias a la asimilación (concepto de acción humana) de nuevos elementos que captan en los procesos de análisis, con el objetivo de enriquecer su expertise. Se modifican según "las experiencias vividas" y en función del conocimiento que adquieren por ensayo-error, lo que les permite perfeccionar su competencia. Sadin considera que este perfeccionamiento consigue que se imponga la autoridad de la IA y la tecnología a la de los humanos y quizás, en el futuro, llegue a desbancar a la intuición humana. Trata a estos programas como "seres".

Para cualquier cuestión sobre la que queramos estudiar que esté relacionada con las personas o la sociedad, es posible encontrar datos que poder analizar en profundidad. Esta situación nos convierte, como individuos, no solo en receptores y posibles usuarios de esos resultados sino como datos mismos de esos estudios (Kearns & Roth, 2020). Coeckelbergh (2021) va un poco más allá con esta reflexión afirmando que los datos ya no son algo pasivo, sino que son justo los datos los que definen lo que ha de hacerse a continuación, son el activo de la acción, el motor.

Si unimos ambas reflexiones, podemos concluir que esta nueva tecnología nos convierte en el centro de las decisiones y las acciones que se van sucediendo. No a nosotros como individuos, sino al rastro que vamos dejando en cada movimiento, con cualquier dispositivo. **La facilidad con la que ahora dejamos nuestros datos en la red ha sido clave para el desarrollo de la IA y de la ciencia de los datos.**

2.3 ¿Podemos hablar de una ética aplicable a la IA?

Si hablamos de una ética aplicada a la IA, esta se encargaría del impacto que el cambio tecnológico tiene en la vida de las personas y también en los cambios que produce en la sociedad y la economía. Los sistemas de IA afectan al modo en el que las personas entienden y se relacionan con el mundo y con otras personas, por eso son tan importantes las reflexiones éticas que de ellas se deriven, teniendo en cuenta todo el proceso de la IA y de cómo ese proceso se conecta con cada una de las personas implicadas. De manera natural, estos avances tecnológicos hacen surgir nuevos dilemas éticos.

Para poder gestionar estos nuevos dilemas, como comentan Olmeda e Ibáñez (2022), la ética no debería evaluar si una tecnología es o no aceptable una vez ya está creada, sino que, ya de inicio, el proceso de creación e implantación de esa tecnología debería crearse sobre la base de un marco ético. Implicar a la ética desde el inicio, a través de una guía clara para cada uno de los procesos a partir de una medición inicial que se use como diagnóstico, llevaría al proyecto a un resultado con impacto positivo.

2.3.1 La relación entre el ser humano y la IA

Si tenemos en cuenta la relación del ser humano con la IA, podemos ver que en función de la posición ética que adoptemos en cuanto a la tecnología (como individuos o como empresas) esto colocará al ser humano en uno u otro lugar.

En el punto anterior hemos hablado de los diferentes enfoques filosóficos de la ética en relación con la tecnología que proponía Singer (uno utilitarista y otro, a priori, más enfocado en los límites del ser humano). Coeckelbergh (2021) se remonta al siglo XVIII y hace un paralelismo con la actualidad para explicar su teoría en relación con la IA y el ser humano vinculado con la tecnología:

- **Ilustrados frente a Románticos:** Por un lado, argumenta que la tensión que existía entre la Ilustración y el Romanticismo se centraba en que los ilustrados y los científicos desafiaron la religión en pro de la razón, ya que pensaban que la ciencia y su evidencia (empirismo) nos enseñaría la verdad del mundo. Los románticos pensaban que era necesario volver al misterio del hombre que el empirismo estaba olvidando. El debate actual podría explicarse de la misma manera: científicos como Dennet o Boden creen que la ciencia puede ponernos sobre la mesa todos los misterios sobre la conciencia y reaccionan contra los filósofos continentales postmodernistas que creen lo contrario.

- **Humanistas frente a transhumanistas:** Recordando la división de Singer, reflexiona acerca de la tensión entre los humanistas y los transhumanistas. Los primeros subrayan el valor intrínseco del ser humano, como por ejemplo los filósofos que defienden los derechos humanos y la dignidad humana como base de la ética de la IA. Los transhumanistas defienden la necesidad de buscar un nuevo tipo de ser humano mejorado gracias a la ciencia que nos hará avanzar como especie.

- **Humanistas frente a posthumanistas:** Los humanistas han sido acusados de haber violentado a los no humanos pensando en el ser humano como ser supremo. Los posthumanistas (como Donna Haraway) critican el antropocentrismo de todas las áreas de conocimiento y están abiertos a explorar las fronteras entre humanos y no humanos. Esta visión podría liberar a la IA de la presión de tener que imitar lo humano como único modelo, lo que quizás nos acercaría al verdadero progreso de ambos.

Como comenta Cortina (2019), en todas estas posibles relaciones entre la tecnología y el ser humano, es aún el ser humano quien dirige todas las acciones. Por eso, no se trata de una ética de los sistemas inteligentes como tal, sino de cómo orientar el uso humano de estos sistemas de forma ética siendo conscientes de qué posición estamos adoptando en función de ella.

2.3.2 Retos de la ética de la IA

La evolución tecnológica puede hacer surgir nuevos dilemas éticos, así como suponer cambios positivos para la sociedad. Pero, igualmente es posible que produzca perjuicios si no se tiene en cuenta la globalidad de los agentes y se estudia en paralelo con la ética.

Según Olmeda e Ibañez (2022), los principios éticos para la IA son "principios que nos permiten preservar los derechos y libertades de las personas, sin frenar la innovación tecnológica" y claramente, están muy influidos por la Declaración de los Derechos Humanos de las Naciones Unidas de 1948. A continuación, vamos a desarrollar algunos de los principios más importantes para la IA, y a conectarlos con la ética y la filosofía.

- **La responsabilidad**

El concepto de responsabilidad es central para el ser humano en cualquier materia. Frente a la incertidumbre que produce el avance de la tecnología, es importante que aparezca en escena la responsabilidad como un principio ineludible. Hans Jonas publica en el año 1979 "El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica" en el que pone ya sobre la mesa la necesidad de responder a los retos de la tecnología mediante la creación de una conciencia que vele por el futuro de la humanidad y de la naturaleza a través de la aplicación del principio de responsabilidad.

Ya hemos mencionado el concepto de "sonambulismo tecnológico" de Winner, pero podemos volver a ponerlo sobre la mesa aquí, ya que éste nos ha hecho creer que la tecnología solucionará los problemas del futuro y eso nos lleva directamente a la creencia de que cuanto más se desarrolle la tecnología, mayor beneficio obtendrá la humanidad. Este utopismo tecnológico defiende de alguna manera la ausencia de cualquier implicación moral en el desarrollo de la tecnología lo que podemos vincular con algunos autores que defienden el utilitarismo del que hemos hablado previamente.

Sin embargo, para Jonas (2014) las consecuencias que la tecnología ha tenido durante el siglo XX le hacen rechazar esa ausencia de implicación moral (imparcialidad axiológica) para despertar esa necesidad de "deber" necesaria, un deber o responsabilidad fijada en el futuro que nos permita darnos cuenta de que no debemos perjudicarnos en pro del desarrollo. Habla sobre una responsabilidad más allá de las consecuencias directas, que nos permita ampliar el horizonte teniendo en cuenta los fines (responsabilidad basada en el compromiso con).

La responsabilidad podría definirse entonces de diferentes maneras:

- Cuando alguien realiza una acción, debe responder de su acto y se le considera responsable del mismo y de sus consecuencias. Esta responsabilidad moral es un principio ético.

- Pero también podemos considerar responsabilidad al hecho en sí, a la acción que se ha de realizar y al cómo la realizamos.

- Además, existe el concepto de responsabilidad en el ejercicio de una profesión que se refiere a cómo un colectivo (o un individuo del colectivo) responde a la legitimación de su conocimiento ante la sociedad (Olmeda e Ibañez, 2022). Cuando los humanos hacemos algo, vinculamos esa capacidad de actuar con una responsabilidad moral y, si lo que haces afecta a tu entorno, eres responsable de esas consecuencias.

- Para Aristóteles, la primera condición de responsabilidad moral era "la condición de control" y argumentaba que la acción se originaba en el agente (si puedes actuar, debes tener la capacidad para asumir las responsabilidades). La segunda condición era epistémica: sólo eres responsable si sabes verdaderamente lo que estás haciendo.

En el ámbito de la inteligencia artificial, el concepto de responsabilidad se refiere a las decisiones y las acciones que se toman mediante el uso de la IA y que implican cuestiones éticas. Un sistema de IA responsable debe tener en cuenta criterios y normas durante todo el proceso y contar con todos los agentes implicados, como ya hemos mencionado.

Para analizar de quién es la responsabilidad en la IA nos basaremos en Terrones citando a Diéguez (2003) que comenta que la tarea de la búsqueda de responsabilidades es compleja ya que todos los procesos se realizan en una cadena a la que siempre redirigen esa responsabilidad y, por tanto, queda diluida.

El mismo Diéguez señala como causas los sistemas en red en los que se teje la tecnología y la infinidad de procesos en los que está sumida, ya que dificultan conocer de dónde viene la responsabilidad. Añade la tecnocracia (los fines preceden a los cuestionamientos) como otra de las causas. Además, Horkheimer y Adorno en "La Modernidad en Dialéctica de la Ilustración" y posteriormente Horkheimer en "Crítica de la razón instrumental" apuntan a que la racionalidad instrumental oculta la realidad dogmática de la tecnología y eso también dificulta la asunción de las responsabilidades.

Según Olmeda e Ibáñez (2022), para poder atribuir esa responsabilidad en estos contextos es importante "vincular la responsabilidad moral al control moral y comprender que la responsabilidad no desaparece, aunque sea compartida".

Aquí citaremos a Cortina (2019) quien comenta que este nuevo mundo plantea cuestiones éticas muy importantes, pero la **principal decisión en relación con el impacto es la decisión que tomemos en cuanto al uso que le demos a la tecnología**, distinguiendo entre:

- Si nuestro objetivo es hacer uso de esos sistemas inteligentes para que nos ayuden a tomar decisiones.

- Si nuestro objetivo es delegar en ellos esas decisiones sobre cuestiones relevantes para la vida de las personas y la naturaleza.

En realidad, en ambos casos debemos tener en cuenta cuál de los dos es nuestro objetivo, aunque evidentemente, en el segundo caso en un grado superlativo.

El estatus moral de la IA es algo que atañe directamente a la filosofía moral y se refiere a dos cuestiones, la primera si la IA puede tener "agencia moral" y la segunda, si es posible que sea un "agente moral completo". Dice Coeckelbergh

(2021) que a la IA se le puede otorgar el papel de "agente moral" en tanto que realiza acciones que tienen consecuencias morales para las personas involucradas.

En los casos que proponía Cortina (2019) en los cuales usamos la IA para tomar decisiones e incluso delegar algunas de ellas, y parafraseando de Aristóteles las condiciones de responsabilidad moral, el problema surge cuando nos damos cuenta de que, efectivamente, la IA realiza acciones que tienen consecuencias éticas, pero no tiene consciencia de ellas y no es capaz de pensar moralmente. Por eso, no puede hacerse responsable de esas decisiones. Como cuenta Coeckelbergh (2021), desde la perspectiva aristotélica, únicamente los humanos pueden realizar acciones voluntarias y reflexionar sobre ellas posteriormente y entonces la solución pasa por hacer responsable al humano de lo que hace la máquina (de la misma manera que nos hacemos responsables como adultos de lo que hacen los niños o los animales).

En palabras de Coeckelbergh (2021), las máquinas son arresponsables, no pueden explicar lo que han hecho y, por tanto, las máquinas no pueden ser "agentes morales".

De esta solución se derivan dos problemáticas, la primera vinculada a encontrar la responsabilidad de un sujeto en esa red compleja, que como decía Diéguez, se diluye, y la segunda, la velocidad con la que la IA es capaz de tomar las decisiones (una velocidad imposible para los humanos).

Hemos ilustrado una de las posibilidades con el punto de vista de Coeckelbergh, pero hay otras posturas filosóficas al respecto de esta cuestión. Como ya hemos visto, algunos expertos consideran que las máquinas no tienen estados mentales o emociones por las cuales otorgarles esa "agencia moral"; pero es precisamente esta cuestión la que se considera un riesgo ya que entonces, ¿es posible que tomen decisiones moralmente correctas en caso de que deleguemos en ellas? ¿Debemos entonces delegar en ellas?

Por otro lado, hay expertos que dotan a la IA de una agencia moral completa al igual que los seres humanos, argumentando que las máquinas son más racionales y objetivas y no se dejan llevar por unas emociones que no poseen. El punto medio está en la "moralidad funcional" que determina de una manera bastante utilitarista que la IA necesita poder evaluar las consecuencias éticas de sus acciones. La cuestión es, ¿es posible enseñar a la IA reglas morales "previas a un suceso acaecido" con las que debe actuar? (Coeckelbergh, 2021)

- **La transparencia y la explicabilidad**

En nuestro día a día, convivimos con millares de modelos predictivos que toman decisiones de manera autónoma y que afectan a nuestras decisiones. Muchos de ellos son algoritmos de caja negra: no sabemos cómo han tomado esa decisión.

Estas decisiones pueden producir perjuicios en la sociedad, por ejemplo, estar sesgadas contra una clase en particular. Es importante asegurarse de que esos algoritmos no están usando datos sensibles ya que, como dice Maclure (2021), las decisiones que son tomadas con estos datos producen desigualdades que se hacen cada vez más patentes en la sociedad. La transparencia y explicabilidad son dos valores que deberían poder asegurarse en el desarrollo y la implementación de la IA.

Gracias a la explicabilidad debería ser posible contar en detalle el porqué de una decisión y entender cómo se ha llegado hasta ella. Esto nos permitiría conocer qué cambiaría en la decisión final si modificáramos una u otra variable, y por tanto explicar a las personas afectadas los porqués de su asignación o no asignación. Este es uno de los puntos clave con los que conseguir una IA confiable y de ello se hace eco la UNESCO (2019) y Coeckelbergh (2021), quienes confirman que, precisamente, la falta de transparencia y de explicabilidad, hace que las personas confíen menos en la tecnología.

Pero, ¿por qué las decisiones no son explicables?

Algunos algoritmos realizan sus razonamientos siguiendo técnicas de cálculo que nada tienen que ver con el razonamiento humano y con los símbolos que solemos usar. Esto hace más compleja la capacidad para explicar las decisiones tomadas por los sistemas de IA en términos humanos, convirtiendo a las decisiones en no transparentes.

Además, ya hemos visto que las redes neuronales de la IA están inspiradas en la actividad de las neuronas y en su conexión a través de sinapsis. El problema de la opacidad de la IA comienza con el aprendizaje automático ya que requiere de gran cantidad de datos, arquitecturas algorítmicas complejas y muchos parámetros que viajan a través de diferentes capas de detrás hacia adelante y al revés en caso de que haya comprobado algún error. Como explica Coeckelbergh

(2021), si la IA usa un árbol de decisión, la decisión será transparente ya que está programada de tal manera que determina su decisión dado un input particular.

Pero, en otros sistemas, sobre todo los que usan aprendizaje automático, y en concreto en el aprendizaje profundo la explicación no es posible ya que el modo por el que la IA decide, no es transparente.

Cuando un ingeniero programa un algoritmo, modifica algunos parámetros para observar cómo eso modifica su rendimiento en cada caso. Y es justo en este momento en el que el ingeniero pierde el rastro de la base del mismo, ya no es posible leer una ruta lógica de la información ya que, al modificarlo, se ha perdido la trazabilidad. Por esto, podemos decir que las decisiones o las predicciones hechas por aprendizaje automático, a menudo no se pueden explicar.

Coeckelbergh (2021) explica que esto constituye un problema también en términos de responsabilidad (que hemos visto previamente) ya que, ni siquiera los creadores de esa IA saben qué está haciendo exactamente y por tanto no pueden responder de sus acciones por lo que, en caso de que alguien se vea afectado por estas decisiones, no podrá recibir una respuesta. A esto lo llaman "caja negra".

Por dejar fijada una definición de los conceptos que estamos tratando, la explicabilidad según Olmeda e Ibañez (2022) "se refiere a la acción de generar el porqué de la predicción de un modelo o del funcionamiento de un modelo teniendo en cuenta la audiencia a la que va dirigida" y la interpretabilidad es "el grado en que un modelo puede ser entendido"

La necesidad de explicabilidad está resultando clave para las empresas. Cuando esas predicciones sin transparencia tienen que ver con áreas cruciales de la vida social, estas pueden tener un impacto directo en las personas, en sus vidas, en sus derechos y oportunidades. Y esto acarrea cuestiones éticas importantes. Más aún si debido a la pérdida de trazabilidad, la decisión tomada por la IA no se puede explicar, y por tanto la empresa no puede justificar e individualizar su decisión.

En relación con la explicabilidad Maclure (2021) distingue al menos dos problemas:

- **De naturaleza tecnocientífica:** debido a la cantidad de datos y parámetros, y al conocimiento requerido para el diseño del algoritmo es complejo explicar por completo el proceso.

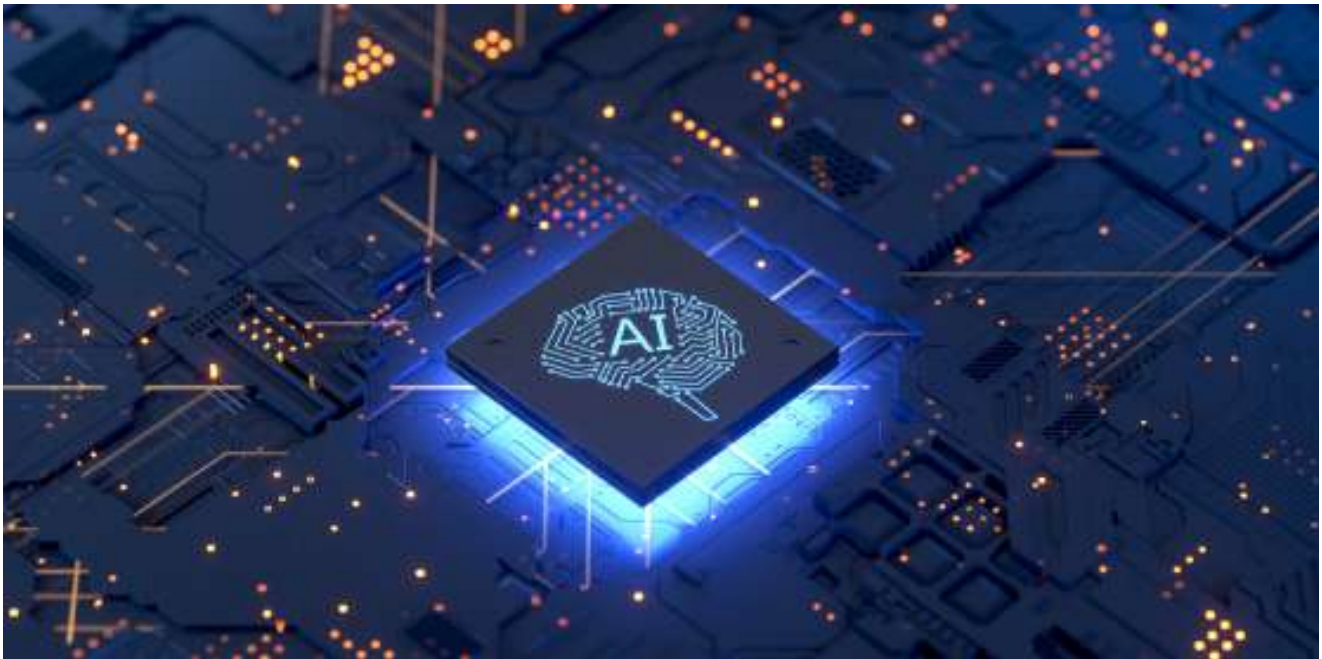
- **Razón pública:** Debido a la cantidad de datos necesarios para el entrenamiento y la capacidad para reaprender del aprendizaje automático, es imposible extraer una secuencia que justifique la decisión. **No existe la forma de traducir el recorrido interno de la máquina con palabras de lenguaje natural.**

En relación con la razón pública, está claro que las personas afectadas deberían poder conocer la razón por la que una decisión fue tomada por parte de un algoritmo. Por tanto, las razones deben ser transparentes y accesibles, y ser compatibles con la justicia siendo examinadas y evaluadas por la autoridad de la razón pública cuyo objetivo, a priori, debería ser "el bien de los ciudadanos". El problema de la opacidad en la IA es un problema ético debido a que las decisiones automatizadas no cumplen con los estándares de la razón pública y por tanto con el principio de legitimidad democrática.

Según Maclure (2021) algunos defensores de la IA apuntan que las profundas redes neuronales de la IA, y también las de los humanos, carecen de transparencia ya que trabajan en dos capas:

- Una primera capa por debajo de la consciencia y lo racional, incorporando la intuición, lo que permite concluir cuestiones sin usar la razón.

- Una segunda capa de los procesos cognitivos que hacen posible el pensamiento racional.



Estos defensores creen que los sesgos cognitivos están presentes en ambas capas también en el pensamiento humano, por lo que tampoco en nosotros el razonamiento es verdaderamente transparente. Dudan entonces de la necesidad de que una IA sea explicable. Para Maclure (2021) este argumento ignora el carácter social del razonamiento y el carácter institucional de los procedimientos relacionados con las tomas de decisiones que inciden en el bienestar, derechos y oportunidades de las personas, lo que no está teniendo en cuenta las diferentes dimensiones sociales e institucionales del razonamiento humano. La cuestión según la autora no es que los razonamientos en las relaciones sociales y en las instituciones sean 100% transparentes, sino que, en estos casos, los criterios de razón pública nos aportan recursos para poder diseñar instituciones epistémicamente más fuertes y tener capacidad de criticarlas cuando nos fallan.

Maclure (2021) defiende la necesidad de que las organizaciones estén legalmente obligadas a demostrar la capacidad de explicar y justificar sus decisiones algorítmicas, sobre todo cuando tienen un impacto directo en el bienestar, los derechos y las oportunidades de las personas. El argumento principal es que la analogía entre la mente humana y las redes neuronales artificiales falla ya que esta última no toma en cuenta todas las dimensiones sociales e institucionales que sí tienen en cuenta los razonamientos humanos. Considera que, hasta que esto no suceda, la IA no debería usarse para delegar decisiones que ahora toman los humanos.

Una de las cuestiones importantes que explica Coeckelbergh (2021) es que, éticamente, que una empresa tenga algoritmos transparentes y explicables no quiere decir que haya tenido que revelar el código de su software. En realidad, lo único que significa es que ha de ser capaz de explicar "cómo funciona" y por qué toma las decisiones que toma. Explicar las decisiones es una exigencia moral y una condición imprescindible para realizar toma de decisiones responsables.

- **Privacidad y protección de datos**

La privacidad es "aquello que no es de propiedad pública, que pertenece a particulares". Es interesante la definición de privacidad que cita Pfeiffer (2008) escrita por Sennet en 1978: "Lo privado es todo lo que está fuera del ámbito de interés público, de los asuntos del Estado, de lo que involucra al conjunto de la sociedad. Lo privado es el ámbito restringido de lo doméstico y lo familiar, de aquellos asuntos del sujeto que no necesariamente deben ser divulgados masivamente"

Para Pfeiffer (2008), los seres humanos son libres y racionales, pero con el tiempo han ido perdiendo, gracias a la capacidad crítica, la convicción de que su racionalidad y su libertad son absolutas. **Cada vez más, existe información personal, consciente e inconsciente, que requiere de límites a la hora de ser compartida** y que no siempre proviene de lo racional; es lo que Nietzsche o Freud llaman "el instinto" (intereses, emociones, necesidades o deseos). Ahora mismo, todo lo que surge del instinto también es compartido y utilizado en forma de "dato", prácticamente sin darnos cuenta.

Para vivir en sociedad han de establecerse pautas o contratos sociales sostenidos desde la base de la libertad y la igualdad y que permiten construir el concepto de derecho negativo (definido como la ausencia de coacción externa a quien desea realizar una acción). El derecho a la intimidad o a la privacidad parte de este derecho a la libertad y como explican Olmeda e Ibáñez (2022) este derecho forma parte de la Declaración Universal de los Derechos Humanos y es de cada individuo, que puede asegurarse la confidencialidad con el fin de garantizar su dignidad.

Pfeiffer (2008) cita a Ortega, para quien "el ser humano es aquel que puede vivir consigo mismo". El derecho a la intimidad es el derecho del sujeto a preservar su información, **el derecho a la privacidad es el reconocimiento por parte de la sociedad de ese espacio íntimo o privado de otros**. La autora afirma también que la intimidad es un derecho "raíz" que da lugar a otros derechos y que, sin intimidad, no hay libertad. Conocer y difundir la intimidad de alguien en contra de su voluntad, comporta la destrucción de esa intimidad.

Hablando de esos datos, para Coeckelbergh (2021) un uso ético de la IA pasa por una recogida y un procesamiento de los datos que respete la privacidad de los seres humanos y su derecho a saber qué va a pasar con ellos, a objetar su recogida o procesamiento, a saber, que se está realizando esa recogida de datos y también si van a formar parte de una decisión tomada por una IA. Sin embargo, en la actualidad se consideran complejos los contextos en los que la IA está actuando, ya que no es tan evidente el uso que se le da a esos datos, ni tampoco para qué se están recogiendo, ni cómo o qué se va a hacer con ellos.

Las aplicaciones que trabajan con el llamado big data implican la recogida de una infinidad de datos de información personal, de hecho, muchas veces estamos ofreciendo datos a la nube sin darnos cuenta y todos ellos, como combinación, son utilizados por los algoritmos. Existe un debate en el que se afirma que el auge de la tecnología ha desdibujado los límites entre los datos personales que se consideran sensibles y los que no. Esto se debe a que, quizás, sea posible obtener datos sensibles a partir del análisis de datos no sensibles. Para Pfeiffer (2008), los datos sensibles son aquellos que afectan a la intimidad de una persona y que permiten identificarla y generar de ella un perfil que poder utilizar después.

Sin embargo, a pesar de la gran cantidad de datos que dejamos en la red y de la cantidad de ellos que se pueden inferir de unos pocos, cuentan Olmeda e Ibañez (2022) que existen en la actualidad muchos problemas reales y con gran i
tienen que ver con la falta de datos que sean realmente relevantes. Citando al MIT Technology Review Insight survey 2020, Olmeda e Ibañez (2022) explican que una de las principales dificultades en el uso de la IA es la disponibilidad, cantidad o calidad de datos necesarios con los que conseguir modelos que sean eficaces. Muchos de esos datos son recogidos de bases de datos sesgadas, lo que dificulta una buena lectura de los resultados.

De hecho, existe un dilema que une privacidad y transparencia que se encuentra enmarcado en la casuística en la que, precisamente por preservar la privacidad de las personas, no es posible tener acceso a suficiente información de calidad que permita tomar una decisión relevante. Esto supone un gran reto para el desarrollo de la IA (Olmeda e Ibáñez, 2020).

- **Sesgo y significado de la vida**

En la actualidad, la toma de decisiones llevadas a cabo por algoritmos guía nuestra vida en muchos ámbitos, tanto laborales como personales. Para Olmeda e Ibáñez (2020), un buen algoritmo debería generalizar sobre la base de un entrenamiento realizado con un conjunto de datos. En función de ese conjunto de datos, es posible que exista un sesgo, lo que indicaría que los datos no están balanceados.

En muchos casos se usa indistintamente sesgo y discriminación, pero existen diferencias entre ambos conceptos. El sesgo es un "error sistemático en el que se puede incurrir cuando, al hacer muestreos, se seleccionan unas respuestas frente a otras"; se trata normalmente de sesgos estadísticos que reflejan la realidad ("esa realidad"), pero, una vez llevados al contexto de los algoritmos de decisiones automáticas, si la decisión es diferente en función de los atributos de la persona, entonces podría considerarse discriminación (Olmeda e Ibáñez, 2020)

Por tanto, cuando una IA toma decisiones, es posible que surja un sesgo: las decisiones pueden ser poco equitativas para uno u otro grupo o uno u otro individuo. Este problema, a menudo, está relacionado con el aprendizaje automático y uno de los peligros es que se perpetúen estos problemas en la sociedad (Coeckelbergh, 2021). Para Zafra (2022) "no existe nada que no se sostenga en la racionalidad de un mundo previamente leído, generado por datos que recogen lo que ya ha pasado" y que proporcionan por tanto una repetición de patrones anteriores.

A menudo estos sesgos no son intencionados, es posible que las personas involucradas en el proceso no prevean estos efectos discriminatorios por no comprender el sistema de IA, no reflexionar con profundidad las consecuencias o no estar cerca de los grupos de interés. La cuestión es que las consecuencias pueden tener un impacto grave en individuos y comunidades, como, por ejemplo, la falta de acceso a recursos o libertades.

El sesgo puede surgir de varias maneras y en cada una de las fases de diseño, prueba y aplicación (Coeckelbergh, 2021). Según Olmeda e Ibáñez (2020) el inicio comienza con una pregunta de negocio a la que vincular un objetivo de la variable a predecir. A partir de él se recopilan los datos y se limpian. El conjunto de datos que se genera, normalmente de datos históricos, (del que se utilizan el 70-80% para entrenar el algoritmo y generar el modelo) sirve para seleccionar los atributos que ayudarán a dar respuesta a la pregunta propuesta. El otro porcentaje de datos restante sirve como test, y si cumple las métricas establecidas, entonces se produce.

Olmeda e Ibáñez (2020) explican los sesgos que pueden ocurrir en cada fase:

FASE 1: Datos.

- Durante la generación de datos existe el sesgo histórico. Se utilizan datos que reflejan la sociedad y que incluyen ya ese sesgo.

- Durante el muestreo existe el sesgo de representación y tiene que ver con la utilización de un número de datos que no incluye un número suficiente de muestras de la población que se quiere evaluar.

- Durante las métricas puede ocurrir el sesgo de medida. Esto se refiere a los atributos seleccionados por los expertos en función del problema a resolver, y en cómo se realizan dichas medidas en los datos.

FASE 2: Diseño del algoritmo.

- En cuanto a la manera de medir la forma en la que el modelo aprende, se podría introducir un sesgo de aprendizaje. Se produce cuando las opciones de modelado hacen mayor el rendimiento entre unos ejemplos de datos y otros.

- En la evaluación se puede producir el llamado sesgo de evaluación, que ocurre en los casos en los que los datos que hemos tomado como referencia no representan a la población. Se produce cuando queremos aplicar diferentes conjuntos de datos para comprobar cómo de bueno es el modelo, lo que provoca una generalización en base a los nuevos datos y muchas veces no es estadísticamente correcto.

- Cuando se utilizan datos de distintos grupos y la asignación o etiquetado de los datos no es consistente para todos, pero se asume que, si lo es, se puede producir un sesgo de agregación, lo que puede llevar a resultados no válidos para todos los grupos.

- Cuando existe una desconexión entre el problema que se quiere resolver y la forma en la que se usa, entonces se produce un sesgo de implementación.

Ya hemos visto que estos sesgos pueden provocar daños éticos e incluso ir en contra de principios y valores básicos de la democracia. Podemos dividir el impacto de los sesgos en los algoritmos en: discriminación social, económica, de acceso libre a la información, privación de la libertad y discriminación y abuso de control (Astobiza, 2017)

Pero, a pesar de saber dónde y cuáles son los sesgos que podemos encontrar en cada fase del proceso, según Coeckelbergh (2021) no es posible determinar si se pueden evitar, si deberían evitarse y a qué coste podría hacerse. Por ejemplo, en caso de que modificar un algoritmo de aprendizaje automático con el objetivo de reducir el sesgo lo hiciera menos preciso, ¿debería hacerse?

Además, y como cuestión superior, no existe consenso entre los filósofos en determinar si es posible generar un algoritmo imparcial ya que no existe un consenso en los conceptos que construirían esa imparcialidad (la justicia y la equidad). Lo que está claro es que debemos tratar de evitar los sesgos lo máximo posible y para ello Coeckelbergh (2021) propone soluciones tecnológicas, sociales, políticas y educativas y utilizar unas u otras en función de la noción de justicia y equidad que se tenga.

2.4 Conclusiones

Nunca habíamos tenido en nuestras manos tanto poder como el que tenemos en la actualidad para transformar el mundo en el que vivimos. Está claro que la tecnología está integrada en nuestro día a día y que se ha convertido en una disciplina transversal que está presente en todos los espacios que habitamos, al igual que debería serlo la ética.

Ya hemos visto cómo la tecnología ofrece grandes oportunidades y posibilidades de crecimiento, pero está sumiendo al mundo en un cambio constante a una velocidad difícil de manejar para otras áreas de conocimiento que deben acompañar dicho crecimiento. El humanismo está vinculado a lo técnico y es importante que pueda adaptarse a la misma velocidad para crear un marco que lo delimite.

Una ética aplicable a la IA debería encargarse del impacto que el cambio tecnológico tiene en la vida de las personas y qué consecuencias produce en la sociedad. Este cambio tiene consecuencias en cómo los humanos se relacionan y entienden el mundo y es necesario tener en cuenta todo el proceso de la IA y cómo se conecta con cada persona, generando una red y unas conexiones que producen nuevos dilemas éticos que hay que gestionar durante todo el proceso de creación e implantación de dicha tecnología a través de un marco ético previamente creado que tenga en cuenta la globalidad de los agentes implicados.

Sería necesario tener en cuenta en este crecimiento algunos principios éticos que nos ayuden a preservar nuestros derechos y libertades como personas. Estos principios éticos son la guía para la acción en los diferentes dilemas éticos que surgen y que surgirán con el crecimiento de la tecnología. Inventariar estos dilemas y tener claros los principios como base a la hora de crear la tecnología, y/o de enfrentarse a alguno de ellos, convertiría a lo técnico y lo humano en un engranaje completo que nos permitiría mirar la tecnología como una oportunidad de crecimiento para el ser humano y para el mundo que le rodea, siempre velando por el futuro de la humanidad y de la naturaleza.

María Cano Bonilla