

La inteligencia artificial y la robótica: sus dilemas sociales, éticos y jurídicos

Artificial intelligence and robotics: its social, ethical and legal dilemmas

Adriana Margarita Porcelli

Universidad Nacional de Luján, Argentina
adporcelli@yahoo.com.ar

Recibido: 26/09/2019

Aceptado: 29/02/2020

Resumen: La creciente digitalización de la actividad humana ha fusionado los mundos físico, digital y biológico de manera que transformará a la humanidad en su esencia misma. Esta nueva realidad utiliza como medio de comunicación Internet y como mensaje los propios metadatos o datos los cuales necesitan aplicaciones y soluciones de inteligencia artificial para ser procesados. Si bien la inteligencia artificial ayudará a resolver los grandes problemas económicos y sociales cada vez más complejos, también plantea numerosos dilemas éticos y la necesaria formulación de una legislación moderna acorde a la realidad a regular. El objetivo del presente trabajo consiste en conceptualizar la inteligencia artificial y analizar las últimas recomendaciones de la Comisión Europea en materia de robótica. La metodología se basó en la actualización del marco conceptual y en el análisis normativo en el marco de la Estrategia Europea para la Inteligencia Artificial para concluir en su necesaria actualización.

Palabras clave: inteligencia artificial, robótica, recomendaciones, Unión Europea

Abstract: The increasing digitalization of human activity has fused the physical, digital and biological worlds in a way that will transform humanity into its very essence. This new reality uses Internet metadata or data itself as a means of communication and as a message which needs artificial intelligence applications and solutions to be processed. While artificial intelligence will help solve the increasingly complex economic and social problems, it also raises numerous ethical dilemmas and the necessary formulation of modern legislation according to the reality to be regulated. The objective of this work is to conceptualize artificial intelligence and analyze the latest recommendations of the European Commission on robotics. The methodology was based on the updating of the conceptual framework and the normative analysis in the framework of the European Strategy on Artificial Intelligence to conclude in its necessary update.

Key words: artificial intelligence, robotics, recommendations, European Union.

Sumario: I. Introducción. II. Marco Conceptual: a) Inteligencia artificial. b) Robótica. c) Ventajas y desventajas de la inteligencia artificial. III. Normativa Europea. IV. Conclusiones, Bibliografía.

I. Introducción

Los economistas afirman que la humanidad se encuentra en el preludio de la Cuarta Revolución Industrial, llamada también Industria 4.0 (término utilizado por primera vez en la Feria de Hanover, Alemania, en el año 2011), continuadora de **los otros tres procesos históricos transformadores**: la Primera Revolución Industrial (entre 1760 y 1830) marcó la transición de la producción manual a la mecanizada, la Segunda- alrededor de 1850- introdujo la electricidad y permitió la manufactura en masa y la Tercera, a mediados del siglo XX, denominada la Revolución Digital, basada en el uso de tecnologías de información para automatizar aún más la producción. Esta Cuarta Revolución Industrial, **no se define por un conjunto de tecnologías emergentes en sí mismas, sino por la completa digitalización de**

las cadenas de valor a través de la integración de tecnologías de procesamiento de datos, software inteligente y sensores. Recurriendo a Internet, a los sistemas ciberfísicos y a las redes virtuales con posibilidades de controlar objetos materiales, se pueden ir modernizando las plantas fabriles hasta transformarlas en fábricas inteligentes (Perasso, 2016).

Dicho en forma más simple, una producción industrial en la que todos los productos y máquinas están interconectados entre sí digitalmente. Las nuevas tecnologías y enfoques están fusionando los mundos físico, digital y biológico de manera que transformarán a la humanidad en su esencia misma.

En ello radica lo novedosos de esta nueva revolución, que a diferencia de las anteriores que se desarrollaron exclusivamente en el mundo físico, la actual conecta ese ámbito físico con el espacio digital, utilizando como medio de comunicación Internet (Internet de las Cosas, IOT, siglas en inglés de *Internet of Things*) y como mensaje los propios metadatos o datos (*Big Data*).

El Comité Económico y Social Europeo, en su dictamen de iniciativa emitido el 31 de mayo de 2017 sobre la “Inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad” señala que en la cuarta revolución industrial, marcada por la convergencia de tecnologías digitales, físicas y biológicas, Internet de las Cosas es el ámbito predilecto de aplicación de las formas más avanzadas de inteligencia artificial y se pone a prueba el principio según el cual los seres humanos tienen en todo momento el control sobre dichos objetos (Comité Económico y Social Europeo, 2017).

En 2017, cerca de 4.000 millones de personas -más de la mitad de la población mundial- utilizaba Internet y un 56% lo hacía con suscripciones a servicios móviles. Por otra parte, el 61% de tales suscripciones operaban sobre redes 3G o 4G y durante el 2017 se descargaron 175.000 millones de aplicaciones y se usaban activamente alrededor de 40 en cada teléfono inteligente. A principios de 2018 se registraban más de 5.000 millones de usuarios únicos de telefonía móvil, de los cuales 57% utilizaba teléfonos inteligentes. En enero de 2018, más de 3.000 millones de personas- el 42% de la población mundial- usaban mensualmente las redes sociales, especialmente mediante dispositivos móviles.

En tanto, el uso de plataformas de comercio electrónico para comprar bienes de consumo creció hasta alcanzar los 1.800 millones de compradores- el 23% de la población mundial- en línea a nivel mundial. Entre las nuevas tecnologías que están impulsando la digitalización, la Internet de las Cosas es una de las que se prevé tendrá mayor impacto, tanto en el desarrollo de bienes y servicios para los consumidores como para usos productivos (Comisión Económica para América Latina y el Caribe, 2018).

Según estimaciones realizadas por la consultora Gartner, en 2020, el número de objetos conectados a Internet será de más de 26.000 millones (excluyendo PCs, *tablets* y *Smartphone*) y que la Internet de las Cosas aportará por sí misma un valor de 1,9 billones de euros a la economía mundial, demostrando la gran importancia estratégica que representará la economía digital en los próximos años (Gartner, 2013).

En sintonía, el estudio del *McKinsey Global Institute*, intitulado “*The Internet of Things: How to capture the value of IoT*” una Internet de las cosas completamente conectada podría añadir hasta \$11 billones a la economía global al año para el 2025 a través de diferentes entornos incluyendo fábricas, ciudades y ámbitos minoristas (McKinsey & Company, 2018).

No es de extrañar que, durante el 2018, siete de las diez empresas más cotizadas en la bolsa fueron tecnológicas. El trio Apple, Amazon y Alphabet (matriz de Google) fue secundado por Microsoft y Facebook para completar el top 5 mundial. El sexto lugar correspondió a Berkshire Hathaway, el brazo inversor de Warren Buffett, seguido de nuevo por dos tecnológicas, en este caso de China: Alibaba y Tencent. Cerraron el top 10 JPMorgan y Johnson & Johnson (Mazo, 2018).

En este contexto, el objetivo del presente artículo consiste, en primer lugar, en conceptualizar y exponer las ventajas y desventajas de la inteligencia artificial y, en segundo lugar, analizar las últimas recomendaciones de la Comisión Europea en materia de inteligencia artificial y robótica. La metodología se basó en la actualización de los conceptos y en el análisis normativo. A tales efectos, el trabajo se compone de dos partes: la primera referida a aspectos tecnológicos y prácticos y las segunda al análisis ético y jurídico.

II. Marco conceptual

a) Inteligencia artificial

En primer lugar, es necesario conceptualizar el término inteligencia. El mismo proviene del latín “*inteligere*”, y se entiende como la facultad de comprender las relaciones entre los hechos y las cosas. No se lo debe confundir con la sabiduría ni con la memoria, siendo ésta la capacidad para recordar una gran cantidad de datos, de manera que una persona puede ser capaz de recordar muchos datos, pero no tener la inteligencia suficiente para utilizarlos a efectos de resolver un problema, lo cual sí sería una manifestación de la inteligencia.

Actualmente, la inteligencia artificial (IA) es uno de los temas que más invita a la reflexión en el campo de la tecnología y de los negocios. Dicho término fue acuñado en 1956, cuando un grupo de científicos (McCarthy, Minsky, Rochester y Shannon) inició, en Dartmouth College en los Estados Unidos, el proyecto de investigación “Inteligencia Artificial”. Inicialmente, su objetivo era que la inteligencia humana pudiera ser descrita de forma tan precisa que una máquina fuera capaz de simularla. Este concepto también fue conocido como “Inteligencia Artificial genérica”, consistente en la inteligencia artificial igual o superior a la capacidad humana promedio.

Durante el verano de ese año, el mismo grupo de científicos—John McCarthy, Marvin L. Minsky, Nathaniel Rochester y Claude E. Shannon—invitaron a diferentes investigadores a la Conferencia de *Dartmouth* en *Dartmouth College* (*Hanover, New Hampshire*) bajo la premisa que cada aspecto del aprendizaje y de la inteligencia humana podía ser tan precisamente descrito que una máquina podría simularlo. En los dos meses de duración del evento, fueron debatidos aspectos tales como el lenguaje natural y su procesamiento, las redes neuronales o la teoría de la computación. En la propuesta se planteó algunos problemas de la inteligencia artificial, por ejemplo:

Si una máquina puede hacer un trabajo, entonces una calculadora automática puede ser programada para simular la máquina. Las velocidades y las capacidades de memoria de las computadoras actuales pueden ser insuficientes para simular

muchas de las funciones superiores del cerebro humano, pero el principal obstáculo no es la falta de capacidad de la máquina, sino nuestra incapacidad para escribir programas aprovechando al máximo lo que tenemos. (McCarthy *et al*, 1955, p.1)

En esta etapa, que se puede caracterizar como originaria, no se puede dejar de mencionar la influencia de los trabajos del prestigioso académico Noam Chomsky que revolucionaron la lingüística teórica al proponer la denominada la gramática generativa, una teoría del lenguaje inspirada en los lenguajes artificiales de la lógica y las matemáticas. Conforme la precitada teoría, con un número finito de reglas se podía generar los infinitos enunciados de una lengua. Así, se opuso al paradigma entonces dominante del conductismo, defendido por el psicólogo de Harvard BF Skinner, basado en la relación estímulo-respuesta y en el uso de técnicas de la estadística para minar y predecir datos. El comportamiento animal, según el conductismo, se reducía a un simple conjunto de asociaciones entre una acción y su posterior recompensa o castigo. Sin embargo, los principios conductistas no podían explicar la riqueza del conocimiento lingüístico, el uso infinitamente creativo del mismo o la rapidez con que los niños lo adquieren con una exposición mínima e imperfecta al lenguaje presentado por su entorno. Según Chomsky, la facultad de idiomas era parte del organismo. En la mente humana ya existe algo innato que será el germen del lenguaje, por ello, los humanos adquieren un sistema cognitivo lingüístico. De acuerdo con este principio, el lenguaje puede ser transmitido genéticamente y los seres humanos poseen una estructura lingüística similar independientemente de sus diferencias sociales y culturales. También distinguía entre la capacidad de una persona para asociar sonidos y significados con determinadas reglas inconscientes y automáticas, y la actuación o ejecución lingüística, que alude a la forma de interpretar y comprender una oración o un lenguaje en particular (Chomsky, 1957).

Así, Chomsky estableció cuatro tipos de lenguaje con sus correspondientes gramáticas y máquinas asociadas en forma jerárquica, conocido como la Jerarquía de Chomsky. Dicha jerarquía se corresponde

con una jerarquía paralela de máquinas: los autómatas formales, que tienen capacidad de resolver distintos tipos de problemas.

De esta manera, influyó en la lingüística computacional, en los sistemas expertos basados en sistemas de diálogos de pregunta y respuesta, con una base de datos primitiva, un motor de búsqueda y resolución de problemas mediante el lenguaje natural que facultara a las máquinas a comunicarse con las personas. El sistema Watson desarrollado por IBM representa uno de los primeros productos del nuevo enfoque de computación cognitiva. Watson no estaba conectado a Internet; accedía a datos almacenados en forma de conocimiento no estructurado, acercándose a la forma en la que el cerebro humano recibe, procesa y toma decisiones con información procedente de los cinco sentidos, pudiendo responder a preguntas en lenguaje natural (Roldán Tudela *et al*, 2018).

Entonces, si la mente humana posee un programa, puede llevar a la conclusión que las máquinas podrían llegar realmente a pensar, exactamente igual que el humano. Bastaría con implementar el tipo adecuado de programas para que puedan pensar en forma real y tener conciencia de lo que están haciendo.

Del análisis de tales propuestas nacieron las dos formas de entender la inteligencia artificial, que se explicarán en los siguientes párrafos: la débil restringida al empleo de máquinas para el estudio de las posibilidades cognitivas del ser humano; mientras que la fuerte se orientaba a ligar los nexos entre la inteligencia artificial y la humana y ver la forma de vincularlas cada vez más.

Uno de los ejemplos más notables y difundidos llegó en el año de 1997, cuando la computadora conocida como *Deep Blue*, creada por IBM y que operaba como inteligencia artificial, derrotó al campeón mundial de ajedrez Gari Kasparov. Este hecho inspiró a la industria tecnológica para ampliar las investigaciones en la temática.

Después de esta breve introducción sobre la lingüística cognitiva que incidieron en los primeros desarrollos de la inteligencia artificial, a continuación, se transcribirán algunas definiciones, ya que no existe una única forma de conceptualizar dicho término.

John McCarthy define la inteligencia artificial como:

La ciencia y la ingeniería de hacer máquinas inteligentes, especialmente

programas informáticos inteligentes. Está relacionado con la tarea similar de usar computadoras para comprender la inteligencia humana, pero la IA no tiene que limitarse a métodos que son biológicamente observables (McCarthy, 2007).

Resulta innegable la naturaleza interdisciplinaria constituida por la convergencia de las ciencias cognitivas: lógica matemática, lingüística, informática, psicología, etnología, sociología, la biología y las neurociencias. En este sentido, Morales Ascencio afirma que:

La IA se ocupa del estudio de los fundamentos del desempeño inteligente del ser humano, la manera como los humanos utilizan el lenguaje, el desarrollo de los procesos de inferencia, la comprensión de los fenómenos, los procesos de aprendizaje, las formas de percepción, la manera como los humanos adquieren, aplican, transfieren conocimientos sobre el mundo; los procesos de planeación, ejecución y evaluación de planes de acción; las formas de comportamiento del hombre con los demás y ante la máquina. (Morales Ascencio, 1997, p. 27)

Por su parte, Andreas Kaplan y Michael Haenlein la definen como

La capacidad de un sistema para interpretar correctamente datos externos, para aprender de dichos datos y emplear esos conocimientos para lograr tareas y metas concretas a través de la adaptación flexible (Kaplan y Haenlein, 2018, p.15).

La inteligencia artificial es el campo científico de la informática que se centra en la creación de programas y mecanismos que pueden mostrar comportamientos considerados inteligentes. En otras palabras, la inteligencia artificial es el concepto según el cual las máquinas piensan como seres humanos. Normalmente, un sistema de inteligencia artificial es capaz de analizar datos en grandes cantidades (*Big Data*), identificar patrones y tendencias y, por lo tanto, formular predicciones de forma automática, con rapidez y precisión (Salesforce, 2017).

También se la define como la rama de la ciencia computacional que investiga y desarrolla programas cuyos productos finales, de ser atribuibles a un ser humano presupondrían procesos mentales inteligentes (Cáceres Nieto, 2010). En otras palabras, un sistema con inteligencia artificial tiene la capacidad de llevar a cabo procesos que realizaría una persona inteligente. Cuando un coche autónomo reconoce la voz de una persona, interactúa con ella y la lleva a su destino, en

realidad está utilizando los conocimientos que tiene adquiridos para resolver un problema.

Es necesario recalcar que la inteligencia artificial tiene el objetivo de construir algoritmos capaces de resolver problemas que los humanos solucionan a diario. Estos algoritmos inteligentes realizan diversas funcionalidades como lectura y procesamiento de datos, aprendizaje automático, traducción de documentos hasta llegar a competir con los mejores en distintos juegos como el alphaGO.

Con el paso del tiempo, la ciencia evolucionó hacia áreas de conocimiento específicas, y fue entonces una combinación entre el reconocimiento de imagen, el procesamiento de lenguaje, las redes neuronales y las acciones orientadas a un único objetivo. Por ejemplo, vehículos autónomos, realidad virtual, sanidad, reconocimiento visual, manipulación y análisis de datos, aparatos domésticos o ciberseguridad. En ocasiones, el mercado se refiere a este tipo de avances como “Inteligencia Artificial débil.”

La Comisión Europea define a la inteligencia artificial de la siguiente manera:

El término «inteligencia artificial» (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción –con cierto grado de autonomía– con el fin de alcanzar objetivos específicos. Los sistemas basados en la IA pueden consistir simplemente en un programa informático (p. ej. asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas) (Comisión Europea, 2018a, p.1)

En enero de 2018, Microsoft publicó su libro *The Future Computed: Artificial Intelligence and its Role in Society*, en el que se plantea que si, bien la inteligencia artificial ayudará a resolver los grandes problemas sociales, habrá desafíos y oportunidades. Por ejemplo, es necesaria la formulación de una legislación moderna, la observancia de principios éticos sólidos, la capacitación para nuevas habilidades e incluso las reformas del mercado laboral. La capacitación para un mundo impulsado por Inteligencia Artificial implica más que ciencia, tecnología, ingeniería y matemática. A medida que las computadoras se comportan más como los humanos, las ciencias sociales y las humanidades se volverán aún más

importantes. Los cursos de idiomas, arte, historia, economía, ética, filosofía, psicología y desarrollo humano pueden enseñar habilidades críticas, filosóficas y éticas que serán fundamentales para el desarrollo y la gestión de las soluciones de inteligencia artificial. Si la Inteligencia Artificial quiere alcanzar su potencial para servir a los humanos, entonces cada ingeniero tendrá que aprender más sobre las ciencias sociales y cada especialidad en las ciencias sociales necesitará aprender más sobre ingeniería (Microsoft, 2018).

Es que el potencial de la Inteligencia Artificial puede llegar a supuestos inimaginables. El científico investigador de Google, Ian Goodfellow, desarrolló una innovadora técnica llamada redes generativas antagónicas (GAN, por sus siglas en inglés de *Generative Adversarial Nets*) que les permite a las máquinas generar contenido original y realista, casi como si fueran humanos. Y lo más impactante es que lo logran a través de entrenamiento no supervisado. Básicamente son dos redes que compiten entre ellas, y mientras compiten, están forzadas a mejorar y superarse. Una de ellas es el generador, que es el que crea las imágenes o cualquier otro dato y la otra es el discriminador, que mira una imagen y trata de adivinar si se trata de una imagen real que proviene de los datos o de una imagen falsa originada por el generador. Es un proceso donde cada una de las redes va mejorando y aprende de su oponente y el generador logra hacer imágenes que engañan al discriminador. En la actualidad este tipo de aprendizaje no supervisado es incipiente y se utiliza para diseñar coronas dentales y, en el futuro, se podría emplear para idear por completo una casa: desde el exterior y hasta sus interiores (Jaimovich, 2018). También el Fondo *Global Goods* está trabajando en un sistema de imágenes por ultrasonido dotado de inteligencia artificial que integra un ecógrafo con aprendizaje profundo capaz de detectar automáticamente un ataque de neumonía y su progresión o la respuesta a un tratamiento (Vecchione, 2018).

Durante los días 29 y 30 de mayo de 2018, Argentina fue sede del Primer Foro sobre Inteligencia Artificial, Internet de las Cosas y Ciudades Inteligentes de América Latina, organizado por la Unión Internacional de Telecomunicaciones (UIT), el Ministerio de Modernización, Argentina, Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), Comisión Económica de las Naciones Unidas para América Latina y el Caribe (CEPAL), Organización

de las Naciones Unidas para el Desarrollo Industrial (ONUDI) , Comisión Interamericana de Telecomunicaciones (CITEL) , Comisión Técnica Regional de Telecomunicaciones (COMTELCA) y Telecomunicaciones de América Latina (Asiet) . Las ciudades consumen la mayoría de los recursos del mundo, causan emisiones masivas de gases de efecto invernadero. En América Latina, albergan a casi el 80 por ciento de la población de la región, una cantidad que probablemente aumentará en el futuro. Esto la convierte en una de las regiones más urbanizadas del mundo. Por tanto, el Foro exploró el papel de las nuevas tecnologías, incluida la conectividad de alta velocidad, resistente y de baja latencia y tecnologías como la computación distribuida, Internet de las Cosas, el aprendizaje automático y la Inteligencia Artificial para abordar los desafíos urbanos y dar forma a ciudades más inteligentes y sostenibles. Como documento final, el Foro produjo la Declaración de Buenos Aires, a través de la cual los participantes pudieron realizar un llamado a la acción, que destaca el valor del análisis de datos en tiempo real (basándose en los datos generados por las tecnologías inteligentes) para apoyar una economía circular eficiente en el uso de los recursos. El procesamiento y la gestión de datos éticos, la construcción de la confianza entre los ciudadanos y los recolectores de datos, entre otros, también se reconocieron como fundamentales para lograr todo el potencial de la inteligencia artificial e Internet de las Cosas. La declaración señala además la importancia de las normas internacionales para apoyar la interconexión e interoperabilidad de los sistemas de las ciudades. La estandarización y la regulación serán una contribución esencial a las protecciones contra las amenazas de seguridad específicas de Internet de las Cosas. También se alienta a incorporar soluciones de la Inteligencia Artificial a los servicios públicos¹.

Las máquinas, para poder aprender necesitan ser alimentadas con datos sobre los usuarios y sobre el entorno general que las rodea. Como no pueden percibir a través de los sentidos, toda la información necesaria para operar es ingresada por los matemáticos por medio de los paquetes de datos en forma de algoritmos.

¹ Véase: Primer Foro sobre Inteligencia Artificial e Internet de las Cosas en Ciudades Inteligentes y Sostenibles en América Latina. “Declaración de Buenos Aires. Inteligencia Artificial e Internet de las Cosas en Ciudades Inteligentes y Sostenibles en América Latina”. (2018). *ITU Foro* Recuperado de: <https://www.argentina.gob.ar/sites/default/files/buenosaires-declaration-spanish-final.pdf>

Dichos algoritmos les permiten a las máquinas poder identificar y clasificar la información mediante los métodos de *machine learning*- aprendizaje automático y de *deep learning*-aprendizaje profundo, más sofisticado. El primero alimenta a la máquina con altas dosis de datos predefinidos y categorizados por el humano para que los reconozca en el futuro. De esta manera, las máquinas son entrenadas para aprender a ejecutar diferentes tareas de forma autónoma. Luego, cuando son expuestas a nuevos datos, ellas se adaptan a partir de los cálculos anteriores y los patrones se moldean para ofrecer respuestas confiables. En lugar de programar reglas en una computadora y esperar el resultado, con *machine learning*, la máquina aprenderá esas reglas por cuenta propia.

El segundo, más preciso, es la parte del aprendizaje automático que, por medio de algoritmos de alto nivel, imita la red neuronal del cerebro humano. Para llegar al nivel de aprendizaje profundo más avanzado, los datos son sometidos a varias capas de procesamiento no lineares que simulan la forma de pensar de las neuronas. De forma simplificada, *deep learning* son esos algoritmos complejos contruidos a partir de un conjunto de diversas capas de “neuronas”, alimentados por cantidades inmensas de datos, *big data*, que son capaces de reconocer imágenes y habla, procesar el lenguaje natural y aprender a realizar tareas extremadamente avanzadas sin interferencia humana. La principal aplicación de estos algoritmos complejos son las tareas de clasificación, en especial, reconocimiento de imágenes.

Para llegar al nivel de aprendizaje profundo más avanzado, los datos son sometidos a varias capas de procesamiento no lineares que simulan la forma de pensar de las neuronas y se basa en la idea del aprendizaje exclusivamente desde el entrenamiento. Es así que, el programador, en lugar de enseñarle al sistema una lista enorme de reglas para resolver un problema, le proporciona directamente un modelo para que pueda, mediante ejemplos, realizar la evaluación.

Concretamente en *deep learning* se construyen algoritmos complejos mediante el empleo de las capas internas (la denominada “caja negra”) de la red neuronal. Cada capa procesa la información y arroja un resultado que se revela en porcentaje. Por ejemplo, se le solicita al sistema que busque en una foto un gato, concluirá que dicha imagen es en un 78% de probabilidad un gato y en un 22% no lo es. La segunda capa que analice la imagen combinará el resultado obtenido por la

primera capa con su propio juicio. De tal forma que la ponderación se modificará y se trasladará a la tercera capa, que también usará este dato para sacar su propia conclusión. Así sucesivamente porque las redes neuronales tradicionales solo contienen dos o tres capas ocultas, mientras que las redes profundas pueden tener hasta 150 capas, de ahí la denominación de profundo. Actualmente se utiliza en programas tales como: traductores inteligentes, reconocimiento de imágenes, asistentes de voz con capacidad de interpretación semántica. Sin embargo, el gran problema que presenta esta técnica de aprendizaje es el fenómeno de la caja negra ya que es imposible conocer lo que sucede en el interior de las capas intermedias de las redes neuronales para justiciar el resultado o decisión tomada. Es así que, solo se conoce la entrada y la salida del mismo, las capas intermedias son tan opacas como el cerebro humano (Torres, 2019).

Es de destacar que *machine learning* comenzó en los años 80 mientras que *deep learning* recién empezó a desarrollarse en el año 2010, con el surgimiento de computadoras poderosas y el aumento de los datos accesibles (Salesforce, 2018).

En base a lo expresado *ut supra*, en la actualidad, se diferencian dos tipos de inteligencia artificial: la débil y la fuerte.

La primera es la inteligencia artificial racional que se centra únicamente en tareas estrechas, capaz de realizar tareas inteligentes pero incapaz de pensar y sería fácil reconocer que se está hablando con una máquina. Se encarga de una tarea automáticamente, haciendo clasificaciones de acuerdo con unos parámetros predeterminados que va puliendo según la cantidad de datos disponible. Se trata de tareas rutinarias y técnicas sobre la base de los gustos del usuario, o clasificar correos electrónicos en deseados o no deseados. En cambio, la fuerte es aquella inteligencia artificial capaz de igualar o exceder la inteligencia de los humanos. Supone capacidad de abstracción, reflexión, afán creativo e improvisación. Al día de hoy es una utopía, pero si se llegase a conseguir supondría un dilema ético, ya que si son capaces de pensar y tener sentimientos deberían tener derechos, pero aun así, no dejarían de ser máquinas.

Quien introdujo esta distinción entre débil y fuerte fue el filósofo John Searle en un artículo crítico publicado en 1980. La inteligencia artificial fuerte implicaría

que una computadora convenientemente diseñada no simula una mente, sino que *es una mente* y por consiguiente debería ser capaz de tener una inteligencia igual o incluso superior a la humana. En cambio, la débil sólo se aplica a un tipo específico de problemas y no la intención de mostrar inteligencia humana en general. Searle en su artículo intenta demostrar que la fuerte es imposible (Searle, 1980)

Nick Bostrom, filósofo sueco de la Universidad de Oxford, considera que es el paso anterior a la creación de una superinteligencia, la que conceptualiza como un intelecto y desempeño cognitivo mucho mayor que los mejores cerebros humanos prácticamente en cualquier campo, incluyendo la creatividad científica, sabiduría general y habilidades sociales (Bostrom, 2014)

Alan Turing propuso, en 1950, un test- denominado comúnmente como el test de Turing-cuya función consistía en discernir si una máquina tenía inteligencia o no. Hoy día es considerado como la prueba de fuego de la inteligencia artificial. La prueba en palabras simples consiste en colocar a un humano a chatear con una máquina diseñada para generar respuestas similares a las de un ser humano, sin saber si de verdad lo hace con una maquina o con una persona. Luego de cinco minutos de charla, el evaluador debe decidir si la charla la sostuvo con un humano o con una máquina. El test de Turing se considera aprobado, si la computadora es capaz de convencer a al menos el 30% de los jueces con quien chatee. Esta prueba no evalúa el conocimiento de la máquina en cuanto a su capacidad de responder preguntas correctamente, solo se toma en cuenta la capacidad de ésta de generar respuestas similares a las que daría un humano (Turing, 1950).

En 1990 se inició un concurso- el Premio Loebner-, consistente en una competición de carácter anual entre programas de computadoras que sigue el estándar establecido en la prueba de Turing. Un juez humano se enfrenta a dos pantallas, una de ellas que se encuentra bajo el control de una máquina y la otra bajo el control de un humano. En el año 2014 y por primera vez desde que se creó la famosa prueba, un programa de computadora fue capaz de pasarla convenciendo al 33% de los jurados que se trataba de un humano. Se trató de Eugene Goostman, un chatbot o programa de computador diseñado para comportarse durante la conversación como un joven ucraniano de 13 años de edad. A pesar de notable logro del equipo desarrollador de Goostman y del hito de superar el famoso test de Turing, la reacción de la comunidad científica no fue

unánime. Se puede diferenciar entre los que celebran el alcance de este logro, y los que cuestionan la veracidad de los resultados obtenidos. Entre los argumentos más sólidos de sus detractores figura el hecho que una máquina consiga engañar a un humano sólo prueba que esa máquina es capaz de imitar la inteligencia, y no que en realidad la posea. A este respecto, ponen en duda incluso la validez del propio test de Turing ya que después de 60 años, se advierte la necesidad de actualizarlo (con una prueba visual) o con una manera distinta y más rigurosa de certificar los avances de la inteligencia artificial. Afirman que, al fin y al cabo, no se trata de una supermáquina, sino de la habilidad de un programa diseñado estrictamente para conversar, pero no para pensar.

En una extensa entrevista a Noam Chomsky publicada por el sitio web *The Atlantic*, en noviembre de 2012, mostró su decepción en cuanto a los avances de la inteligencia artificial expresando que su desarrollo era fruto de un craso error ya que su proceso de aprendizaje era más afín al conductismo de Skinner. La “nueva inteligencia artificial”, afirmó, centrada en el uso de técnicas de aprendizaje estadístico para extraer y predecir mejor los datos, no arrojaría principios generales sobre la naturaleza de los seres inteligentes o sobre la cognición. No hay forma que las computadoras complejas pudieran manifestar propiedades tales como la capacidad de elección. Jugar al ajedrez puede ser reducido a un mecanismo y cuando una computadora juega al ajedrez no lo hace del mismo modo que lo efectúa una persona; no desarrolla estrategias, no hace elecciones, simplemente recorre un proceso mecánico².

En efecto, los sistemas dotados de inteligencia artificial son capaces de resolver eficazmente una única tarea, una vez contextualizada. Sin embargo, no poseen la habilidad de sumar aprendizajes como los humanos. Por ejemplo, pueden saber jugar muy bien al ajedrez, pero ser incapaces de encontrar un tumor en una imagen. Y, si aprenden a identificar tumores, se olvidarán de la manera de jugar al ajedrez. Es lo que se conoce como el “olvido catastrófico”. Dicho concepto alude a la dificultad de enseñar nuevas habilidades al sistema para realizar nuevas tareas sin perder las funciones previamente aprendidas.

² La entrevista complete se puede leer en: Katz, Y. (1, noviembre de 2012) “Noam Chomsky on Where Artificial Intelligence Went Wrong” *The Atlantic*. Recuperado de: https://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/?single_page=true

Conforme las peculiaridades de cada una de las inteligencias artificiales anteriormente explicadas, Huawei (2018) -empresa privada multinacional china de alta tecnología especializada- establece las siguientes diferencias:

Inteligencia Artificial Débil	Inteligencia Artificial Fuerte
Existe en la actualidad	Solo en las películas (ej. Wall E)
Orientada a problemas concretos	Resuelve problemas abiertos
Reactivo: espera al incentivo humano	Proactivo: toma la iniciativa
Rígidos, un esquema	Flexible
Depende de la programación humana	Se autoprograma
Pocas redes neuronales	Muchas redes neuronales, en conflicto
No razonan, computan	Imitan el comportamiento humano
Aprenden de ejemplos similares	Aprenden de las personas
No reemplazan al humano	Tareas similares a las humanas
Tareas repetitivas	Aprenden tareas nuevas
No se pueden adaptar a los cambios	Adaptabilidad a nuevos escenarios

Fuente: Huawei, 2018

En esta misma línea de pensamiento y a los efectos de clarificar lo anteriormente expuesto, Arend Hintze, profesor de Biología Integrada y Ciencias de la Computación de la Universidad de la Michigan, clasifica la inteligencia artificial en cuatro tipos, según el nivel de complejidad:

1) Máquinas reactivas, la expresión más básica de la inteligencia artificial: este tipo de máquinas son las más fáciles de comprender para los seres humanos ya que es la que más se acerca al funcionamiento tradicional de una computadora al que se le proporciona una serie de datos y debe procesarlos para obtener un resultado satisfactorio. Cuenta con una base de datos en la que se incluyen toda la información que puede necesitar para desarrollar su tarea, así como los conocimientos para usar y combinarla para lograr el resultado esperado. El gran exponente fue **Deep Blue** y AlphaGo, explicados en el párrafo precedente.

2) Inteligencia artificial con memoria limitada: es la que se está expandiendo en la actualidad. Su nombre no hace referencia a la cantidad de memoria física que utilizan estos dispositivos, sino a su capacidad para registrar experiencias pasadas para aplicarlas en sus decisiones. Se utilizan en aplicaciones tan variadas como la industria automotriz, la inteligencia artificial integrada en dispositivos móviles o los asistentes por voz. Estos sistemas utilizan el aprendizaje automático para, en base a unos conocimientos base, añadir nueva información a su base de datos para mejorar su funcionamiento.

Por ejemplo, los autos **autónomos** poseen una serie de datos preestablecidos, como calles, señales u obstáculos. Además, se le añaden otros datos en constante cambio como son la velocidad y distancia de otros vehículos, incluso los peatones. De momento, tampoco se almacenan como recuerdos, aunque de cara al futuro podría ser útil en una misma ciudad.

3) Máquinas basadas en la teoría de la mente: la teoría de la mente es un principio psicológico en el que se sostiene que todos los seres vivos pueden tener emociones y pensamientos que afectan tanto a su comportamiento como a las decisiones que tomen. Esta capacidad de procesar emociones y exponerlas al entorno es uno de los grandes pilares en los que se basan las sociedades humanas, por lo que supone el primer paso en la integración de la inteligencia artificial como parte de esa sociedad y no como un mero dispositivo. Este planteamiento base supone un funcionamiento mucho más humano en el desarrollo de procesos de inteligencia artificial ya que el agente de inteligencia artificial debe ser capaz de comprender las emociones de su interlocutor y de mostrar las propias. Por el momento esta categoría se encuentra solo en fase experimental, pero ya existen procesos capaces de detectar estados de ánimo en las personas por los gestos de su cara, los matices de la voz, gritos, insultos y otros indicadores de conducta humana. Esta capacidad para obtener información sobre el estado psicológico de los interlocutores con los que la inteligencia artificial debe interactuar es un paso, pero todavía queda un largo camino por recorrer para conseguir que un agente de inteligencia artificial sea capaz de procesar ese estado de ánimo y mostrar algo parecido a lo que un humano podría reconocer como empatía real.

4) Inteligencia artificial con conciencia de sí misma: esta fase se encuentra en un estado meramente teórico. Es el nivel máximo ya que sería el más próximo a los seres vivos. Las máquinas comprenderían sus sentimientos y son capaces de verse a sí mismas con perspectiva en su entorno, de manera interna y siento capaces de predecir comportamientos y sentimientos ajenos (Hintze, 2016).

Stuart Russell y Peter Norvig diferencian cuatro enfoques de inteligencia artificial: 1) sistemas que piensan como humanos, que son capaces de tomar decisiones autónomas y resolver problemas, y también tienen capacidad de aprendizaje; 2) sistemas que actúan como humanos e imitan su comportamiento; 3) sistemas que piensan racionalmente, capaces de inferir una solución a un caso a partir de una información sobre un contexto dado y 4) sistemas que emulan la forma racional del comportamiento humano, como los sistemas inteligentes o expertos (Russell y Norvig, 2004).

Desde otro ángulo, se pueden identificar los siguientes casos en los que la inteligencia artificial se utiliza para los negocios.

1) Aprendizaje automático: generalmente, este concepto se confunde con el de inteligencia artificial débil. Es en este campo en donde los avances más importantes de este tipo de inteligencia se están llevando a cabo en la actualidad. En términos prácticos, es la ciencia que se encarga de hacer que las computadoras realicen acciones sin necesidad de programación explícita. La idea principal es que se les puede proporcionar datos a los algoritmos de aprendizaje automático y luego usarlos para saber cómo hacer predicciones o guiar decisiones.

Algunos ejemplos de algoritmos de aprendizaje automático incluyen los siguientes: diagramas de decisiones, algoritmos de agrupamiento, algoritmos genéticos, redes Bayesianas.

2) Aprendizaje profundo: consiste en una técnica de aprendizaje automático que utiliza redes neuronales para realizar tareas de clasificación. Algunos ejemplos de aplicaciones prácticas del aprendizaje profundo son las siguientes: identificación de vehículos, peatones y placas de matrícula de vehículos autónomos, reconocimiento y diferenciación de imágenes, traducción y procesamiento de lenguaje natural.

3) Descubrimiento de datos inteligentes: es el próximo paso en soluciones de inteligencia empresarial. La idea consiste en permitir la automatización total del ciclo de la inteligencia empresarial: la incorporación y preparación de datos, el análisis predictivo y los patrones y la identificación de hipótesis.

4) Análisis predictivo: el concepto principal de análisis predictivo o modelado significa que se puede utilizar un número de variables (ingresos, código postal, edad, etc.) combinadas con resultados (por ejemplo, buen o mal pagador) para generar un modelo que proporcione una puntuación (un número entre 0 y 1) que representa la probabilidad de un evento (por ejemplo, pago, migración de clientes, accidente, etc.). Gráficamente sería el contrato de seguro para un auto, donde se presentan variables que influyen en su riesgo. Detrás de las preguntas del asegurador se encuentra un modelo predictivo que informa sobre la probabilidad de que ocurra un accidente con base en su edad, código postal, género, marca de auto, etc. Es el mismo principio que se emplea en los modelos predictivos de crédito para identificar a los buenos y malos pagadores. Por lo tanto, los casos de uso en los negocios son amplios: modelos de crédito, modelos de segmentación de clientes (agrupamiento), modelos de probabilidad de compra y modelos de migración de clientes, entre otros.

b) Robótica

La robótica es un componente de la inteligencia artificial, y como ciencia, es la que estudia el diseño y construcción de máquinas capaces de desempeñar las tareas del ser humano mediante procesos mecanizados y programados.

Teniendo en cuenta su objetivo, Jones *et al* la definen en los siguientes términos:

La robótica consiste en el diseño de sistemas. Actuadores de locomoción, manipuladores, sistemas de control, sensores, fuentes de energía, software de calidad--todos estos subsistemas tienen que ser diseñados para trabajar conjuntamente en la consecución de la tarea del robot. (Jones *et al*, 1998, p.1)

Estas definiciones se asocian con el concepto de robot. Conforme la *International Federation of Robotics*, en línea con la *International Organization for Standardization ISO 8373:2012, Robots and robotic devices — Vocabulary*, un robot industrial es un manipulador multiusos, reprogramable y controlado automáticamente, programable en tres o más ejes, que pueden estar fijos o móviles para uso en aplicaciones de automatización industrial. (International Federation of Robotics, 2018, p.1)

También se lo ha conceptualizado como la herramienta interconectada, interactiva, cognitiva y física capaz de percibir su entorno, razonar sobre acontecimientos, hacer o revisar planes y controlar sus acciones (Holder, *et. al*, 2018).

García-Prieto Cuesta considera que un robot es una máquina, provista de cierta complejidad tanto en sus componentes como en su diseño o en su comportamiento, y que manipula información acerca de su entorno para así interactuar con él (García-Prieto Cuesta, 2018, p. 38).

Por su parte, Barrio Andrés (2018) apunta que es prudente referirse al robot *strictu sensu* al que define como “aquel objeto mecánico que capta el exterior, procesa lo que percibe, y a su vez, actúa positivamente sobre el mundo” (Barrio, 2018, p. 70).

De lo antedicho, surge inequívocamente que un robot es una máquina que puede variar desde un programa de *software* hasta un artefacto con forma de humanoide, en la medida que desempeñe tareas sin el control o intervención humana.

Para realizar dichas tareas están equipados con efectores como piernas, ruedas, articulaciones sensores y pinzas con el objeto de percibir y transmitir fuerzas físicas al entorno. Hoy en día, utilizan diversos tipos de sensores, incluyendo cámaras y ultrasonidos para medir el entorno, y giroscopios y acelerómetros para medir el propio movimiento del robot. Desde un punto de vista técnico y como características distintivas de los mismos se pueden mencionar: la capacidad de recoger datos mediante sensores; procesar los datos en bruto; planificar y cumplir acciones mediante conocimientos e informaciones adquiridas, generalmente, en función de objetivos prefijados pudiendo, además presentar capacidad de

comunicación con un operador, con otros robots o con una red externa; y la de aprendizaje. (Palmerini, 2017)

La mayoría de los robots actuales se basan en una de las siguientes tres categorías.

1) Manipuladores, o brazos robóticos: están físicamente anclados en su lugar de trabajo, por ejemplo, en una línea de ensamblaje o en la estación espacial internacional. El movimiento de los robots manipuladores normalmente requiere un desplazamiento en cadena de las articulaciones para posicionar a los efectores en cualquier lugar del entorno de trabajo. Los manipuladores son los robots industriales más extendidos con más de un millón de unidades instaladas en el mundo. Algunos manipuladores móviles se utilizan en hospitales para asistir a cirujanos.

2) Robots móviles: se desplazan por su entorno utilizando ruedas, piernas o mecanismos similares. Son muy utilizados para distribuir comidas en hospitales, mover contenedores a los muelles de carga y tareas similares. Otros tipos de robots móviles incluyen los vehículos aéreos sin tripulación, los drones, frecuentemente usados para tareas de vigilancia, fumigación y operaciones militares; también los vehículos submarinos sin tripulación que exploran el fondo marino.

3) El tercer tipo es un híbrido: un robot móvil equipado con manipuladores. Esto incluye a los robots humanoides, cuyo diseño se asemeja al torso humano.

Siguiendo el razonamiento de Sánchez-Urán Azaña y Grau Ruiz, se puede clasificar a los robots conforme diferentes criterios según se tenga en cuenta su complejidad, sus componentes, su aplicación, y la influencia de la inteligencia artificial en la robótica, que le permite adquirir mayor capacidad en las tareas específicas y un cierto grado de impredecibilidad de su comportamiento. Lo antedicho ha ocasionado un debate sobre la denominada “falacia del Androide”. En líneas generales, si bien se acepta que hay que entender las capacidades técnicas de los actuales robots, también se debe entender lo que nunca deberían hacer, vale decir, pensar en ellos como algo más que las máquinas que realmente son. Una primera clasificación, atendiendo a su aplicación, distingue entre robot industrial y robot de servicio. Dentro de este último se subclasifica en: de servicio personal,

utilizado para una tarea no comercial y robot de servicio profesional destinado a una tarea comercial, por lo general operado por una persona debidamente capacitada que es designada para iniciar, monitorear y detener el funcionamiento previsto de un robot. Y una segunda, teniendo en cuenta la interacción del robot con los humanos y el nivel de intervención humana en la actividad del robot, entre robots dependientes o no autónomos y robots independientes, autónomos o inteligentes. Con respecto a estos últimos, en los párrafos siguientes se identificarán sus características conforme la normativa europea (Sánchez-Urán Azaña y Grau Ruiz, 2018).

En cuanto al grado de autonomía, se puede diferenciar entre los robots no autónomos que realizan tareas automatizadas; los autónomos y la inteligencia artificial aplicada a los robots que perciben el ambiente externo por sí mismos, sin necesidad de órdenes pre programadas externas y con capacidad de discernir entre diferentes circunstancias que pueden acontecer a su alrededor. Estos sistemas deben obedecer tanto las órdenes de otros sistemas de inteligencia artificial como las humanas que interactúan con él. No existe ningún robot hoy con estas características.

Un desafío a sortear por los desarrolladores de los robots es la paradoja de Moravec, la cual consiste en que las máquinas son capaces de hacer cosas que son difíciles para los seres humanos, como jugar al ajedrez, pero incapaces de aprender, habilidades psicomotrices o perceptivas que hasta un bebé posee (López de Mántaras, citado por Paniagua, 2019).

En base a los conceptos explicados en los párrafos precedentes, sintéticamente se pueden enumerar las siguientes características de la inteligencia artificial y la robótica:

- 1) Eficacia en los cálculos tanto matemáticos, estadísticos. Su poder es abrumador.
- 2) No posee una conciencia: esta es de las características principales. Y es que a día de hoy no hay una inteligencia artificial que tenga una conciencia propia. Están simplemente hechos para obedecer órdenes.
- 3) Inteligencia limitada: hay máquinas con memoria limitada. Por este limitante, los humanos no pueden ser reemplazados. Un ejemplo claro, es que un robot no

puede entender en la parte emocional lo que quiere expresar una persona.

4) Carecen de emoción alguna: por lo menos hasta hoy, no tiene empatía, no tiene sociabilidad, el sentido común de las cosas les cuesta entenderlo.

5) Peligrosidad: depende de su uso. Si la inteligencia artificial es usada por criminales, el peligro rondará. Pero si es usada para mejorar y cambiar el mundo, quizás la vida en la tierra sea mejor y más justa

6) Obedientes: en la actualidad sólo sirven para obedecer. Así que dependiendo de la tarea que se le asigne a sí mismo, está dispuesta a ejecutarla.

7) Nada fuera de lo programado: los algoritmos reconocidos en estas inteligencias no son capaces de realizar instrucciones para las cuales no fueron programados inicialmente. El auto aprendizaje es significativo, sin embargo, sus funcionalidades son limitadas.

c) Ventajas y desventajas de la inteligencia artificial y la robótica

Los beneficios de la inteligencia artificial y la robótica son innumerables. Se destaca la importancia en la consecución de los Objetivos de Desarrollo Sostenible. Tanto así que puede utilizarse para combatir la pobreza extrema y mejorar la calidad de vida en áreas remotas de formas muy diversas. Puede ayudar a identificar las causas de la pobreza y a detectar las regiones más necesitadas. De esta forma se podrá combatir la pobreza en regiones específicas mediante la mejora de las tierras agrícolas y la agricultura. Además puede ayudar con la distribución de la ayuda en áreas más pobres y devastadas por la guerra o los desastres naturales.

La pobreza y la falta de educación están altamente vinculadas. La inteligencia artificial posee el potencial de crear sistemas educativos personalizados y sumamente eficaces que puedan adaptarse a las necesidades de los estudiantes. Puede utilizarse para automatizar actividades básicas en educación, como la calificación y ayudar a los habitantes a aprender nuevas habilidades para apoyar a las comunidades.

En la atención médica, se utiliza el poder de las computadoras para analizar y dar sentido a una gran cantidad de datos electrónicos sobre los pacientes, como

las edades, las historias clínicas, el estado de salud, los resultados de las pruebas, las imágenes médicas, las secuencias de ADN y muchas otras. Estos análisis ayudaran a los médicos a tomar mejores decisiones y más eficaces, mejorando la esperanza de vida de los pacientes.

En el ámbito laboral reduce los costos y salarios adicionales, aumentando la efectividad y productividad de la empresa. La automatización de las actividades posibilita que las empresas mejoren su desempeño al reducir los errores y mejorar la calidad y la velocidad; en algunos casos puede llevar a lograr resultados que vayan más allá de la capacidad humana. Como ya lo ha hecho a lo largo de la historia, la automatización también contribuye a mejorar la productividad. En épocas donde el crecimiento de la productividad es mediocre, puede dar el impulso necesario para el crecimiento económico y la prosperidad, así como ayudar a compensar el impacto de la disminución de la población en edad de trabajar en muchos países.

En el último informe de PwC *Fourth Industrial Revolution for Earth* se identificaron más de ochenta aplicaciones de inteligencia artificial que solucionan los retos que afronta el planeta, a saber, el cambio climático, la biodiversidad y la conservación, la salud de los océanos, la seguridad de los recursos hídricos, la pureza del aire y el riesgo de catástrofes. Dentro de las aplicaciones se puede enumerar:

- a) Vehículos eléctricos autónomos y conectados,
- b) Redes eléctricas distribuidas,
- c) Agricultura y sistemas alimentarios inteligentes,
- c) Meteorología, climatología y pronósticos,
- d) Respuesta inteligente a las catástrofes,
- e) Ciudades inteligentes, conectadas y habitables, diseñadas por Inteligencia Artificial,
- f) Un planeta digital transparente,
- g) Aprendizaje por refuerzo para avanzar en el estudio de la tierra. (PwC, 2018)

La inteligencia artificial y la robótica podrán desempeñar tareas peligrosas, aburridas o difíciles para el ser humano. En concreto pueden ayudar con tareas del hogar como la limpieza, así como también desarrollar vehículos autónomos.

Puede realizar tareas que el hombre nunca hubiera podido concretar o bien porque afecta la integridad del investigador o por falta de tecnología. Por ejemplo, en la exploración del espacio exterior.

En el área de la minería cabe resaltar ventajas como la seguridad en la excavación y la extracción de combustible. En la búsqueda de nuevos descubrimientos en la tierra, se resalta la posibilidad de examinar el océano y otros sitios donde el ser humano no ha podido llegar.

No obstante la importancia que ha generado la inteligencia artificial, existen críticas en contra de su implementación, entre las cuales principalmente se señalan las siguientes:

Este tipo de máquinas pueden utilizarse con fines poco éticos e incluso ilegales en cualquier parte del mundo y en diferentes áreas, ya sea de tipo militar, espionaje industrial, tráfico y comercio ilegal de datos personales, etc. Una práctica que va en aumento es su utilización para campañas políticas.

Específicamente, en el área de defensa nacional, se desarrollaron sistemas de combate capaces de detectar y clasificar sus objetivos y de dirigir el lanzamiento de misiles antiaéreos. Por ejemplo, los sistemas de misiles tierra-aire S-400 en Rusia y el sistema estadounidense de información Aegis, que controla el armamento de las embarcaciones de guerra. A lo largo de la zona desmilitarizada, en la frontera con la República Popular Democrática de Corea, la República de Corea ha desplegado varios robots militares SGR-A1 encargados de la vigilancia que pueden disparar automáticamente ante la presencia del enemigo, pero no contra las personas que tienen las manos en alto. Ninguno de estos sistemas es utilizado por los militares en modo automático, vale decir sin la intervención humana, pero se teme que una mayor autonomía se traduzca en sistemas que hagan posible que una máquina tenga la capacidad de seleccionar un objetivo y atacarlo sin intervención humana, lo que algunos denominan “robots asesinos”. (Sychev, 2018)

Es por ello que un grupo de expertos en robótica, inteligencia artificial, científicos e investigadores, entre los que se puede mencionar a Stuart Russell, Elon Musk, Stephen Hawking, Noam Chomsky, Peter Norving y Mustafa

Suleyman (fundador de la compañía DeepMind Technologies, hoy propiedad de Google) publicaron, en *Future of Life Institute*, una carta solicitándole a las Naciones Unidas la prohibición del desarrollo y uso de armas mortíferas controladas por software autónomo. Señalaron que las armas letales autónomas ocasionarán conflictos armados de una trascendencia nunca vista y en una escala de tiempo inimaginable para los seres humanos. Igualmente podrían convertirse en armas terroríficas de déspotas en contra de poblaciones inocentes y manipuladas de maneras desastrosas. Además, cuando un ser humano se encuentra en una situación de conflicto, toma decisiones en las que intervienen, entre otras, sus actitudes morales, sentimientos y emociones. La observación directa del sufrimiento de los demás produce, además, un efecto disuasivo en los soldados, aun cuando, entre los militares profesionales, la compasión y la sensibilidad terminan por desaparecer. Consideraron que era como abrir la caja de Pandora con consecuencias impredecibles³.

Dichas críticas provienen del temor que origina la manera en que funcionan las máquinas inteligentes, pues poseen memoria y capacidad autónoma de aprendizaje, lo que a su vez les permite ampliar sus conocimientos y bancos de datos; por ellos existe la posibilidad que ganarán relevancia mientras más aprendan de su entorno e impactando el modo de vida de los habitantes. Por tanto, es de suma importancia establecer pautas éticas respecto al desarrollo y uso de productos y servicios basados en inteligencia artificial.

Es ilustrativo mencionar la polémica que generó la utilización del algoritmo COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) en Wisconsin, y en diez Estados de Norteamérica. El mismo determina la probabilidad de reincidencia y aconseja al juez a la hora de tomar la decisión acerca de la duración de la pena de prisión de un condenado. Los códigos para poder acceder al algoritmo y verificarlo son de carácter privado y utiliza variables como el sexo o la edad, o por su entorno económico-social. De un pormenorizado estudio realizado por un grupo de periodistas, se verificó que el sistema solo acertó un 60% de las veces, y tan sólo un 20% en las predicciones

³ La carta se puede leer en el siguiente link: Russell, *et al* (28, julio de 2015) “Autonomous Weapons: An Open Letter From AI & Robotics Researchers” *The Future Life Institute*. Recuperado de: <https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1>

de reincidencia con violencia. De este completo y exhaustivo análisis estadístico se extrajeron conclusiones bastante peligrosas, como ser que los acusados negros fueron catalogados por el algoritmo con mayor riesgo de reincidencia del que realmente tenían. En cambio, los blancos fueron catalogados con menor riesgo de reincidencia del que realmente tenían (Larson *et al.*, 2016).

Además, los sistemas requieren mucho tiempo y dinero: el capital que requiere estas grandes tecnologías solo se puede lograr por gigantes corporaciones o países desarrollados.

Por ser software, necesitan de constantes actualizaciones, ocasionando dependencia tecnológica.

Aumento del desempleo: la automatización del trabajo y su reemplazo por máquinas inteligentes, producirá una disminución de una cantidad considerable de puestos de trabajo. Esto traerá consigo una necesidad creciente de que las personas sean reeducadas profesionalmente. Ejemplo de esto se aprecia en lo declarado en el año 2017 por Ángel Gurría, secretario general de la Organización para la Cooperación y el Desarrollo Económico (OCDE), quien señaló que aproximadamente 9% de los empleos de los países miembros de ese organismo tienen posibilidades de ser automatizados, y por consecuencia, el riesgo para personas con bajos niveles educativos de perder su empleo es mayor.

Según las investigaciones del Instituto de Investigaciones de la Multinacional McKinsey, en el año 2030 la tecnología de la inteligencia artificial impactará la economía global principalmente en el sector laboral, lo que conllevará a que el 20% de la población económicamente activa a nivel mundial sea sustituida por alguna clase de máquina inteligente (McKinsey Global Institute, 2017).

III. Normativa Europea

El 19 de abril de 2016 la Comisión Europea publicó el documento *Advancing the Internet of Things in Europe*. resaltando que las tecnologías digitales estaban evolucionando a un ritmo tan rápido que debía supervisarse constantemente el marco jurídico. A tales efectos, expresó la necesidad de examinar el marco normativo de las innovaciones digitales para esclarecer, entre otros temas, la

propiedad y el uso de los datos generados en un contexto industrial; los sistemas que actúan de modo autónomo, como los vehículos sin conductor o los drones y las aplicaciones y demás software no empotrado, vale decir no incluido en un medio tangible, temáticas que representan un desafío para las normas de seguridad y responsabilidad (Comisión Europea, 2016).

El 6 de diciembre de 2018, el Comité Económico y Social Europeo aprobó el dictamen de iniciativa intitolado “Dictamen del Comité Económico y Social Europeo sobre Confianza, privacidad y seguridad de los consumidores y las empresas en el internet de las cosas”. Parte de la premisa que uno de los pilares del internet de las cosas radica en la toma de decisiones automática sin intervención humana y que la interconectividad de personas y objetos, representa oportunidades para los ciudadanos y las empresas. Sin embargo, reconoce la necesidad de dotar de un marco jurídico adecuado estableciendo una serie de garantías y controles para asegurar una implantación satisfactoria del mismo que garantice los derechos de los consumidores. El mismo avanza sobre todos los riesgos a la seguridad, privacidad, intimidad y de carácter ético o contrarios a los derechos humanos fundamentales que esta tecnología conlleva. Y solicita a las instituciones europeas y a los Estados miembros, entre otras medidas, a velar por la protección de la seguridad y la privacidad; adoptar medidas de seguimiento y control; cubrir los vacíos legales en las tecnologías disruptivas; delimitar la responsabilidad en la cadena de suministro del producto; fomentar las iniciativas de normalización europeas e internacionales; tener en cuenta a los grupos vulnerables y establecer mecanismos efectivos y colectivos de solución de controversias (Comité Económico y Social Europeo, 2018).

En marzo de 2012, la Comisión Europea habilitó una partida presupuestaria para financiar el “Proyecto Robolaw”, denominación abreviada de “*Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics.*” El objetivo principal del precitado proyecto consistió en elaborar un informe detallado con todas las cuestiones éticas y legales que planteaban los robots, así como ofrecer principios para guiar a los legisladores a la hora de regular el tema. Los resultados de dicha investigación, presentados en octubre de 2014, han constituido la base de la Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho

civil sobre robótica (Palmerini, 2017).

En su introducción, además de mencionar al monstruo de Frankenstein, el mito de Pigmalión, el Golem de Praga, entre otros, la precitada Resolución reconoce que la humanidad se encuentra en los albores de una nueva revolución industrial, en la cual los robots, *bots*, androides y otras formas de inteligencia artificial cada vez más sofisticadas, probablemente afecten a la sociedad en su conjunto. En consecuencia, considera necesario que el legislador pondere las consecuencias jurídicas y éticas, sin obstaculizar con ello la innovación y que disponga de una serie de normas en materia de responsabilidad, transparencia y rendición de cuentas que reflejen los valores humanistas intrínsecamente europeos y universales. Al mismo tiempo, alerta que tal normativa no debería afectar al proceso de investigación, innovación y desarrollo en el ámbito de la robótica. Solicita a la Comisión Europea que presente una propuesta de Directiva relativa a las normas de legislación civil en materia de robótica, siguiendo las recomendaciones detalladas que figuran en el Anexo de dicha Resolución. Entre las principales líneas de trabajo a futura consideración legislativa, se pueden enumerar:

1) En primer lugar, parte de la necesidad de establecer una definición europea común de robots autónomos inteligentes, incluidas las de sus subcategorías, conforme las siguientes características: la capacidad de adquirir autonomía mediante sensores y/o mediante el intercambio de datos con su entorno (interconectividad) y el análisis de dichos datos; la capacidad de aprender a través de la experiencia y la interacción; la forma del soporte físico del robot y la capacidad de adaptar su comportamiento y acciones al entorno.

2) Formular reglas relativas a la responsabilidad civil al entender que podría ser insuficiente el marco jurídico vigente en el ámbito de la responsabilidad contractual para cubrir los daños causados por la nueva generación de robots, en la medida en que se les puede dotar de capacidades de adaptación y aprendizaje que entrañan cierto grado de imprevisibilidad en su comportamiento, ya que un robot podría aprender de forma autónoma de sus experiencias concretas e interactuar con su entorno de un modo imprevisible y propio únicamente a ese robot.

3) La sanción de un código de conducta ético voluntario como base para la regulación de los impactos sociales, ambientales y de salud humana de la robótica, teniendo en cuenta los principios recogidos en la Carta sobre Robótica establecida en ese mismo Anexo. En dicha Carta pone énfasis en las fases de investigación y desarrollo de la trayectoria tecnológica pertinente (proceso de concepción, análisis ético, controles de auditoría, entre otros.). Dicho código de conducta debería tener como objetivo no solo la necesidad de abordar la necesidad de cumplimiento de determinadas normas éticas por parte de investigadores, profesionales, usuarios y diseñadores, sino también de introducir un procedimiento para la resolución de los dilemas éticos y permitir que estos sistemas puedan funcionar de una manera éticamente responsable.

4) Posteriormente, presenta un modelo de Código de Conducta Ética voluntario para los Ingenieros en Robótica y un Código Deontológico para los Comités de Ética de la Investigación. El primero ofrece un conjunto de principios generales y directrices para las medidas que adopten todas las partes interesadas. Invita a todos los investigadores y diseñadores a actuar de forma responsable respetando la dignidad, intimidad, autonomía y la seguridad de las personas, vale decir los derechos humanos internacionalmente reconocidos. Solicita la estrecha colaboración entre todas las disciplinas para garantizar la investigación en robótica en Europa de un modo seguro, ético y eficaz. Con referencia a los organismos públicos y privados de financiación de la investigación en dicho ámbito, les aconseja que exijan la realización y presentación de una evaluación del riesgo para cada propuesta de financiación en la materia. También se refiere a los investigadores, los que deben respetar los siguientes principios: a) beneficencia : los robots deben actuar en beneficio del hombre; b) principio de no perjuicio o maleficencia: no hacer daño, en virtud del cual los robots no deberían perjudicar a las personas; c) autonomía: la capacidad de tomar una decisión con conocimiento de causa e independiente sobre los términos de interacción con los robots; d) justicia en la distribución de los beneficios asociados a la robótica y la asequibilidad de los robots utilizados en el ámbito de la asistencia sanitaria a domicilio y de los cuidados sanitarios en particular. La responsabilidad siempre incumbe a los seres humanos, no a los robots.

Especial mención tienen los siguientes principios: a) principio de precaución, ya que los investigadores deben anticiparse a los posibles impactos de sus resultados sobre la seguridad y adoptar las precauciones debidas, fomentando el progreso en beneficio de la sociedad y del ambiente; b) el de participación ya que los ingenieros en robótica deben garantizar la transparencia y el respeto al derecho legítimo de acceso a la información de todas las partes interesadas; c) el de rendición de cuentas sobre las consecuencias sociales y medioambientales y el impacto sobre la salud humana que la robótica puede conllevar para las generaciones presentes y futuras; d) el de seguridad en la medida en que los diseñadores en robótica deben respetar la integridad física, la seguridad, la salud y los derechos de -las personas y divulgar con prontitud los factores que pudieran poner en peligro a la población o al medio ambiente; e) reversibilidad, la que considera como una condición necesaria de la posibilidad de control, como un concepto fundamental en la programación de robots para que se comporten de manera segura y fiable. Un modelo de reversibilidad indica al robot qué acciones son reversibles y, en su caso, el modo de revertirlas. La posibilidad de deshacer la última acción o secuencia de acciones, permite al usuario anular las acciones no deseadas y volver a la fase inocua de su trabajo; f) el de privacidad: la información privada debe conservarse en total seguridad y solo se utilizará de forma adecuada. El consentimiento consciente de la persona tiene que solicitarse y recabarse con anterioridad a cualquier interacción hombre-máquina. A tal efecto, los diseñadores en robótica tienen la responsabilidad de desarrollar y aplicar procedimientos para garantizar el consentimiento válido, la confidencialidad, el anonimato, el trato justo y el respeto de la legalidad y g) maximizar beneficios y reducir al mínimo los daños en todas las fases, desde su concepción hasta su difusión. Es conveniente evitar cualquier daño a los participantes o a los seres humanos que participen en los experimentos, ensayos o estudios en el ámbito de la investigación. En caso de aparición de riesgos inevitables, normalmente no deberían ser superior a los existentes en la vida cotidiana, es decir, las personas no han de estar expuestas a riesgos mayores a aquellos a los que están expuestas en su vida diaria y reposar en los principios de proporcionalidad y de precaución. Por su parte, el Código Deontológico establece la necesidad de un proceso de revisión ética independiente de la propia investigación y a cargo de un Comité de

Ética de la Investigación el que debería tener un carácter multidisciplinar: incluir a hombres y mujeres, estar constituido por miembros con una amplia experiencia y conocimientos en el ámbito de la investigación en robótica.

5) La creación de una Agencia Europea de Robótica e Inteligencia Artificial y un Registro Europeo de los robots inteligentes. Al mismo tiempo destaca que el desarrollo de la tecnología robótica debe orientarse a complementar las capacidades humanas y no a sustituirlas.

6) Elaborar un estatuto de persona electrónica.

7) Estudiar nuevos modelos de empleo y analizar la viabilidad del actual sistema tributario y social con la llegada de la robótica.

8) Garantizar la interoperabilidad de los robots autónomos conectados a la red autónoma que interactúan entre sí.

9) Finalmente, detalla la licencia para los diseñadores y para los usuarios. En esta última, se autoriza a los usuarios a utilizar un robot sin miedo de perjuicio físico o psicológico, a que efectúe las tareas para las que haya sido expresamente concebido, respetando la fragilidad humana, tanto física como psicológica, así como las necesidades emocionales de los seres humanos, a la vida privada de las personas, incluida la desactivación de videomonitores durante procedimientos íntimos, y no están autorizados a modificar los robots para utilizarlos como armas (Parlamento Europeo, 2017)

El Anexo también menciona la trazabilidad de los robots avanzados, pero siempre bajo el control de los seres humanos sobre las máquinas inteligentes (Valente, 2019)

En la Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones, “Inteligencia artificial para Europa” del 25 de abril de 2018 se destaca que la aparición de la inteligencia artificial, en particular, el complejo ecosistema que la hace posible y la adopción autónoma de decisiones, exige una reflexión acerca de la idoneidad de algunas de las normas establecidas en materia de seguridad y de responsabilidad civil. Por ejemplo, los robots avanzados y los productos del internet de las cosas pueden no comportarse con arreglo a lo previsto cuando el sistema entró en funcionamiento por primera vez. Teniendo en

cuenta los desafíos planteados por las tecnologías digitales emergentes, reconoce la necesidad de revisar los marcos jurídicos en materia de seguridad, las normas sobre responsabilidad por los daños causados por productos defectuosos, así como la Directiva sobre máquinas a fin de constatar si resultan adecuados para el fin previsto o si existen lagunas que deben colmarse. Considera de vital importancia que las personas puedan controlar los datos generados por la utilización de estas herramientas y saber si están comunicándose con una máquina o con otro ser humano, así como la forma de garantizar que las decisiones del sistema puedan verificarse o corregirse. Es que, en muchos casos, la toma de decisiones automática, sin intervención humana, afectó los derechos de los consumidores, ocasionó riesgos éticos o contrarios a los principios y derechos humanos fundamentales. En consecuencia, la Comisión puso en marcha la estrategia europea sobre inteligencia artificial centrada en el ser humano (Comisión Europea 2018a). Y para concretizar dicha estrategia desarrolló junto con los Estados miembros un plan coordinado sobre la inteligencia artificial (Comisión Europea, 2018b).

Es por eso que Elon Musk, Stephen Hawking y cientos de otros investigadores, líderes tecnológicos y científicos han respaldado una lista de 23 principios correctores elaborados en la conferencia organizada por el Future of Life Institute en enero de 2017 en California, para guiar el desarrollo de la inteligencia artificial en una dirección productiva, ética y segura. En la misma línea se adscribe la Comunicación de la Comisión al Parlamento Europeo, al Comité Económico y Social Europeo y al Comité de las Regiones: Generar confianza en la inteligencia artificial centrada en el ser humano, del 8 de abril de 2019. La Comunicación pretende ser la fase piloto para la implementación práctica de la orientación ética para el desarrollo y uso de la inteligencia artificial. Comienza destacando que la Unión se fundamenta en un sólido marco normativo, que constituirá la referencia mundial para la inteligencia artificial centrada en el ser humano. Entre las normativas más significativas enumera el Reglamento general de protección de datos personales (RGPD), el Reglamento relativo a la libre circulación de datos no personales, el Reglamento de Ciberseguridad y el Reglamento sobre la privacidad y las comunicaciones electrónicas propuesto.

No obstante, este tipo de inteligencia conlleva nuevos retos, ya que permite a las máquinas aprender, tomar decisiones y ejecutarlas sin intervención humana.

Ahora bien, las decisiones adoptadas mediante algoritmos pueden dar datos incompletos y no fiables, pueden ser manipulados por ciberataques, sesgados o simplemente estar equivocados. Aplicar de forma irreflexiva la tecnología a medida que se desarrolla produciría resultados problemáticos, violatorios de los derechos humanos. Por tanto, las aplicaciones de inteligencia artificial no solo deben ajustarse a la ley, sino también respetar principios éticos y garantizar que su implementación evite daños involuntarios. En cada una de las fases de su desarrollo debe estar garantizada la diversidad en cuanto al género, el origen racial o étnico, la religión o las creencias, la discapacidad y la edad. A los efectos de la elaboración de directrices éticas basadas en el marco regulador existente y que sean aplicadas por desarrolladores, proveedores y usuarios de la inteligencia artificial, la Comisión creó un grupo de expertos de alto nivel sobre la Inteligencia Artificial, cuyas conclusiones fueron presentadas ante dicho cuerpo en marzo de 2019.

Las directrices propugnan que, para lograr una inteligencia artificial fiable, son necesarios tres componentes: a) conforme a la ley, b) respetar los principios éticos y c) sólida. Sobre estos tres componentes y los valores europeos de respeto a la dignidad y a los derechos humanos, las directrices señalan siete requisitos esenciales que las aplicaciones inteligentes deben respetar para ser confiables, a saber:

1) Intervención y supervisión humana: debe lograrse a través de mecanismos de gobernanza, como participación humana (*human-in-the-loop*), la supervisión humana (*human-on-the-loop*), o el control humano (*human-in-command*).

2) Solidez y seguridad técnica: los algoritmos deben ser lo suficientemente seguros para ser resilientes, fiables y sólidos para resolver errores o incoherencias durante todas las fases del ciclo vital del sistema inteligente. Asimismo, deben tener la fortaleza para repeler los ataques abiertos como tentativas más sutiles de manipular datos o los propios algoritmos y deben garantizar un plan de contingencia en caso de problemas. Sus decisiones deben ser acertadas o, como mínimo, reflejar su nivel de acierto, y sus resultados, reproducibles

3) Privacidad y gestión de datos: deben garantizarse la privacidad, el pleno control y la protección de datos en todas las fases. Los registros digitales del

comportamiento humano pueden permitir que los sistemas inteligentes inferan no solo las preferencias, la edad y el sexo de las personas, sino también su orientación sexual o sus opiniones religiosas o políticas. Es necesario asegurar al consumidor que no se utilizarán para perjudicarles o discriminarles.

4) Transparencia: debe garantizarse la trazabilidad de los sistemas; se debe registrar y documentar tanto las decisiones tomadas como la totalidad del proceso que dio lugar a las decisiones. Es importante comunicar adecuadamente las capacidades y limitaciones del sistema a las distintas partes interesadas afectadas.

5) Diversidad, no discriminación y equidad: apunta a garantizar la accesibilidad mediante un enfoque de diseño universal para lograr la igualdad de acceso para las personas con discapacidades.

6) Bienestar social y medioambiental: debe tomarse en cuenta su impacto sobre el ambiente y sobre otros seres sensibles. Todos los seres humanos, incluso las generaciones futuras, deberían beneficiarse de la biodiversidad. Debe fomentarse la sostenibilidad y la responsabilidad ecológica de los sistemas inteligentes y las soluciones que abordan deben tener en cuenta los Objetivos de Desarrollo sostenible de las Naciones Unidas.

7) Rendición de cuentas: es imprescindible la posibilidad de auditar estos sistemas por parte de auditores internos y externos y la disponibilidad de los informes de evaluación. Los potenciales impactos negativos de los sistemas deben señalarse, evaluarse, documentarse y reducirse al mínimo. En caso de producir efectos adversos injustos, se debe prever mecanismos accesibles que garanticen una reparación adecuada.

Estos requisitos están pensados para ser aplicados a todos los sistemas de inteligencia artificial, pero la intensidad va a depender del contexto específico de aplicación. Por ejemplo, la aplicación inteligente que califica a un libro de lectura como inadecuado es mucho menos peligrosa que otra que diagnostique erróneamente un cáncer y, por tanto, puede estar sujeta a una supervisión menos estricta.

Es de destacar que estas directrices no son vinculantes, no tienen fuerza de ley, son *soft law*. No obstante, muchas disposiciones vigentes del Derecho de la Unión reflejan uno o varios de estos requisitos esenciales. Por ejemplo, las

normas de seguridad, de protección de los datos personales, de privacidad o de protección del medio ambiente (Comisión Europea, 2019).

El 12 de febrero de 2019, el Parlamento Europeo dictó una Resolución referida a la política industrial global europea en materia de inteligencia artificial y robótica. En la misma tuvo en cuenta su Resolución del 16 de febrero de 2017, referida a las recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica-comentada *ut supra*- y demás Resoluciones sobre la materia, así como la estrategia europea sobre inteligencia artificial centrada en el ser humano de la Comisión Europea.

Dicho documento recoge las aspiraciones y el plan estratégico de Europa para diseñar y desarrollar acciones concretas hacia el nuevo liderazgo en inteligencia artificial y robótica. No desconoce la importancia de las alianzas internacionales con aquellos países fuera de la Unión que tengan un nivel avanzado mayor, sino más bien, la promueve. Parte de la base que la inteligencia artificial y la robótica llevan ya una serie de años formando parte de la vida cotidiana, creando nuevos modelos empresariales y desempeñando un papel clave en la transformación de las sociedades y en muchos sectores de la economía, como la industria, la sanidad, la construcción y el transporte. En particular, que la inteligencia artificial es una de las tecnologías estratégicas para el siglo XXI y que alrededor de una cuarta parte de todos los robots industriales y la mitad de todos los de servicios profesionales existentes en el mundo están producidos por empresas europeas. Es por ello que la Unión ya cuenta con un importante activo en el que debería basar su política industrial europea. Sin embargo, debe competir con las enormes inversiones realizadas por terceros países, especialmente los Estados Unidos y China.

También alude a que, en su Resolución del 16 de febrero de 2017, el Parlamento instaba a la Comisión a que propusiera un marco jurídico coherente en relación con el desarrollo de la robótica, incluidos los sistemas autónomos y los robots autónomos inteligentes, siempre sobre la base del pleno respeto de los derechos humanos y en particular de los principios de protección de datos, privacidad y seguridad, preservando la dignidad, la autonomía y la autodeterminación de las personas. Esta realidad, junto con el crecimiento exponencial en los avances tecnológicos, hace necesaria la revisión de las normas y procesos existentes y, en

su caso, modificarse para tener en cuenta la inteligencia artificial y la robótica. En dicha revisión, considera que se debe tener en cuenta la ciberseguridad ya que es vital para garantizar que los datos no sean manipulados de forma malintencionada o usados indebidamente en forma perjudicial para los ciudadanos o las empresas. Asimismo, como el uso de la inteligencia artificial por sí solo no garantiza la verdad ni la equidad, todos los implicados en el diseño, desarrollo y comercialización deben incorporar desde el inicio-*ab initio*- características de seguridad y ética, reconociendo de ese modo que deben estar preparados para aceptar la responsabilidad jurídica por lo que respecta a la calidad de la tecnología que producen. En particular, en el ámbito sanitario, la utilización de la inteligencia artificial debe basarse siempre en el principio de responsabilidad según el cual la máquina es operada por el ser humano.

En el precitado documento, destaca que la política industrial global europea en materia de inteligencia artificial y robótica debe basarse en los siguientes parámetros:

1) Una sociedad inteligente, vale decir con el apoyo de la inteligencia artificial y la robótica. En primer lugar, destaca que, en esa nueva sociedad, la mano de obra, al igual que en anteriores revoluciones tecnológicas, se verá afectada ya que se sustituirán algunos puestos de trabajo, pero también se crearán otros nuevos que transformarán la vida y práctica profesional. Por tanto, insta a los Estados miembros a que se centren en la reconversión de los trabajadores de las industrias más afectadas por la automatización de las tareas, en el desarrollo de las capacidades de los trabajadores en los nuevos desafíos laborales y en la elaboración de programas de alfabetización digital en los centros escolares como uno de los factores más importantes de la inteligencia artificial. A ese respecto, también reclama a la Comisión y a los Estados miembros que desarrollen y apliquen estrategias de formación y reciclaje en materia de capacidades digitales para todas las edades. Asimismo, un mayor uso de la robótica y la inteligencia artificial reducirá la exposición de los seres humanos a condiciones perjudiciales y peligrosas.

Especial atención merece el uso malintencionado de la inteligencia artificial y los derechos fundamentales ya que podría constituir una amenaza para la seguridad digital, la física y la pública. Menciona varios malos usos de la

inteligencia artificial, como por ejemplo, para llevar a cabo ataques a gran escala, muy selectivos y muy eficientes, para lanzar campañas de desinformación y, en general, para reducir el derecho de las personas a la autodeterminación, poniendo en riesgo a la democracia. Por ello, le solicita a la Comisión que proponga un marco que penalice las prácticas de manipulación de la percepción cuando los contenidos personalizados o las fuentes de noticias provoquen sentimientos negativos y una deformación de la percepción de la realidad con consecuencias negativas, por ejemplo, en los resultados electorales o sobre cuestiones sociales como la migración. Otro riesgo a tener en cuenta se relaciona con la elaboración de perfiles que clasifican a las personas, sufriendo discriminación en función de tal clasificación y su estigmatización. Expresa su gran preocupación por el empleo de aplicaciones de inteligencia artificial, incluido el reconocimiento facial y vocal, en los programas de vigilancia emocional, es decir, el seguimiento de las condiciones mentales de los trabajadores con el fin de obtener un incremento de la productividad y preservar la estabilidad social.

2) La vía tecnológica hacia la inteligencia artificial y la robótica: en cuanto a investigación y desarrollo, recuerda que Europa cuenta con una comunidad líder mundial en investigación y que representa el 32 % de las instituciones de investigación de la inteligencia artificial en todo el mundo. Subraya que se debe invertir no solo en la tecnología y en el desarrollo de la innovación, sino también en los ámbitos sociales, éticos y de responsabilidad relacionados con la inteligencia artificial; estima que todos los modelos que se desplieguen deben ser éticos desde su concepción y que deben respetar los derechos humanos. También destaca que un entorno de investigación competitivo también es fundamental y reconoce que las remuneraciones de los investigadores de la Unión siguen siendo muy inferiores a las de sus homólogos de los Estados Unidos y China, reconociéndose que esa es la principal razón que los lleva a abandonar Europa. Además, señala la importancia de invertir más en este campo para seguir siendo competitivos, pero que no permitan que se financie la inteligencia artificial con fines armamentísticos ni en el desarrollo de la conciencia artificial.

3) Política industrial: si bien la inteligencia artificial y la robótica ya tienen aplicaciones industriales arraigadas desde hace tiempo, los avances en este

ámbito están aumentando y ofrecen amplias y diversas aplicaciones en todas las actividades humanas. Por ello cualquier marco regulador debe ser flexible y permitir la innovación y el libre desarrollo de nuevas tecnologías. Destaca que el sector público es prioritario y que también puede beneficiarse en varios aspectos de la inteligencia artificial y la robótica, por lo cual, deben invertir en programas de educación y formación en la materia para ayudar a sus empleados. La adopción por el sector público de inteligencia artificial fiable puede contribuir a la reforma de la administración pública en la toma de decisiones y a mejorar tales servicios. En materia de salud, enfatiza que el contacto humano es un aspecto esencial de la atención a las personas. Sin embargo, cuando la inteligencia artificial se combina con el diagnóstico humano, la tasa de error tiende a ser considerablemente inferior respecto al diagnóstico únicamente humano. Asimismo, destaca la contribución de la inteligencia artificial y la robótica a la innovación de prácticas y técnicas preventivas, clínicas y de rehabilitación en el sector sanitario, haciendo especial referencia a los beneficios que aportan a los pacientes con discapacidad. No obstante, la utilización de datos en dicho sector debe ser objeto de un seguimiento atento y ético, y no debe obstaculizar en modo alguno el acceso a la protección o a la seguridad social. Con respecto a la energía, reconoce que en la actualidad se están utilizando soluciones a base de sensores para gestionar el consumo de energía en las viviendas, lo que se tradujo en un ahorro considerable tanto energético como económico. Sin embargo, con dicho sector cada vez más digitalizado, las redes energéticas se amplían y están más expuestas a ciberamenazas. En consecuencia, pide a los Estados miembros y a la Comisión que acompañen la transformación digital en los sectores energéticos con medidas que mejoren la ciberseguridad. Y sobre el transporte, resulta innegable la capacidad de la inteligencia artificial y la robótica para mejorar en gran medida los sistemas de transporte gracias a la introducción de trenes y vehículos de motor autónomos. Sin embargo, la presencia de los vehículos autónomos conllevará riesgos en cuanto a la confidencialidad de los datos y a fallos técnicos, y trasladará la responsabilidad del conductor al fabricante, lo que hará que las compañías de seguros tengan que cambiar la forma en que incorporan el riesgo en la suscripción de sus pólizas. Otro inconveniente identificado en el documento consiste en que la comunicación de voz se utiliza cada vez más en la interacción con los vehículos pero que estas funciones solo

están disponibles en algunas lenguas europeas, por lo que debe garantizarse que todos los europeos puedan utilizar este recurso en su lengua materna. Sobre la agricultura y la cadena alimentaria, observa que la inteligencia artificial tiene el potencial de transformación disruptiva del actual sistema alimentario hacia un modelo para el futuro más diverso, resiliente, adaptado a las regiones y saludable. Asegura que la inteligencia artificial cumple un rol fundamental desde la etapa de producción agrícola hasta el consumo para ayudar a abordar los problemas de seguridad alimentaria, predecir las hambrunas y los brotes de enfermedades de origen alimentario, reducir la pérdida y el despilfarro de alimentos y ayudar a mejorar la gestión sostenible de la tierra, el agua y otros recursos ambientales fundamentales para la salud de los ecosistemas.

La ciberseguridad es un aspecto importante de la materia, por lo que solicita la efectivización de una serie de medidas, entre ellas: la rápida aplicación del Reglamento de Ciberseguridad, la elaboración de un plan de acción para la ciberseguridad por parte de la Agencia de Seguridad de las Redes y de la Información de la Unión Europea (ENISA), reforzar la base industrial en cuanto componente estratégico, inversión en independencia tecnológica, que la Unión desarrolle sus propias infraestructuras, centros de datos, sistemas en la nube y componentes, como procesadores gráficos y chips y que estudie el uso de aplicaciones de ciberseguridad basadas en cadenas de bloques o blockchain para mejorar la resistencia y la fiabilidad a través de modelos de cifrado de datos sin intermediación.

4) Marco jurídico para la inteligencia artificial y la robótica: reitera la solicitud a la Comisión a fin que evalúe periódicamente la legislación actual con el objeto de garantizar que sea adecuada, que respete, al mismo tiempo, los valores fundamentales de Europa y que intente modificar o sustituir las nuevas propuestas cuando así no lo sea. Al mismo tiempo considera que dicha legislación, con participación pública de todos los sectores, debería abordar con cautela una regulación integral de todos los tipos de inteligencia artificial, partiendo de la débil y no solo sistemas de aprendizaje profundo, hasta el nivel en el que tiene sentido para el sector industrial.

Por otra parte, y bajo el subtítulo *“Un mercado interior para la inteligencia artificial”* recuerda que la robótica se utiliza cada vez más en los vehículos

autónomos, como los automóviles autónomos y los drones civiles. Esto ocasionó que algunos Estados miembros adopten leyes en este ámbito concreto, lo que podría dar lugar a un mosaico de legislaciones nacionales obstaculizando, de esta manera, el desarrollo de los vehículos autónomos. Por tanto, insta a la elaboración de un conjunto único de normas que encuentre el equilibrio adecuado entre los intereses de los usuarios, las empresas y otras partes interesadas y los potenciales riesgos para ellos, evitando la sobrerregulación en los sistemas de robótica.

Si bien reconoce la existencia de un número significativo de normas sobre cuestiones como la seguridad, la fiabilidad, la interoperabilidad y la protección, establece como prioridad la promoción y desarrollo de normas comunes en materia de robótica e inteligencia artificial.

En materia de ética y responsabilidad, los consumidores están cubiertos por el marco regulador vigente, a saber, la Directiva de servicios, la Directiva sobre las cualificaciones profesionales y la Directiva sobre comercio electrónico. Resalta que los seres humanos siempre deben ser los responsables, en última instancia, de la toma de decisiones, especialmente en el caso de los servicios profesionales, como las profesiones médicas, jurídicas y de contabilidad. Manifiesta su preocupación en el tema de los datos personales y derecho a la intimidad. En consecuencia, solicita a la Comisión que vele para que todo futuro marco reglamentario garantice la privacidad y la confidencialidad de las comunicaciones, la protección de los datos de carácter personal, incluidos los principios de legalidad, equidad y transparencia, la protección de datos desde el diseño y por defecto, la limitación de la finalidad y de la conservación, la exactitud y la minimización de datos. En materia de responsabilidad, felicita a la Comisión la creación de un grupo de expertos sobre responsabilidad y nuevas tecnologías con el fin de proporcionar conocimientos especializados sobre la aplicabilidad de la Directiva relativa a la responsabilidad por productos defectuosos a los productos tradicionales, las nuevas tecnologías y los nuevos retos sociales. Sin embargo, hasta la fecha no se ha presentado ninguna propuesta legislativa, retrasando la actualización de las normas sobre responsabilidad, poniendo en peligro la seguridad jurídica tanto para los comerciantes como para los consumidores.

5) Aspectos éticos: como en el anterior Reglamento, pide la creación de una carta ética de buenas prácticas para la inteligencia artificial y la robótica que

deben seguir las empresas y expertos. Insiste en que las normas éticas deben garantizar el desarrollo de una inteligencia artificial centrada en el ser humano, la rendición de cuentas y la transparencia de los sistemas algorítmicos de toma de decisiones, así como unas normas claras en materia de responsabilidad y equidad, siempre basándose en el control humano en todo momento. Este marco ético de referencia debe basarse en los principios de beneficencia, no maleficencia, autonomía y justicia, así como en los derechos humanos universalmente reconocidos, por ejemplo, la dignidad humana, la igualdad, la justicia y la equidad, la no discriminación, el consentimiento informado, la vida privada y familiar y la protección de datos, así como en otros principios y valores inherentes al Derecho de la Unión, como la no estigmatización, la transparencia, la autonomía, la responsabilidad individual, y la responsabilidad social, sin olvidar las actuales prácticas y códigos éticos.

6) Gobernanza: considera necesaria una cooperación reforzada entre los Estados miembros y la Comisión para garantizar normas transfronterizas coherentes y un marco de política de datos armonizado que fomenten la colaboración entre las empresas europeas. Además, anima a la creación de una agencia europea reguladora para la inteligencia artificial y la toma de decisiones algorítmica, encargada de establecer una matriz de evaluación de riesgos, investigar el uso de sistemas algorítmicos en caso de sospecha de violación de los derechos humanos, asesorar a otras agencias reguladoras sobre los sistemas algorítmicos, mejorar la eficacia del mecanismo de responsabilidad civil, auditar las evaluaciones algorítmicas de impacto de las aplicaciones del sector privado e investigar los presuntos casos de violaciones de derechos por los sistemas algorítmicos de toma de decisiones, tanto para los casos de decisión individual como para los modelos estadísticos. En materia internacional, reconoce los trabajos en curso de la Organización Internacional de Normalización, la creación del Observatorio de las Políticas de Inteligencia Artificial de la OCDE y le solicita a la Comisión que trabaje para garantizar la máxima coherencia entre los actores internacionales y que defienda los principios éticos europeos a escala mundial (Parlamento Europeo, 2019)

Como lo ha señalado la Unión Europea en todos los documentos precitados, la inteligencia artificial es una tecnología estratégica que se está desplegando y

utilizando a gran ritmo, en todo el mundo. Sin perjuicio de ello, también conlleva nuevos desafíos para el futuro y su aplicación plantea problemas jurídicos y éticos relevantes. Por tanto, es necesario un marco legal y ético para que la inteligencia artificial no actúe fuera de control y que se transforme en un riesgo para los seres humanos. Las decisiones adoptadas mediante algoritmos pueden dar datos incompletos y, por lo tanto, no fiables, que pueden ser manipulados por ciberataques o pueden estar sesgados o simplemente estar equivocados. Aplicar de forma irreflexiva la tecnología a medida que se desarrolla produciría, por tanto, resultados problemáticos, así como la renuencia de los ciudadanos a aceptarla o utilizarla.

Los sistemas inteligentes deben desarrollarse de manera que las personas sean su centro y así se gane la confianza del público. Esto implica que las aplicaciones de inteligencia artificial no solo deben ajustarse a la ley, sino también respetar principios éticos y garantizar que su implementación evite daños involuntarios.

Y con respecto a este tema, es necesario analizar si el sistema jurídico actual de responsabilidad civil puede ser efectiva y eficientemente traspolado al tema de los sistemas robóticos inteligentes. La Unión Europea es clara al respecto, es necesario complementar y dictar un nuevo marco ético y jurídico que responda a la realidad de la inteligencia artificial.

Sin embargo, se debe ofrecer algo más que una simple lista de principios éticos. Es necesario una orientación acerca de la forma de poner en práctica esos principios en los sistema socio-técnicos. Las orientaciones a tener en cuenta son el respeto a la autonomía humana, la prevención del daño, la equidad y la explicabilidad. Explorar y afrontar las tensiones que pueden surgir entre estos principios y especial atención a las situaciones que afecten a los grupos más vulnerables, como los niños, las personas con discapacidad y otras que se hayan visto históricamente desfavorecidas o que se encuentren en riesgo de exclusión.

Así como las situaciones caracterizadas por asimetrías de poder o de información como las que pueden producirse entre empresarios y trabajadores o entre empresas y consumidores.

IV. Conclusiones

A lo largo del presente artículo se ha conceptualizado la inteligencia artificial y la robótica, así como sus aplicaciones. La humanidad está cursando la Cuarta Revolución Industrial y como todo cambio genera desconcierto y se elevan algunas voces apocalípticas.

Elon Musk, el empresario fundador de pay pal considera a la inteligencia artificial como una amenaza para la humanidad y que por eso es que puede resultar tan peligrosa como las ojivas nucleares. En la misma línea, el filósofo británico **Nick Bostrom**, de la Universidad de Oxford, que compara nuestro destino con el de los caballos, cuando fueron sustituidos por los automóviles y los tractores, vaticinando un holocausto nuclear y que las máquinas reemplazaran a las personas naturales. En cambio, Jack Ma, dueño de Alibaba y el segundo hombre más rico de China, es bastante optimista sobre los efectos que puede tener la inteligencia artificial sobre los humanos por lo que no hay nada que temer (BBC News Mundo, 2019)

Muchas veces, el inconsciente colectivo visualiza la inteligencia artificial como robots humanoides, los identifica con máquinas con forma humana y hasta con sentimientos, fomentado por la literatura y las películas de ciencia ficción. Por ejemplo, El Hombre Bicentenario, Yo, Robot, de Isaac Asimov y más recientemente Her. Sin embargo, hasta la fecha los científicos no pudieron crear una máquina que sienta como el ser humano. Conforme la filosofía antropológica, el hombre es multidimensional, posee intelecto, pero también posee una esfera emotiva, volitiva y socio-política. Esta nueva tecnología, hasta el presente no es más que algoritmos desarrollados que le permiten a un programa informático analizar y procesar una enorme cantidad de datos, tomar decisiones automatizadas y hasta predecir determinados acontecimientos si se producen determinadas variables. Pero no puede ir más allá de lo programado.

Sin embargo, como la velocidad en la innovación tecnológica crece de manera exponencial, los profesores Erik Brynjolfsson y Andrew McAfee, del *Massachusetts Institute of Technology* (MIT), hicieron famosa la referencia a “la segunda era de las máquinas”, al afirmar que el mundo está en un punto

de inflexión en que el efecto de estas tecnologías digitales se manifestará con una fuerza sin precedentes a través de la automatización y la creación de cosas (Brynjolfsson y Andrew McAfee, 2014)

No se trata sólo de una nueva era o nueva revolución que traerá consigo una nueva tecnología, sino de algo mucho más grande: la concentración de revoluciones a un ritmo tan rápido que literalmente los humanos no serán capaces de asimilar: la singularidad tecnológica. Dicho concepto no es nuevo, el matemático John von Neumann se lo planteó en los años 50, pero recién en 1983 es popularizado por el matemático y escritor Vernor Vinge

La singularidad tecnológica es una hipótesis en el campo de la prospectiva que sugiere que la velocidad tan acelerada a la que progresa la tecnología provocará que la inteligencia artificial tarde o temprano exceda la capacidad intelectual de los humanos y por ende el control sobre ella. Esto cambiará para siempre la civilización o terminará con ella. Se refiere a la posibilidad de crear mentes artificiales que puedan evolucionar por sí mismas hasta alcanzar lo que se ha venido en denominar superinteligencia, una mente superior a la humana e incomprensible para ella. Consta de tres fases: fase I: las computadoras alcanzarán un nivel de computación similar al cerebro humano, la II los robots inteligentes tendrán la capacidad para mejorarse a sí mismos, iniciando un proceso de mejora exponencial y la III dicho crecimiento exponencial será tan rápido que la inteligencia artificial alcanzada sería imposible de entender por el ser humano.

Stephen Hawking se enrola dentro de los científicos que afirman dicha posibilidad ya que consideró que los seres humanos están limitados por una evolución biológica lenta, por lo que no podrían competir con la inteligencia artificial y serían superados por ella.

Dicho concepto se relaciona con la nueva generación de computación: la computación cuántica que se encuentra en fase experimental en los laboratorios de Google. El 23 de octubre de 2019, el gigante tecnológico anunció que sus investigadores habían logrado realizar un cálculo que las supercomputadoras más grandes no podrían completar en menos de 10.000 años. Y lo hicieron en 3 minutos con 20 segundos logrando un hito en el que los científicos habían estado trabajando desde los años ochenta: la supremacía cuántica que permitiría

que nuevos tipos de computadoras puedan hacer cálculos a velocidades que son inconcebibles con la tecnología actual (Yaffe-Bellany, 2019).

Sin embargo, todavía no deja de ser una hipótesis y el verdadero riesgo en la actualidad se presenta frente a la utilización y manipulación de la tecnología que violenten los derechos humanos y los principios del Estado de Derecho. El desarrollo de tecnología puntera suele estar reservado a países desarrollados que realizan grandes inversiones en el área tecnológica y que además son los primeros en beneficiarse económica y militarmente. El intelectual israelí Yuval Noah Harari afirma que podría derivar en el “colonialismo de datos”, entendiendo por tal una nueva y desigual manera de interactuar entre Estados en la que las empresas recolectarían datos de países con legislaciones de privacidad menos desarrolladas para procesarlos en los países donde se dispone de los mayores adelantos en inteligencia artificial y aplicar ahí los beneficios (Harari, 2019)

La inteligencia artificial y la robótica, como toda tecnología, no es buena ni mala en sí misma, sino la utilización que los seres humanos hagan de ella. Se puede utilizar para manipular y distorsionar la información, para ciberataques, vigilancia electrónica, discriminación racial o de género. Pero presenta un gran potencial en beneficio de la humanidad.

En consecuencia, según la finalidad del humano al programar los algoritmos se puede obtener resultados para el bien. Por ejemplo el ocio, análisis de imágenes de alta resolución procedentes de satélites, drones o escáneres médicos, aumentar la productividad agrícola, ayudar a los médicos a identificar enfermedades, a los discapacitados, crear ciudades inteligentes, predecir catástrofes naturales, ayudar contra el desperdicio de los alimentos, registrar el cambio climático, colaborar con los trabajadores en los puestos de trabajo realizando tareas monótonas, repetitivas o altamente riesgosas, crear nuevos puestos de trabajos calificados, favorecer la educación inclusiva a través de la metodología educación a distancia y avanzar en los proyectos de investigación para poder llegar a lugares vedados al ser humano, como a las profundidades del océano, al sol o a Marte. Por tal motivo es que el derecho debe dictar normas claras y éticas en cuanto a la utilización de la inteligencia artificial. Y como trasciende las fronteras de los Estados, es el derecho internacional el que debe sentar las bases jurídicas y éticas que sirvan de guía para todos los Estados. A eso apunta la Estrategia Europea para la Inteligencia

Artificial centrada en el ser humano y en Latinoamérica debemos comenzar a pensar y repensar la normativa para adaptarla a los cambios tecnológicos futuros.

Como se analizó en el presente artículo, tanto para la Unión Europea como para la comunidad internacional en general, la inteligencia artificial y la robótica ocupan un papel protagónico en la agenda mundial ya que es una disciplina transversal que atraviesa todos los aspectos de la vida cotidiana de la sociedad internacional. Y como presenta tantas formas de utilización, algunas de ellas violentan los derechos fundamentales de las personas, se hace imperioso su regulación normativa, las que deben tener una base ética, pero cuidando de no desalentar su desarrollo.

Si bien, a nivel internacional existen diferentes sistemas jurídicos nacionales derivados de diversas concepciones antropológicas producto de las diferentes culturas, idiosincrasia y cosmovisiones, se puede identificar una serie de consideraciones comunes en los modos de realizar esta aproximación a la ética y jurídica de la inteligencia artificial y la robótica, a saber:

1) Diversidad, no discriminación y equidad: debe realizarse por el bien de la humanidad y beneficiar al mayor número de personas, minimizando el riesgo a la exclusión y a cualquier tipo de discriminación. También debe garantizar la accesibilidad para lograr la igualdad de acceso para las personas con discapacidades.

2) Los estándares de seguridad deben ser altísimos. Para ello, es necesario un control ético y finalista de investigación, transparencia y cooperación en su desarrollo.

3) Los investigadores y diseñadores tienen una responsabilidad fundamental: toda la investigación y desarrollo de la inteligencia artificial y la robótica debe estar caracterizada por la transparencia, la reversibilidad y trazabilidad de los procesos. Se debe registrar y documentar tanto las decisiones tomadas como la totalidad del proceso que dio lugar a las decisiones.

4) En general, se considera insuficiente el existente marco jurídico vigente en el ámbito de la responsabilidad contractual para cubrir los daños causados por la nueva generación de robots basados en inteligencia artificial. Por tal motivo, se propone la formulación de reglas relativas a la responsabilidad civil teniendo como base que los seres humanos siempre deben ser los responsables, en última

instancia, de la toma de decisiones, especialmente en el caso de los servicios profesionales

5) Control humano, intervención y supervisión humana, vale decir que en todo momento sean las personas naturales las que tomen la decisión sobre el obrar de los sistemas robóticos o basados en inteligencia artificial.

6) La sanción de un código de conducta ético voluntario compuesto por una serie de normas éticas que deben garantizar el desarrollo de una inteligencia artificial centrada en el ser humano, la rendición de cuentas y la transparencia de los sistemas algorítmicos de toma de decisiones, así como unas normas claras en materia de responsabilidad y equidad, siempre basándose en el control humano.

6) En todo momento se debe respetar los principios de beneficencia, no maleficencia, neutralidad tecnológica, autonomía, justicia, precaución, rendición de cuentas, reversibilidad, así como en los derechos humanos universalmente reconocidos, por ejemplo, la dignidad humana, la igualdad, la justicia y la equidad, la no discriminación, el consentimiento informado, privacidad, la vida privada y familiar y la protección de datos. Resulta necesario aclarar que Oriente, particularmente China y Corea del Norte, no participan de la misma concepción sobre los derechos humanos que Occidente, en especial lo referido a la privacidad.

7) La creación de Comités de Ética de la Investigación cuya función sea la revisión ética independiente de la propia investigación, el que debería tener un carácter multidisciplinar: incluir a hombres y mujeres, estar constituido por miembros con una amplia experiencia y conocimientos. Y la sanción de un Código Deontológico para los Comités de Ética de la Investigación.

8) Registros de los robots inteligentes.

9) La elaboración de un marco que penalice las prácticas de manipulación de la percepción, la elaboración de perfiles que clasifican a las personas, sufriendo discriminación en función de tal clasificación y su estigmatización y los programas de vigilancia emocional.

10) En la medida de lo posible, la sanción de un conjunto único internacional de normas que encuentre el equilibrio adecuado entre los intereses de los usuarios, las empresas y otras partes interesadas evitando la sobrerregulación en los sistemas de robótica e inteligencia artificial. Dicho marco regulador debe ser flexible, permitir la innovación y el libre desarrollo de nuevas tecnologías.

11) Que desde los Estados se desarrollen políticas que fomenten las capacidades de los trabajadores en los nuevos desafíos laborales y en la elaboración de programas de alfabetización digital en los centros escolares.

12) Gestionar el riesgo, que significa que cuanto más grave sea el riesgo potencial, más estrictos deberán ser los sistemas de control y gestión del riesgo.

13) No desarrollo de la robótica o de la inteligencia artificial para realizar armas de destrucción autónomas, no a los denominados “robots asesinos”.

14) Incertidumbre ya que los avances en estos campos son inciertos, en ámbitos y alcances que en ciertos casos son inimaginables. Por ello, las regulaciones y marcos deben someterse a continua revisión y repensarse en el medio plazo cuando otros avances se hayan hecho realidad.

El dilema consiste en determinar la manera regular estas nuevas tecnologías, si el sistema de responsabilidad civil actual es suficiente para solucionar los problemas, los daños y riesgos que pueden ocasionar la utilización de las mismas o es necesario realizar algún tipo de adaptación o directamente elaborar uno nuevo. De los documentos de la Unión Europea se puede colegir que es necesario establecer un marco jurídico propio y específico, así como se pueden adaptar algunas normativas en materia de responsabilidad civil. En este sentido, el Grupo de Expertos en Responsabilidad y Nuevas Tecnologías de la Unión Europea, publicó, en 2019 un documento intitulado *Liability for Artificial Intelligence and other emerging technologies*. Es un documento muy interesante ya que sienta varias reglas en cuanto al factor de atribución de responsabilidad y del cual se pueden elaborar algunas conclusiones en materia de factores de atribución.

En este sentido, la persona que opera este tipo de tecnología que, sin embargo, conlleva un mayor riesgo de daño a otros, por ejemplo, robots inteligentes en espacios públicos, debe estar sujeta a una responsabilidad estricta por los daños resultantes de su funcionamiento. Pero en el caso que la tecnología no represente un mayor riesgo de daño para los demás se le aplicará la responsabilidad subjetiva esto, se responsabilizará en la medida en que no cumpla con los deberes de seleccionar, operar, monitorear y mantener adecuadamente la tecnología en uso. En cambio, si la tecnología tiene un cierto grado de autonomía, a su operador se le podría aplicar la responsabilidad directa del principal por el hecho del dependiente.

En cuanto a los productores, parece adecuado aplicar el sistema de responsabilidad objetiva por los defectos que aparecieran en los productos, inclusive luego de que éstos hayan sido puestos en el mercado, más aún cuando tuvieran el control sobre las actualizaciones y las mejoras del dispositivo. Ello facilita la carga probatoria, dado que los mismos se encuentran en una mejor posición para afrontar costos y dificultades técnicas que pudieran interferir en el proceso. Para evitar paralizar la tecnología, en caso de mal uso del consumidor, dicha circunstancia, liberaría al productor. Se aconseja establecer que los fabricantes contraten un seguro de responsabilidad civil obligatorio para cubrir los riesgos por los daños ocasionados a las víctimas de dichas tecnologías. Cuando una tecnología en particular aumenta las dificultades para probar la existencia de un elemento de responsabilidad más allá de lo que se puede esperar razonablemente, las víctimas deben tener derecho a la facilitación de la prueba.

Se debe establecer un registro obligatorio de dichas tecnologías y la falta de esta obligación o de proporcionar un acceso razonable a los datos registrados, ocasionará la inversión de la carga de la prueba para no perjudicar a la víctima. En el caso de los robots inteligentes con cierta autonomía, no se los puede responsabilizar por los daños que pueden causar, sino que la responsabilidad debe ser atribuible a personas humanas o jurídicas titulares del software o de la máquina que actuó y produjo el daño. En caso que la titularidad sea colectiva, la responsabilidad sería solidaria.

Otro aspecto a considerar es si este análisis jurídico debe enmarcarse dentro de una nueva rama jurídica autónoma, el Derecho de los Robots (*Robot Law*) para dar respuesta a estos insólitos desafíos y situaciones disruptivas, tal como sucedió con el Derecho del Trabajo o el Derecho Ambiental. En esta línea de pensamiento se enrola Barrio Andrés (2018) que propone un Derecho de los Robots que a la vez adapte el Derecho general y, en la medida de lo necesario, genere un Derecho Nuevo. Las cuestiones de responsabilidad adquieren un papel capital, pues los paradigmas tradicionales no resultan adecuados ante sistemas que modifican su propio código, pueden conducir a soluciones no previstas por ninguna mente humana y son asimismo objeto de crecientes ciberataques.

Del mismo modo surgen tres nuevos riesgos, a saber: el riesgo de autonomía, que tiene su origen en “decisiones” autónomas tomadas por los agentes de

inteligencia artificial, el de asociación, que se debe a la estrecha cooperación entre las personas y los sistemas más avanzados y finalmente el riesgo de red que se produce cuando los sistemas funcionan en estrecha integración con otros sistemas en la nube.

Al respecto, el Dr. Gerónimo Brenna disiente ya que considera que el derecho informático debería incorporar al derecho de los robots como una de sus ramas y que lo que se debe reconocer como una rama autónoma del Derecho es el Derecho Informático como una de las ramas del derecho

Ya sea que se lo considere como una sub rama del Derecho Informático como una nueva rama del derecho autónoma, el marco jurídico que se le otorgue a los robots requerirá una actualización permanente.

En todo caso, el debate recién está comenzando y va tomando diferentes aristas conforme evoluciona la tecnología. El hecho que la inteligencia artificial se encuentre todavía en una fase incipiente ha propiciado algunos experimentos fallidos.

Uno de los ejemplos más relevantes fue lo que sucedió con Tay, un programa de inteligencia artificial diseñado por Microsoft para entablar conversaciones como si fuera una joven de entre 18 y 24 años. El objetivo era que Tay aprendiera de la interacción con los usuarios, pero recibió un ataque coordinado de muchísimos usuarios que empezaron a entrenarla con expresiones que eran ofensivas e intolerantes y que luego repetía en otras conversaciones. Así es que, en cuestión de minutos, publicó en Twitter frases como “odio a las feministas, deberían morir todas y pudrirse en el infierno” o “Hitler habría hecho un trabajo mejor que el mono (Barack Obama) que tenemos ahora”, lo que provocó que Microsoft cerrara la cuenta de la tuitera Tay a las 16 horas de abrirla. Así reflexiona Carme Torras, experta en roboética:

Muchas de las nuevas máquinas autónomas no dejan de ser electrodomésticos más versátiles que nos liberan tiempo de trabajo para dedicarlo a otras cosas. No dejan de ser máquinas. Quienes debemos ser éticos somos los humanos. (Torras, citada por Ortí, 2019)

Bibliografía

- Barrio Andrés, M. (dir.) (2018) *Derecho de los Robots*. Madrid: Wolters Kluwer, 2018.
- BBC News Mundo (3, septiembre de 2019) “Los multimillonarios Jack Ma y Elon Musk discrepan sobre cuál es la mayor amenaza para la humanidad” *La Nación*. Recuperado de: <https://www.lanacion.com.ar/tecnologia/los-multimillonarios-jack-ma-elon-musk-discrepan-nid2283854>
- Bostrom, N. (2014) *Superinteligencia: Caminos, peligros, estrategias*. Zaragoza: Tell
- Brynjolfsson, E. y McAfee, A. (2014) *The Second Machine Age. Work, Progress, and Prosperity in a Time of Brilliant Technologies*. Nueva York: W. W. Norton & Company.
- Cáceres Nieto, E. (2010). *Sistema EXPERTIUS: Sistema experto para la ayuda a la decisión judicial en la tradición romano-germánica*. XIV Congreso Iberoamericano de Derecho e Informática. IJJ-UNAM. Conferencia realizada en Monterrey, N.L, México.
- Chomsky, N. (1957). *Estructuras sintácticas*. Madrid: Siglo XX.
- Comisión Económica para América Latina y el Caribe. CEPAL (2018) *Datos, algoritmos y políticas. La redefinición del mundo digital*. Naciones Unidas: Santiago
- Comisión Europea (2016) “Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones: “Digitalización de la industria europea Aprovechar todas las ventajas de un mercado único digital” 19 de abril de 2016. *COM (2016) 180 final*
- Comisión Europea (2018a) “Comunicación de la Comisión al Parlamento Europeo, al Comité Económico y Social Europeo y al Comité de las Regiones: Inteligencia Artificial para Europa”. *COM(2018) 237 final*. 25 de abril de 2018. Bruselas.

- Comisión Europea (2018b) “Comunicación de la Comisión al Parlamento Europeo, al Comité Económico y Social Europeo y al Comité de las Regiones: Plan Coordinado sobre la Inteligencia Artificial”. *COM(2018) 795 final*. 7 de diciembre de 2018. Bruselas.
- Comisión Europea (2019) “Comunicación de la Comisión al Parlamento Europeo, al Comité Económico y Social Europeo y al Comité de las Regiones: Generar confianza en la inteligencia artificial centrada en el ser humano”. *COM(2019) 168 final*. 8 de abril de 2019. Bruselas.
- Comité Económico y Social Europeo (2017) Dictamen del Comité Económico y Social Europeo sobre la “Inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad” *Diario Oficial de la Unión Europea*. C 288. 60º año. 31 de mayo de 2017, pp. C 288/1- C 4288/9.
- Comité Económico y Social Europeo (2018) Dictamen del Comité Económico y Social Europeo sobre «Confianza, privacidad y seguridad de los consumidores y las empresas en el internet de las cosas» *Diario Oficial de la Unión Europea*. C 440. 61º año. 6 de diciembre de 2018, pp. C 440/8- C 440/13.
- Expert Group on Liability and New Technologies (2019) *Liability for Artificial Intelligence and other emerging technologies*. European Union
- García- Prieto Cuesta, J. (2018) “¿Qué es un robot?”, en Barrio Andrés, M. (dir.), *Derecho de los Robots*. Madrid: Wolters Kluwer.
- Gartner (2013) “Says Personal Worlds and the Internet of Everything Are Colliding to Create New Markets”. Recuperado de: <http://www.gartner.com/newsroom/id/2621015>
- Harari, Y. (2019) “Who Will Win the Race for AI?” *Foreign Policy magazine*. Recuperado de: <https://foreignpolicy.com/gt-essay/who-will-win-the-race-for-ai-united-states-china-data/>
- Hintze, A. (13, noviembre de 2016) “Understanding the four types of AI, from reactive robots to selfaware beings” *The Conversation*.

- Recuperado de: <http://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616>
- Holder, Chris *et al.* (2016) “Robotics and law: Key legal and regulatory implications of the robotics age (Part I of II)” *Computer Law & Security Review* 32 pp. 383/402.
- Huawei (2018) “Inteligencia artificial débil vs fuerte: ¿hasta dónde llega una y otra?” *Huawei*. Recuperado de: <https://iahuawei.xataka.com/inteligencia-artificial-debil-vs-fuerte-donde-llega-otra-infografia/>
- International Federation of Robotics (2018) *The impact of Robots on Productivity, Employment and Jobs*. Frankfurt: International Federation of Robotics.
- Jaimovich, D. (19, septiembre de 2018) “Quién es el científico que lidera la revolución creativa de las máquinas” *Infobae*. Sección Tecno. Recuperado de: <https://www.infobae.com/america/tecnologia/2018/09/19/quien-es-el-cientifico-que-lidera-la-revolucion-creativa-de-las-maquinas/>
- Jones *et al* (1998) *Mobile robots: Inspirations to implementation. Second Edition*. A K Peters Ltd
- Kaplan, A. y Haenlein, M. (2018) “Siri, Siri in my Hand, who’s the Fairest in the Land? On Intelligence” *Illustrations and Implications of Artificial Intelligence*. *Business Horizons*, 62, (1), pp. 15-25.
- Katz, Y. (1, noviembre de 2012) “Noam Chomsky on Where Artificial Intelligence Went Wrong” *The Atlantic*. Recuperado de: https://www.theatlantic.com/technology/archive/2012/11/noam-chomsky-on-where-artificial-intelligence-went-wrong/261637/?single_page=true
- Larson, J., *et al* (23, mayo de 2016). “How we analyzed the COMPAS recidivism algorithm” *ProPublica*. Recuperado de: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>
- Mazo, E. (18, agosto de 2018) “Las 100 empresas con mayor valor bursátil” *Expansión*. Recuperado de: <https://www.expansion.com/empresas/2018/08/04/5b65d2ce268e3e405d8b4577.html>

- McCarthy, J. (11, noviembre de 2007). “What Is Artificial Intelligence” *Universidad de Stanford. Sección Basic Questions*. Recuperado de: <http://www-formal.stanford.edu/jmc/whatisai/node1.html>
- McCarthy *et al.* (31, agosto de 1955) “A Proposal for the Dartmouth Summer Research project on Artificial Intelligence”. *Universidad de Stanford* Recuperado de: <https://web.archive.org/web/20080930164306/http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- McKinsey & Company (2018) The Internet of Things: How to capture the value of IoT. Recuperado de: <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/The%20Internet%20of%20Things%20How%20to%20capture%20the%20value%20of%20IoT/How-to-capture-the-value-of-IoT.ashx>
- McKinsey Global Institute (2017) *Un Futuro que Funciona: Automatización, Empleo y Productividad*. McKinsey Global Institute
- Microsoft (2018) *The Future Computed: Artificial Intelligence y and its Role in Society*. Redmond, Washington: Microsoft Corporation.
- Morales Ascencio, B. (1997) “La lingüística en el contexto de la inteligencia artificial” *Forma y Función*, 10, pp. 25-50
- Ortí, A. (16, junio de 2019) ”Manual de ética para robots” Magazine. Recuperado de: <http://www.magazinedigital.com/historias/reportajes/manual-etica-para-robots>
- Palmerini (2017). “Robótica y derecho: sugerencias, confluencias, evoluciones en el marco de una investigación europea” *Revista de Derecho Privado* n° 32, pp. 53-97.
- Parlamento Europeo (2017) “Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL))” Recuperado de: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_ES.pdf
- Parlamento Europeo (2019) “Resolución del Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global europea

- en materia de inteligencia artificial y robótica (2018/2088(INI))” Recuperado de: http://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_ES.pdf
- Paniagua, E. (2019) *Future Trend Forum. Inteligencia Artificial*. Madrid: Fundación Innovación Bankinter
- Perasso, V. (12, octubre de 2016) “Qué es la cuarta revolución industrial (y por qué debería preocuparnos)” BBC Mundo. Recuperado de <http://www.bbc.com/mundo/noticias-37631834>.
- Primer Foro sobre Inteligencia Artificial e Internet de las Cosas en Ciudades Inteligentes y Sostenibles en América Latina. “Declaración de Buenos Aires. Inteligencia Artificial e Internet de las Cosas en Ciudades Inteligentes y Sostenibles en América Latina”. (2018). *ITU Foro* Recuperado de: <https://www.argentina.gob.ar/sites/default/files/buenosaires-declaration-spanish-final.pdf>
- PwC (2018) *Fourth Industrial Revolution for Earth*. PwC. Recuperado de: <https://www.pwc.com/gx/en/sustainability/assets/ai-for-the-earth-jan-2018.pdf>
- Roldán Tudela, J.M. *et al* (2018) *La Inteligencia Artificial aplicada a la Defensa*. España: Instituto Español de Estudios Estratégicos
- Russell, *et al* (28, julio de 2015) “Autonomous Weapons: An Open Letter From AI & Robotics Researchers” *The Future Life Institute*. Recuperado de: <https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1>
- Russell, S. y Norvig, P. (2004) *Inteligencia Artificial. Un Enfoque Moderno*. Segunda Edición. Madrid: Pearson Educación S.A
- Salesforce (2017). *Inteligencia Artificial para CRM*. México: Salesforce Latinoamérica.
- Salesforce (2, julio de 2018) “Machine Learning y Deep Learning: aprende las diferencias.” *Salesforce*. Recuperado de: <https://www.salesforce.com/mx/blog/2018/7/Machine-Learning-y-Deep-Learning-aprende-las-diferencias.html>
- Sánchez-Urán Azaña, M. Y. y Grau Ruiz, M.A. (6, abril de 2018) “El impacto de la Robótica, en especial la Robótica Inclusiva, en el

- trabajo: aspectos jurídico-laborales y fiscales.” *Plataforma Digital Interuniversitaria sobre el Futuro del Trabajo*. OIT. Recuperado de: <https://iniciativaoitinteruniversitariafuturodeltrabajo.com/ver-articulos/item/el-impacto-de-la-robotica-en-especial-la-robotica-inclusiva-en-el-trabajo-aspectos-juridico-laborales-y-fiscales>
- Searle, J. (1980). “Minds, brains, and programs”. *Behavioral and Brain Science*, vol. 3, (3), pp. 417-457.
- Sychev, V. (2018) “La Amenaza de los Robots Asesinos” *El Correo de la UNESCO. Inteligencia Artificial. Promesas y Amenazas*, número 3, pp. 25-28
- Torres, M. (2019) *Derechos y desafíos de la Inteligencia Artificial* Buenos Aires: CyTA
- Turing, A. (1950). *Computing machinery and intelligence*. Oxford University Press on behalf of the Mind Association.
- Valente, L.A. (2019) “La Persona Electrónica” *Revista Anales de la Facultad de Ciencias Jurídicas y Sociales. Universidad Nacional de La Plata. UNLP*, número 49 (16). Recuperado de: <https://revistas.unlp.edu.ar/RevistaAnalesJursoc/article/view/8483/8163>
- Vecchione, M. (2018)” La tecnología para el bien – Un enfoque novedoso” *ITUNews Magazine. Inteligencia Artificial para el Bien en el Mundo*, pp.11-15
- Yaffe-Bellany, D. (24 de octubre de 2019)” *Computación cuántica explicada en unos minutos” The New York Times, sección ciencia y tecnología*. Recuperado de: <https://www.nytimes.com/es/2019/10/24/espanol/ciencia-y-tecnologia/computacion-cuantica-google.html>

Como citar el artículo: Porcelli A, (2020). La Inteligencia Artificial y la Robótica: sus dilemas sociales, éticos y jurídicos. *Derecho Global, Estudios sobre Derecho y Justicia*, VI (16) pp 49-105 <https://DOI.org/10.32870/dgedj.v6i16.286>

