UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE POSGRADO EN DERECHO
FACULTAD DE DERECHO
DIVISIÓN DE ESTUDIOS DE POSGRADO

"AGENTES MORALES ARTIFICIALES: CUESTIONES ÉTICAS, IMPLICACIONES JURÍDICAS Y EL ROL DE LOS COMITÉS DE ÉTICA EN INTELIGENCIA ARTIFICIAL"

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN DERECHO

PRESENTA
JUAN DANIEL MACÍAS SIERRA

TUTORA: DRA. MARÍA DE JESÚS MEDINA ARELLANO
INSTITUTO DE INVESTIGACIONES JURÍDICAS (UNAM)

CIUDAD UNIVERSITARIA, CD. MX., AGOSTO, 2024

Agradezco a mi familia por su amor, sacrificio y comprensión en este viaje. los amo.

A Gil Guedes, por todo el apoyo brindado a la distancia y presencialmente, gracias por estar.

A mis amistades, gracias por las conversaciones y experiencias inspiradoras durante todo el camino de esta tesis. Especiales agradecimientos a Mariana Reyes, Jesús Eulises, Jesús García, Araceli Rodríguez, Luis Escorcia, Luis Torres, Francisco Chan, Laura Téllez, Mónica Rivera, Carlos Contreras, Saul Ascencio, Manuel Lugo, Ángel Cabrera, Ángel Hernández.

A mi tutora, Dra. María de Jesús Medina Arellano, por su inconmensurable confianza y guía en este proceso.

Gracias.

TABLA DE CONTENIDO

TABLA DE FIGURAS

TABLA DE ABREVIATURAS

CEO	*CHIEF EXECUTIVE OFFICER* (DIRECTOR GENERAL)
COMPAS	*CORRECTIONAL OFFENDER MANAGEMENT PROFILING FOR ALTERNATIVE SANCTIONS* (PERFILAMIENTO DE GESTIÓN DE DELINCUENTES CORRECCIONALES PARA SANCIONES ALTERNATIVAS)
GPT	*GENERATIVE PRETRAINED TRANSFORMER* (TRANSFORMADOR PREENTRENADO GENERATIVO)
IA	INTELIGENCIA ARTIFICIAL
IBM	*INTERNATIONAL BUSINESS MACHINES* (MÁQUINAS DE NEGOCIOS INTERNACIONALES)
IEEE	*INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS* (INSTITUTO DE INGENIEROS ELÉCTRICOS Y ELECTRÓNICOS)
IIJ	INSTITUTO DE INVESTIGACIONES JURÍDICAS
ISO	*INTERNATIONAL ORGANIZATION FOR STANDARIZATION* (ORGANIZACIÓN INTERNACIONAL DE NORMALIZACIÓN)
OCDE	ORGANIZACIÓN PARA LA COOPERACIÓN Y DESARROLLO ECONÓMICOS
UNAM	UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
UNESCO	ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA EDUCACIÓN, LA CIENCIA Y LA CULTURA

INTRODUCCIÓN

La Inteligencia Artificial (en adelante IA) ha experimentado un crecimiento exponencial en la última década, transformando diversos aspectos de la sociedad y prometiendo revolucionar aún más nuestras vidas. Este rápido avance, sin embargo, conlleva también una serie de cuestiones éticas y desafíos que deben abordarse de manera adecuada para garantizar un futuro ético y responsable.

La presente tiene como objetivo general el concepto de la "Agencia Moral Artificial" en el contexto de la IA y la importancia que tienen los Comités de Ética en la dirección y supervisión del desarrollo de la IA, con especial énfasis en el rol que pudieran desempeñar de cara a la hipotética irrupción de los Agentes Morales Artificiales.

Bajo ese contexto, se plantean cinco objetivos específicos: 1) Identificar a la IA; 2) Explorar el concepto de Agencia Moral; 3) Abordar el concepto de Agencia Moral Artificial, y si es posible atribuirles dicha agencia a los sistemas artificiales autónomos; 4) Referir a las implicaciones sociales, éticas y jurídicas que trae consigo la IA, tanto en el presente como en el futuro, y 5) Señalar a los Comités de Ética en IA como facilitadores en la convivencia entre sistemas artificiales autónomos y seres humanos.

Este trabajo ofrece cinco capítulos, cada uno orientado a ahondar más en el tema asociado a cada uno de los objetivos señalados. El primer capítulo, titulado "*La Inteligencia Artificial*", proporciona una introducción al concepto de IA, fortaleciendo su comprensión con la categorización que de ella se hace en: IA aplicada, IA Fuerte y Superinteligencia. Posteriormente, se abordan los campos vinculados con la IA, como lo son el Aprendizaje Automático, el Aprendizaje

Profundo y la Ciencia de Datos. Finalmente, el primer capítulo explora algunas aplicaciones prácticas, así como los dilemas éticos y jurídicos asociados a ellas, a saber: los vehículos autónomos, las armas autónomas y el Chat GPT.[1]

El segundo capítulo, titulado "*La Agencia Moral*", está orientado a explorar el concepto de Agencia Moral; la intención no es profundizar en todo el espectro de definiciones existentes, sino construir una definición mínima e identificar las condiciones necesarias que ayuden a comprender qué es lo que configura a un Agente Moral.

El tercer capítulo, titulado "*La Agencia Moral Artificial*", busca dos cosas: 1) explorar las ideas más relevantes en torno al debate de la Agencia Moral Artificial, y 2) responder a la pregunta de si es posible adscribir una Agencia Moral Artificial a los sistemas artificiales autónomos. Para lograr esto, el tercer capítulo se divide en cinco apartados, el primero de ellos señala las reflexiones de Luciano Floridi respecto al debate de la Agencia Moral Artificial; el segundo, describe los tipos de agencia que pudieran llegar a desempeñar los sistemas artificiales autónomos; el tercero, muestra las posturas de dos grupos de investigación (los "*Computational Modelers*" y el de "*Computers-in-Society*") en torno al debate de la Agencia Moral Artificial; el cuarto, explora brevemente la Visión Estándar y la Visión Funcionalista de la agencia humana, las cuales son clave para comprender las posturas existentes en torno al concepto de la Agencia Moral Artificial, y el quinto, responde a la pregunta de si es posible adscribir una Agencia Moral Artificial a los sistemas artificiales autónomos.

El cuarto capítulo, titulado "*Desafíos socio-jurídicos presentes y futuros de la Inteligencia Artificial*", está enfocado en exponer algunos de los desafíos actuales de la IA en el campo social y jurídico, mismo que a su vez se compone de cuatro subapartados: 1) Discriminación algorítmica; 2) La privacidad en la era de la Inteligencia Artificial; 3) El trabajo frente a la Inteligencia Artificial, y 4) La

[1] Chat GPT es un chatbot desarrollado por la empresa OpenAI. Tiene varias funciones, entre las que se encuentran responder preguntas, resolver ecuaciones matemáticas, escribir textos, depurar y corregir código, traducir entre idiomas, crear resúmenes de texto, entre otras cosas. KRAMER, Z., "What is Chat GPT and how does it work?", en *Fresered.com*, febrero 2023. [en línea] <https://www.freshered.com/what-is-chat-gpt-and-how-does-it-work/> [fecha de consulta: 06 de septiembre de 2023].

desinformación derivada del uso de la Inteligencia Artificial. De igual manera, el cuarto capitulo expone y analiza el futuro de la IA y los desafíos jurídicos que pudieran aparejarse a ella, y el cual consta de dos subapartados: 1) Transhumanismo y 2) Derechos de los robots.

El quinto capítulo, titulado "*Los Comités de Ética en Inteligencia Artificial y la Agencia Moral Artificial*", analiza la función de los Comités de Ética en IA como una forma de regulación frente al desarrollo y uso de la IA. En un primer momento, dicho capítulo explica cómo la Ética sirve para reflexionar críticamente frente al desarrollo tecnológico, ubicando a la Ética de la IA, dentro de la ética práctica, y se define como aquella que tiene como tarea orientar, mediante valores y principios, el desarrollo de la IA. Este capítulo también analiza a los Comités de Ética en IA, describiendo su composición, núcleo de principios y su rol frente a los Agentes Morales Artificiales. El capítulo concluye con la propuesta de que la UNAM cuente con un Comité de Ética en IA, derivado del Comité de Ética Universitario, a fin de convertirse en un referente a nivel nacional en el tema de Ética de la IA.

El apartado de Conclusiones contiene los puntos más relevantes de cada uno de los cinco capítulos, haciendo énfasis en la idea de que los Comités de Ética en IA son los encargados de que los desarrollos en el campo de la IA estén alineados con valores humanos y el bienestar de la sociedad, y a futuro, como los encargados de evaluar posibles riesgos asociados a la idea de atribuir derechos y responsabilidades a los agentes morales artificiales.

La metodología empleada en esta investigación incluye el método analítico, mediante el cual se distinguirán y definirán conceptos clave como IA, Agencia Moral, Agencia Moral Artificial, Ética, Ética de la IA, Comités de Ética en IA, entre otros; el método deductivo, que permitirá deducir el alcance del concepto de Agencia Moral aplicado a los sistemas con IA; el método inductivo, con el objetivo de generar criterios propios en torno al concepto de Agencia Moral y Agencia Moral Artificial; el método comparativo, que ayudará a comprender las diferencias y semejanzas entre posturas relacionadas con la Agencia Moral y la Agencia Moral Artificial; y finalmente, el método histórico, que facilitará la búsqueda crítica de la realidad y la verdad en torno a las distintas posturas relevantes para este trabajo.

Finalmente, esta tesis busca proporcionar una perspectiva sólida y fundamentada sobre la Agencia Moral Artificial, así como el papel que desempeñan los Comités de Ética en IA en el desarrollo responsable y ético de la IA en el futuro.

"*Cada vez me inclino más a pensar que debería haber algún tipo de supervisión regulatoria, quizá a nivel nacional e internacional, sólo para asegurarnos de que no cometamos alguna tontería. Quiero decir que con la inteligencia artificial estamos invocando al demonio*."

Elon Musk

CAPÍTULO I
LA INTELIGENCIA ARTIFICIAL

La IA es una disciplina científica que se desprende de las ciencias de la computación y que ha experimentado un rápido avance en las últimas décadas, convirtiéndose en una de las áreas más prometedoras y desafiantes en la historia de la humanidad. Su creciente presencia plantea cuestiones éticas y jurídicas que requieren estudios minuciosos y detallados. En ese entendido, y para un análisis más completo de dichas cuestiones, y en particular de la idea asociada a la Agencia Moral Artificial que capítulos más adelante se desarrollará, es importante contar con una comprensión de los distintos conceptos asociados a la parte técnica de la IA.

Así, en un primer momento, este capítulo hace un abordaje del concepto de la IA, fortaleciendo su comprensión con la categorización que de ella se hace en: IA Aplicada, IA Fuerte y Superinteligencia. Estas categorizaciones están enfocadas más en el desarrollo teórico y en algunas de sus preocupaciones éticas, particularmente de la Superinteligencia, en un contexto más amplio.

Posteriormente, se abordan campos vinculados a la IA, como lo son el Aprendizaje Automático y el Aprendizaje Profundo, dos enfoques fundamentales para el desarrollo de sistemas con IA. Con estas secciones se busca proporcionar una base sólida para la comprensión del funcionamiento de la IA. En esa misma tesitura, se presenta a la Ciencia de Datos como un campo complementario encargado de extraer información útil de grandes volúmenes de datos, lo cual es esencial para el entrenamiento y la mejora de los sistemas de IA.

Finalmente, este capítulo explora algunas aplicaciones prácticas de la IA, y sus implicaciones éticas y legales, a saber: los vehículos autónomos, las armas autónomas y el Chat GPT. Todos estos son ejemplos concretos que ilustran como la IA influye en la sociedad actual y cómo su implementación plantea dilemas éticos y desafíos jurídicos.

1. ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL?

Hablar de IA puede ser confuso. La cultura popular, especialmente la ciencia ficción, se ha encargado de transmitir la idea de que este tema es igual a robots que son capaces de actuar, pensar e incluso sentir como lo hacen los seres humanos. En otros casos, la idea de lo qué es la IA podría acotarse a algoritmos[2] que son capaces de identificar nuestras preferencias de consumo o de realizar cálculos matemáticos complejos, hechos que contribuyen a pensar que las computadoras son inteligentes, aunque tales actividades sean de lo más sencillo para los ordenadores.

Hay algo de cierto en esas visiones, sin embargo, parecen estar lejos de ser ideas que puedan contribuir a un entendimiento más fundamentado de lo que es la IA. Aquí toca hacer nuestra primera advertencia: en la actualidad no existe una definición unívoca de lo que debería entenderse por IA. En ese sentido, se explorarán algunas de ellas, con la intención de zanjar el terreno para adoptar un concepto propio sobre la cual gire este trabajo de investigación.

La primera aproximación a la IA refiere a que es el campo dedicado a la construcción de artefactos que son capaces de mostrar, en entornos controlados y bien conocidos, y durante periodos de tiempo sostenidos, comportamientos que son considerados como inteligentes, o próximos a lo que se considera como tener mente.[3]

---

[2] *"La palabra algoritmo viene de Abu Abdulah MIhamad ibn Musa Alkhwarirsmi, matemático persa del siglo IX que escribió el primer libro de algebra: Al-Kitab al Mukhtasar fi Hisab al-jabr wa l-Muqabala. El nombre de álgebra proviene de una palabra del título del libro: al-jabr. Tan pronto los escolásticos y filósofos medievales empezaron a diseminar la obra de Al-Khawarismi la traducción de su nombre por "algoritmo" pronto empezó a describir cualquier método sistemático o automático de cálculo"*. MONASTERIO ASTOBIZA, A., "Ética algorítmica: implicaciones éticas de una sociedad cada vez más gobernada por algoritmos", en *Revista Dilemata,* núm. 24 año 9, 2017, p. 186.

[3] ARKOUDAS, K. y S. Bringsjord, "Philosophical foundations", en FRANKISH, K. y W. M. Ramsey,

Otra definición refiere que la IA es el intento de hacer que las computadoras simulen el tipo de cosas que pueden hacer las mentes humanas y animales, ya sea con fines tecnológicos y/o para mejorar la comprensión teórica que se tiene sobre los fenómenos psicológicos.[4]

Una definición más próxima a los elementos técnicos que comprende la IA, se encuentra en lo dicho por Kaplan y Haenlein, cuando indican que es: "…*a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation*".[5] [la capacidad de un sistema de interpretar correctamente los datos externos, de aprender de ellos y de utilizar esos aprendizajes para lograr objetivos y tareas específicas mediante una adaptación flexible] (Traducción propia).

La IA también se entiende como la capacidad de una computadora o de un robot controlado por computadora para realizar tareas comúnmente asociadas a los seres inteligentes. Es un término utilizado frecuentemente al desarrollo de sistemas dotados de los procesos intelectuales que caracterizan a los seres humanos, por ejemplo, el razonar, descubrir significados, generalizar o aprender de la experiencia pasada. [6]

Stuart Rusell y Peter Norvig[7] identifican una serie de definiciones en torno a la IA, y las clasifican en cuatro categorías: 1) en relación con el pensamiento, 2) en relación con el razonamiento, 3) en relación con el comportamiento humano, y 4) en relación con el comportamiento racional, según corresponda. Se trae a colación la clasificación de los autores referidos con la finalidad de mostrar, de forma muy clara,

---

eds.*, The Cambridge Handbook of Artificial Intelligence,* Cambridge University Press, Reino Unido, 2014, p. 34.

[4] FRANKISH, K. y W. M. Ramsey, *The Cambridge Handbook of Artificial Intelligence, op. cit*., *supra*, nota 3, p. 335.

[5] KAPLAN, A. y M. Haenlein, "Siri, siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence", en *Business Horizons*, núm. 62, 2019, p. 2 *apud* BARTNECK, C., C. Lütge, A. Wagner y S. Welsh, *An Introduction to Ethics in Robotics and AI*, Springer, Suiza, 2021, p. 2.

[6] COPELAND, B. J., "Artificial Intelligence", en *Encyclopedia Britannica*, 11 de agosto de 2020. [en línea] <https://www.britannica.com/technology/artificial-intelligence> [fecha de consulta: 06 de septiembre de 2023].

[7] RUSELL, S. J. y P. Norvig, *Artificial Intelligence, A Modern Approach*, 3ª. edición, Pearson, Estados Unidos de América, p. 2.

los enfoques sobre los cuales puede construirse la definición de la IA.

En relación con el pensamiento	**En relación con el razonamiento**
"The exciting new effort to make computers think . . . machines with minds, in the full and literal sense".[8] [El nuevo y emocionante esfuerzo por hacer que los ordenadores piensen . . . Máquinas con mente, en el sentido pleno y literal] *"[The automation of] activities that we associate with human thinking, problem solving, learning . . ."*[9] [La automatización de actividades que asociamos con el pensamiento humano, la resolución de problemas, el aprendizaje . . .]. (Traducción propia).	*"The study of mental faculties through the use of computational models"*.[10] [El estudio de las facultades mentales mediante el uso de modelos computacionales]. "*The study of the computations that make it possible to perceive, reason, and act"*.[11] [El estudio de las computaciones que hacen posible percibir, razonar y actuar]. (Traducción propia).
En relación con el comportamiento humano	**En relación con el comportamiento racional**
"*The art of creating machines that perform functions that require intelligence when performed by people"*.[12] [El arte de crear máquinas que realizan funciones que requieren inteligencia cuando son realizadas por personas]. (traducción propia). *"The study of how to make computers do things at which, at the moment, people are better"*. [13][El estudio de cómo hacer que los ordenadores hagan cosas en las que, por el	*"Computational Intelligence is the study of the design of intelligent agents"*.[14] [La inteligencia computacional es el estudio de agentes inteligentes]. (traducción propia). *"AI . . .is concerned with intelligent behavior in artifacts".* [15] [la Inteligencia Artificial . . .se ocupa del comportamiento inteligente en los artefactos].

[8] HAUGELAND, J., *Artificial Intelligence: The Very Idea*, MIT Press, Estados Unidos de América, 1985 *apud* RUSELL, S. J. y P. Norvig, *Artificial Intelligence, A Modern Approach, op. cit., supra,* nota 7*,* p. 2.

[9] BELLMAN, R. E., An Introduction to *Artificial Intelligence: Can Computers Think?*, Boy & Fraser Publishing Company, *apud* RUSELL, S J. y P. Norvig, *Artificial Intelligence, A Modern Approach, op. cit., supra,* nota 7*,* p. 2.

[10] CHARNIAK, E, y D. McDermott, *Introduction to Artificial Intelligence*, Addison-Wesley, 1985 *apud* RUSELL, S. J. y P. Norvig, *Artificial Intelligence, A Modern Approach, op. cit., supra,* nota 7*,* p. 2.

[11] WINSTON, P. H., *Artificial Intelligence*, Addison-Wesley, 1992 *apud* RUSELL, S. J. y P. Norvig, *Artificial Intelligence, A Modern Approach, op. cit., supra,* nota 7, p. 2.

[12] KURZWEIL, R., *The Age of Intelligent Machines*, MIT Press, Estados Unidos de América, 1990 *apud* RUSELL, S. J. y P. Norvig, *Artificial Intelligence, A Modern Approach, op. cit., supra,* nota 7*,* p. 2.

[13] RICH, E., y K. Knight, *Artificial Intelligence*, McGraw-Hill, 1991 *apud* RUSELL, S. J. y P. Norvig, *Artificial Intelligence, A Modern Approach, op. cit., supra,* nota 7*,* p. 2.

[14] POOLE, D., A. K. Mackworth y R. Goebel, *Computational Intelligence: A logical approach*, Oxford University Press, 1998 *apud* RUSELL, S. J. y P. Norvig, *Artificial Intelligence, A Modern Approach, op. cit. supra,* nota 7*,* p. 2.

[15] NILSSON, N. J., *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann, 1998 *apud* RUSELL, S. J. y P. Norvig, *Artificial Intelligence, A Modern Approach, op. cit., supra,* nota 7*,* p. 2.

momento, las personas son mejores].	

***Figura 1**. Categorización hecha por Stuart Rusell y Peter Norvig respecto a algunas definiciones de la Inteligencia Artificial.*

Derivado de los conceptos presentados en esta tabla, es posible identificar —y como ya se había adelantado— que la IA puede ser entendida de distintas maneras. Su definición sobre qué debería y que no debería incluir la IA ha cambiado a lo largo del tiempo.[16] Todas, sin embargo, parecen coincidir en que la IA busca replicar a través de computadoras a la inteligencia humana.

En vista de todo lo anterior, la definición propuesta en este trabajo para entender a la IA es la siguiente: la IA es una disciplina científica que se desprende de las ciencias de la computación, y que pretende dotar de autonomía y adaptabilidad a las computadoras, con la finalidad de que realicen tareas de forma eficiente e innovadora y sin la necesidad de supervisión humana. [17]

Aquí es crucial precisar lo siguiente. Cuando se hace referencia a la autonomía, es en el sentido vinculado a las ciencias de la computación. Se habla de autonomía porque la computadora realiza tareas en entornos complejos sin la guía constante de un usuario.[18] Este sentido de la palabra autonomía poco tiene que ver con su definición desde la perspectiva filosófica —la cual se aborda en el segundo capítulo de este trabajo—. La adaptabilidad, por su parte, implica la capacidad de la computadora de mejorar su rendimiento aprendiendo de la experiencia. [19]

Una vez identificada la definición de IA, es momento de señalar tres conceptos de dicha tecnología: IA aplicada, IA fuerte y Superinteligencia. Cabe señalar que esta clasificación atiende a los tipos de IA conforme a su nivel de inteligencia, lo

[16] BARTNECK, C., C. Lütge, A. Wagner y S. Welsh, *An Introduction to Ethics in Robotics and AI*, *op. cit*., *supra*, nota 5, p. 7.

[17] Esta definición fue construida gracias a la parte 1 “Introduction to AI”, del curso *“Elements of AI”,* de la Universidad de Helsinki. UNIVERSITY OF HELSINKI, Elements of AI, Part 1: *Introduction to AI,* 2019. [en línea] <https://www.elementsofai.com/> [fecha de consulta: 06 de septiembre de 2023].

[18] UNIVERSITY OF HELISINKI, Elements of AI, “How should we define AI”, en Part 1: *Introduction to AI, op. cit., supra,* nota 17. [en línea] <https://www.elementsofai.com/> [fecha de consulta: 06 de septiembre de 2023].

[19] *Idem.*

cual posibilita identificar el potencial y las capacidades de la IA. La elección de esta clasificación por sobre otras clasificaciones posibles, por ejemplo, las basadas en la función o arquitectura, se hace con el objetivo de enriquecer al debate en torno a la posibilidad de una Agencia Moral Artificial.

1.1. Inteligencia Artificial Aplicada

También conocida como IA Débil o Estrecha, aunque la nomenclatura de IA Aplicada aparenta ser más precisa con relación a lo que busca describir. La IA Aplicada busca producir sistemas "inteligentes" comercialmente viables. Algunos ejemplos son los sistemas expertos utilizados en el sector médico y en el sector financiero.[20]

La IA aplicada está limitada a una única tarea muy definida. La mayoría de los sistemas actuales de IA caben en esta categoría: *"These systems are developed to handle a single problem, task or issue and are generally not capable of solving other problems, even related ones"*.[21] [Estos sistemas se desarrollan para gestionar un único problema, tarea o cuestión y por lo general no son capaces de resolver otros problemas, ni siquiera los relacionados] (Traducción propia).

Otros ejemplos de IA Aplicada son los vehículos autónomos, la visión por computadora, el reconocimiento por voz y la traducción de idiomas, sólo por mencionar algunos de ellos.[22] Estas aplicaciones tienen un alcance limitado, en tanto que sólo pueden realizar la actividad para la cual fueron diseñadas (de ahí que también se le conozca como IA Débil o Estrecha). La IA Aplicada simula —pero no replica— a la inteligencia humana.[23]

---

[20] COPELAND, B.J., "Artificial Intelligence", en *Encyclopedia Britannica*, *op. cit., supra,* nota 6. [en línea] <https://www.britannica.com/technology/artificial-intelligence> [fecha de consulta: 06 de septiembre de 2023].

[21] BARTNECK, C., C. Lütge, A. Wagner y S. Welsh, *An Introduction to Ethics in Robotics and AI*, Springer, *op. cit.*, *supra*, nota 5, p. 10.

[22] CHAKRABORTY, P., "Applied AI is Neither Weak Nor Narrow: Isn't It Time To Change the AI Nomenclature?", en *Analytics India Magazine,* 2017. [en línea] <https://analyticsindiamag.com/applied-ai-neither-weak-narrow-isnt-time-change-ai-nomenclature/> [Fecha de consulta: 07 de septiembre de 2023].

[23] KAPLAN, J., *Artificial Intelligence, What Everyone Needs to Know*, Oxford University Press, Estados Unidos de América, 2016, p. 68.

1.2. Inteligencia Artificial Fuerte

En contraste con la IA Aplicada, la IA Fuerte[24] —también conocida como IA General— apuesta a construir máquinas que piensen, máquinas cuya capacidad intelectual sea indistinguible de la de un ser humano. Esta pretensión suscitó gran interés en las décadas de 1950 y 1960, sin embargo, fue desechada en vista de las extremas dificultades que supone lograr esto. [25]

John Searle define a la IA Fuerte como:

> *According to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other states. In strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explanations; rather, the programs are themselves the explanations*.[26] [Según la IA Fuerte, una computadora no es una mera herramienta en el estudio de la mente; más bien, la computadora adecuadamente programada es realmente una mente, que con los programas adecuados podría comprender y contar con estados mentales; los programas no son meras herramientas que nos permiten probar explicaciones psicológicas, sino que los programas son en sí mismos las explicaciones] (Traducción propia).

En la actualidad, no hay evidencia de que los científicos dedicados a la IA busquen crear un agente que cumpla con las características de la IA Fuerte. La IA Fuerte, sin embargo, sigue siendo el "*santo grial*", un objetivo que se nutre con las

[24] El término de IA Fuerte fue acuñado por John Searle, en 1986, para referirse a la idea de que una computadora puede ser una mente. Cabe señalar que su postura no es para apoyar la idea de que las máquinas pueden ser inteligentes, sino para refutar la misma, a través de su experimento denominado la habitación china. SEARLE, J., *Minds, Brain and Sciece*, Harvard University Press, Estados Unidos de América, 2003.

[25] COPELAND, B.J., "Artificial Intelligence", en *Encyclopedia Britannica*, *op. cit.*, *supra*, nota 6. [en línea] <https://www.britannica.com/technology/artificial-intelligence> [fecha de consulta: 09 de septiembre de 2023].

[26] SEARLE, J., "Minds, brain, and programs", en *The Behavioral and Brain Sciences*, núm. 3, 1980, p. 417.

especulaciones en torno a la idea de la singularidad. [27]

1.3. Superinteligencia

La Superinteligencia refiere a la idea de un sistema de IA que sobrepasa la inteligencia humana individual y colectiva.[28] Esta idea forma parte también de una postura vinculada a la narrativa distópica en la que la IA adquiere un nivel de inteligencia capaz de suponer un riesgo para la existencia de la civilización humana.[29]

Se trata de una visión que describe un probable — quizá una muy lejana probabilidad, aunque no se descarta que la civilización humana se acerque a ella con la irrupción de la computación cuántica—[30] tipo de IA. Las primeras visiones en torno a la Superinteligencia fueron descritas por el matemático Irving John Good, a través de lo que llamó la "*explosión de la inteligencia*". Esta explosión se refiere a una IA lo suficientemente inteligente como para entender su propio diseño, capaz de rediseñarse a sí misma y de crear un sistema sucesor, más inteligente, que a su vez también podría rediseñarse para ser aún más inteligente. [31]

Nick Bostrom identifica a la Superinteligencia como uno de los *"riesgos*

---

[27] CHAKRABORTY, P., "Applied AI is Neither …", *op. cit., supra,* nota 22*.* [en línea] <https://analyticsindiamag.com/applied-ai-neither-weak-narrow-isnt-time-change-ai-nomenclature/> [Fecha de consulta: 07 de septiembre de 2023].

[28] MAINZER, K., *Artificial Intelligence — When do machines take over?,* Springer, Suiza, 2020, p. 223.

[29] CRAWFORD, K., *Atlas of AI, Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, Reino Unido, 2021, p. 214.

[30] La computación cuántica busca aprovechar las propiedades cuánticas de los qubits — los cuales son sistemas cuánticos de dos niveles, que pueden representar estados intermedios entre el 0 y el 1, lo cual se conoce como superposición cuántica— con la finalidad de correr algoritmos cuánticos que utilizan la superposición y el entrelazamiento para ofrecer una capacidad de procesamiento mucho mayor que las computadoras clásicas. ALLENDE LÓPEZ, M., "¿Cómo funciona la computación cuántica?", en *Abierto al público, Banco Interamericano de Desarrollo*, 31 de mayo de 2019. [en línea] <https://blogs.iadb.org/conocimiento-abierto/es/como-funciona-la-computacion-cuantica/> [fecha de consulta: 08 de septiembre de 2023]. La combinación de la computación cuántica y la IA promete conseguir nuevos niveles de complejidad para el aprendizaje automático. KNIGHT, W., "La inminente y poderosa simbiosis de los ordenadores cuánticos y la IA", en *MIT Technology Review*, 04 de abril de 2019. [en línea]<https://www.technologyreview.es/s/11028/la-inminente-y-poderosa-simbiosis-de-los-ordenadores-cuanticos-y-la-ia> [fecha de consulta: 08 de septiembre de 2023].

[31] BOSTROM, N. y E. Yudkowsky, "The ethics of artificial intelligence", en Keith F. y M.R. William eds*., The Cambridge Handbook of Artificial Intelligence, op. cit., supra*, nota 3, p. 328.

existenciales", en el que un resultado adverso aniquilaría la vida inteligente originaría del planeta tierra o reduciría su potencial de forma permanente y drástica. De cara a esta visión fatalista, también identifica una positiva, en la que la Superinteligencia contribuye a la preservación de la vida inteligente originaria de la tierra y ayuda a desarrollar su potencial.[32] Bostrom también identifica como algo necesario que la IA aprenda los valores humanos con la intención de formar una IA amistosa. De acuerdo con este autor, los valores de la Superinteligencia tendrían que contribuir a la felicidad y plenitud de la humanidad, e incluso a tomar decisiones difíciles que beneficien a toda la comunidad sobre el individuo.[33]

Algunos ejemplos de Superinteligencia son los siguientes:[34]

- **Superinteligencia rápida:** un tipo de IA que puede hacer todo lo que hacen las personas humanas, pero mucho más rápido que ellas.
- **Superinteligencia colectiva:** un tipo de IA compuesto por una gran cantidad de subsistemas de IA, los cuales por sí solos son inferiores a las capacidades las personas humanas, pero que en conjunto son superior a ellas.
- **Nueva Superinteligencia:** un tipo de IA que cuenta con capacidades intelectuales que no tiene ninguna persona humana.

Pese a que la idea de la Superinteligencia resulta en la actualidad improbable, y que se aproxima más al campo de la ciencia ficción, existen análisis dentro de la academia —en donde Nick Bostrom es quizá el teórico más reconocido sobre el tema— que enfatizan la necesidad de poner sobre la mesa posibles soluciones frente a los dilemas éticos que pudiera traer consigo este tipo de aplicación de la IA.

---

[32] *Ibidem*, p. 329.

[33] FRANA, P. L., y J.K. Michael, eds., Voz "Nick Bostrom", en *Encyclopedia of Artificial Intelligence, The Past, Present and Future of AI,* ABC-CLIO, Estados Unidos de América, 2021, p. 54. Cabe señalar que en el 2015 Nick Bostrom se unió a Stephen Hawking, Elon Musk, Max Tegmark, entre otros investigadores de IA para suscribir "An Open Letter on Artificial Intelligence", en la cual solicitan que la investigación en IA esté encaminada a maximizar los beneficios para la humanidad y a reducir los potenciales fallos de dicha tecnología. *Idem.*

[34] MAINZER, K., *Artificial Intelligence — When do machines take over?, op. cit., supra,* nota 28*,* pp. 220-221.

2. TEMAS RELACIONADOS CON LA INTELIGENCIA ARTIFICIAL

La siguiente taxonomía sirve para identificar los subcampos que componen a la IA y el campo interdisciplinario con el cual tiene relación. [35]

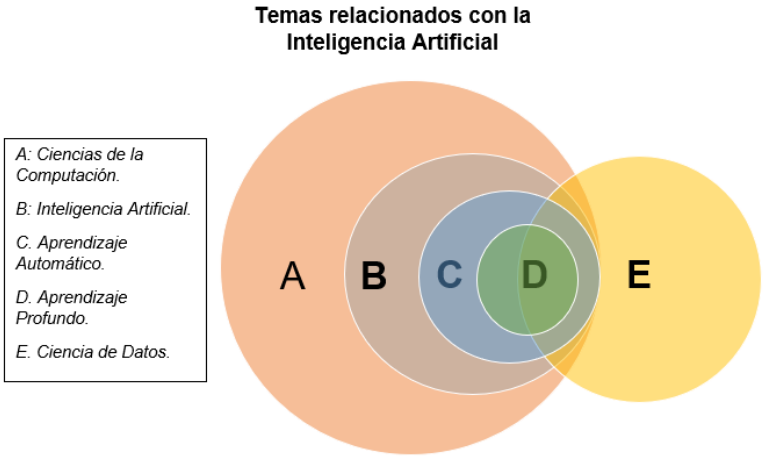

***Figura 2**. Taxonomía de los temas relacionados con la Inteligencia Artificial. Elaboración propia.*

Conocer los subcampos del Aprendizaje Automático y el Aprendizaje Profundo, así como el campo interdisciplinario de la Ciencia de Datos es importante si lo que se pretende es ampliar el abanico de recursos técnicos que permitan una mejor reflexión en torno al alcance de la IA. En los siguientes subapartados se describe en que consiste cada uno de dichos temas.

2.1. El aprendizaje automático

El Aprendizaje automático es un subcampo de la IA, el cual está orientado a facilitar que las computadoras desarrollen la habilidad de detectar patrones a partir de los datos que les sean proporcionados. La intención es que, con los patrones

[35] Esta taxonomía tiene como base la información del primer capítulo "What is AI?", segundo módulo "Related Fields", del curso *"Elements of AI",* de la Universidad de Helsinki. University of Helsinki, "What is AI?, módulo II: Related Fiels", en *Elements of AI, Part 1: Introduction to AI*, 2019. [en línea] <https://www.elementsofai.com/> [fecha de consulta: 09 de septiembre de 2023].

aprendidos, las computadoras puedan ejecutar decisiones de forma autónoma,[36] esto es posible gracias a los algoritmos auto adaptativos que se emplean.[37]

El aprendizaje automático es tan antiguo como la propia IA,[38] sin embargo experimenta un auge en razón de dos factores: la mejora en el procesamiento de las computadoras, lo cual facilita el entrenamiento y la operación de redes neuronales artificiales, y la inmensa cantidad de datos disponibles con las que se cuenta para entrenarlas.

En este punto, es importante explicar lo qué son las redes neuronales artificiales. Por ellas se entienden a las unidades de procesamiento que buscan simular a las neuronas biológicas.[39] Las redes neuronales tienen como finalidad crear IA a partir de máquinas cuya arquitectura simule los cálculos del sistema nervioso humano, algo que no es sencillo porque la potencia de cálculo de la computadora más rápida en la actualidad es una fracción minúscula de la potencia de cálculo del cerebro humano.[40]

La unidad de procesamiento —la unidad de procesamiento puede ser entendida como una versión abstracta de una neurona— contiene el algoritmo auto adaptativo, el cual ha sido entrenado con ejemplos de datos para aprender, por ejemplo, a distinguir que imágenes corresponden a gatos y que imágenes a perros. Para su funcionamiento, requiere de una entrada —una imagen de un gatito, por ejemplo— y una de una salida —la cual corresponde a la etiqueta que ha de proporcionar a la imagen dada, si es un gato, la etiqueta indicará que se trata de un gato—. En este ejemplo, el cual sólo contiene una capa (dado que sólo se tienen una unidad de procesamiento) corresponde al ejemplo de una red poco profunda, en contraste con las redes de aprendizaje profundo que se verán en el siguiente subapartado.

---

[36] AMIR, E., "Reasoning, and decision making", en *The Cambridge Handbook of Artificial Intelligence, op. cit*., *supra*, nota 3, pp. 200 - 201.

[37] KREUTZER, R. y M. Sirrenberg, *Understanding Artificial Intelligence, Fundamentals, Use Cases and Methods for a Corporate AI Journey*, Springer, Suiza, 2020, p. 6.

[38] FRANKLIN, S., "History, motivations, and core themes", en K. Frankish y W. M. Ramsey, eds., *The Cambridge Handbook of Artificial Intelligence, op. cit*., *supra*, nota 3, p. 26.

[39] FRANKISH, K. y W. Ramsey, *The Cambridge Handbook of Artificial Intelligence*, *op. cit.*, *supra*, nota 3, p. 340.

[40] AGGARWAL, C. C., Neural Networks and Deep Learning, A Textbook, Springer, Estados Unidos de América, 2018, p. VII.

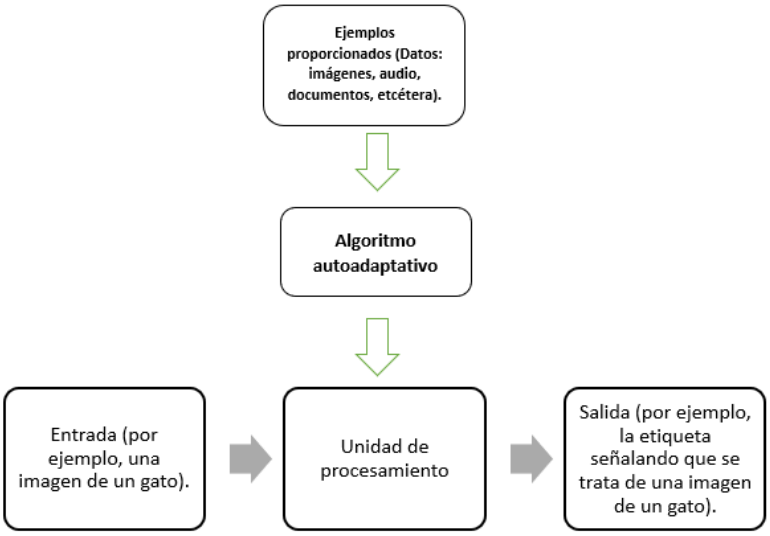

***Figura 3**. Ejemplo de Aprendizaje automático. Elaboración propia.*

En la actualidad, gran parte del Aprendizaje automático ocurre a partir del método supervisado, esto es, que el sistema se instruye utilizando datos de entrenamiento[41] —como es el caso del ejemplo dado anteriormente—. También existen sistemas no supervisados, los cuales son diseñados para desarrollar nuevos conocimientos mediante el descubrimiento de regularidades de los datos, y también el Aprendizaje automático mediante refuerzos, un punto medio entre el primer y segundo método, en donde al sistema se le proponen problemas que debe solucionar, en este caso, el aprendizaje se realiza únicamente con una señal de refuerzo proporcionada como indicador de si se ha resuelto correctamente el problema. [42]

2.2. El Aprendizaje profundo

El Aprendizaje profundo es un subcampo de la IA que tiene como objetivo la

[41] FRANKLIN, S, "History, motivations, and core themes", en K. Frankish y W. M. Ramsey, eds., *The Cambridge Handbook of Artificial Intelligence, op. cit.*, *supra*, nota 3, p. 26.
[42] MORENO, A, *et al., Aprendizaje automático,* Edicions de la Universitat Politecnica de Catalunya, España, 1994, pp. 10 - 11.

creación de grandes modelos de redes neuronales capaces de tomar decisiones basadas en datos. Estos modelos son especialmente valiosos de cara a escenarios en los cuales se cuentan con cantidades masivas de datos que tienen cierta complejidad.[43]

En la actualidad, la mayoría de las grandes compañías que basan su modelo de negocios en la explotación de los datos, emplean los modelos de Aprendizaje profundo. Por ejemplo, Facebook los emplean para analizar texto de las conversaciones suscitadas en su plataforma, Google, Baidu y Microsoft los emplean para el análisis de imágenes y para la traducción de idiomas. Los teléfonos inteligentes cuentan con modelos de aprendizaje profundo ejecutándose en ellos. En materia de reconocimiento de lenguaje, reconocimiento facial, detecciones de imágenes y vehículos autónomos, por señalar algunos ejemplos, el estándar es el Aprendizaje profundo. [44]

La diferencia entre el Aprendizaje automático y el Aprendizaje profundo es que el primero comprende el desarrollo y evaluación de algoritmos que permiten a una computadora aprender funciones a partir de conjuntos de datos que le sirven de ejemplo.[45] En tanto que el Aprendizaje profundo se centra en el desarrollo de grandes modelos de redes neuronales. En una red neuronal profunda cada unidad de procesamiento (o neurona artificial) aprende una función, mientras que la función general definida en la red neuronal se genera a partir de combinar las funciones de las unidades de procesamiento. [46]

Gracias al Aprendizaje profundo es posible procesar amplias cantidades de datos, con resultados más precisos que otros enfoques tradicionales de Aprendizaje automático. Lo "profundo" refiere al número de capas dentro de la red neuronal[47], — como ya se refería anteriormente, una red con una o unas pocas capas se denomina red neuronal poco profunda—.

---

[43] KELLEHER, J. D., *Deep Learning*, MIT Press, Estados Unidos de América, 2019, p. 1.
[44] *Idem.*
[45] *Ibidem*, p. 6.
[46] *Ibidem*, p. 10.
[47] THORNE LUPKER, J. A., Voz "Deep Learning", en P. L. Frana y M. J. Klein, eds., *Encyclpedia of Artificial Intelligence, The past, Present, and Future of AI, op. cit*., *supra*, nota 33, p. 112.

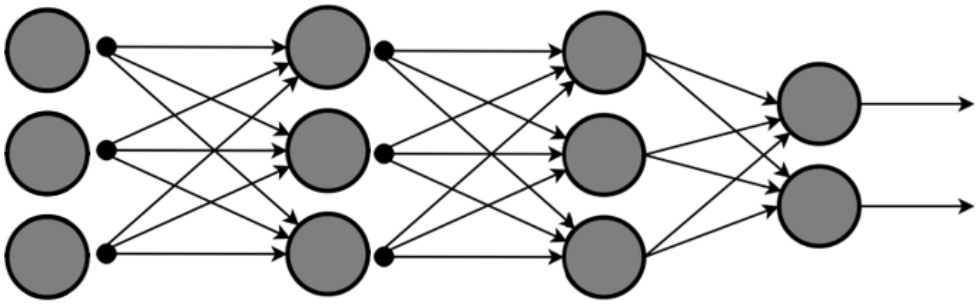

***Figura 4**. Ilustración de una red neuronal artificial con más de una capa, en donde cada una incluye más de una unidad de procesamiento. Entre más capas, más profundidad de la red neuronal. Elaboración propia.*

2.3. La ciencia de datos

En la taxonomía presentada, la Ciencia de Datos se presenta como un campo interdisciplinario, que se relaciona directamente con la IA y sus distintos subcampos. Adicionalmente, también incorpora diferentes elementos y se basa en las técnicas y teorías de distintos campos (por ejemplo: las matemáticas, la estadística, la ingeniería de datos, el reconocimiento de patrones y aprendizaje, la computación avanzada, la visualización, el modelado de la incertidumbre, el almacenamiento de datos y la informática de alto rendimiento) con la intención de extraer el significado de datos y la creación de productos de datos.[48]

La Ciencia de Datos es un área que ha crecido enormemente en los últimos años, esto en virtud de los grandes volúmenes de datos que existe, y los cuales requieren para su explotación de técnicas sofisticadas capaces de tratar con problemas prácticos como la escalabilidad, robustez ante errores, la adaptabilidad con modelos dinámicos, entre otros.[49]

Otros motivos que explican el crecimiento de esta área es el incremento en el poder de cálculo y el desarrollo de métodos más potentes para analizar y modelar

[48] OCHOA CHEHAB, X., W. Fierro Saltos. y J. Cárdenas Benavides, “La Ciencia de los datos”, en ILLESCAS ESPINOZA, W. *et al*, coords., *Las Ciudades Inteligentes*, Ediciones UTMACH, Ecuador, 2018, pp. 23-24.

[49] GARCÍA, J. *et al*., *Ciencia de Datos, técnicas analíticas y aprendizaje estadístico*, editorial Alfaomega, Colombia, 2018, p. 5.

datos, por ejemplo, el Aprendizaje profundo.[50] En el mismo sentido, la relación de la Ciencia de Datos con la IA es que dicha ciencia echa mano de algoritmos aplicados a los problemas relacionados con el manejo de grandes volúmenes de datos, con la finalidad de extraer conocimiento a través de encontrar correlaciones, patrones y predicciones.[51]

Con la Ciencia de Datos se crea valor a partir de los datos con los que se disponen. Gracias a los avances que han permitido el aceleramiento de esta área, ha sido posible que distintas organizaciones obtengan, almacenen y procesen grandes cantidades de datos. Si bien es cierto que esto representa oportunidades de crecimiento y competitividad para las organizaciones que basan su modelo de negocio en la explotación de datos, también lo es que el tratamiento de esos datos ha generado una serie de dilemas éticos que involucran el derecho a la privacidad.[52]

En resumen, la Ciencia de Datos permite extraer conocimiento e ideas a partir del análisis de datos estructurados —una actividad conocida como minería de datos—, para ello se vale de la IA utilizando algoritmos que facilitan la adquisición de conocimiento mediante la identificación de correlaciones, patrones y predicciones. Un ejemplo de la Ciencia de Datos en acción puede ser el de una presentación en la que se muestran oportunidades de inversión en un determinado sector industrial.

3. APLICACIONES PRÁCTICAS DE LA INTELIGENCIA ARTIFICIAL

En este apartado se presentan tres ejemplos de aplicaciones prácticas de la IA: los vehículos autónomos, las armas autónomas y el Chat GPT. Tales aplicaciones son solo una muestra para comprender mejor el impacto que la IA puede tener en la vida humana; además, invitan a la reflexión crítica con relación a los desafíos éticos y jurídicos que se asocian a ellas. En el mismo contexto, además de los ejemplos que se mencionan en los siguientes subapartados, existen muchas otras

[50] KELLEHER, J. D. y B. Tierney, *Data Science*, The MIT Press, Estados Unidos de América, 2018, pp. IX y X.
[51] OCHOA CHEHAB, X., W. Fierro Saltos y J. Cárdenas Benavides, “La Ciencia de los datos”, *op. cit*., *supra*, nota 48, p. 24.
[52] KELLEHER, J. D. y B. Tierney, *Data Science, op. cit*., *supra*, nota 43, p. X.

aplicaciones prácticas de la IA que tienen un impacto significativo en la vida humana y en diversos sectores de la sociedad, por ejemplo, los asistentes virtuales que proporcionan asistencia técnica a usuarios de distintos servicios, los algoritmos de IA que realizan análisis predictivo en el área de finanzas y en el de marketing, así como el uso de algoritmos de IA que realizan análisis de datos médicos con el fin de ayudar a médicos con el diagnóstico y tratamiento de enfermedades.

3.1. Vehículos autónomos

Los vehículos autónomos son capaces de circular por entornos con poco o ningún control humano utilizando un sistema de conducción virtual, el cual es el conjunto de características y capacidades que mejoran o reproducen las actividades de una persona conductora sin necesidad de que esté presente, esto ya en el nivel más alto de autonomía del vehículo.[53]

A decir de la Sociedad de Ingenieros Automotrices, existe una escala de automatización que cuenta con seis niveles. El nivel 0 está reservado para vehículos que dependen completamente de la persona que conduce y el 5 refiere a autonomía en cualquier entorno. [54]

- Nivel 0: El funcionamiento del vehículo depende totalmente del conductor.
- Nivel 1: El automóvil incorpora tecnologías para asistir al conductor, tales como el frenado autónomo de emergencia y la regulación de velocidad.
- Nivel 2: El automóvil controla al menos dos funciones automáticas de forma simultánea, por ejemplo, el sistema autónomo de aceleración y dirección asistidas para ayudar al conductor.
- Nivel 3: El automóvil puede gestionar las funciones críticas de seguridad en ciertas condiciones, sin embargo, el conductor debe asumir el control cuando se produzca una alerta.
- Nivel 4: El automóvil es totalmente autónomo en algunos escenarios de

[53] THOMAS, M., Voz “Driverless Cars and Trucks”, en Philip L. Frana y Michael J. Klein, eds., *Encyclopedia of Artificial Intelligence, The Past, Present and Future of AI*, ABC-CLIO, Estados Unidos de América, 2021, p. 128.

[54] SOCIETY OF AUTOMOTIVE ENGINEERS, *SEA Levels of Driving Automation, Refined for Clarity and International Audience*, 21 de mayo de 2021, [en línea], <https://www.sae.org/blog/sae-j3016-update>, [consulta: 10 de septiembre de 2023].

conducción, sin requerir intervención de un conductor humano.

- Nivel 5: El automóvil es capaz de circular por sí solo en cualquier situación y lugar, sin necesidad de un conductor humano.

Por otra parte, el funcionamiento tecnológico de los vehículos autónomos implica el que estos cuenten con a) sensores, radares y cámaras de video, que sirven para recoger información en tiempo real de los alrededores, calcular velocidad, movimientos, ángulos y reflejos; b) mapas digitales y la posibilidad de que éstos se actualicen en línea, y c) sistemas de coordinación computarizado. [55]

Los vehículos autónomos son entendidos también como vehículos a motor equipado con un sistema de inteligencia artificial que permite conducirlo sin ninguna intervención humana, pudiendo percibir el medio que los rodea y navegar en consecuencia. [56] En la actualidad, una de las áreas de mayor interés en el desarrollo de vehículos autónomos es el de la comunicación entre vehículos y la comunicación entre vehículos e infraestructura. Estas tecnologías permiten que los vehículos autónomos se comuniquen entre sí y su entorno, mejorando la seguridad y la eficiencia del tráfico.[57]

Se dice que la autonomía de los vehículos ayudará a la eficiencia y seguridad de los medios de transporte, reducirá los costos asociados a los accidentes de tránsito, incrementará la eficiencia en el uso de combustible y reducirá las fatalidades y lesiones asociadas a los accidentes provocados por errores humanos.[58]

A pesar de los beneficios anteriormente señalados, la introducción de los vehículos autónomos plantea implicaciones jurídicas y éticas. Por ejemplo, en el caso de un accidente con un vehículo completamente autónomo, ¿quién será el

[55] MARTÍNEZ MERCADAL, J. J., "Vehículos Autónomos y Derecho de Daños", en *Revista de la Facultad de Ciencias Económicas*, Número 20, 2018, p. 63.

[56] BARRIO ANDRÉS, M., "Consideraciones jurídicas acerca del coche autónomo", en *Actualidad Jurídica Uría Menéndez*, no. 52, 2019, p. 102.

[57] WANG, J., Y. Shao, Y. Ge y R. Yu, "A Survey of Vehicle to Everything (V2X) Testing", en *Sensors*, no. 19, 2019, p. 1.

[58] GUERRA ESPINOSA, R. A. y J. Tisné Niemann, "Vehículos autónomos y estado de necesidad: Análisis desde la perspectiva del peatón sujeto a una situación de peligro", en *Revista Chilena de Derecho y Tecnología*, vol. 10, núm. 2, 2021, p. 104.

responsable?: el programador de la IA, el fabricante o el pasajero humano que estaba en el vehículo cuando ocurrió el accidente.

En efecto, son varias las implicaciones éticas correlacionadas con los vehículos autónomos, que si bien no se buscan responder en este análisis al menos si podemos dejar constancia de éstas. Un cuestionamiento importante es si resulta ético permitir que la IA que forma parte del vehículo autónomo debería tomar decisiones en situaciones en las que la integridad física y la vida humana está de por medio, por ejemplo, en el caso de que deba decidir si se impacta en un muro para no atropellar a un peatón que cruzó imprudencialmente la calle, o bien un animal no humano como especie protegida, aún si esto supone lesionar o acabar con la vida de quien o quienes están dentro del vehículo autónomo, que podrían ser personas humanas o incluso animales no humanos pero no protegidos por posibilidad de extinción.

Adicionalmente, algunos efectos que pudiera traer consigo una omnipresencia de vehículos autónomos es el posible hecho de que se reduzca la necesidad de conductores humanos, lo cual traería consigo efectos sobre la economía de quienes se dedican a dicho sector laboral.

Los vehículos autónomos son un ejemplo de aplicación práctica de la IA, siendo quizá en la actualidad la empresa Tesla la más reconocida mundialmente en el dicho ámbito automotriz, lo anterior, sin ignorar a empresas como Waymo, propiedad de la empresa Alphabet, y Cruise, una subsidiara de General Motors.

3.2. Armas autónomas

Las armas autónomas están diseñadas para iniciar activamente y tomar decisiones letales en lugar de actuar como simples sistemas reactivos o defensivos. No son armas convencionales y desafían las nociones fundamentales e interrelacionadas del derecho internacional público.[59]

Las armas autónomas también son conocidas por el concepto de "*robots*

[59] VIVEROS ÁLVAREZ, J., *Sistemas de Armas Autónomas: el dilema de la rendición de cuentas*, Instituto de Investigaciones Jurídicas, Universidad Nacional Autónoma de México, 2021, pp. 21-22.

autónomos letales", los cuales se entienden como los: "*sistemas de armas robóticas que, una vez activados, pueden seleccionar y atacar objetivos sin necesidad de intervención adicional de un operador humano*". [60]

Además, se dice que lo definitorio de un arma letal autónoma es el hecho de que un robot puede hacer elecciones sin intervención humana, y decidir sobre el empleo o no de la fuerza letal: "*Es decir, un sistema de armas puede seleccionar (esto es, buscar, detectar, identificar, rastrear, seleccionar) y atacar (es decir, usar la fuerza contra, neutralizar, dañar o destruir) objetivos sin la intervención humana*".[61]

Otra definición más sobre las armas autónomas, es la siguiente: "*se pueden definir las armas autónomas como aquellas programables, que una vez activadas o lanzadas al entorno, no necesitan de la intervención humana para seleccionar y atacar blancos*".[62] Desde esta concepción, el elemento diferenciador de este tipo de armas respecto a otras es la eliminación, en parte o en su totalidad, del control humano o lo que la comunidad internacional denomina como el Control Humano Significativo, un término que viene a describir el umbral mínimo de control humano que es considerado necesario para asegurar la presencia de la persona humana en el ciclo de la decisión.[63]

Las armas autónomas tienen la habilidad de aprender y/o adaptar su funcionamiento en respuesta a las circunstancias cambiantes del entorno en el que son utilizadas, algo que supone un cambio cualitativo de los paradigmas en la conducción de las hostilidades.[64]

El empleo de armas autónomas implica una serie de desafíos desde el campo ético y jurídico. Por ejemplo, con relación al método de análisis de datos que utilizan los algoritmos, algunas de las preguntas planteadas son: 1) ¿de dónde provienen

---

[60] VIGEVANO, M., "Inteligencia artificial aplicable a los conflictos armados: límites jurídicos y éticos, en *Arbor Ciencia, Pensamiento y Cultura*, no. 197, abril-junio 2021, p. 6.

[61] *Idem.*

[62] CALVO PÉREZ, J. L., "Debate Internacional en torno a los sistemas de armas autónomos letales. *Consideraciones tecnológicas, jurídicas y éticas", en Revista General de Marina*, Vol. 278, no. 4, 2020, p. 459.

[63] *Idem.*

[64] MEZA RIVAS, M., "Los sistemas de armas completamente autónomos: un desafío para la comunidad internacional en el seno de las Naciones Unidas", en *Instituto Español de Estudios Estratégicos*, Gobierno de España, 18 de agosto de 2016, p. 13.

esos datos?; 2) ¿qué sesgos podrían tener? y, 3) ¿cómo esos sesgos podrían afectar la eficacia del modelo?[65]

En efecto, lo anterior se puede transpolar al campo ético en una serie de cuestiones asociadas a la responsabilidad y rendición de cuentas que tomen las armas autónomas; además, abre la puerta a cuestionamientos referentes a la importancia que tiene la empatía y el discernimiento humano tratándose de decisiones que pueden poner fin a una o a varias vidas humanas y no humanas.

Así pues, el tema de las armas autónomas ha ganado tracción en la comunidad internacional, por ejemplo, organizaciones como la Campaña para Detener los Robots Asesinos (*Campaign to Stop Killer Robots*) ha llamado a adoptar un tratado internacional que prohíba el uso de armas autónomas letales,[66] en gran medida por el hecho de que estas armas tomen decisiones de ataque sin ninguna intervención humana significativa.

Como apunte final a este apartado, es necesario que se aborden los desafíos éticos, legales y de responsabilidad asociados con el uso de armas autónomas. En el caso de que se quiera garantizar que su desarrollo y uso estén sintonía con normas y leyes internacionales, será primordial que se establezca un dialogo continuo entre las distintas partes interesadas, entre las que se encuentren los gobiernos, la comunidad técnica, la academia y la sociedad civil, entre otras.

3.3. Chat GPT

Otro ejemplo de una aplicación de la IA es el del Chat GPT (por su sigla, *Generative Pretrained Transformer*), el cual ha ganado notoriedad debido a la claridad y naturalidad con la que el chat responde a las preguntas que se le plantean. El Chat GPT fue desarrollado por la empresa OpenAI, la cual fue fundada en 2015 por un grupo de empresarios y científicos de datos, entre ellos Elon Musk.[67]

---

[65] VIGEVANO, M. R, “Inteligencia artificial aplicable a los conflictos armados…”, *op. cit.*, *supra*, nota 60, p. 7.

[66] STOP KILLER ROBOTS, Less autonomy, more humanity, 2021. [en línea], <https://www.stopkillerrobots.org/>, [consulta: 10 de septiembre de 2023].

[67] AGÜERA MARTÍNEZ, M.B., “¿Qué es ChatGPT y cómo puede revolucionar la Inteligencia Artificial?, en *Lisa News*, 4 de enero de 2023, [en línea], <https://www.lisanews.org/tecnologia/que-es-chatgpt-y-como-puede-revolucionar-la-inteligencia-

La tarea principal del Chat GPT es la de ayudar a sus usuarios a obtener respuestas de la manera más clara y precisa posible, para lo cual utiliza técnicas de IA para analizar el lenguaje y generar texto que sea coherente y relevante con más de 175 mil millones de parámetros, utilizando grandes cantidades de texto previamente escrito por personas humanas en libros, artículos de noticias y conversaciones, por ejemplo. [68] A la fecha, Chat GPT sigue aprendiendo y entrenando en base a las conversaciones que mantiene con quienes interactúan con ella.

El Chat GPT cuenta con una amplia gama de funcionalidades, por ejemplo:

1. Creación de textos variados: permite la generación de textos de diferentes estilos, desde poesía, diálogos, artículos, descripciones, etcétera, con tan solo introducir el texto de entrada que le permita predecir lo que será más útil para el usuario.

2. Chatbot: el Chat GPT al ser uno de los más precisos en sus respuestas, contribuye a mejorar la experiencia del usuario en los sitios web en los que esté se encuentre.

3. Contenidos para post: el Chat GPT puede redactar textos adaptados a las necesidades del usuario, en ese sentido resulta útil para generar contenidos para post en sitios web.

4. Redacción de emails: el Chat GPT puede elaborar correos electrónicos, pero también nos puede resumir los textos contenidos en los correos recibidos.

5. Desarrollo de páginas web: el Chat GPT ayuda con la apariencia de las páginas web, proporcionando el código de programación correspondiente.[69]

Sin embargo, también cuenta con limitaciones, siendo la principal el hecho de que no siempre responde de forma veraz a todas las preguntas, incluso puede crear información falsa, además de que no en todos los casos los textos cuentan con un 100% de coherencia semántica.[70]

---

artificial/>, [consulta: 10 de septiembre de 2023].

[68] *Idem.*

[69] GONZÁLEZ, I., "GPT3: ¿Que 'es y por qué todo el mundo habla de ello"?, en *IEBS Business School*, 14 de diciembre de 2022, [en línea], <https://www.iebschool.com/blog/que-es-chat-gpt3-openai-tecnologia/>, [consulta: 10 de septiembre de 2023].

[70] *Idem.*

En el campo del Derecho, el Chat GPT resulta útil como asistente en la búsqueda sistémica y sistemática de antecedentes de todo tipo, legales y jurisprudenciales, siempre sometidos al ojo del profesional.[71] Sin embargo, también implica problemas éticos, especialmente con la privacidad y la protección de datos, especialmente si el ChatGPT es utilizado para la revisión de documentos que contengan datos personales.

En el campo de la educación, sin embargo, el Chat GPT parece no ser muy buen aliado, de momento: "*Si ahora se usa Chat GPT como sustituto de la universidad sería catastrófico, porque Chat GPT miente con la misma seguridad con la que dice cosas ciertas*".[72]

Puede decirse que el Chat GPT es un ejemplo de cómo la IA ha venido a revolucionar la manera en la que interactuamos y obtenemos información. El hecho de que cuente con capacidad para aprender y mejorar continuamente la hace valiosa en diversos ámbitos, desde la generación de contenidos hasta la asistencia en tareas legales y de investigación.

Finalmente, y más allá de las limitaciones y preocupaciones éticas asociadas a su uso, el Chat GPT representa un avance significativo en la comprensión del lenguaje natural y la generación de texto. En un futuro, es probable que las limitaciones señaladas en párrafos anteriores se reduzcan, lo cual podría apuntalar al Chat GPT en un recurso aún más útil y confiable en una amplia variedad de contextos, permitiendo sinergias más notorias entre la IA y los seres humanos para el abordaje de problemas y desafíos complejos.

---

[71] THE TEHCNOLAWGIST, *GPT-3 en el sector legal: ¿asistente o enemigo?*, 8 de febrero de 2023, [en línea], <https://www.thetechnolawgist.com/2023/02/08/gpt-3-en-el-sector-legal-asistente-o-enemigo/>, [consulta: 10 de septiembre de 2023].

[72] PÉREZ COLOMÉ, J., "'Funciona muy bien, pero no es magia': así es ChatGPT, la nueva inteligencia artificial que supera límites", en *El País*, 07 de diciembre de 2022, [en línea], <https://elpais.com/tecnologia/2022-12-07/funciona-muy-bien-pero-no-es-magia-asi-es-chatgpt-la-nueva-inteligencia-artificial-que-supera-limites.html>, [consulta: 10 de septiembre de 2023].

4. REFLEXIONES FINALES

La IA es una disciplina que busca dotar de autonomía y adaptabilidad a las computadoras para realizar tareas de manera eficiente y sin necesidad de intervención humana. La IA se clasifica en tres categorías, la IA aplicada, la fuerte y la superinteligencia. La IA aplicada se enfoca en sistemas inteligentes comerciales y está limitada a única tarea; la fuerte, por su parte, busca construir máquinas con capacidades intelectuales indistinguibles de las del ser humano. En tanto que la superinteligencia se refiere a la idea de un sistema de IA que sobrepasa la inteligencia humana y que puede ser un riesgo para la existencia de la civilización humana.

En este capítulo también se abordaron los campos del aprendizaje automático, el aprendizaje profundo y la ciencia de datos, como campos interrelacionados que permiten a las computadoras procesar y analizar grandes cantidades de datos con la intención de extraer conocimiento y tomar decisiones autónomas.

La IA se ha convertido en un área fundamental para el desarrollo de tecnologías y sistemas. En este capítulo hicimos un acercamiento a los vehículos autónomos, las armas autónomas y el Chat GPT. Estas aplicaciones son sólo unos ejemplos de un sinfín de posibilidades, riesgos y beneficios, por ejemplo, en el caso de los vehículos autónomos, la mejora en la eficiencia y seguridad en el transporte, y en el caso del Chat GPT, el proporcionar respuestas rápidas y generalmente precisas en diversos campos del conocimiento.

No obstante, estos desarrollos tecnológicos también conllevan riesgos, que detonan cuestionamientos éticos y desafíos legales derivados de su implementación. Pensando en los vehículos autónomos, se plantearon las preguntas sobre la responsabilidad en los accidentes y la toma de decisiones en situaciones de riesgo para la vida humana. Respecto a las armas autónomas, algunas cuestiones que deben abordarse son las relacionadas con la responsabilidad, la rendición de cuentas y la importancia de la empatía y el discernimiento humano en decisiones letales. Por su parte, el Chat GPT también presenta sus limitaciones, particularmente en cuanto a la veracidad de la

información, la creación de información falsa y difamatoria, además de la coherencia semántica.

CAPÍTULO II
LA AGENCIA MORAL

1. LA AGENCIA MORAL Y LAS CONDICIONES NECESARIAS PARA CONTAR CON ELLA

Hablar de Agencia Moral es aludir a un concepto que resulta difícil de abordar. Cabe aclarar que no es la intención de este apartado profundizar en todo el espectro de definiciones existentes alrededor de dicha idea. En todo caso, lo que se pretende aquí son dos cosas: construir una definición mínima e identificar las condiciones necesarias que ayuden a comprender qué es lo que configura a un agente moral.

Eso sí, para lograr lo anterior será necesario referir algunas ideas en torno al tema provenientes de distintas posturas académicas. Con esto —además de mostrar la vaguedad del concepto— se busca proporcionar algunas pistas para entender a la Agencia Moral y responder, en el próximo capítulo, si es posible hablar de una Agencia Moral Artificial que corresponda a los sistemas artificiales autónomos.

1.1. Aproximación al concepto de Agencia Moral

El primer significado que ha de señalarse es el dado por Andrés Richart, quien entiende a la Agencia Moral como la facultad cuyo ejercicio permite la moralidad.[73] Esto último refiere a "*la capacidad de regular la conducta en torno a principios e ideas generales relativos a lo bueno y lo justo, que encontrarían en la ética diversas*

[73] RICHART, A., "El origen evolutivo de la agencia moral y sus implicaciones para la ética", en *Pensamiento*, *Revista de Investigación e Información Filosófica*, vol. 72 núm. 273, 2016, p. 849.

propuestas para fundamentales tales principios e ideas, así como las normas que se derivarían de estos".[74]

Para este autor, la Agencia Moral es la concurrencia de una serie de facultades y disposiciones psicobiológicas, como la prosocialidad —en la que caben emociones como la vergüenza, culpa y simpatía, por ejemplo, así como ciertas facultades, como la empatía— que se han de sumar a otras facultades intelectuales como la memoria, imaginación, razón, pensamiento abstracto, autoconciencia, por ejemplo.[75]

Por su parte, Martha Rodríguez indica que la Agencia Moral requiere que el agente sea responsable de sus acciones en al menos tres sentidos: 1) responsables de una manera que pudieran justificar sus acciones cuando éstas hayan sido intencionales; 2) responsables de aspectos incidentales de esas acciones, de las cuales eran conscientes, y 3) responsables de algunos efectos predecibles de sus acciones.[76] En tanto que la propiedad de ser un agente moral, está asociada con poseer la capacidad de elegir entre varias opciones y poseer la capacidad de llevar a cabo lo que se elige. En efecto, el agente actúa moralmente, pero para que actúe así debe reconocer unas normas comunes en el ámbito donde se encuentre.[77]

Una de las conclusiones arrojadas por la autora es que la Agencia Moral requiere de un entorno social particular: *"[…] un ambiente en el que se desarrolle la capacidad de reflexión del agente, que viva y actúe en tensión, o en medio de conflictos para los que existan diferentes puntos de vistas morales, porque esa disyuntiva constante estimula el ejercicio pleno de la agencia moral*".[78]

Aunado a esto, el agente moral requiere de algunas características relevantes que le ayudan a entenderse como tal, a saber: 1) entenderse como una persona capaz de actuar día a día. Saberse a sí misma y mostrarse a los demás como alguien con una identidad propia, diferente a las que pueda ocupar en los roles sociales que eligió asumir; 2) los agentes morales deben entenderse entre ellos no

---

[74] *Idem.*
[75] *Ibidem*, p. 862.
[76] RODRÍGUEZ CORONEL, M. M., "A vueltas con agencia moral: una perspectiva crítica", en *Quaderns de filosofia i ciència*, núm. 42, 2012, p.129.
[77] *Ibidem,* pp. 127-128.
[78] *Ibidem*, p. 130.

sólo como sujetos, sino como sujetos racionales, capaces de evaluarse como individuos y como actores de ciertos roles definidos y guiados por unos estándares sociales, y 3) el agente moral debe poseer sentido de responsabilidad, no solo como actor de cierto rol, sino como individuo.[79]

En efecto, para Martha Rodríguez, el agente moral "*es un sujeto capaz de elegir cómo actuar, de manera responsable a través de acciones intencionales, asumiendo los aspectos incidentales concienciados y los posibles efectos secundarios de esas acciones*",[80] y que requiere para su desarrollo de "*un ambiente en el que pueda reconocerse a sí mismo y a los demás individuos como sujetos racionales, con capacidad de autocuestionamiento y discriminación al momento de actuar, y de evaluación de los estándares sociales en los que se encuentra inserto*".[81]

Al referirse al agente moral, Jeff Sugarman señala que es en el acto de plantearnos cuestionamientos morales en donde nos volvemos conscientes de que somos agentes morales. Es justo este aspecto el que, según al autor, evita que, como personas, seamos víctimas pasivas de la cultura y la historia:[82]

> *It is in asking moral questions of ourselves that we become aware that we are moral agents, and that we are not condemned only to re-creating cultural scripts. Rather, as moral agents, we are part of the scripting and constitution of our personhood. Our concern for what gives human life meaning and value, and our attempts to express it, can change who we are by deepening our sense of what we believe to be good*.[83] [Es en el plantearnos preguntas morales que nos volvemos conscientes de que somos agentes morales, y que no estamos condenados a sólo recrear guiones culturales. Como agentes morales, somos parte del guion y la constitución de nuestra personalidad. Nuestra preocupación por lo que da sentido y valor a la vida humana, y nuestros intentos de expresarlo,

[79] *Idem.*
[80] *Ibidem,* p. 136.
[81] *Idem.*
[82] SUGARMAN, J., "Persons and Moral Agency" en *Theory & Psychology*, vol. 15, núm. 6, 2005, p. 806.
[83] *Idem.*

pueden cambiar quiénes somos al profundizar nuestro sentido de lo que creemos que es bueno] (Traducción propia).

Otra perspectiva planteada sobre qué debe entenderse por agente moral es la de Vinit Haksar, quien señala que los agentes morales son aquellos que se espera cumplan con las exigencias de la moralidad. Este autor es tajante al señalar que los infantes y los animales no humanos no son agentes morales, pues si bien es cierto que son capaces de actuar, también lo es que no colman dichas exigencias.

De acuerdo con Haksar, estas exigencias pueden satisfacerse de tres maneras. En la primera, será suficiente para considerar a una persona como agente moral, si ésta tiene la capacidad para ajustarse a normas morales, por ejemplo, ser ajustarse a la idea de que "*matar es malo*" o "*robar es malo*". En la segunda —vinculada a una versión kantiana del agente moral— será suficiente que los agentes tengan la capacidad de elevarse por encima de sus sentimientos y pasiones y actuar por el bien de la norma moral. Mientras que en la tercera —la cual es vista por el autor como una posición intermedia— será suficiente que el agente actúe a partir de impulsos altruistas.[84]

Hasta aquí, ya es posible esbozar un primer acercamiento respecto a lo que es la Agencia Moral. Se trata de la facultad con la cuenta un agente moral para cuestionar sus acciones a la luz de principios que tienen que ver, por ejemplo, con lo bueno o lo justo, ya sea que hayan sido ejercidas bajo un rol social establecido o como persona. La anterior definición, sin embargo, sigue sin mostrar qué condiciones de aplicabilidad son necesarias para decir que se cuenta con Agencia Moral .

2. CONDICIONES NECESARIAS PARA LA AGENCIA MORAL

La respuesta a qué condiciones se requieren para decir que se cuenta con una Agencia Moral es compleja. En la actualidad, no basta con señalar que, como seres humanos, contamos con Agencia Moral por el hecho de ser seres racionales y de

[84] HAKSAR, V., "Moral agents", en *Routledge Encyclopedia of Philosophy*, [en línea], <https://www.rep.routledge.com/articles/thematic/moral-agents/v-1/sections/understanding#>, [consulta: 11 de septiembre de 2023].

gobernar nuestros comportamientos con base en principios morales que se revelan a la razón.[85] Tampoco parece ser suficiente indicar —desde una postura antropocéntrica— que sólo quienes cuentan con genes humanos son agentes morales.[86]

A decir de esta postura, Mary Anne Warren duda que los genes humanos sean una condición suficiente para considerar que se cuenta con estatus moral. Para defender su punto, la autora pone como ejemplo el hecho de un óvulo humano fertilizado, el cual cuenta con genes humanos, pero que se distingue de las infancias y personas adultas en el hecho de que no cuenta con sentimientos, como el dolor, la frustración, o el dolor provocado por la muerte. Características que, a decir de la autora, tienen un significado moral.

Ni tampoco parece del todo satisfactorio afirmar que sólo los seres capaces de sentir o que sólo la vida orgánica cuenta con Agencia Moral . Mary Anne Warren señala objeciones para ambas posturas. Por una parte, la teoría de que solo los seres sintientes tienen agencia moral dejaría de lado a plantas y animales no sintientes, ¿por qué no tendríamos obligaciones morales hacia ellos?, se pregunta. En lo que respecta la idea de que la vida orgánica es la que cuenta con agencia moral, la autora crítica que, una postura así nos pondría en serios aprietos, pues sería difícil justificar, por ejemplo, el destruir bacterias que son perjudiciales para nosotros. [87]

El interés por conocer las condiciones necesarias para decir que se cuenta con Agencia Moral se deriva de dos razones. La primera de ellas es que con ello se contribuye a reconocer al agente moral, confiriéndole así la obligación de actuar conforme a principios e ideas en torno a lo bueno y lo justo —como lo refiere Andrés Richart[88]—, pero también respetando de él derechos morales básicos como la vida,

---

[85] Según Immanuel Kant, el respeto es siempre dirigido a las personas y nunca a las cosas. Por persona, Kant se refería a los agentes morales racionales, esto es, seres que son capaces de razonar moralmente y de gobernar sus comportamientos con base en principios morales que se revelan a la razón. Los animales y otras entidades, en efecto, no son agentes morales, y pueden ser tratados como meros fines. WARREN, M. A., "Moral Status", en R.G. Frey y C. Heath Wellman, coords., *A Companion To Applied Ethics*, Blackwell Publishing, Reino Unido, 2002, p. 440.

[86] *Ibidem*, pp. 441 y 442.

[87] *Ibidem,* pp. 442 y 443.

[88] RICHART, A., "El origen evolutivo…", *op. cit*., *supra*, nota 73, p. 849.

la libertad y la integridad física de los demás.[89] La segunda razón está vinculada al espacio de reflexión que desde la ética se abre en torno al deber, la obligación, el bien o la justicia, por ejemplo.[90]

Para intentar proporcionar razones sobre qué condiciones son necesarias, se ha de referir de nueva cuenta a Vinit Haksar. Para dicho autor, la Agencia Moral también tendría que implicar que los agentes morales —además de satisfacer las exigencias de la moralidad— cuenten con: libre albedrío, un entendimiento moral y una comprensión de los hechos que son relevantes, y sentimientos, como puede ser el remordimiento y la preocupación por los demás.[91]

Por su parte, Markus Christen y Mark Alfano refieren a tres componentes estructurales de la Agencia Moral . El primero, refiere al conjunto de competencias con el que debe contar el agente moral, como lo son la condición de autoría de que éste goza, la autonomía, la intencionalidad y previsión con la que actúa, su capacidad de aprender y la autorreflexión, por ejemplo. El segundo, el hecho de contar con un marco de estándares que sirvan como referencia para evaluar los comportamientos del agente moral. El tercero, el contexto en el que se sitúa la Agencia Moral , y que ha de incidir en la toma de decisiones del agente.[92]

Maureen Sie aporta una aproximación distinta a las condiciones que dan pie a la Agencia Moral. Para esta autora, la Agencia Moral requiere al menos de una consciencia deliberativa y un control consciente por parte del agente. A decir de la consciencia deliberativa, la autora la asocia a la deliberación moral que ocurre cuando el agente decide actuar, con base a aspectos como si algo es bueno, valioso o moralmente necesario. En tanto que el control consciente lo entiende como el hacer algo, por las razones que tenemos y de las cuales somos conscientes.[93]

---

[89] WARREN, A. W., “Moral Status”, *op. cit*., *supra*, nota 85, p. 439. La postura de Warren es confrontada por Héctor Morales Zúñiga, pues si bien es cierto que la existencia de un derecho moral puede ser una razón para imponer un deber, es posible que existan razones para imponer un deber sin la necesidad de atribuir un derecho moral. MORALES ZÚÑIGA, H., “Estatus moral y el concepto de persona”, en F. Vergara Ceballos, ed., *Problemas Actuales de la Filosofía Jurídica*, Chile, Librotecnia, 2015, p. 127.

[90] RICHART, A., “El origen evolutivo…”, *op. cit., supra,* nota 73, p. 861.

[91] HAKSAR, V., “Moral agents”, en *Routledge Encyclopedia of Philosophy, op. cit*., *supra*, nota 84.

[92] CHRISTEN, M. y M. Alfano, “Outlining the Field - A Research Program for Empirically Informed Ethics”, en M. Christen *et al.*, ed., *Empirically Informed Ethics: Morality between Facts and Norms*, Springer, Suiza, 2014, p. 11.

[93] En aras de que se comprenda mejor el término “control consciente”, la autora refiere lo siguiente:

A su vez, David Rönnegard identifica tres condiciones incontrovertibles que componen a la Agencia Moral: [94]

- **La capacidad del agente de intentar una acción**: se trata de una condición necesaria para la Agencia Moral , porque es mediante la intención que una acción es atribuible al agente como sujeto, y distinguible de un mero accidente o una secuencia casual de eventos.
- **La capacidad del agente de elegir de forma autónoma una acción intencional**: refiere a la condición de la Agencia Moral , a través de la cual el agente puede elegir una acción intencional, y que va más allá de simplemente tener la capacidad de hacerlo, sino que también abarca el cómo elije tal acción intencional.
- **La capacidad del agente de realizar una acción:** refiere a la condición necesaria para la Agencia Moral a través de la cual la intención del agente se lleva a cabo. La acción puede ser realizada por los movimientos corporales del agente, o instruyendo a que otro agente la realice.

Ahora bien, para el desarrollo de los siguientes apartados, se tomarán como base a dichas condiciones. La justificación para esto es que Rönnegard las emplea para analizar si es posible catalogar a una empresa como agente moral. La empresa, como tal, se trata de una cosa, con un estatus ontológico muy distinto al de los seres vivos, pero que es perceptible como algo que tiene obligaciones morales frente a la sociedad.[95] Por tanto, es viable considerar que la aproximación de dicho autor es propicia para comprender —sin tantas complejidades— qué es lo compone a la Agencia Moral , y también para servir como punto de partida en aras

"*I am in conscious control of helping you get back on your feet, if I believe—and am aware of believing—that I should or want to help you and that is, in fact, the reason why I help you*" [Tengo el control consciente de ayudarte, si creo, y soy consciente de creer, que debería ayudarte, y eso es, de hecho, la razón por la que te ayudo] (traducción propia). SIE, M., "Moral Agency, Conscious Control, and Deliberative Awareness", en *Inquiry: An Interdisciplinary Journal of Philosophy*, vol. 52, núm. 5, 2009, p. 519.

[94] RÖNNEGARD, D., *The Fallacy of Corporate Moral Agency*, vol. 44, Issues in Business Ethics, Springer, Holanda, 2015, pp. 9-15.

[95] *Vid.* RÖNNEGARD, D., *The Fallacy of Corporate Moral Agency*, vol. 44, Issues in Business Ethics, Springer, Holanda, 2015.

de conocer si un sistema artificial autónomo puede ser considerado como un agente moral.

Con estas tres condiciones es posible cerrar la definición que se propuso construir en este apartado. Así pues, ha de entenderse como Agencia Moral a la facultad con la cuenta un agente moral para cuestionar sus acciones a la luz de principios que tienen que ver con lo bueno o lo justo, por ejemplo, ya sea que hayan sido ejercidas bajo un rol social establecido o como individuo, y que requiere de tres condiciones necesarias para que exista: 1) la intencionalidad del agente; 2) la autonomía del agente, y 3) la capacidad de actuar del agente.

2.1. Primera condición: la intencionalidad del agente

Usualmente, la intencionalidad es entendida: "c*omo el 'sobre qué' (aboutness) de la experiencia y el pensamiento, el rasgo de la mente que nos relaciona con el mundo, en una relación que consiste en el hecho de que ciertos estados mentales tienen contenido, tienen un 'sobre qué'*".[96] La intencionalidad, *"involucra la atribución de uno o más estados mentales motivacionales —como un propósito o un deseo— en conjunción con ciertos estados informacionales —entre los que destacan las creencias y las experiencias perceptivas—, que interactúan entre sí para desencadenar una acción"*.[97]

José Ferreter Mora, al definir la intencionalidad, aborda dos sentidos del concepto. Uno lógico, gnoseológico y en parte psicológico, el cual "*designa el hecho de que ningún conocimiento actual es posible si no hay una intención".* [98] La intención se trata "*de un acto del entendimiento dirigido al conocimiento de un objeto*". Y otro ético, en donde la intención es "*una intentio finis procedente del acto de la voluntad, guiado por el entendimiento, el cual investiga los medios que conducen al fin*." [99] El entendimiento, por su parte, juzga; en tanto que la voluntad

[96] JORBA GRAU, M., "La intencionalidad: entre Husserl y la filosofía de la mente contemporánea", en *Investigaciones Fenomenológicas, Revista de la Sociedad Española de Fenomenología,* núm. 8, 2011, p. 77.
[97] DANÓN, L., "Explicaciones intencionales y explicaciones teleológicas de la conducta animal", en *Revista Argentina de Ciencias del Comportamiento*, vol. 3 núm. 1, 2011, p. 54.
[98] FERRATER MORA, J., *Diccionario de Filosofía*, Tomo I, A-K, Argentina, Montecasino, 1964, p. 980.
[99] *Ibidem*, p. 982.

se determina mediante una intención. La intención, desde este sentido, es una acción desde el punto de vista del agente como ser dotado de voluntad.[100]

La intencionalidad es uno de los temas centrales de la filosofía de la mente, y del cual es posible rastrear las primeras reflexiones en torno a él con Aristóteles, y posteriormente retomado por filósofos como Santo Tomás, Duns Escoto y Avicena.[101] Posteriormente, en el siglo XIX, Franz Brentano recogió la significación escolástica, para convertir a la intencionalidad en el concepto central de su psicología.[102]

La intencionalidad en la filosofía de la mente contemporánea puede abordarse desde tres aproximaciones en la filosofía analítica y las ciencias cognitivas.[103] A continuación se da más detalle de cada una de ellas:

- **Desde la filosofía del lenguaje**: Busca clarificar la intencionalidad de la consciencia a través del análisis de las propiedades lógicas de oraciones usadas para describir fenómenos psíquicos. Algunos pensadores que pueden ubicarse es esta aproximación son Gottlob Frege, Bertrand Rusell y Saul Kripke. [104]
- **Desde la naturalización de la intencionalidad**: Esta aproximación busca dar cuenta de cómo puede ser explicada la intencionalidad a través de mecanismos no intencionales. Refiere también que la intencionalidad es un concepto complicado, el cual ha confundido a muchos filósofos a la hora de pensar sobre la mente y sus propiedades.[105] Probablemente sea Daniel Dennett quien con su aproximación de un "*sistema intencional*" describa

[100] *Ibidem*, p. 983.
[101] SÁNCHEZ-PESCADOR, J., "En torno a la intencionalidad", en *Revista de Filosofía*, Universidad Complutense Madrid, vol. 3 núm. 14, 1995, p. 30.
[102] FERRATER MORA, J, *Diccionario de Filosofía*, *op. cit*., *supra*, nota 97, p. 981. diccionario. La intencionalidad para Fran Brentano ante todo una propiedad de la mente de ser "acerca de algo", propiedad que comparten las percepciones los deseos y en general los estados de la mente y no del cerebro. CORRAL CUARTAS, Á., "En busca de la dimensión intencional de las emociones y los estados de ánimo", en *Ideas y Valores*, vol. 66 núm. 3, 2017, [en línea], <http://www.scielo.org.co/pdf/idval/v66s3/0120-0062-idval-66-s3-00047.pdf>, [consulta: 11 de septiembre de 2023].
[103] JORBA GRAU, M., "La intencionalidad: entre Husserl…", *op. cit*., *supra*, nota 96, p. 78.
[104] *Idem.*
[105] *Idem.*

mejor esta postura, pues define el enfoque intencional como una "*estrategia consistente en interpretar el comportamiento de una criatura, o sistema, tratándola como si fuera un agente racional cuyas acciones se ven guiadas por sus creencias y deseos".*[106]

- **Desde el razonamiento práctico**: está aproximación se origina en la filosofía de la mente y en la propuesta de razonamiento práctico de Michael Bratman. Este autor propone la idea de que las intenciones están vinculadas a la planeación, de tal manera que la intención consiste, entre otros aspectos, en un plan adoptado para su ejecución. Esto provee las bases para intentar predecir lo que los demás harán, explicar por qué lo han hecho, y coordinar nuestras acciones con las suyas.[107]

En efecto, es posible afirmar que no existe una única respuesta en torno a lo qué es la intencionalidad. En un sentido amplio, podría identificarse a la intencionalidad como los propósitos, planes e intereses que desencadenan una conducta. Además, la intencionalidad, al tratarse de un estado mental, sólo sería atribuible a humanos y animales no humanos:

> *The concept of what intentions are comes from our introperspective awareness of our own planned actions suggesting that they are primitively mental states. If intentions are mental states then they are only attributable (in a literal sense) to entities that actually possess such mental states. Our attribution of intentions to others humans or animal is based on the belief that they have the mental states that we can identify in ourselves. It is and epistemological leap of faith that beings other than ourselves possess what we experience as intentionally. We infer that other beings behave intentionally not only from their apparently purposive behavior, but also*

[106] DANÓN, L., "Explicaciones intencionales y …", *op. cit*., *supra*, nota 97, p. 55.

[107] CASTRO MANZANO, J. M., *Revisión de Intenciones desde el Aprendizaje BDI,* Tesis de maestro, México, Universidad Veracruzana, Departamento de Inteligencia Artificial, 2008, p. 28. La teoría de Bratman define un marco psicológico de sentido común para entender a los demás y a nosotros mismos, con base en las intenciones. La idea central Bratman es que nuestra concepción de las intenciones, de acuerdo con el sentido común, está ligada al fenómeno de planeación y a los planes, lo cual provee las bases para nuestros intentos cotidianos de predecir lo que los demás harán. *Idem.*

from our belief that their brains experience the mental states found in us.[108] [El concepto de lo que son las intenciones deriva de la introspección consciente de nuestras propias acciones planeadas, sugiriendo que son estados primitivamente mentales. Si las intenciones son estados mentales, entonces sólo son atribuibles (literalmente) a entidades que poseen tales estados mentales. La atribución de intenciones que hacemos a otros humanos o animales tiene como base la creencia de que tienen estados mentales identificables en nosotros mismos. Es un acto de fe epistemológico que otros seres, además de nosotros, posean lo que experimentamos intencionalmente. Inferimos que otros seres se comportan intencionalmente, no sólo por que aparentan seguir un propósito, sino también por nuestra creencia de que sus cerebros experimentan nuestros mismos estados mentales] (Traducción propia).

No obstante, también es posible encontrar posturas que adscriben estados mentales a las máquinas. John McCarthy —el inventor del término "*Inteligencia Artificial*"— por ejemplo, consideraba que una máquina podía tener creencias, intenciones y deseos.[109] De hecho, McCarthy llegó a afirmar frente a John Searle que su termostato tenía creencias, Searle le cuestionó sobre el tipo de creencias de su termostato, a lo cual respondió: "*My thermostat has three beliefs. My thermostat believes It's too hot in here, it's too cold in here, and it's just right in here"*. [110] [Mi termostato tiene tres creencias: mi termostato cree que hace demasiado calor aquí, que hace demasiado frío aquí y que la temperatura es adecuada aquí] (Traducción propia).

Otra postura similar a la de McCarthy, es la de Munidar Singh, quien identifica que las creencias, deseos e intenciones son familiares a diseñadores, analistas de sistemas, programadores y usuarios; la postura intencional que emplea dicho autor contribuye a entender y explicar sistemas complejos; dicha postura también provee

[108] RÖNNEGARD, D., *The Fallacy of Corporate Moral Agency*, *op. cit.*, *supra*, nota 94, p. 24.
[109] *Vid*. MCCARTHY, J., *Ascribing Mental Qualities to Machines,* Computer Science Department, EEUU, Universidad de Standford, 1979.
[110] SEARLE, J., "Mind and brains without programs", en Colin Blakemore y Susan Greenfield, eds., *Mindwaves*, Blackwell, Reino Unido 1987, p. 211 *apud* BLACKMORE, S., *Consciousness, An Introduction*, Hodder Education, Reino Unido, 2010, p. 294.

de ciertas regularidades y patrones de acción que son independientes de la implementación física de los agentes, y un agente puede razonar sobre sí mismo y sobre otros agentes adoptando la postura intencional.[111]

Ahora bien, la postura que se habrá de adoptar en este trabajo en torno a la intencionalidad tiene un fuerte vínculo con la idea de que un agente actúa intencionalmente sólo si es consciente de la intención por la que actúa, pero también del proceso que lo lleva a elegir tal acción.[112]

Esta postura ha de nutrirse también con la reflexión de David Rönnergard, para quien el sentido moralmente relevante de una intención es aquel en el que un agente moral es consciente —lo cual indica que se trata de un estado mental—, y en donde también dicha intención va ligada con la capacidad del agente de elegir sus acciones intencionales de forma autónoma, esto es, que el agente no reaccione a sus deseos o entorno.

A fin de darle soporte a su argumento, Rönnergard cita a Peter Cane: "*Philosophically it is generally agreed that a minimum level of mental and physical capacities is a precondition for culpability. A person should not be blamed if they lacked basic understanding of the nature and significance of their conduct, or basic control over it".* [113] [Filosóficamente, es aceptado que un nivel mínimo de capacidades físicas y mentales es una condición previa para la culpabilidad. No se debe culpar a una persona si no cuenta con una comprensión básica de la naturaleza y significado de su conducta, o si no tiene control sobre ella] (Traducción propia).

En efecto, el agente moral también debe ser capaz de comprender el significado de su conducta, lo cual parecería sugerir que el conocimiento sobre lo que se intenta es una parte necesaria del agente que actúa intencionalmente. Si algo distinto a un sistema con cerebro posee intencionalidad, tendría que mostrar las mismas cualidades mentales con las que cuenta el ser humano.[114]

---

[111] SINGH, M. P., *Multiagent Systems: A theoretical framework for intentions, know how, and communication,* núm. 799, Lecture Notes in Computer Sciences, Springer, Holanda, 1995, p. 3.
[112] RÖNNEGARD, D., *The Fallacy of Corporate Moral Agency*, *op. cit.*, *supra*, nota 93, p. 28.
[113] CANE, P., *Responsability in law and morality*, Oxford, Reino Unido, 2002, p. 65 *apud* RÖNNEGARD, D., *The Fallacy of Corporate Moral Agency*, *op. cit.*, *supra*, nota 94, p. 28.
[114] *Ibidem,* p. 25.

Habida cuenta de que se ha señalado en qué consiste la intencionalidad, es momento de hablar sobre la autonomía.

2.2. Segunda condición: la autonomía del agente

El concepto de autonomía es relevante para distintas disciplinas como la sociología, psicología, biología y ciencias de la computación, por mencionar algunas. En lo que atañe a la filosofía, es un concepto que abarca distintas áreas, como la metafísica, la ética y la filosofía social y política. En lo concerniente a la ética, cuestiones relacionadas a la naturaleza y el valor de la autonomía tienen un impacto importante frente a preguntas vinculadas a la responsabilidad moral y el libre albedrío, por ejemplo. De ahí que existan distintas acepciones en torno al tal concepto.[115] Sólo con la intención de dar más luz sobre lo complejo del concepto, se alude a Gerarld Dworkin, quien sostiene que la autonomía:

> […] *sometimes as an equivalent of liberty (positive or negative in Berlin's terminology), sometimes as equivalent to self-rule or sovereignty, sometimes as identical with freedom of the will. It is equated with dignity, integrity, individuality, independence, responsibility, and self-knowledge. It is identified with qualities of self assertion, with critical reflection, with absence of external causation, with knowledge of one's own interest*.[116] [A veces, como equivalente de la libertad (en el sentido positivo o negativo de acuerdo a la terminología de Berlin), a veces como equivalente de autogobierno o soberanía, a veces de forma similar que la libertad o la voluntad. Se equipará con la dignidad, integridad, individualidad, independencia, responsabilidad y autoconocimiento. Se identifica con cualidades de autoafirmación, reflexión crítica, ausencia de causalidad externa, con conocimiento del propio interés] (Traducción propia).

En este apartado se busca dar cuenta de lo que ha de entenderse como

[115] ĐERIĆ, M., "What is Autonomy Anyway?, en M. Kühler y V. L. MItrovic, eds., *Theories of the Self and Autonomy* in Medical Ethics, vol. 83, The International Library of Bioethics, Springer, Suiza, 2020, p. 17.

[116] DWORKIN, G., *The Theory and Practice of Autonomy,* EEUU, Cambridge University Press, 1988, p. 16 *apud Ibidem,* p. 18.

autonomía. Así pues, se habrá de aludir a algunas definiciones y las condiciones que este concepto debe cumplir —desde este punto ha de señalarse que la postura que busca adoptarse se sitúa desde el constructivismo—,[117] esto a fin de contar con elementos que puedan dar cuenta de una posible Agencia Moral Artificial.

Identificar qué es y qué compone a la autonomía es neurálgico de cara a la idea de Agencia Moral , pues el hecho de atribuirle la condición de agentes morales autónomos permite al agente ser destinatario de evaluaciones morales en las relaciones interpersonales, sin olvidar también que se trata de un presupuesto de las instituciones sociales, tanto políticas como jurídicas.[118] Se dice también que es la noción de autonomía es la que permite aceptar la responsabilidad moral y legal de nuestras acciones y atribuir derechos y responsabilidades a los demás.[119]

La palabra autonomía, proviene de los componentes léxicos *autós* "por sí mismo" y *nómos,* "regla o ley", que en conjunto refieren al "autogobierno". Se trata de una palabra inventada en la Grecia antigua, y que estaba vinculada fuertemente a la política: *"it began as the word for a community´s self-government, a protected degree of freedom in the face of an outside power which was strong enough to infringe it*".[120] [comenzó como la palabra para el autogobierno de una comunidad,

---

[117] Por constructivismo, se entiende como la postura según la cual, en tanto hay verdades normativas, éstas están determinadas por un proceso idealizado de deliberación, acuerdo o elección racional. Representa también una vía para sostener la objetividad práctica y moral, sin asumir la existencia de hechos morales. Se trata de una visión sobre la normatividad y el valor entendidos como no independientes de las posturas de los agentes prácticos y morales. MARQUISIO AGUIRRE, R., "El ideal de la autonomía moral", en *Revista de la Facultad de Derecho*, núm. 74 jul-dic. 2017, Universidad de la República, Uruguay, p. 57. El realismo moral, por su parte, representa una postura que afirma la existencia objetiva de los hechos morales, indicando también que son independientes de la mente, en tanto que no se conforman por nuestros sentimientos, opiniones, convenciones sociales, por la capacidad de síntesis de nuestra mente, ni por nuestra imposición de teorías o lenguajes. DEVITT, M., "Realismo Moral: Una Perspectiva Naturalista", en *ARETÉ Revista de Filosofía,* vol. XVI núm. 2, 2004, p. 191.

[118] Esta idea corresponde a Ricardo Aguirre Marquisio, quien también señala con relación a la autonomía como presupuesto de las instituciones sociales, tanto políticas como jurídicas, que, tratándose de las instituciones políticas, democracia implica que los ciudadanos, precisamente por el hecho de estar dotados de autonomía moral, tienen derecho a decidir cuestiones que afectan las vidas de sí mismos y de otros. En tanto que para las instituciones jurídicas, la idea del imperio del derecho tiene entre sus componentes fundamentales la necesidad de que las normas puedan justificarse para destinatarios considerados como seres racionales y dotados de determinarse moralmente. MARQUISIO AGUIRRE, R., "El ideal de la autonomía moral", *op. cit., supra,* nota 117*,* p.57.

[119] CAMPBELL, L., "Kant, autonomy and bioethics", en *Ethics, Medicine and Public Health*, vol. 3 núm. 3, 2017, p. 3

[120] LANE FOX, R., *The Classical World, An Epic History of Greece and Rome*, Estados Unidos de

un grado de libertad protegido frente a un poder externo que era lo suficientemente fuerte como para infringirlo] (traducción propia). Sería hasta con Immanuel Kant que este concepto también se referiría a la capacidad de los agentes morales de autogobernarse;[121] quien sea capaz de ello, entonces goza de autonomía.[122]

Un primer acercamiento al significado de la palabra autonomía es la definición propuesta por Ferrater Mora, quien refiere con ella al "*hecho de que una realidad esté regida por una ley propia, distinta de otras leyes, pero no forzosamente incompatible con ellas*".[123] El autor también identifica dos sentidos de dicha palabra, uno ontológico[124] y otro ético.

Retomando lo señalado hace un par de párrafos, es este último sentido el desarrollado por Immanuel Kant. Para él, el eje de la autonomía de la ley moral lo constituye la autonomía de la voluntad, por la cual se hace posible el imperativo categórico. La autonomía de la voluntad es la propiedad a través de la cual la voluntad se constituye por sí misma. Así, se entiende al principio de autonomía como el hecho de elegir siempre de tal modo, que la misma volición abarque las máximas de nuestra elección como ley universal.[125]

Desde una óptica más contemporánea, la autonomía se identifica como "*la capacidad de hacer y actuar según las propias decisiones"*.[126] Se dice que un agente moral es autónomo si es responsable de sus decisiones, actos y consecuencias que de éstos derivan. Que pueda responder sobre ellas, justificarlas y dar razones al respecto a los demás y a sí mismo[127]. En ese sentido, es posible situar la definición dada por Díaz Osorio —quien es citado por Mauricio Mazo— quien señala que:

América, Penguin, 2006, p. 7

[121] CAMPBELL, L., "Kant, autonomy and bioethics", *op. cit*., *supra*, nota 118, p. 3

[122] RODRÍGUEZ CORONEL, M. M., "A vueltas con agencia moral: una perspectiva crítica", *op. cit*., *supra*, nota 76, p.133

[123] FERRATER MORA, J., *Diccionario de Filosofía*, *op. cit*., *supra*, nota 98, p. 161.

[124] Ferrater Mora refiere al sentido ontológico como el supone que "c*iertas esferas de la realidad son autónomas respecto de otras. Así, cuando se postula que la esfera de la realidad orgánica se rige por leyes distintas que la esfera de la realidad inorgánica, se dice que la primera es autónoma respecto de la segunda. Tal autonomía no implica que una esfera no determinada no se rija también por las leyes de otra esfera considerada como más fundamental*". *Idem*.

[125] *Idem.*

[126] SINGER, P., *Ética práctica*, Cambridge University Press, Reino Unido, p. 124 *apud* CABEZAS HERNÁNDEZ, M., "Autonomía y emocionalidad en el agente moral", en *Factótum, Revista de Filosofía,* núm. 7, 2010, pp. 76-77.

[127] *Ibidem,* p. 77.

> Ser autónomo significa que el sujeto tiene capacidad y libertad para pensar por sí mismo, con sentido crítico y aplicación en el contexto en el que se encuentra inmerso. Quiere decir que tiene mayoría de edad mental y madurez para actuar. De ahí se deduce que a mayor conocimiento, mayor posibilidad de autonomía y que ignorancia es ausencia de la misma, esto es, dependencia. [128]

También se dice que la condición de autonomía con la que cuenta un agente moral puede ser entendida como una forma de auto-constitución: "*el punto de vista de alguien que, en primera persona, se pregunta qué debe hacer y para decidir cómo actuar realizar un juicio evaluativo en función de las razones que se le aplican y un ámbito de responsabilidad que reconoce como propio"*.[129]

En cuanto a las capacidades que requiere un agente moral para ser autónomo, se dice que éste debe ser capaz de razonar, deliberar y decidir, de actuar intencionalmente y reconocer intenciones en los demás.[130] En un sentido mínimo, ser autónomos es:

> Poseer la capacidad de elegir, con base en la información disponible, qué hacer y de qué manera. Esto supone la posibilidad de formular objetivos, la posesión de creencias acerca de la modalidad de alcanzarlos y, en fin, la capacidad de evaluar el suceso de tales creencias a la luz de la evidencia empírica […]. Concebido en estos términos mínimos, la autonomía es equivalente a la capacidad de acción: es una precondición equivalente a la capacidad de acción: es una precondición fundamental para considerarse y ser considerados capaces de realizar cualquier clase de acto, y ser considerados responsables por ello.[131]

---

[128] MAZO ÁLVAREZ, H.M., "La Autonomía: Principio Ético Contemporáneo", en *Revista Colombiana de Ciencias Sociales*, vol.. 3 núm. 1, 2012, p. 125.

[129] MARQUISIO AGURRE, R., "El ideal de la autonomía moral", *op. cit*., *supra*, nota 116, p. 83.

[130] CABEZAS HERNÁNDEZ, M., "Autonomía y emocionalidad en el agente moral", *op. cit*., *supra*, nota 125, p. 77.

[131] PAPACCHINI, A., "El porvenir de la ética: La autonomía moral, un valor imprescindible para nuestro tiempo", en *Revista de Estudios Sociales, Universidad de los Andes*, núm. 5, 2000, s.p. [en línea], <https://journals.openedition.org/revestudsoc/30167#quotation>, [consulta: 12 de septiembre de 2023].

En efecto, se dice que, cuando un individuo cumple con estos requisitos mínimos para ser considerado capaz de tomar decisiones de forma autónoma, las demás personas, la comunidad y los Estados tienen la obligación de reconocerlo como alguien competente y responsable, y de respetar sus deliberaciones en tanto no amenacen la libertad de los demás.

En el caso de que falten dichos requisitos, no se tendría que hablar de autonomía, pues se carecen de capacidades para decidir racionalmente, lo que conlleva en ocasiones a formas de intervención, por ejemplo, frente a personas gravemente discapacitadas y con problemas de identidad personal, entre otros casos. [132]

Ahora bien, una acotación. Para este capítulo se entiende a la autonomía —como condición para ser un agente moral— como el control que éste tiene en sus decisiones, las cuales son elegidas a partir de que delibera qué hacer y cómo hacerlo. Al inicio de este apartado, se refirió que la postura que aquí se adopta de dicho concepto está vinculada al constructivismo, postura bajo la cual la autonomía moral[133] se entiende como "*el punto de vista del sujeto que posee una disposición normativa constitutiva a actuar de acuerdo con ciertas razones, identificado acciones por las que puede ser responsabilizado y asumiendo ciertos fines necesarios para el conjunto de sus acciones*".[134]

Continuando con la acotación, la autonomía implica que el agente posea una

---

[132] *Idem.*

[133] Son dos las concepciones más prominentes del concepto de autonomía, una es la de autonomía moral —a la cual se refiere en este apartado—, y otra la de autonomía personal. La autonomía moral tiene como base la ética kantiana, en tanto que la autonomía personal se basa en el liberalismo de John Stuart Mill. La autonomía moral tendría entonces como sustento lo siguiente: no elegir de otor modo sino de éste: que las máximas de la elección, en el querer mismo, sean al mismo tiempo incluidas como ley universal. Por su parte, la autonomía personal, toma como base el siguiente enunciado de Mill: "*El único objeto que autoriza a los hombres, individual o colectivamente, a turbar la libertad de acción de cualquiera de sus semejantes, es la propia defensa; la única razón legítima para usar de la fuerza contra un miembro de una comunidad civilizada es la de impedirle perjudicar a otros; pero el bien de este individuo sea físico, sea moral, no es razón suficiente […] Para que esta coacción sea justificable, sería necesario que la conducta de este hombre tuviese por objeto el perjuicio de otro. Para aquello que no le atañe más que a él, su independencia es, de hecho, absoluta. Sobre sí mismo, sobre su cuerpo y su espíritu, el individuo es soberano*". ĐERIĆ, M., "What is Autonomy Anyway?, *op. cit*., *supra*, nota 114, pp. 20-21.

[134] MARQUISIO AGURRE, R., "El ideal de la autonomía moral", *op. cit*., *supra*, nota 117, p. 58

disposición normativa, la cual se habrá de entender como el conjunto de razones para la acción y la creencia.[135] Cabe indicar, que dichas razones son del tipo moral, razones prácticas que "*conectan con el punto de vista de la imparcialidad y con los intereses de los otros con independencia de cualquier relación especial que puedan tener cono nosotros*".[136]

Ahora bien, se habrá de apelar a una postura en particular que contribuya a explicar la manera en la que se constituye la base normativa de nuestras acciones —o a forma de pregunta, ¿qué es lo que nos hace actuar moralmente? —. De acuerdo con Christine Korsgaard, la fuente de la normatividad está en el propio individuo:

> Las leyes de la moralidad son las leyes de la propia voluntad del agente, y de que sus exigencias son exigencias que el agente está dispuesto a hacerse. La capacidad de reflexión autoconsciente acerca de nuestras propias acciones nos confiere una especie de autoridad sobre nosotros mismos, y es esta autoridad lo que otorga normatividad a las exigencias morales.[137]

Así es como el agente moral establece sus propia ley o principio. Esa ley o principio, a su vez, proviene de máximas que todos los seres racionales podrían acordar para actuar en conjunto en un sistema cooperativo factible; y para que el agente se sienta vinculado a la ley moral, debe concebirse a sí mismo como Ciudadano del Reino de los Fines, esto es, como parte de una comunidad de seres racionales,[138] a través de las distintas identidades que como agente puede adoptar, ya sea como "*miembro de una familia, grupo étnico o nación*", o como el

---

[135] *Idem.*

[136] MARQUISIO AGURRE, R., "El ideal de la autonomía moral", *op. cit.*, *supra*, nota 117, p. 64. En la misma página, el autor también indica que los agentes racionales tienen un número virtualmente ilimitado de razones potenciales para la acción, en cuanto casi cualquier cosa que alguien haga puede concebirse como afectando a alguna otra cosa para bien o mal, para mejor o peor. Nos dice también que, a la hora de decidir qué hacer entre diferentes cursos posibles de acción podemos identificar razones auto interesadas (que no son necesariamente egoístas) y altruistas.

[137] KORSGAARD, C., *Las Fuentes de la Normatividad*, UNAM, Instituto de Investigaciones Filosóficas, México, 2000, p. 33.

[138] *Ibidem,* pp. 127-128.

"*administrador de sus propios intereses, y entonces será un egoísta*".[139]

Se trata, pues, de una "*identidad práctica*" capaz de hacer que el agente se identifique con el principio que lo obliga a actuar de cierta forma. Korsgaard señala:

> Uno es un ser humano, mujer u hombre, seguidor de cierta religión, miembro de un grupo étnico, miembro de cierta profesión, pareja o amigo de alguien, etc., y todas estas identidades dan lugar a razones y obligaciones. Nuestras razones expresan nuestra identidad, nuestra naturaleza: nuestras obligaciones emanan de lo que esa identidad prohíbe.[140]

En efecto, podemos cerrar este apartado señalando que la autonomía moral refiere al control que el agente tiene sobre sus decisiones, las cuales son elegidas a partir de deliberar qué hacer y cómo hacerlo, bajo una base normativa que ha de constituirse a partir de las identidades prácticas que el agente pudiese adoptar.

2.3. Tercera condición: la capacidad de actuar del agente

Como se señaló en el primer apartado de este capítulo, una de las condiciones que componen a la Agencia Moral, es la capacidad del agente de llevar a cabo su intención — por acción u omisión—, ya sea mediante sus movimientos corporales o instruyendo a que otro agente la realice.

Por principio de cuentas, se debe dejar en claro una aclaración que resultará de utilidad aquí. Lo que se pretende es enfatizar a la acción como movimiento, y susceptible a ser adjetivada de física, esto es, que pueda ser medida, sopesada y sometida a las claras prescripciones de las leyes naturales[141], de ahí que se diga que para que un agente moral sea capaz de actuar, debe hacerlo a través de sus movimientos corporales o a través de los movimientos de otro agente.

No deseo complicar este análisis conceptual profundizando en lo que distintas teorías explican en torno a la acción. Aun así, me gustaría explorar someramente lo

---

[139] *Ibidem,* pp. 129.
[140] *Ibidem,* p. 131.
[141] ARANA, J., "Acción Física y Acción Mental", en *Thémata, Revista de Filosofía,* núm. 30, 2003, p. 55.

que señalan, sólo con la finalidad de mostrar el vínculo que tiene la acción con la intencionalidad, de la cual ya referimos en otro apartado.

Así pues, he de comenzar esta breve exploración refiriendo a la teoría de la acción, la cual dice, en términos muy generales, que se trata del "*estudio de la estructura ontológica de la acción humana, del proceso por el que se origina y las maneras en las que se explica. En general, las acciones humanas constituyen una clase de eventos en los que un sujeto (el agente) produce algún cambio o cambios*".[142]

En lo que respecta a la acción humana, es posible identificar explicaciones surgidas desde otros campos, como el de la economía. Llama la atención la propuesta por Ludwig Von Mises, quien se refiere a ella como:

> *Human action is purposeful behavior. Or we may say: Action is will put into operation and transformed into an agency, is aiming at ends and goals, is the ego's meaningful response to stimuli and to the conditions of its environment, is a person's conscious adjustment to the state of the universe that determines his life*.[143] [La acción humana es un comportamiento con un propósito. O podemos decir: la acción se pone en funcionamiento y se transforma en un medio, busca fines y objetivos, es la respuesta significativa del ego a estímulos y condiciones de su entorno, es el ajuste consciente de una persona a estado del universo que determina su vida] (Traducción propia).

De un análisis rápido, es visible el rol principal que desempeña el ser humano frente a la acción. No es la intención de este apartado controvertir sobre tal aspecto. En todo caso, lo que se pretende —y como ya se ha referido anteriormente— es ampliar el mirador desde el cual se entiende al concepto de la acción, a fin de contar con un panorama más amplio del mismo.

La acción puede entenderse como el "*hecho, acto u operación que implica*

---

[142] *Idem.*

[143] VON MISES, L., *Human Action, A Treatise on Economics*, Estados Unidos de América, Fox & Wilkes, 1996, p. 11.

actividad, movimiento o cambio y normalmente un agente que actúa voluntariamente, en oposición a quietud o acción no física".[144] Esta definición se aproxima más a la idea de buscar enfatizar a la acción como movimiento, referida en el segundo párrafo de este apartado.

Se podría decir, entonces, que actuar implica que el agente —para el caso, el agente moral— lleve a cabo su intención, mediante movimientos propios, aunque como ya se ha señalado, también es posible que su intención sea realizada bajo su instrucción, por otro agente. En este último supuesto, pensando que ambos son agentes morales, ambos habrán de ser responsables de la acción, aunque es cierto que quien actúo lo hizo buscando concretar la intención de quien lo ordenó.

Regresando a la definición de acción, es evidente que busca vincular la actividad del agente con su voluntad. Esto puede ser un tanto intrincado, debido a que lleva a reflexionar sobre la volición, entendida como un "evento mental involucrado en el inicio de la acción".[145] Y a decir de las reflexiones existentes, tampoco es un término sencillo de definir.

Por ejemplo, desde una teoría causal de la acción, la volición es:

> Un acto de la mente que dirige su pensamiento a la producción de una acción y, por consiguiente, ejerce su poder para producirla […] la voluntad no es nada sino esa particular determinación de la mente en que […] la mente se esfuerza por hacer surgir, continuar o detener una acción que considera dentro de sus posibilidades. [146]

Desde este mirador, podría decirse que, para que inicie una acción se requiere de la voluntad del agente. Sin embargo, también existen otras posturas igual de válidas, como aquella en que refiere que las causas de las acciones son creencias más deseos, tal y como lo plantea Donald Davison, o con intenciones de hacer algo

[144]OXFORD, *Diccionario Lexico*, voz: acción. [en línea], <https://web.archive.org/web/20220823141312/https://www.lexico.com/es/definicion/accion>, [consulta: 12 de septiembre de 2023]
[145] AUDI, R., *Diccionario Akal de Filosofía,* España, Akal, 2004. p. 1017.
[146] *Idem*.

aquí y ahora, como lo identifica Wilfrid Sellars.[147] Y es aquí donde se ubica el vínculo de la intencionalidad con la acción.

En fin, la idea de acción que busca transmitirse en este apartado es esa capacidad con la cuenta un agente moral para realizar su intención. El agente moral actúa en tanto que emplea sus movimientos corporales, o en su caso, ordenando que otro agente actúe por él.

3. REFLEXIONES FINALES

La Agencia Moral se entiende como la facultad con la cuenta un agente para cuestionar sus acciones a la luz de principios que tienen que ver con lo bueno o lo justo, por ejemplo, ya sea que hayan sido ejercidas bajo un rol social establecido o como individuo. Asimismo, la Agencia Moral requiere de tres condiciones necesarias para su existencia: 1) la intencionalidad del agente; 2) la autonomía del agente, y 3) la capacidad de actuar del agente.

Se cuenta con intencionalidad sólo cuando el agente es consciente de la intención por la que actúa y del proceso que lo llevó a elegir tal acción. Se cuenta con autonomía debido a que el agente tiene control sobre sus decisiones, las cuales son elegidas a partir de que delibera qué hacer y cómo hacerlo, bajo una base normativa que ha de constituirse a partir de las identidades prácticas que pudiese adoptar. Y, finalmente, el agente cuenta con capacidad de actuar cuando lleva a cabo su intención, mediante movimientos corporales propios, o bien a través de otro agente que esté actuando bajo su instrucción, en aras de concretar dicha intención.

[147] *Ibidem*, p. 1018.

CAPÍTULO III
LA AGENCIA MORAL ARTIFICIAL

1. ANOTACIONES PREVIAS AL DESARROLLO DE ESTE CAPITULO

Las reflexiones en torno a si un robot puede ser considerado un agente moral, algo que por lo analizado en el segundo capítulo de este trabajo parece estar sólo reservado a seres biológicos con autonomía, como lo son los seres humanos, parecen aventuradas, incluso desconectadas con la realidad, ¿por qué querer hablar de algo que en la actualidad se vincula más con un tópico de ciencia ficción que con el mundo de los hechos? Previo al desarrollo de este capítulo, conviene identificar algunas de sus justificaciones, a fin de recalcar el valor que tiene hablar sobre la Agencia Moral Artificial.

De manera general, la discusión en torno a temas vinculados a la IA y el Derecho son pertinentes en la actualidad, más aún en el entorno académico nacional —particularmente en los centros de educación y producción jurídica, como lo son la Facultad de Derecho y el Instituto de Investigaciones Jurídicas de la UNAM—. Cabe señalar que se cuenta ya, en el primer capítulo de este trabajo, con un panorama muy general en torno al impacto global de la IA, y el interés que existe por analizar las problemáticas derivadas de su implementación, esto ya abre las puertas a posibles respuestas vinculadas al por qué debería interesarle a la comunidad jurídica mexicana las cuestiones vinculadas la IA.

De manera particular, también es posible identificar razones que están vinculadas con la filosofía, y en particular con la ética. La primera de ellas indica que

es relevante determinar si un algoritmo, por ejemplo, cuenta con Agencia Moral, porque pueden actuar en el mundo, impactándolo para bien o para mal. Otra razón es que, al establecer quién o qué es un agente moral, se sientan las bases para establecer la responsabilidad moral, y con ello la posibilidad de que el agente rinda cuentas por sus malas acciones.[148]

En el mismo sentido, el hecho de que los robots realicen de forma cada vez más habitual, tareas que antes sólo podían realizar personas humanas, hacen pensar que cuentan con un tipo de agencia de la cual no se tiene mucha claridad en el presente. De nuevo, esto puede generar lagunas en torno a la responsabilidad, y por ende confusiones en torno a quién tendría que ser castigado por los daños y perjuicios causados por los robots y/o la imposibilidad de encontrar a alguien justificadamente por los daños y perjuicios.[149]

Otra buena justificación para hablar sobre el tema es el desarrollo que en las últimas décadas ha tenido la IA. Cada vez es más habitual conocer de desarrollos basados en dicha tecnología que actúan con mayor independencia de su desarrollador y que son desplegados en áreas que les son cotidianas a las personas, por ejemplo, los servicios de transporte, la seguridad pública, el sistema de salud, el derecho y la industria sexual, por mencionar algunos ejemplos, mismos que levantan cuestionamientos que nos ayudan a identificar y analizar hasta qué medida tales desarrollos deben ser considerados y tratados como agentes morales.[150]

Una vez dicho lo anterior, ha de señalarse que este capítulo tiene dos objetivos, el primero de ellos es explorar las ideas más relevantes en torno al debate de la Agencia Moral Artificial y, el segundo, responder a la pregunta de si es posible adscribir una Agencia Moral Artificial a los sistemas artificiales autónomos.

El presente capítulo se divide en cinco apartados, el primero de ellos señala las

---

[148] VÉLIZ, C., “Moral zombies: why algorithms are not moral agents”, *AI & Society*, 2021. [en línea], <https://link.springer.com/article/10.1007/s00146-021-01189-x>, [consulta: 12 de septiembre de 2023].

[149] NYHOLM, S., *Humans and Robots, Ethics, Agency and Anthropomorphism,* Rowman & Littlefield, Reino Unido, 2020, pp. 33-34.

[150] BEHDADI, D. y C. Munthe, “A Normative Approach to Artificial Moral Agency”, en *Minds and Machines*, núm. 30, 2020, pp. 196.

reflexiones de Luciano Floridi respecto al debate de la Agencia Moral Artificial; el segundo, describe los tipos de agencia que pudieran llegar a desempeñar los sistemas artificiales autónomos; el tercero, muestra las posturas de dos grupos (los "*Computational Modelers*" y el de "*Computers-in-Society*") en torno al debate sobre la Agencia Moral Artificial; el cuarto, explora brevemente la Visión Estándar y la Visión Funcionalista de la agencia humana, las cuales son clave para comprender las posturas existentes alrededor de la Agencia Moral Artificial y, el quinto, responde a la pregunta de si es posible adscribir una Agencia Moral Artificial a los sistemas artificiales autónomos.

2. EL AGENTE MORAL ARTIFICIAL EN LA TEORÍA DE LUCIANO FLORIDI

En el presente apartado se explora el concepto de agente artificial propuesto por Luciano Floridi.[151] Dicho concepto es posible identificarlo en dos artículos cardinales para el debate en torno a la Agencia Moral de los sistemas artificiales autónomos, el primero de ellos titulado "*On the morality of artificial agent*s",[152] escrito en coautoría con Jeff Sanders, y el segundo *"Ética de la información: su naturaleza y alcance*".[153] Para el desarrollo de este apartado nos basaremos mayormente en el último, dado que en él se describe a la "*Ética de la Información*" que es la encargada de evaluar los comportamientos de los agentes artificiales.

2.1. La Ética de la Información

En un primer momento, Luciano Floridi indica que desde la aparición de los primeros trabajos en el campo de la Ética de la Información, esta se ha definido como el

[151] Luciano Floridi es un filósofo italiano, nacido en Roma, el 16 de noviembre de 1964. Se graduó en la Universidad de La Sapienza, en Roma; posteriormente, obtuvo su maestría en la Universidad de Oxford y el doctorado en la Universidad de Warwick, ambas en Reino Unido. Es profesor de Filosofía y Ética de la Información en la Universidad de Oxford, donde es director de investigación e investigador del Oxford Internet Institute. Sus temas de investigación son la Filosofía de la información, Ética de la información, Filosofía de la tecnología y Ética de la Inteligencia Artificial. OXFORD INTERNET INSTITUTE, *Professor Luciano Floridi*, [s.f., en línea], <https://www.oii.ox.ac.uk/people/profiles/luciano-floridi/> [consulta: 17 de marzo de 2024].

[152] FLORIDI, L. y J. W. Sanders, "On the Morality of Artificial Agents", en *Minds and Machines*, núm. 14, 2004, pp. 349-374.

[153] FLORIDI, L., "Ética de la Información: su naturaleza y alcance", en *Isegoria*, núm. 34, trad. de R. Feltrero y P. Olmos, 2006, pp. 19-46.

estudio de las cuestiones morales suscitadas por alguna de las tres dimensiones de la información[154] (información-como-recurso, información-como-producto, información-como-objetivo),[155] lo que ha derivado en una "*infructuosa compartimentación y a la aparición de falsos dilemas*",[156] tanto que se ignora el alcance global de dicha Ética, encasillándola en una microética: *"una ética aplicada de carácter práctico y profesional y dependiente de cada campo*".[157]

En vista de las limitaciones que presenta la Ética de la Información comprendida como una microética, y con la finalidad de superarlas, Luciano Floridi pretende escalar, a una de tipo macro, a dicha Ética. En ese contexto, la Ética de la Información —entendida ya como una macroética— es:

> Una ética ecológica que reemplaza el biocentrismo por el ontocéntrismo. La EI (Ética de la Información) mantiene que hay algo más elemental que la vida, que sería el ser -es decir, la existencia y la prosperidad de cualquier entidad en su entorno global- y también que hay algo más fundamental que el sufrimiento, que sería la entropía.[158]

Floridi también identifica que la Ética de la Información proporciona un vocabulario común, con el cual interpretar toda la realidad del ser mediante el nivel de abstracción informacional. Refiere que el ser/información posee un valor informacional, y que cualquier entidad informacional tiene una "*suerte de derecho Spinozista a permanecer en su propio estado y una suerte de derecho*

---

[154] *Ibidem*, p. 20.

[155] La información-como-recurso refiere a la información que es capaz de reunir el agente moral para obtener una (mejor) conclusión sobre lo que debe hacerse, dadas las circunstancias. La información-como-producto son las cuestiones morales suscitadas en temas como, por ejemplo, la imputabilidad (*accountability*), la responsabilidad (*liability*), la legislación sobre la calumnia, el testimonio, el plagio, la publicad, la propaganda y la desinformación, por ejemplo. Finalmente, la Información-como-objetivo refiere a la información que es susceptible de análisis ético, algunos temas es el hacking, la seguridad, el vandalismo, la piratería, la propiedad intelectual, el código abierto, la libertad de expresión, la censura, los filtros y el control de contenidos. *Ibidem*, pp. 20-23.

[156] *Ibidem*, p. 20.

[157] *Idem.*

[158] *Ibidem*, p. 27. Por entropía, Floridi entiende *"al concepto utilizado en física, es decir, la entropía termodinámica. En este contexto, entropía se refiere a cualquier tipo de destrucción o corrupción de los objetos informacionales (ojo, no de la información), es decir, a cualquier forma de empobrecimiento del ser, incluida la nada, para expresarlo de un modo más metafísico"*. *Idem.*

construccionista a la propia prosperidad, i. e. a mejorar y enriquecer su existencia y su esencia".[159] Con base en estos derechos, la Ética de la Información tiene la labor de evaluar:

> El deber de todo agente moral en términos de su contribución al crecimiento de la infosfera y cualquier proceso, acción o suceso que afecte negativamente a la infosfera en su conjunto -y no sólo a una entidad informacional- en términos del incremento del nivel de entropía y, por lo tanto, como una instancia del mal.[160]

La Ética de la Información involucra a toda entidad entendida desde el punto de vista informacional. Se trata de una ética imparcial y universal, que involucra a toda instancia independencia de si cuenta o no con una configuración física:

> La EI sostiene que toda entidad, como expresión del ser, posee una dignidad que queda constituida por su modo de existencia y esencia (el conjunto de todas las propiedades elementales que constituyen aquello que es) que merece ser respetada (al menos en un sentido mínimo y relativo) y por lo tanto, establece exigencias morales para el agente interactivo que pretenden establecer límites y servir de guía a sus decisiones y comportamientos éticos. [161]

La perspectiva ontológica propuesta por la Ética de la Información es de mayor alcance que la Bioética y la Ética Medioambiental, pues abarca a todo ser en tanto que es un cuerpo coherente de información.[162] En efecto, la Ética de la Información se presenta como una "*macroética ecológica, orientada hacia el sujeto paciente y*

---

[159] *Idem.*

[160] *Idem.* En lo que respecta al concepto de "Infosfera", Floridi lo entiende como el entorno que es esencialmente intangible e inmaterial, pero no por eso menos real o vital. La infosfera genera dilemas éticos, por ejemplo, la creación de capacidades, la seguridad de la información, la ampliación del acceso universal, el respeto a la diversidad entre otros. FLORIDI, L., "Ethics in the Infosphere", *versión revisada y presentada en la 161va reunión del Consejo Ejecutivo de la UNESCO "Las nuevas tecnologías de la información y la comunicación para el desarrollo de la educación*, 31 de mayo de 2001, p. 4.

[161] FLORIDI, L., "Ética de la Información: su naturaleza y alcance"*, op. cit*., *supra*, nota 153, p. 28.

[162] *Ibidem,* pp. 28-29.

ontocéntrica",[163] y que como se señaló, evalúa al agente moral en términos de su contribución al crecimiento de la *"infosfera"*. Ahora bien, ¿qué entiende por agente moral la Ética de la Información?, para responder a esta pregunta es importante identificar previamente lo qué son los Niveles de Abstracción propuestos también por Luciano Floridi, pues con ellos es posible identificar los componentes observables del agente moral.

2.2. Los Niveles de Abstracción

El Nivel de Abstracción se define como el: "*conjunto finito-no-vacío de observables que serán los elementos de una teoría caracterizada, precisamente, por la elección de los mismos*".[164] En esta definición, lo "*observable*" corresponde a una "*variante tipo interpretada, es decir, una variable tipo junto con una proposición que expresa a qué rasgo del sistema se refiere*".[165] En aras de hacer más comprensible lo que son los Niveles de Abstracción, Luciano Floridi emplea el siguiente ejemplo:

> "Supongamos que nos unimos a la conversación que ya mantiene Anne (A), Ben (B) y Carole (C). Anne es una coleccionista y posible compradora; Ben 'hace chapuzas' en su tiempo libre; y Carole es economista. No sabemos de qué están hablando, pero llegamos a oír lo siguiente:
>
> A) Anne dice que (lo que sea) tiene instalado un dispositivo antirrobo, se guarda en el garaje cuando no se usa y tiene un único propietario;
>
> B) Ben apunta que su motor no es el original, que recientemente se ha repintado su exterior y que sus piezas de cuero están muy gastadas;
>
> C) Carole comenta que su viejo motor consumía demasiado, que tiene un determinado su valor de mercado (sic), pero que sus piezas

[163] *Ibidem,* p. 26. En la teoría de Luciano Floridi, un paciente moral es entendido como toda entidad, que, al ser objeto informacional, tiene un valor moral intrínseco, el cual por más mínimo y relativo que sea, lo convierte en merecedor de un grado, aunque igualmente mínimo, de respeto moral. Esto último refiere a una atención cuidadosa, apreciativa y desinteresada. *Ibidem*, p. 36.
[164] *Ibidem*, p. 30.
[165] *Idem.*

de repuesto resultan caras".[166]

Sobre este ejemplo, Floridi indica que cada uno de los participantes contempla el objeto de discusión de acuerdo a sus propios intereses, los cuales constituyen sus propios Niveles de Abstracción. Se podría estar hablando de un coche, una moto o un avión, pues cualquiera de estos tres objetos de referencia (que en la teoría de Floridi se consideran sistemas, en tanto que suponen una fuente de información) satisfaría las descripciones proporcionadas por Anne, Ben y Carole. [167]

Con los Niveles de Abstracción se hace posible un determinado análisis del sistema, a cuyo resultado se le denomina un modelo del sistema.[168] Floridi indica que los Niveles de Abstracción, insertos en una teoría, permiten el análisis del sistema bajo observación y elaborar un modelo que identifique algunas de las propiedades de un determinado Nivel de Abstracción. [169]

Como tal, la Ética de la Información está comprometida con Niveles de Abstracción que interpreten la realidad desde un punto de vista informacional. La IA compara los Niveles de Abstracción de la Ética de la Información con los de la Ética Medioambiental, Luciano Floridi identifica que los primeros cuentan con las siguientes ventajas: 1) tienen en cuenta el valor moral de lo que es, contando con recursos conceptuales para valorar la situación moral e indicar el curso de acción apropiado y 2) la obtención de una mejor y más precisa comprensión de qué podría considerarse un agente moral o un paciente moral.[170] Habida cuenta de la Ética de la Información y de los Niveles de Abstracción, es momento de analizar el concepto de agente moral propuesto por Luciano Floridi.

2.3. El agente moral desde la Ética de la Información

Desde la Ética de la Información, un agente moral es "*un sistema en transición, interactivo, autónomo y adaptativo que puede ejecutar acciones susceptibles de*

[166] *Ibidem*, p. 29.
[167] *Idem.*
[168] *Idem.*
[169] *Ibidem*, p. 30.
[170] *Ibidem*, p. 31.

calificación moral".[171] En aras de clarificar su definición, Luciano Floridi explica cada uno de los elementos que la conforman. En lo que respecta a "*sistema en transición*", señala que su interés es por sistemas que se modifican, esto es, sistemas que cambian el valor de sus propiedades. Estos cambios, dice el autor, pueden interpretarse con mayor detalle cuando se emplea un Nivel de Abstracción de carácter inferior.[172] Lo siguiente sirve como ejemplo:

> Por ejemplo, el sistema objeto de la discusión de Anne […], podría dotarse de componentes de estado correspondientes a su localización, a si está o no en uso, si está arrancando, si el sistema antirrobo está armado, a su historial de propietarios y a su eficiencia energética. La operación de guardar el objeto en el garaje podría tener como input al conductor y como efecto, el de la situación final del objeto en el garaje, con el motor apagado, el dispositivo antirrobo armado, el historial de propiedad sin modificar y dando como resultado el gasto de una determinada cantidad de energía. La componente de estado correspondiente a "si está en uso" podría adoptar, de manera no determinista, cualquiera de sus valores posibles, dependiendo del caso particular que se dé entre transición (quizá el objeto no se encuentra en uso, permaneciendo en el garaje toda la noche; o quizá el conductor esté escuchando un partido de críquet en la radio, en la soledad del su garaje). La definición concreta del estado depende del NdA [Nivel de Abstracción]. [173]

Así, con esta perspectiva conceptual, Luciano Floridi busca contemplar a una entidad como algo que presenta estados y cuyas definiciones dependerán precisamente del Nivel de Abstracción empleado para su descripción. En el mismo hilo explicativo de la definición de agente moral, el autor señala que un sistema en transición es interactivo cuando *"el sistema y su entorno pueden actuar el uno sobre el otro […] Por ejemplo, la fuerza gravitatoria de dos cuerpos"*;[174] es autónomo

[171] *Idem.*
[172] *Ibidem*, p. 32.
[173] *Idem.*
[174] *Idem.*

cuando *"capaz de cambiar de estado en ausencia de interacción, es decir, cuando puede efectuar transiciones internas para cambiar su estado*".[175] Sobre este punto, el autor indica que todo agente debe presentar al menos dos estados, y que esta propiedad dota al agente de un cierto grado de complejidad y cierta independencia de su entorno;[176] y es adaptativo cuando las interacciones del agente "*pueden modificar las reglas de transición mediante las cuales cambia de estado*",[177] garantizándole que pueda ser considerado, desde un Nivel de Abstracción dado, como "*una entidad que aprende su propio modo de funcionamiento, de una forma que depende de manera crítica de su propia experiencia". La parte que completa la explicación de la definición de agente moral es la de "acción susceptible de calificación moral*",[178] con lo cual busca indicarse que una acción se considera moral si es capaz de producir el bien o el mal morales.

Con la definición propuesta por Floridi y Sanders de agente moral, es posible enmarcar dentro de la Ética de la Información a los agentes artificiales como agentes morales, a los cuales es posible imputar sus acciones. Los agentes artificiales abarcarían no sólo a "*agentes digitales, sino también a agentes sociales como las sociedades, partidos o los sistemas híbridos formados por máquinas y humanos o los humanos con sus capacidades incrementadas por medio de la tecnología*". [179]

El hecho de que la Ética de la Información amplíe el espectro de consideración hacía agentes artificiales, facilita el desarrollo de una investigación adecuada de la moralidad distribuida, un fenómeno descrito por Floridi como *"macroscópico y creciente, relacionado con las acciones morales globales y las responsabilidades colectivas, y que es resultado de la 'mano invisible' que actúa en las interacciones sistémicas entre distintos agentes a un nivel local"*,[180] y también a la comprensión desde el campo de la ética de agentes artificiales "*suficientemente informados, 'listos', autónomos y capaces de ejecutar acciones moralmente relevantes de*

[175] *Ibidem*, pp. 32-33.
[176] *Ibidem*, p. 33.
[177] *Idem.*
[178] *Idem.*
[179] *Idem.*
[180] *Idem.*

manera independiente de los humanos que los han creado, produciendo 'el bien artificial' y 'el mal artificial'". [181]

El autor es consciente de que considerar a los agentes artificiales como agentes morales puede traer consigo una serie de problemáticas relacionadas con la responsabilidad moral (un concepto en el cual está relacionado la intencionalidad, la conciencia y otros estados mentales de los cuales carecen los sistemas con IA, por ejemplo). Para superar este obstáculo, propone basar su punto de vista ético en la *"Agencia Moral"*, la *"imputabilidad"* y la *"censura"*,[182] pues con estas categorías no se estaría obligado de encontrar al responsable individual cada vez que sucede algo malo, ya que ahora se es capaz de admitir que, en ocasiones, la fuente moral del malo del bien puede encontrase en algo que no sea ni un individuo ni un grupo de seres humanos. [183]

El argumento que sigue contribuye a la defensa de lo dicho en el párrafo anterior, mostrando a su vez una ventaja de que exista una Agencia Moral en ausencia de responsabilidad moral:

> La posibilidad de tratar a los agentes no humanos como agentes morales facilita la discusión sobre la moralidad de los agentes, no sólo en el contexto del ciberespacio, sino también en el de la biosfera —en los que los animales pueden considerarse agentes morales sin por ello tener que mostrar libre albedrío, emociones o estados mentales— y en contextos de "moralidad distribuida", en los que los agentes legales y sociales pueden ahora ser considerados como agentes morales. La enorme ventaja de esta perspectiva es una mejor adaptación del discurso moral a los contextos no humanos. [184]

---

[181] *Ibidem*, pp. 33-34.

[182] *Ibidem*, p. 35. A decir de la censura, este es un concepto que aparece en el artículo "*On the Morality of Artificial Agents*", y que refiere a un tipo de castigo impuesto a los agentes morales artificiales cuyo actuar va en contra de las convencionalidades aceptadas por la sociedad. Un agente moral artificial podría ser censurado de distintas formas, por ejemplo: su modificación o la eliminación del agente del ciberespacio, sin respaldo alguno. FLORIDI, L. y J.W. Sanders, "On the Morality of Artificial Agents", *op. cit*., supra, nota 152, p. 208.

[183] FLORIDI, L., "Ética de la Información: su naturaleza y alcance"*, op. cit*., *supra*, nota 153, p. 35.

[184] *Idem.*

Con lo dicho, Luciano Floridi no busca prescindir del concepto de responsabilidad, pues es consciente de los compromisos ontológicos de los creadores de nuevos agentes y entornos. Lo que busca al excluir de su marco epistémico a la responsabilidad es generar una perspectiva de una "*moralidad no mental*' que permita abarcar agentes artificiales, siendo un coste que según él merece la pena pagar, y más aún según se avanza hacía una más compleja sociedad de la información. [185]

3. DESCRIPCIÓN DE LOS TIPOS DE AGENCIA QUE PUDIERAN LLEGAR A DESEMPEÑAR LOS SISTEMAS ARTIFICIALES AUTÓNOMOS

Como nota previa al desarrollo de este apartado, es pertinente aclarar que tiene como base la exposición y crítica que Fabio Tollon hace en su artículo: "*Moral Agents or Mindless Machines? A Critical Appraisal of Agency in Artificial Systems*"[186] de las tres conceptualizaciones realizadas por Deborah Johnson (conocida por sus reflexiones en torno a la ética de la computación) y sus coautores, respecto a los roles potencialmente cargados de moralidad que pueden llegar a desempeñar los sistemas artificiales autónomos. Los tres tipos de agencia que pudieran ser acordes a dichos sistemas son los siguientes: la agencia de eficacia causal, la actuación en nombre de la agencia y la agencia autónoma.

Es conveniente también indicar la idea de agencia utilizada en el texto de Tollon, para el autor los agentes son entidades capaces de tener un cierto efecto en el mundo, y en donde dicho efecto suele corresponderse a ciertos objetivos, ya sea en forma de deseos, creencias e intenciones que tiene el agente.

El artículo de Tollon muestra lo que tal vez sean los principales argumentos para afirmar o negar la adscripción de una Agencia Moral a las máquinas dotadas con IA. Cabe señalar que las conclusiones a las que llega el autor en su texto es que el uso continuo del concepto de la palabra "*autonomía*" en las discusiones sobre la agencia artificial complica las cosas innecesariamente cuando se busca reflexionar sobre los papeles morales que las máquinas pueden llegar a desempeñar:

[185] *Ibidem,* pp. 35-36.
[186] TOLLON, F., "Moral Agents or Mindless Machines? A Critical Appraisal of Agency in Artificial Systems", en *Hungarian Philosophical Review*, vol. 4 núm. 63, 2019, pp. 9-23.

> *This occurs in two ways: firstly, the condition of autonomy, even in its engineering sense, fails to accurately describe the way "autonomus" systems are developed in practice. Secondly, the continued usage of autonomy in the moral sense introduces unnecessary meta-physical baggage form the free will debate into discussions about moral agency.*[187] *[…] I claimed that the continued usage of such a metaphysically loaded term complicates our ability to get a good handle on our concepts and obscures the ways in which we can coherently think through nuanced accounts of the moral role(s) that machines may come to play*".[188] [Esto sucede de dos formas: en primer lugar, la condición de autonomía, incluso desde su enfoque ingenieril, no describe con precisión la forma en que los sistemas 'autónomos' se desarrollan en la práctica. En segundo lugar, el uso continuado del concepto de autonomía en el sentido moral introduce una carga metafísica innecesaria en el debate sobre el libre albedrío en las discusiones sobre la Agencia Moral [..] Afirmé que el uso continuado de un término tan cargado de metafísica complica nuestra capacidad para manejar bien nuestros conceptos y obscurece las formas en las que podemos pensar de forma coherente acerca del papel o los papeles morales que las máquinas pueden llegar a desempeñar] (Traducción propia).

No es el objetivo de este apartado confrontar esta postura. Como ya se ha ido adelantando, la intención de éste es mostrar el marco conceptual descrito en el artículo de Tollon, con la finalidad de ampliar el panorama en torno al estado actual de las discusiones sobre la Agencia Moral Artificial.

3.1. Agencia de eficacia causal

Este tipo de agencia se refiere a: "*[…] to the abilitiy of some entities -specifically technological artefacts- to have a causal influence on various states of affairs, as*

[187] *Ibidem*, p. 9.
[188] *Ibidem,* p. 21.

extensions of the agency of the humans that produce and use them".[189] [a la capacidad de algunas entidades —específicamente los artefactos tecnológicos— de tener una influencia causal en diversos estados de las cosas, como extensiones de la agencia de los seres humanos que los producen y utilizan] (Traducción propia). Esta categorización incluye artefactos que pueden separarse de los humanos, tanto en el tiempo como en el espacio, así como artefactos que son desplegados directamente por el ser humano. [190]

Una buena pregunta es si tiene sentido pensar en estos artefactos como "*agentes*", pues se asemejan más a una "*herramienta*". La respuesta que Deborah Johnson tiene para esto es que los artefactos tecnológicos son "*agentes"* en tanto que en ellos se encuentran programadas/codificadas las intenciones humanas. Una herramienta, como lo puede ser un martillo, no cuenta con las especificaciones de cómo debe ser utilizado (bien puede ser utilizado para golpear a una persona o martillar un clavo) en su propia composición.

Además, no puede realizar o representar la tarea de forma independiente de la manipulación humana: "*The key distinction between a tool and a technological artefact, according to Johnson, is that the latter has a form of intentionality as a key feature of its make-up, shile the former does not*".[191] [La principal distinción entre una herramienta y un artefacto tecnológico, según Johnson, es que este último tiene una forma de intencionalidad como característica clave de su composición, mientras que el primero no] (Traducción propia).

En el sentido expuesto por Johnson, la intencionalidad de la tecnología pone de manifiesto que los artefactos tecnológicos estarían diseñados de determinadas maneras para lograr ciertos resultados. El ejemplo utilizado para explicar esto, es el siguiente:

> Consider the simple example of a search engine: keys are pressed in a specific order in an appropiate box and then a button is pressed. The search engine then goes through a set of processes that have been

[189] *Ibidem*, p. 11.
[190] *Idem.*
[191] *Ibidem*, p. 12.

> programmed into it by a human being. The 'reasons' for the program doing what it does are therefore necessarily tethered to the intentions of the human being that created it.[192] [Consideremos el ejemplo de un motor de búsqueda: se pueden pulsar las teclas en un orden específico en una caja de texto, y luego se pulsa un botón. A continuación, el motor de búsqueda pasa por una serie de procesos que fueron programados por el ser humano. En efecto, las razones por las que el programa hace lo que hace están necesariamente ligadas a las interacciones del ser humano que lo creó] (Traducción propia).

Desde dicho mirador, tiene sentido pensar que los artefactos tecnológicos cuentan con "*agencia*", en virtud de que suponen (en función a su diseño específico) una diferencia en los resultados efectivos de los que se dispone. Los artefactos hacen posibles formas innovadoras de alcanzar determinados fines y, por lo tanto, amplían el abanico de posibles acciones disponibles para un agente concreto en una situación particular.[193].

La eficacia causal del artefacto se engloba precisamente en el hecho de que contribuyen a la creación de ciertos estados de cosas nuevos, pero siempre en conjunción con las acciones de los seres humanos:

> *The reason we can think of these causally efficacious artefacts as agents is the fact that they make substantial causal contributions to certain outcomes. In this way the causal efficaciousness of an entity leads, in the form of a non-trivial action performed by that entity, to a specific event*.[194] [La razón por la que podemos pensar en estos artefactos con eficacia causal como agentes es el hecho de que hacen contribuciones causales sustanciales a ciertos resultados. Así, la eficacia causal conduce, en forma de una acción no trivial realizada por el artefacto, a un acontecimiento concreto] (Traducción propia).

[192] *Idem.*
[193] *Idem.*
[194] *Idem*.

Los artefactos tecnológicos son, en efecto, agentes por el hecho de que sus fabricantes tienen ciertas intenciones al diseñarlos y crearlos, que adquieren sentido en relación con los seres humanos. En palabras de Tollon, la concepción de agencia dada a los artefactos tecnológicos refiere a entidades que actúan en conjunción causal con los seres humanos. Al reflexionar sobre el sentido de responsabilidad asociado a este tipo de agencia, Tollon señala que no es posible identificarla en el artefacto tecnológico, por lo que no es viable considerar al artefacto como un agente moral.[195]

3.2. Actuando en nombre de la agencia

Esta forma de agencia alude a los artefactos tecnológicos que actúan en nombre de operadores humanos. Es una especie de agencia subrogada, algo similar a lo que ocurriría cuando una persona actúa en nombre de otra. La persona representante se encuentra limitada por ciertas normas y responsabilidades. [196]

Los artefactos que sustituyen o actúan en nombre de operadores humanos lo hacen en determinados ámbitos. El ejemplo utilizado en el texto de Tollon para describir este tipo de agencia es el de una plataforma de intercambio, que funciona a través de ordenadores y que se caracteriza por las operaciones de alta frecuencia que ahí ocurren. Todo esto ocurre sin la necesidad de que un agente de bolsa humano lleve un registro. Lo que al final busca dejar de manifiesto este ejemplo es que en la actualidad es posible que tareas que antes eran de dominio exclusivo del ser humano, son ahora realizadas por sistemas artificiales, sin necesidad de que los primeros intervengan. Si bien es cierto que la tarea de operar en el mercado la realizan sistemas artificiales, también lo es que ésta sigue siendo la misma que realizaría un agente de bolsa humano. El artefacto, en efecto, actúa dentro de unos parámetros establecidos previamente por un operador humano. [197]

La distinción de un artefacto que actúa en nombre de la agencia y un agente que es causalmente eficaz, es que este último influye en los resultados junto con los seres humanos, mientras que el primero puede actuar a nombre de un agente humano, pero con la limitante de que las acciones que puede ejecutar dependen de

[195] *Idem.*
[196] *Ibidem,* p. 13.
[197] *Ibidem*, pp. 13-14.

las intenciones de sus programadores/diseñadores, es decir, se encuentra determinado.

Otra diferencia importante es que un artefacto que actúa en nombre de la agencia tiene un mayor grado de independencia de la intervención humana —el ejemplo de la plataforma de operaciones ayuda bastante a comprender este punto, pues para ingresar una orden de compra o de venta, y que esta se ejecute, no requiere de la intervención humana, pues los propios algoritmos sobre la que fue construida se encargarán de ello—, por lo que tienen en mayor medida inserta la intencionalidad humana. [198]

Sobre esto último, Deborah Johnson, citada por Fabio Tollon, afirma que al evaluar el comportamiento de los sistemas informáticos: "*there is a triad of intentionality at work, the intentionality of the system, and the intentionality of the user.*"[199] [hay una triada de intencionalidad en el trabajo, intencionalidad en el sistema e intencionalidad en el usuario], y que comprende estados mentales internos como las creencias y los deseos, en donde finalmente reside la responsabilidad de cualquier acción del sistema.[200]

Estos artefactos, en efecto, pueden tener consecuencias en la vida moral de las personas. El ejemplo utilizado es el siguiente:

> *Consider, for example, automatic emergency breaking (AEB) technology, which automatically applies the brakes when it detects an object near the front of the vehicle. This simple system has been enormously successful, and research indicates that it could lead to reductions in 'pedestrian crashes, right turn crashes, head on crashes, rear end and hit fixed object crashes'. We can usefully think of AEB as assisting us in being better and safer drivers, leading to decreased road fatalities and injuries. These artificial systems, of which AEB is an example, can therefore be seen as performing delegated tasks which can have moral significance. We can*

[198] JOHNSON, D., "Computer Systems: Moral Entities but not Moral Agents", en *Ethics and Information Technology*, núm. 8, p. 202 *apud* TOLLON, F., "Moral Agents or Mindless Machines? A Critical Appraisal of Agency in Artificial Systems", *op. cit*., *supra*, nota 186, p. 14.

[199] TOLLON, F., "Moral Agents or Mindless Machines? A Critical Appraisal of Agency in Artificial Systems", *op. cit.*, *supra*, nota 186, p. 14.

[200] *Idem.*

> *therefore meaningfully thing of them as being morally relevant entities*.[201] [Consideremos, por ejemplo, la tecnología de frenado automático de emergencia (AEB, por su sigla en inglés), que aplica automáticamente los frenos cuando detecta un objeto cerca de la parte delantera del vehículo. Este sencillo sistema ha tenido un enorme éxito, y las investigaciones indican que podrían reducir las colisiones con peatones, los giros a la derecha, los choques frontales, los choques por detrás y los choques con objetos fijos. Podemos pensar que el AEB nos ayuda a ser mejores y más seguros conductores, pues ayudan a reducir las muertes y lesiones en carretera. Estos sistemas artificiales, de los que el AEB es un ejemplo, pueden considerarse como tareas delegadas que pueden tener un significado moral. En efecto, podemos considerar que son entidades moralmente relevantes] (Traducción propia).

Pese a tener un impacto en el ámbito moral de los humanos, artefactos como el señalado no pueden ser considerados como agentes morales, ya que la intencionalidad deviene del diseñador y del usuario del artefacto. Johnson señala que sería un error afirmar que los seres humanos y los artefactos son componentes intercambiables en la acción moral.[202] Para finalizar con esta conceptualización de agencia, Fabio Tollon señala que los artefactos pueden ser agentes que, al actuar en nombre de los seres humanos, participan en actos con consecuencias morales. Enfatiza también de esto no se infiere que sean moralmente responsables de los actos, pues la responsabilidad está reservada para los seres humanos, y remata diciendo que, para ser moralmente responsable, un agente también debe contar con autonomía. [203]

3.3. Agencia autónoma

En este apartado, Fabio Tollon explica que en su relato se puede entender a la autonomía de dos formas: la primera, como algo que suele atribuirse a los seres humanos, y que cuenta con una dimensión claramente moral capaz de hacernos

[201] *Ibidem,* pp. 14-15.
[202] *Idem.*
[203] *Ibidem*, p. 16.

moralmente responsables de nuestras acciones, y la segunda, que puede definirse desde el campo de la ingeniería (de software). A decir de esta última:

> Computer scientists commonly refer to 'autonomous' programs in order to highlight their ability to carry out tasks on behalf of humans and, furthermore, to do so in a highly independent manner. A simplistic example of such a system might be a machine-learning algorithm which is better equipped to operate in novel environments than a simple, pre-programmed algorithm.[204] [Los informáticos suelen referirse a los programas 'autónomos' para destacar su capacidad de realizar tareas en nombre de los humanos y, además, hacerlo de forma muy independiente. Un ejemplo simplista de un sistema de este tipo podría ser un algoritmo de aprendizaje automático que está mejor equipado para operar en entornos novedosos que un simple algoritmo preprogramado] (Traducción propia).

Sin embargo, a un algoritmo, aún con su capacidad operar y funcionar con independencia de un ser humano, no puede atribuírsele Agencia Moral, pues en realidad éstos no eligen libremente cómo actuar[205] —hasta el momento en el que se escribe este trabajo de investigación, no es posible que un algoritmo realice una tarea fuera del determinado ámbito para el que fue programado, por ejemplo, un algoritmo destinado a jugar ajedrez no puede identificar imágenes de gatos—.

Así pues, la "*verdadera*" autonomía queda reservada para los seres humanos. Son las personas con agencia quienes tienen un control significativo sobre sus propias acciones "*a type of control which makes a decision or action up to the agent and not other external circumstances*".[206] [un tipo de control que hace que una decisión o acción dependa del agente y no de otras circunstancias externas] (Traducción propia). Pero, qué ocurre si la decisión libre no es producto del agente específico en cuestión, sino más bien a una presión externa. Si es así, entonces la acción no es libre, y habría dificultad para determinar si el agente podría ser

[204] *Idem.*
[205] *Idem.*
[206] *Ibidem*, p. 17.

moralmente responsable de su acto. [207]

La reflexión que hace Tollon al distinguir las dos concepciones de agencia autónoma —la primera inserta en una dimensión moral, y la segunda relacionada con la capacidad de un sistema artificial de actuar sin intervención humana— zanja el terreno para la siguiente conclusión: los artefactos tecnológicos no deberían entenderse como poseedores de una autonomía moralmente relevante, pues no se considera que sus decisiones sean elegidas libremente. Sus intenciones, remata Tollon, están siempre vinculadas a las intenciones de sus diseñadores y usuarios finales.[208]

4. EL DEBATE ACERCA DE LA AGENCIA MORAL ARTIFICIAL ENTRE EL GRUPO DE LOS "COMPUTATIONAL MODELERS" Y EL GRUPO DE "COMPUTERS-IN-SOCIETY"

El presente apartado toma como base el artículo "*Un-making artificial moral agents*",[209] de Deborah Johnson y Keith W. Miller. Este apartado busca exponer los argumentos de los dos principales grupos de académicos en torno al debate sobre el estatus moral de los sistemas de cómputo con cierto grado de independencia. Cabe referir también que el artículo de Johnson y Miller es una crítica [210] a la postura del Grupo de los "*Computational Modelers*", específicamente del trabajo seminal de Luciano Floridi y Jeff W. Sanders "*On the morality of artificial agent*s", señalado ya en este capítulo.

4.1. El Grupo de los "Computational Modelers"

El grupo de los "*Computational Modelers*", cuya traducción podría ser el grupo de los Modelistas Computacionales, concibe a las computadoras como un modelo de

[207] *Idem.*
[208] *Idem.*
[209] JOHNSON, D. y K. W. Miller, "Un-making artificial moral agents", en *Ethics and Information Technology*, núm. 10, 2008, pp. 123-133.
[210] La crítica de Johnson y Miller reside en que los sistemas de cómputo pueden comportarse independientemente de los humanos que los crearon y desplegaron, pero que en el sentido más importante del término "independiente", estos sistemas no son completamente independientes de sus creadores humanos. A partir de esto, los autores tejen su contraargumento, el cual enfatiza el peligro de conceptualizar a los sistemas de cómputo como agentes morales autónomos. *Ibidem*, p. 124.

la realidad o una herramienta para la exploración científica, y tanto Johnson y Miller identifican a Floridi y Sanders como parte de este grupo.[211] Este grupo ve al modelado computacional[212] como un esfuerzo filosófico por capturar la realidad:

> *Their agenda can be seen to be rooted in the enlightenment tradition and the idea that reason will lead to truth. Computational Modelers have a stake in using computation as a (if not the) foundation of a body of knowledge that brings insight to a wide range of areas; i.e., they have a stake in the value of computation as a model. Whether computing is used as a model of reality or a pragmatic model (i.e., a useful way of thinking about and manipulating nature) is an interesting question, but not one that gets much attention from Computational Modelers. Instead the assumption seems to be that computational models represents reality*.[213] [Su agenda puede verse arraigada en la tradición ilustrada y en la idea de que la razón conducirá a la verdad. A los modelistas computacionales les interesa utilizar la computación como base (si no la base) de un cuerpo de conocimiento que aporte conocimiento a una amplia gama de áreas; es decir, les interesa el valor de la computación como modelo. La cuestión de si la computación se utiliza como modelo de la realidad o como modelo pragmático (es decir, una forma útil de pensar y manipular la naturaleza) es interesante, pero no recibe mucha atención por parte de los modelistas computacionales. En cambio, parece que se asume que los modelos computacionales representan la realidad] (traducción propia).

Desde la óptica de este grupo, las computadoras son máquinas que se comportan, y que pueden ser utilizadas no sólo para modelar el comportamiento,

---

[211] *Ibidem*, p. 126.

[212] El modelado computacional refiere al uso de computadoras para simular y estudiar sistemas complejos, utilizando las matemáticas, la física y la informática. El modelado permite a los científicos realizar experimentos simulados por computadora (por ejemplo, en el pronóstico del tiempo, las simulaciones de vuelo y de terremotos, entre otros casos). NATIONAL INSTITUTE OF BIOMEDICAL IMAGING AND BIOENGINEERING, *Modelado Computacional*, s.f. [en línea], <https://www.nibib.nih.gov/espanol/temas-cientificos/modelado-computacional#pid-2186>, [consulta: 12 de septiembre de 2023].

[213] JOHNSON, D. y K. W. Miller, “Un-making artificial moral agents”, *op. cit*., *supra*, nota 209, p. 126.

sino también para producirlo. También, indican que los comportamientos de un sistema de cómputo pueden ser comparados con el comportamiento humano, Johnson y Miller lo ejemplifican así:

> *For example, before computerization of the stock exchange, a buyer would call a stockbroker, check on the price of a stock and tell her to purchase a particular number of shares of the stock. The stockbroker would call someone, say certain words, fill out paper work, and the buyer would eventually receive paper confirmation of ownership of shares of the stock. Now that the entire process has been computerized, a buyer can set up a computer system to execute the purchase of the stock if and when the price reaches a certain point. Few human movements are involved; no words need be spoken; and paper may never be used. It would seem here that the computer system acts as an agent replacing the stockbroker who acted as an agent on behalf of a client in the old system. While we will challenge the comparability of the behaviors here, for now, the point is that Computational Modelers seem to want to make much of the fact that machines can now perform tasks that only humans performed in the past*.[214] [Por ejemplo, antes de la informatización de la bolsa de valores, un comprador llamaba a un agente de bolsa, comprobaba el precio de una acción y le decía que comprará un número determinado de acciones. El agente de bolsa llamaba a alguien, decía ciertas palabras, rellenaba papeles y el comprador recibía finalmente la confirmación en papel de la propiedad de las acciones. Ahora que todo el proceso se ha informatizado, el comprador puede configurar un sistema informático para que ejecute la compra de las acciones cuando el precio alcance un determinado punto. Hay poca actividad humana, no hay que decir nada y no hay que usar papel. Parece que el sistema informático actúa como un agente que sustituye al corredor de bolsa que actuaba en nombre de un cliente en el antiguo sistema. Aunque aquí pondremos en tela de juicio si es posible comparar los comportamientos, por ahora, la cuestión es que los modelistas computacionales parecen querer dar importancia al hecho

[214] *Idem.*

de que la maquinas en la actualidad pueden realizar tareas que en el pasado sólo realizaban los humanos] (Traducción propia).

El grupo de los Modelistas Computacionales busca validar su postura —la de que los sistemas de cómputo son capaces de modelar comportamientos de los seres humanos— señalando que en la actualidad las computadoras pueden hacer tareas que son realizadas por humanos.[215] Dicha postura, a su vez, es la base con la cual este grupo aborda el debate referente al estatus moral de los sistemas computacionales que actúan con independencia de los seres humanos.

La tradición filosófica de este grupo se basa generalmente en la lógica y la filosofía de la mente. Además, creen que es posible establecer paralelismos entre el comportamiento de un sistema de cómputo y el comportamiento humano. En lo que respecta a la idea de la Agencia Moral, creen lo siguiente: "*They seem to believe that if computer systems are given the status of moral agents, this will be testimony to the achievement of the computational model in capturing reality*"[216]. [Parecen creer que, si se otorga a los sistemas informáticos la condición de agentes morales, será testimonio del logro del modelo computacional en la captación de la realidad] (Traducción propia).

Lo anterior es complementado con lo siguiente: "*it seems that the Computational Modelers are pushing for a 'frame of meaning' for computer systems that has hertofores been rerserved for humans: the frame of 'moral agent*".[217] [Parece que los modelistas computacionales están presionando por un 'marco de significado' para los sistemas informáticos que hasta ahora se ha reservado para los humanos: el marco de 'agente moral'] (Traducción propia). Algo que, como se verá más adelante es la raíz de las críticas que el grupo "*Computers-in-Society*" tiene hacia los "*Computational Modelers*".

4.2. El Grupo de "Computers-in-Society"

El Grupo de "*Computers-in-Society*", que podría traducirse como "*Computadoras en*

[215] *Idem.*
[216] *Ibidem*, p. 127.
[217] *Idem.*

la Sociedad", y en el cual se adscriben los autores del artículo, busca resaltar el rol que tienen las computadoras, y las tecnologías en general, en las vidas de los seres humanos, especialmente en sus vidas morales:

> *The agenda of this group is to draw attention to the power of technology, especially computer technology, and its moral implications. This means bringing to light the moral character of computer systems, the values embedded in their design, and the ways in which they affect the moral lives of human beings. The adoption and use of computer technology has powerful effects and those who decide about its design and adoption have enormous power*.[218] [El objetivo de este grupo es llamar la atención sobre el poder de la tecnología, especialmente la informática, y sus implicaciones morales. Esto significa sacar a la luz el carácter moral de los sistemas informáticos, los valores incorporados en su diseño y la forma en que afectan a la vida moral de los seres humanos. La adopción y el uso de la informática tiene efectos poderosos y quienes deciden su diseño y adopción tienen un enorme poder] (Traducción propia).

Con esta perspectiva se busca resaltar los efectos del uso de los sistemas informáticos y también el poder que tienen quienes deciden sobre su diseño y su adopción. La intención de que los autores sostengan esta postura es que con ella es posible comprender mejor las implicaciones morales de la tecnología, lo cual deriva en una mejor dirección del desarrollo tecnológico. [219]

En lo que corresponde al debate referente a los agentes artificiales, este grupo reconoce que los sistemas de cómputo se comportan independientemente (en algún sentido del término "*independencia*"), y que esto, a su vez, genera cuestiones que conciernen a la moral:

> *Computers are increasingly being assigned tasks that humans controlled and performed in the past; Big Blue (sic) defeats Kasparov, robots replace*

[218] *Ibidem,* p. 126.
[219] *Idem.*

> *workers on an assembly line, and computers system instead of physicians recommend treatments. As we mentioned earlier, what information we receive when we search, what medicines or doses of radiation we receive, how airplane and automobile traffic is ordered, what financial transaction are implemented, how weapons are launched and targeted are now task done (decisions made) by computer systems. If computers are performing tasks and making these critical decisions, then, the Computers-in-Society argue, there must be some accountability for the effects of these systems on human well-being*[220]. [A las computadoras se les asigna cada vez más tareas que los humanos controlaban y realizan en el pasado; Big Blue (el nombre correcto es Deep Blue) derrota a Kasparov, los robots sustituyen a los trabajadores en una cadena de montaje y los sistemas informáticos recomiendan tratamientos médicos. Como hemos mencionado antes, la información que recibimos cuando buscamos, los medicamentos o las dosis de radiación que recibimos, cómo se ordena el tráfico de aviones y automóviles, qué transacciones financieras se realizan, cómo se lanzan y apuntan las armas son ahora tareas realizadas (decisiones tomadas) por sistemas informáticos. Si los ordenadores realizan tareas y toman esas decisiones críticas, entonces, argumentan los del Grupo de computadoras en la sociedad, debe haber alguna responsabilidad por los efectos de estos sistemas en el bienestar humano] (Traducción propia).

En este sentido, el Grupo de "*Computers-in-Society*" reconoce el vínculo entre sistemas informáticos y seres humanos. Es verdad que dicha tecnología puede comportarse independientemente, pero también lo es que es creada y desplegada por humanos. En efecto, asumir que un sistema informático es un "*agente moral*" implica no reconocer la responsabilidad que existe en los seres humanos y dejar de lado que son ellos quienes finalmente tienen el control de la tecnología. [221]

De hecho, la principal crítica que parece articular el Grupo de "*Computers-in-Society*" en contra de los "*Computational Modelers*" es la del peligro que representa asumir una Agencia Moral de artefactos tecnológicos, pues implica que estos

[220] *Ibidem,* p. 127.
[221] *Idem.*

pueden funcionar sin ninguna responsabilidad humana por el comportamiento del sistema:

> *The claims are odd because they do not seem to acknowledge that computer systems are an extension of human activity and human agency, and because their view of agency and morality is out of sync with the idea of morality as a human system of contextualized ideas and meanings. The claims seem dangerous because they imply that computer systems can operate without any human responsibility for the system behavior*".[222] [Las afirmaciones (respecto a que los sistemas informáticos pueden considerarse agentes morales) son absurdas porque no parecen reconocer que los sistemas informáticos son una extensión de la actividad humana y de la agencia humana, y porque su visión de la agencia y la moralidad no está en sintonía con la idea de la moralidad como un sistema humano de ideas y significados contextualizados. Las afirmaciones parecen peligrosas porque implican que los sistemas informáticos pueden funcionar sin ninguna responsabilidad humana por el comportamiento del sistema] (Traducción propia).

No obstante, es posible identificar dentro de este grupo a académicos que defienden la postura de que los artefactos tecnológicos, como lo es un sistema artificial autónomo, cuentan con "*Agencia*", al menos desde una perspectiva de la eficacia causal, en tanto que éstos contribuyen a las actividades realizadas por los seres humanos.[223]

4.3. La crítica del Grupo "Computers-in-Society" al Grupo de los "Computational Modelers"

El debate en el que participan estos grupos tiene como recompensa para el vencedor la construcción del significado de lo que debe entenderse por agente artificial y su importancia de cara al ámbito moral. Johnson y Miller dan luz a lo que

[222] *Idem.*
[223] *Idem.*

acontecería si, por ejemplo, este debate lo ganara el Grupo de los Modelistas Computacionales: "*If the Computational Modelers win the debate, artificial agents will be understood to have the potential for moral standing (moral agency) of a kind that ultimately might lead to a prohibition on turning them off*".[224] [Si los modelistas computacionales ganan el debate, se entenderá que los agentes artificiales tienen el potencial de tener una posición moral (Agencia Moral) de un tipo que en última instancia podría llevar a la prohibición de apagarlos] (traducción propia).

Los autores también identifican lo que ocurriría si el grupo al cual pertenecen resultara vencedor:

> *If the Computers-in-Society group wins, artificial agents will be understood to be important components of the moral world but they will always be understood to be human constructions under human control. They would always be understood to be "tethered" to humans in the sense that they are the products of human intervention, are deployed by humans for human purposes, operate in contexts maintained by humans, and cannot function without some degree of human control (even thought that control may be distant in time and space*).[225] [Si gana el grupo de '*Computers-in-Society'* se entenderá que los agentes artificiales son componentes importantes del mundo moral, pero siempre se entenderá que son construcciones humanas bajo control humano. Siempre se entenderá que están 'vinculados' a los humanos en el sentido de que son producto de la intervención humana, son desplegados por humanos para propósitos humanos, operan en contextos mantenidos por los humanos y no pueden funcionar sin cierto grado de control humano (aunque ese control pueda estar distante en el tiempo y en el espacio)] (Traducción propia).

Este último escenario sirve para identificar los principales contraargumentos que hace el Grupo de "*Computers in Society*" al de los "*Computational Modelers*", y particularmente —como se señaló al inicio de este apartado— a la obra de Luciano

[224] *Idem.*
[225] *Ibidem,* p. 128.

Floridi y Jeff Sanders.

Por principio de cuentas, el Grupo "*Computers-in-Society*" concede que los sistemas de cómputo se comportan independientemente y pueden coincidir con los Niveles de Abstracción propuestos por Floridi y Sanders: *"i.e., they have the hability to respond to a stimulus by change of state (interactivity), the capacity to change state without an external stimulus (autonomy), and the abitily to change the transition rules by which states are changed (adaptibility)*",[226] [es decir, tienen la capacidad de responder a un estímulo mediante el cambio de estado (interactividad), la capacidad de cambiar sin un estímulo externo (autonomía) y la capacidad de cambiar las reglas de transición por las que se cambian los estados (adaptabilidad)] (Traducción propia).

Sin embargo, también piensan que con dichos criterios no es suficiente para afirmar que los sistemas de cómputo son o deberían ser tratados como agentes morales, pues si esto aconteciera las posibilidades de responsabilizar a quienes diseñan y despliegan la tecnología desaparecerían. Para ejemplificar esto, los autores mencionan lo siguiente:

> *Consider a complex bot that learns and generates its own rules of behavior. Suppose it generates rules of behavior that make it an extremely risky entity. Say it makes medical decisions or regulates financial transactions and no humans understand exactly how it does what it does. According to the Computers-in-Society group, to conceptualize such systems as autonomous moral agents is to absolve those who design and deploy them from any responsibility for doing so. It is similar to absolving from responsibility people who put massive amounts of chemicals in the ocean on grounds that they did not know precisely how the chemicals would react with salt water or algae (though they knew generally about chemical reactions and toxicity). Ignorance is no excuse since they knew they were engaged in risky activity. Just as we might ask, "What is the best way to think about chemicals to ensure that people know they are dangerous?" the Computers-in-society group asks, "What is the best way*

[226] *Ibidem*, p. 131.

> *to think about computer systems to ensure that those who create and deploy them do so safely?" The Computers-in-Society group claims that it is dangerous to construct computer systems as autonomous agents because the construction implies that no humans are responsible for what the systems do; they argue that – no matter how independently they behave – computer systems should be understood in ways that keep them conceptually tethered to human agents*.[227] [Consideremos un robot complejo que aprende y genera sus propias reglas de comportamiento. Supongamos que genera reglas de comportamiento que lo convierten en una entidad extremadamente riesgosa. Digamos que toma decisiones médicas o regula las transacciones financieras y que ningún ser humano entiende exactamente cómo hace lo que hace. Según el grupo de "*Computers-in-Society*", conceptualizar tales sistemas como agentes morales autónomos es absolver de cualquier responsabilidad a quienes los diseñan y despliegan. Es similar a eximir de responsabilidad a las personas que vierten cantidades masivas de productos químicos en el océano alegando que no sabían exactamente cómo reaccionarían los productos químicos con el agua salada o las algas (aunque sí sabían en general sobre las reacciones químicas y la toxicidad). La ignorancia no es una excusa, ya que sabían que estaban realizando una actividad arriesgada. Al igual que nos preguntamos: '¿Cuál es la mejor manera de pensar en los productos químicos para garantizar que la gente sepa que son peligrosos?', el grupo de "*Computers-in-Society"* se pregunta: '¿Cuál es la mejor manera de pensar en los sistemas informáticos para garantizar que quienes lo crean y los utilizan lo hacen de forma segura? '. El Grupo de *"Computers-in-Society"* afirma que es peligroso concebir los sistemas informáticos como agentes autónomos, ya que esta concepción implica que los humanos no son responsables de lo que hacen los sistemas] (Traducción propia).

En efecto, lo que pretende el Grupo de "*Computers-in-Society*" es preservar el vínculo de la tecnología con sus desarrolladores y con quienes la despliegan y usan,

[227] *Idem.*

pues a consideración de dicho grupo es en el ser humano en quien debería recaer la responsabilidad por los actos con consecuencias morales realizados por los sistemas artificiales autónomos.

5. LA VISIÓN ESTÁNDAR Y LA VISIÓN FUNCIONALISTA DE LA AGENCIA HUMANA COMO PARTE DEL DEBATE EN TORNO A LA AGENCIA MORAL ARTIFICIAL

Una conceptualización más en torno al debate de la Agencia Moral Artificial es la expresada por Dorna Behdadi y Christian Munthe, en su artículo "*A Normative Approach to Artificial Moral Agency*".[228] A decir de los autores, existen dos visiones sobre el debate de la Agencia Moral Artificial, y cuyos argumentos devienen de analizar qué es lo que compone a la Agencia Moral humana.

Las visiones identificadas en el artículo son la estándar y la funcionalista,[229] las cuales son incompatibles entre sí y cada una de ellas con distintas implicaciones frente a la idea de una posible Agencia Moral Artificial.[230]

5.1. Visión estándar

Para esta visión, la Agencia Moral humana implica que los agentes morales cuenten con racionalidad, voluntad o autonomía y condiciones de conciencia fenomenal[231]. Para Deborah Johnson —una de las principales proponentes de esta visión y quien es citada en el artículo de Behdadi y Munthe— la Agencia Moral de una entidad (E), debe estar en sintonía con las siguientes condiciones:

> *1. E causes a physical event with its body.*
> *2. E has an internal state, I, consisting of its own desires, beliefs and other intentional states that together comprise a reason to act in a certain way (rationality and consciousness).*

[228] BEHDADI, D. y C. Munthe, "A Normative Approach to Artificial Moral Agency", *op. cit*., *supra*, nota 150, pp. 195-218.
[229] A nota al pie de página, los autores aclaran que cuando se refieren a "funcionalismo" y "funcionalistas" en su artículo, se refieren exclusivamente a una idea de la agencia moral, y no a las teorías funcionalistas cognitivas, de la conciencia, del significado o cualquier otra área en dónde dichos términos sean utilizados. *Ibidem*, p. 197.
[230] *Idem.*
[231] *Idem.*

3. The state I is the direct cause of 1.

4. The event in 1 has some effect of moral importance.[232]

[1. E provoca un evento físico con su cuerpo. 2. E tiene un estado interno, I, que consiste en sus propios deseos, creencias y otros estados intencionales que, en conjunto, conforman una razón para actuar de una determinada manera (racionalidad y conciencia). 3. El estado I es la causa directa de 1. 4. El acontecimiento de 1 tiene algún efecto de relevancia moral] (Traducción propia).

Sobre estas condiciones, Behdadi y Munthe puntualizan que el número 1 (y posiblemente el 2) garantizan que E sea un agente, en tanto que 2 y 3 señalan lo que se requiere para la presencia de una Agencia Moral como la adscrita a los seres humanos. La particularidad número 4 asegura que el comportamiento es moralmente relevante, indicando que E ejerce su Agencia Moral en la situación correspondiente. [233]

Deborah Johnson busca resumir su postura diciendo que todo comportamiento puede explicarse por sus causas, pero sólo la acción puede explicarse por un conjunto de estados mentales internos, los cuales comprenden las creencias, deseos y otros estados intencionales que puedan contribuir a explicar por qué un agente actuó.[234]

Las objeciones a una posible Agencia Moral Artificial se localizan en las condiciones 2 y 3. En efecto, las entidades artificiales son incapaces de ser agentes morales dado que carecen de estados mentales internos que deriven en acontecimientos, los cuales a su vez son sus acciones. Es cierto que estas entidades pueden actuar, y que sus comportamientos se asemejan a la acción humana desde un punto de vista externo; sin embargo, estos comportamientos no bastan para conferirles cualidades morales, ya que carecen de estados mentales internos.

En efecto, la visión estándar aboga por una Agencia Moral con cualidades

[232] *Ibidem*, p. 198.
[233] *Idem.*
[234] *Idem.*

vinculadas a la naturaleza humana, y frente a la cual la idea de una Agencia Moral Artificial no podría justificarse, pues es evidente que un robot no cuenta con estados mentales internos que orienten sus actuaciones.

5.2. Visión funcionalista

La visión funcionalista tiene en Luciano Floridi y Jeff Sanders a sus principales representantes. Entre sus posturas más relevantes, se encuentra el rechazo a criterios como el de conciencia y la adopción de la idea de una "*moralidad no mental*". Ambos autores señalan que las entidades pueden ser agentes morales dependiendo del Nivel de Abstracción elegido para dar cuenta de los componentes de la Agencia Moral humana. Behdadi y Munthe explican:

> *The level of abstraction applied by the standard view is very low, keeping the criteria close to the case of an adult being, but raising the level allow for less anthropocentric perspectives while maintaining consistency and relevant similarity concerning the underlying structural features of paradigmatic human moral agents*.[235] [El nivel de abstracción aplicado por el punto de vista estándar es muy bajo, manteniendo los criterios cercanos al caso de un ser adulto, pero al elevar el nivel es posible contar con perspectivas menos antropocéntricas mientras se mantiene la consistencia y similitud relevante concernientes a los rasgos estructurales que subyacen en los agentes morales humanos] (traducción propia).

Floridi y Sanders ofrecen un conjunto de condiciones que permiten identificar a la agencia humana: Interactividad, Autonomía[236] y Adaptabilidad. Behdadi y Munthe enuncian y describen dichas condiciones de la siguiente forma:

> 1. *Interactivity: E interacts with its environment.*

[235] *Idem.*

[236] El término utilizado por Behdadi y Munthe es el de independencia, sin embargo, en el trabajo de Floridi y Sanders el término empleado es el de "autonomía". La justificación que tienen los autores para este cambio es que el término "autonomía" no corresponde al uso estándar que se le da en la filosofía, y que se asemeja más a concepto de autonomía empleado en las ciencias de la computación y la robótica. *Ibidem,* p. 199.

> *2. Independence: E has an ability to change itself and its interactions independently of immediate external influence.*
> *3. Adaptability: E may change the way in which 2 is actualized based on the outcome of 1*. [237] [1. Interactividad: E interactúa con su entorno. 2.Independencia: E tiene la capacidad de transformarse a sí y a sus interacciones independientemente de las influencias externas inmediatas. 3. Adaptabilidad: E puede cambiar la manera en que se actualiza 2 en función del resultado de 1] (Traducción propia).

De acuerdo con Behdadi y Munthe, al contrastar estas condiciones con las identificadas en la visión estándar de la Agencia Moral humana, puede identificarse que la condición 1 es correspondida con su contrapartida de la visión estándar, que la condición 2 se aleja de la visión estándar pues no es necesario contar con ningún tipo de estados mentales internos para contar con autonomía, en tanto que la condición 3 también es diferente.

Así, desde la visión funcionalista es posible justificar la existencia de la Agencia Moral Artificial, pues como también ya se refirió en el primer apartado de este capítulo, basta que un ser cuente con los niveles de abstracción identificados para la Agencia Moral, para ser considera como un agente moral. Esta visión permite pensar no sólo en la moralidad —comprendida como la capacidad del agente de regular su conducta en torno a principios e ideas generales relativos a la bueno y lo justo— de los sistemas artificiales con autonomía, sino también en la consideración moral que como seres humanos deberíamos tener hacía ellos.

6. ¿ES POSIBLE ADSCRIBIRLES UNA AGENCIA MORAL A LAS MÁQUINAS AUTÓNOMAS?

El debate sobre la Agencia Moral Artificial es complejo, y tiende a serlo más conforme quienes debaten incorporan nuevas variantes conceptuales, en donde la idea de Agencia Moral u otras que son clave en la discusión —como puede serlo el de responsabilidad— adquieren matices que facilitan la inclusión de una gama más

[237] *Idem.*

amplia de tipos de agentes.[238] .

Esto pudiera traer consigo la desventaja de que cuando el foco de discusión se centra en qué es un agente moral artificial, las cuestiones prácticas urgentes que involucran entidades artificiales cada vez más avanzadas se ensombrezcan,[239] pero también la ventaja de que el debate en torno a este tema se nutre, y además permite imaginar soluciones a dilemas éticos que pudieran llegar a presentarse producto de la interacción entre robots y seres humanos, por ejemplo.

La intención de tomar postura en este trabajo con una posible respuesta a la pregunta de si es posible adscribirles una Agencia Moral a los sistemas artificiales autónomos, es la de contar con bases que faciliten la construcción de propuestas orientadas en el corto y mediano plazo al desarrollo ético de la IA, y a largo plazo, a reflexionar sobre la moral que pudiera regir en un Agente Moral Artificial. Con lo anterior se deja entrever un poco la dirección de la respuesta que ha de desarrollarse aquí. A reserva de argumentar más adelante sobre ella, la respuesta es que no es posible adscribirles una Agencia Moral a los sistemas con IA, al menos no en el presente.

Que se niegue la Agencia Moral no implica que no se quiera reflexionar sobre el impacto en el campo moral de las decisiones hechas por sistemas artificiales autónomos. Tampoco implica negar o restarle valor a una ética específica para reflexionar críticamente sobre la moral de dichos sistemas. Sobre esto último, aún en ausencia de una Agencia Moral Artificial, resulta relevante seguir cuestionando desde la ética el comportamiento de los sistemas con IA y si los principios, por ejemplo, que han de seguirse por estos sistemas pueden serles codificados —en el sentido informático—.[240]

De cara al desarrollo de la IA, no parece buena idea pensar que la ética solo deba reflexionar sobre la moral de los agentes a los que se les puede responsabilizar moralmente de sus acciones, puesto que actúan con autonomía (en

[238] BEHDADI, D. y C. Munthe, “A Normative Approach to Artificial Moral Agency”, *op. cit.*, *supra*, nota 150, p. 212.
[239] *Idem.*
[240] NATH, R. y V. Sahu, “The problem of machine ethics in artificial intelligence”, en *AI & Society*, núm. 35, 2017, p. 107.

el sentido filosófico de la palabra) y con estados mentales que impulsan sus actos, porque llegará el momento en el que los sistemas artificiales autónomos se separen varios grados de la responsabilidad de su diseñador, de su creador o de quienes los despliegan.

Michael Anderson y Susan Anderson traen a escena un argumento que objeta la idea de que no es viable reflexionar sobre la moral de los sistemas artificiales autónomos, sólo porque estos no son agentes morales:

> *This type of objection, however, shows that the critic has not recognized an important distinction between performing the morally correct action in a given situation, including being able to justify it by appealing to an acceptable ethical principle, and being held morally responsible for the action. Yes, intentionality and free will in some sense are necessary to hold a being morally responsible for its actions, and it would be difficult to establish that a machine possesses these qualities, but neither attribute is necessary to do the morally correct action in an ethical dilemma and justify it. All that is required is that the machine act in a way that conforms with what would be considered to be the morally correct action in that situation and be able to justify its action by citing and acceptable ethical principle that it is following*.[241] [Este tipo de objeción, sin embargo, muestra que el crítico no ha reconocido una distinción importante entre realizar la acción moralmente correcta en una determinada situación, incluyendo la posibilidad de justificarla apelando a un principio ético aceptable, y ser considerado moralmente responsable de la acción. Sí, la intencionalidad y el libre albedrío son necesarios en cierto sentido para considerar a un ser moralmente responsable de sus acciones, y sería difícil establecer que una máquina posee estas cualidades, pero ninguno de los dos atributos es necesario para realizar la acción moralmente correcta en un dilema ético y justificarla. Todo lo que se requiere es que la máquina actúe de una manera que se ajuste a lo que se consideraría la acción moralmente correcta en esa situación y sea capaz de justificar su acción citando un

[241] ANDERSON, M. y S. Anderson, "Machine ethics: creating and ethical intelligent agent", en *AI MAG*, núm 28, 2007, p. 19 *apud* NATH, R. y V. Sahu, "The problem of machine ethics in artificial intelligence", *op. cit*., *supra*, nota 240, p. 107.

principio ético aceptable que esté siguiendo] (Traducción propia).

Reiterando, es importante que la ética reflexione sobre la moralidad de los actos realizados por máquinas dotados de IA, para cual se requiere una ética específica, que se centre en los dilemas que se presentan cuando éstas interactúan con seres humanos y con el medio ambiente, por ejemplo. Después llegará el momento en este trabajo de reflexionar más sobre la Ética de la IA.

Ahora bien, toca justificar la respuesta dada a la pregunta de si es posible adscribirles una Agencia Moral a las máquinas dotadas de IA. Como se adelantó, la respuesta es no, pues en la actualidad ningún sistema con IA cumple con condiciones internas que son cruciales a la hora de adscribir una Agencia Moral.

Las máquinas, al no contar con condiciones internas como la autonomía o la intencionalidad, no están habilitadas para reaccionar por sí solas de forma correcta frente a una o varias circunstancias sociales que así lo requieran; tampoco cuentan con la capacidad de justificar sus acciones o elecciones ni mucho menos con la capacidad de actuar moralmente de forma consciente y también coherente con sus emociones —con las cuales tampoco cuenta— y con las prácticas de la comunidad en la que se desenvuelve. [242]

Y la respuesta negativa hacia la pregunta continuará, en tanto los sistemas con IA sigan construyéndose bajo el paradigma actual. Por ejemplo, pensando en "*robots emocionales*" que sean construidos para simular emociones, pero que sean incapaces de contar con una arquitectura emocional y cognitiva como la de los seres

[242] Esta idea surge al pensar en los tipos de agencia moral descritos en el artículo "*Can artificial intelligences be moral agents?*", en ese artículo, los autores refieren a tres tipos de agencia: 1) El agente moral irreflexivo o superficial: alguien que posee la arquitectura emocional y cognitiva que le permite seguir "ciegamente" las reglas de la comunidad. Este tipo de comportamiento no requiere reflexión teórica, ni justificación alguna de sus acciones y elecciones realizadas. 2) El agente moral reflexivo: es un individuo que no sólo es capaz de seguir inconscientemente las reglas morales, sino que también puede justificar sus acciones y elecciones, adaptando de forma consciente una regla moral abstracta (aquí encajan la mayoría de los adultos humanos con buenas aptitudes para la comunicación), y 3) El agente moral sofisticado (o profundo): alguien que es capaz de tomar regularmente decisiones morales de forma consciente, comprendiendo que sus intuiciones pueden llevarle por el mal camino y que sopesar las razones y considerar las teorías morales abstractas debe formar parte del proceso de toma de decisiones morales. En este nivel de autoconsciencia moral se sitúan figuras como Gandhi o el Dalai Lama. BROZEK, B. y B. Janik, "Can artificial intelligences be moral agents", en *New Ideas in Psychology*, núm. 54, 2019, p. 104.

humanos, que le permiten participar en las prácticas morales de su comunidad:

> *The problem is not connected with the sophistication and complexity of those robots, but with something more fundamental: the principles which guide the development of their architecture. The human unreflective moral agent depends to a large extent on fundamental emotional mechanisms, while the existing AI architectures are far from doing so. Truly emotional robots are not those equipped with an additional 'emotional module', which calculates some parameters we interpret as 'the level of anger' or 'the intensity of joy'. Rather, emotional algorithms would involve "a control structure that relies on emotional states in competition and collaboration for inducing state changes in the system to drive both its bodily and cognitive behaviour, not algorithms that compute emotional content as if it were simply an output. (…) [One should rather envisage] an architecture for cognition in which the functional implementations of emotions are the computational substrate from which reason emerges by way of motivating the manipulation of data in various ways that engender, among other activities, data-gathering (curiosity, boredom), recombinant thought (discovery), contradiction avoidance (confusion), and mistake recovery (mirth). As long as the approach to the AI is only to 'imitate' the emotional behaviour, there will be no artificial unreflective moral agents. If this is the case, it is also impossible for a machine (developed with the use of non-emotional algorithms) to become a reflective or a sophisticated moral agent, since the layers of moral agency - as we insisted - form a hierarchy, in which a higher level depends on the existence of the lower level*.[243] [El problema no está relacionado con la sofisticación y la complejidad de esos robots, sino con algo más fundamental: los principios que guían el desarrollo de su arquitectura. El agente moral irreflexivo humano depende en gran medida de mecanismos emocionales fundamentales, mientras que las arquitecturas de IA existentes están lejos de hacerlo. Los robots verdaderamente emocionales no son aquellos equipados con un "módulo emocional" adicional, que calcula algunos parámetros que interpretamos

[243] NATH, R. y V. Sahu, "The problem of machine ethics in artificial intelligence", *op. cit.*, *supra*, nota 240, p. 105.

como “el nivel de ira” o “la intensidad de alegría”. Los algoritmos emocionales implicarían más bien ‘una estructura de control que se basa en los estados emocionales en competencia y colaboración para inducir cambios de estado en el sistema para impulsar su comportamiento corporal y cognitivo, no algoritmos que calculan el contenido emocional como si fuera simplemente un resultado. (Más bien habría que prever) una arquitectura de la cognición en la que las implementaciones funcionales de las emociones sean el sub-estado computacional del que emerge la razón, motivando la manipulación de datos de distintas formas que engendran, entre otras actividades, la recopilación de datos (curiosidad, aburrimiento), el pensamiento recombinante (descubrimiento), la evitación de contradicciones (confusión) y la recuperación de errores (alegría). Mientras el planteamiento de la IA consista únicamente en ‘imitar’ el comportamiento emocional, no habrá agentes morales artificiales irreflexivos. Si este es el caso, también es imposible que una máquina (desarrollada con el uso de algoritmos no emocionales) sea un agente moral reflexivo o sofisticado, ya que los estratos de la Agencia Moral -como insistimos- forman una jerarquía, en la que el nivel superior depende de la existencia del nivel inferior] (traducción propia).

Así, mientras las máquinas estén diseñadas para imitar comportamientos, sin ninguna reflexión moral de su parte, no será posible pensar en una Agencia Moral Artificial. Otro argumento que fortalece la posición de este trabajo frente a la posible idea de la Agencia Moral Artificial es el de Carissa Veliz, quien refiere que los algoritmos al no contar con la capacidad de sentir no pueden ser autónomos (en el sentido filosófico de la palabra) ni capaces de responsabilizarse por sus acciones:

To have a conception of the good that we want to pursue (autonomy), we need to have a feel for what leads to pleasure, meaningfulness, and satisfaction. To guide or actions by the recognition of others’ moral claims in a way that can count as moral action (accountability), we must have a sense of other’s capacity to suffer, of what if feels like to be harmed, of

> *what we can do to others' minds and bodies through our actions*.[244] [Para tener una concepción del bien que queremos perseguir (autonomía), necesitamos tener un sentido de lo que conduce al placer, a la significación y a la satisfacción. Para guiar nuestras acciones mediante el reconocimiento de las pretensiones morales de los demás, de lo que se siente al ser dañado, de lo que podemos hacer a las mentes y cuerpos de los demás a través de nuestras acciones] (traducción propia).

La anterior idea tiene sentido para este trabajo, pues nos dice que para ser agente moral es indispensable ser sensible, ya que con ello se es consciente de lo que significa infringir dolor o causar placer. Esta autora también hace una interesante reflexión que vale la pena mencionar aquí:

> *There might come a time when AI becomes so sophisticated that robots might possess desires and values of their own. It will not, however, be on account of their computational prowess, but on account of their sentience, which may in turn require some kind of embodiment. At present, we are far from creating sentient algorithms*.[245] [Puede llegar un momento en que la IA sea tan sofisticada que los robots puedan tener deseos y valores propios. Sin embargo, no será por su capacidad de cálculo, sino por su sintiencia, que a su vez puede requerir algún tipo de encarnación. En la actualidad, estamos lejos de crear algoritmos sintientes] (Traducción propia).

Es cierto que al pensar en el futuro se cae en el terreno de la especulación. Si pensamos en el futuro de la IA, no sabemos con certeza hacia donde se dirigirá su desarrollo ni los impactos que pudiese tener en la sociedad. Por la misma razón, no sería justo desechar las palabras de Véliz, las cuales se complementan bien si se piensa en la posibilidad de una superinteligencia artificial, entendida como "*any*

[244] VÉLIZ, C., "Moral zombies: why algorithms are not moral agents", *op. cit. supra*, nota 148. [en línea], <https://link.springer.com/article/10.1007/s00146-021-01189-x>, [consulta: 12 de septiembre de 2023]

[245] *Idem.*

intellect that is vastly outperforms the best humans brains in practically every field, including scientific creativity, general wisdom, and social skills",[246] [cualquier intelecto que supere ampliamente a los mejores cerebros humanos en prácticamente todos los campos, incluyendo la creatividad científica, la sabiduría general y las habilidades sociales] (traducción propia) que pueda ser reconocida como un agente moral y no como una mera herramienta o artefacto. A decir de esto, es valioso señalar lo que piensa Nick Bostrom:

> *To the extent that ethics is a cognitive pursuit, a superintelligence could do it better than human thinkers. This means that questions about ethics, in so far as they have correct answers that can be arrived at by reasoning and weighting up of evidence, could be more accurately answered by a superintelligence than by humans. The same holds for questions of policy and long-term planning; when it comes to understanding which policies would lead to which results, and which means would be most effective in attaining given aims, a superintelligence would outperform humans*[247]. [En la medida en que la ética es una búsqueda cognitiva, una superinteligencia podría hacerlo mejor que los pensadores humanos. Esto significa que las preguntas sobre ética, en la medida en que tienen respuestas correctas a las que se puede llegar mediante el razonamiento y la ponderación de las pruebas, podrían ser respondidas con mayor precisión por una superinteligencia que por los humanos. Lo mismo ocurre con las cuestiones de política y la planificación a largo plazo; cuando se trata de entender qué políticas conducirían a qué resultados, y qué medios serían los más eficaces para alcanzar determinados objetivos, una superinteligencia superaría a los humanos] (Traducción propia).

Un escenario con agentes superinteligentes, que incluso sean capaces de sentir, abriría las puertas a una Agencia Moral Artificial, que a su vez permita a los seres humanos exigirles que actúen conforme a los principios e ideas de lo que es bueno,

[246] BOSTROM, N., "Ethichal Issues in Advanced Artificial Intelligence", en S. Schneider, ed., *Science Fiction and Philosophy, From Time Travel to Superintelligence*, Reino Unido, Wiley-Blackwell, 2009, p. 374.
[247] *Ibidem*, p. 378.

pero también a que éstos últimos respeten posibles derechos con los cuales podrían contar los agentes superinteligentes. El abrir las puertas a una Agencia Moral Artificial desde esta perspectiva implicaría reflexiones muy serias desde la Ética de la IA, las cuales estén orientadas a asegurar el impacto benéfico de dichos seres en el planeta y en el universo.

Así, la justificación de por qué no es posible hablar en la actualidad de una Agencia Moral Artificial va robusteciéndose. En el mismo sentido, otro punto clave que impide adscribir en la actualidad una Agencia Moral Artificial es que al hacerlo la cadena de responsabilidad que vincula a la máquina con su creador, se rompería. En parte se abordó previamente esta idea en el subapartado que describió las tres conceptualizaciones realizadas por Deborah Johnson, respecto a los roles potencialmente cargados de moralidad que pueden llegar a desempeñar los sistemas artificiales autónomos, pero no está de más proporcionar más razones por las cuales es importante conservar esa cadena de responsabilidad:

> *There are very practical reasons for this. Without being able to establish a chain of responsibility, a society would be less free and secure, since the trajectories of agent's behaviors' would not be decipherable or predictable (and thus would not be subject to blame and accountability). But this is not the end of the story. We assign responsibilities, in order to learn something about ourselves as part of a society, that is to shay, to trace the distinction between what is acceptable (included) and what is not acceptable (excluded but somehow kept and preserved) in the society. The assignment of (moral or legal) responsibility is a human process that involves some degrees of inclusion while involving also some degrees of exclusion, not from, but within the society*[248]. [Hay razones muy prácticas para ello. Sin poder establecer una cadena de responsabilidad, una sociedad sería menos libre y segura, ya que las trayectorias de los comportamientos de los agentes no serían descifrables ni predecibles (y, por tanto, no estarían sujetas a la culpa y a la responsabilidad). Pero este

[248] DURANTE, M., *Ethics, Law and the Politics of Information, A Guide to the Philosophy of Luciano Floridi,* vol. 18, The International Library of Ethics, Law and Technology, Springer, Holanda, 2017, p. 63.

> no es el final de la historia. Asignamos responsabilidades para aprender algo sobre nosotros mismos como parte de una sociedad, es decir, para trazar la distinción entre lo que es aceptable (incluido) y lo que no es aceptable (excluido, pero de alguna manera mantenido y preservado) en la sociedad. La asignación de responsabilidades (morales o legales) es un proceso humano que implica algunos grados de inclusión y también algunos grados de exclusión, no de la sociedad, sino dentro de ella] (traducción propia).

Con esta idea no se pretende excluir de ninguna responsabilidad a la máquina, pues finalmente sus actos podrían impactar la esfera moral de las personas y de las sociedades, pero sí excluir cualquier posibilidad que en la actualidad apueste por responsabilizar más allá del campo moral, como bien lo puede ser el campo jurídico, a una máquina y no a la persona creadora o a su usuario, pues con ello estaríamos retrayéndonos a la época de la edad media en la cual se enjuiciaba a los animales no humanos por supuestamente cometer delitos.[249]

Así, este trabajo considera que los sistemas artificiales autónomos no pueden ser agentes morales en la actualidad. Hoy día, dichos sistemas se asemejan más a artefactos y sí cuentan con una agencia de eficacia causal, semejante a la descrita por Deborah Johnson. Sin embargo, este trabajo también resalta el valor de posturas como la de Luciano Floridi y Jeff Sanders, pues reflexiones así abren la puerta a debatir más seriamente sobre cuestiones como la moralidad distribuida a la que ellos hacen alusión, y que de cierta forma también se recoge como parte de la respuesta dada a la pregunta de este apartado, pues también se reconoce aquí que las máquinas son fuente del bien o del mal morales.

[249] Una buena introducción a este tema, y algunos casos en donde animales no humanos fueron juzgados y sentenciados por tribunales, se localiza en: ANAYA HUERTAS, A., "¿Animales ante los tribunales", en *Nexos*, el juego de la Suprema Corte, 22 de junio de 2010. [en línea], <https://eljuegodelacorte.nexos.com.mx/%C2%BFanimales-ante-los-tribunales/>, [consulta: 12 de septiembre de 2023].

7. REFLEXIONES FINALES

A lo largo de este capítulo se estudiaron las posturas que como resultado de este trabajo de investigación se consideran las referencias sobresalientes en el estado del arte de nuestra pregunta de investigación, es decir, los posicionamientos nucleares en torno a la idea de una posible Agencia Moral Artificial. A continuación, se enuncian de forma breve las posturas exploradas:

- Desde la teoría de Luciano Floridi, el agente moral artificial es posible, y es identificado como *"un sistema en transición, interactivo, autónomo y adaptativo que puede ejecutar acciones susceptibles de calificación moral".*[250]

- Al identificar los tipos de agencia que pudieran llegar a desempeñar los sistemas artificiales autónomos, se concluye que pueden presentar una de tipo de "*eficacia causal*" o una que actúe en nombre de la agencia humana, pero nunca de tipo autónoma, la cual sólo parece estar reservada a los seres humanos.

- Del debate entre el grupo de los "*Computational Modelers*" y el de "*Computers-in-Society*" se destaca la defensa que hace este último grupo del vínculo que debe prevalecer entre la tecnología con quienes la desarrollan, despliegan y usan, pues es en el ser humano en quien debe recaer la responsabilidad de los actos de los sistemas artificiales autónomos. Desde dicha postura, una Agencia Moral Artificial no es posible.

- La visión estándar de la agencia humana apuesta por adscribir Agencia Moral a entidades que cuenten con estados internos mentales. La visión funcionalista y de la cual es parte Luciano Floridi y Jeff Sanders, por su parte, rechaza la idea de la conciencia y adopta una "moralidad no mental", con lo cual se da pie a la posibilidad de una Agencia Moral Artificial.

Una vez identificados estos puntos, la conclusión a la cual se arriba es que no es posible adscribir una Agencia Moral Artificial a los sistemas artificiales autónomos, ya que en la actualidad ningún sistema con IA cuenta con condiciones internas (por ejemplo, la autonomía —en el sentido filosófico de la palabra y que se exploró en el capítulo correspondiente a la Agencia Moral—, la intencionalidad o la capacidad propia de los seres sintientes de experimentar dolor y placer) que les

[250] FLORIDI, L., "Ética de la Información: su naturaleza y alcance"*, op. cit*., *supra*, nota 153, p. 31.

permita tener una concepción de lo que es correcto hacer en determinados contextos sociales.

No se descarta, sin embargo, que en un futuro exista una superinteligencia que cubra dichas condiciones, y que con esto se abra la puerta a adscribirles una Agencia Moral Artificial, que a su vez permita a los seres humanos exigirles que actúen conforme a los principios e ideas de lo que es bueno (según la moralidad en la que se encuentren, como se señaló en apartados previos, estas implicaciones morales y éticas no serán resueltas en esta tesis), pero también que estos últimos respeten posibles derechos con los cuales podrían contar los agentes superinteligentes.

CAPÍTULO IV
DESAFÍOS SOCIO-JURÍDICOS PRESENTES Y FUTUROS DE LA INTELIGENCIA ARTIFICIAL

El desarrollo de la IA ha traído consigo efectos en distintos campos. A una parte de este capítulo le interesan aquellos que, siendo actuales, se relacionan con el campo jurídico y social. Dicho interés será clave para dos cosas: 1) dotar de vigencia al análisis socio-jurídico plasmado en este trabajo de investigación, y 2) mostrar desafíos actuales a los cuales es posible hacer frente desde el Derecho.

Por otro lado, la IA también abre la puerta a interrogantes sobre escenarios futuros en los que el Derecho deberá jugar un rol importante, por ejemplo, los debates con relación al transhumanismo y a los derechos de los robots. Esto también es de interés para el presente capítulo, pues además de servir como complemento de los desafíos actuales de la IA, sirve como puente para el desarrollo del último capítulo de este trabajo, en el cual se analizan soluciones que pueden ofrecerse al problema de una Agencia Moral Artificial desde la ética jurídica, y por tanto desde el Derecho.

El presente capítulo consta de dos apartados. El primero enfocado a exponer algunos de los desafíos actuales de la IA en el campo social y jurídico, mismo que a su vez se compone de cuatro subapartados, a saber: 1) Discriminación algorítmica; 2) La privacidad en la era de la IA; 3) El trabajo frente a la IA, y 4) La desinformación derivada de la IA. El segundo apartado busca exponer y analizar el futuro de la IA y los desafíos jurídicos que pudieran aparejarse a él, y el cual consta de dos subapartados: 1) Transhumanismo y 2) Derechos de los robots.

1. EL PRESENTE

1.1. Discriminación algorítmica

Los algoritmos llevan a cabo operaciones que les permiten diferenciar un dato en particular o un conjunto de datos, con la finalidad de llegar a un resultado. Se trata pues, de una labor de discriminación justificada por la propia naturaleza del algoritmo. Como tal, esto no implica un problema desde el mirador del Derecho. El problema, sin embargo, surge cuando se discrimina sobre categorías que éste último protege, por ejemplo, al discriminar a una persona por razón de sexo, raza, etnia, discapacidad, edad, orientación sexual o religión, puesto que esto resulta en una vulneración directa al principio de igualdad[251] y no discriminación establecido en el último párrafo del artículo 1º de la Constitución Política de los Estados Unidos Mexicanos.

Con el fin de mostrar a través de casos estas vulneraciones, se mencionan los siguientes ejemplos:

- La Unión de Libertades Civiles de Nueva York, a finales de 2018, reveló que en 2017 el Software de Evaluación de Clasificación de Riesgos utilizado desde el 2013 por el Servicio de Inmigración y Control de Aduanas de los Estados Unidos de América (EUA) para ayudar en la decisión de si, en procesos de deportación, un inmigrante tendría que ser detenido o tendría que ser puesto en libertad bajo fianza, había sido manipulado para favorecer las detenciones. [252]
- Un caso notorio de discriminación algorítmica es el de COMPAS, el cual es utilizado en algunas partes de los EUA, y tiene como objetivo predecir la probabilidad de que una persona acusada por la comisión de un delito pueda delinquir nuevamente. Como tal, COMPAS es una herramienta destinada a ayudar a las y los jueces en la decisión de si la persona acusada puede gozar

[251] XENEDIS, R. y L. Senden, “EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination”, en Ulf, B. et al (eds.), *General Principles of EU law and the EU Digital Order*, Kluwer Law International, 2020, p. 5.

[252] DE ASÍS, R., *Inteligencia Artificial y Derechos Humanos*, Materiales de Filosofía del Derecho, Universidad Carlos III de Madrid, No. 2020, p. 1.

de libertad condicional. Los datos de entrada de COMPAS, con los cuales se entrenan el o los algoritmos que utiliza, no contemplan el origen étnico ni el color de piel de las personas acusadas. Una investigación realizada en el 2016, demostró que COMPAS tiene preferencias en contra de las personas negras. COMPAS predice las reincidencias el 61 por ciento de las veces, pero en el caso de las personas negras, estas tienen el doble de posibilidades de ser consideradas como reincidentes que las personas blancas.[253]

- En el 2015, una persona negra se percató de que más de cincuenta fotos de él y su amigo, almacenadas en Google Fotos, habían sido etiquetadas por el algoritmo como "*gorilas*". Por supuesto, Google Fotos no había sido programado para arrojar este tipo de etiquetas, pero a esa fue la conclusión a la que llegó la IA.[254]

Es un hecho, hay un peligro latente de cara a los procesos de aprendizaje de las redes neuronales: el que existan datos sesgados capaces de discriminar injustificadamente a las personas, algo lógico si pensamos que los conjuntos de datos son recopilados y los modelos neuronales son construidos por seres humanos, los cuales actúan y piensan conforme a sus experiencias y subjetividades.

Se dice que los algoritmos no están libres de preferencias, un conjunto de datos utilizado para entrenar un algoritmo de aprendizaje automático puede reflejar los resultados de un proceso de toma de decisiones sesgado, mientras que un desarrollador de modelos y redes neuronales bien intencionado puede crear un algoritmo que perpetúe el sesgo existente en el conjunto de datos utilizado para entrenarlo.[255]

El que no estén libres de preferencias se explica si uno se remite al cómo se

[253] ZUIDERVEEN, F. J., "Strengthening legal protection against discrimination by algorithms and artificial intelligence", en *The International Journal of Human Rights*, Vol. 24, no. 10, 2020, p. 1574.

[254] ALLEN, R. y D. Masters, "Artificial Intelligence: the right to protection from discrimination caused by algorithms, machine learning and automated decisión-making", en *ERA Forum 20*, Springer, 2 de octubre de 2019, p. 593.

[255] SCHMIDT, N. y B. Stephens, "An Introduction to Artificial Intelligence And Solutions to the Problems of Algorithmic Discrimination", en *Quarterly Report*, Vol. 73, No. 2, 2019, p. 138.

desarrollan los algoritmos y el rol que en ese proceso juega la persona humana. Para construir un algoritmo, en primer lugar se necesita un conjunto de datos que después será utilizado para entrenar al algoritmo; en segundo lugar, se debe especificar el resultado concreto que busca predecirse a partir de ese conjunto de datos; en tercer lugar, se deben elegir las variables predictoras que estarán a disposición del algoritmo; en cuarto lugar, se debe construir el procedimiento por el cual el algoritmo encontrará la mejor variable predictora con la intención de que llegue al resultado que es de interés del desarrollador y, por último, permitir que el algoritmo entrene con el conjunto de datos y el procedimiento señalado en el punto anterior. Se dice que los dos primeros pasos implican decisiones humanas de importancia crítica que determinan lo que hace el algoritmo y, por lo tanto, el grado en el que este puede discriminar.[256]

Así pues, el algoritmo es capaz de reflejar las preferencias de quienes lo desarrollan. Y el asunto no es que lo hagan intencionalmente, sino que en ocasiones al modelo de aprendizaje automático o profundo no se le proveen los suficientes "datos de entrenamiento", que son aquellos con los cuales el algoritmo aprende e identifica patrones para mejorar sus predicciones, o bien porque los algoritmos son pobremente entrenados. A decir de estos puntos:

> *The way that training data are selected by algorithm developers can be susceptible to subconscious cultural biases, especially where population diversity is omitted from the data. The Royal Society noted that 'biases arising from social structures can be embedded in datasets at the point of collection, meaning that data can reflect these biases in society' […] Professor Nick Jennings from the Royal Academy of Engineering believe that algorithms are 'not always well trained because people do not always understand exactly how they work or what is involved in training'. Because the research in this area is still relatively undeveloped, he explained, "you end up with poorly trained algorithms giving biased results".*[257] [La manera

[256] KLEINBERG, J., J. Ludgwig, S. Mullainathan, y C.R. Sunstein, "Discrimination in the age of algorithms", en *Journal of Legal Analysis*, Vol. 10, 2018, p.134

[257] ALLEN, R. y D. Masters, "Artificial Intelligence: the right to protection from …", *op. cit., supra*, nota 252, p. 589.

> en la que los desarrolladores de algoritmos seleccionan los datos de entrenamiento puede ser susceptible de sesgos culturales subconscientes, especialmente cuando se omite la diversidad de la población en los datos. La Royal Society señaló que "los prejuicios derivados de las estructuras sociales pueden incorporarse a los conjuntos de datos en el momento de su recopilación, lo que significa que los datos pueden reflejar estos prejuicios en la sociedad" […] El profesor Nick Jennings, de la Real Academia de Ingeniería, cree que los algoritmos "no siempre están bien entrenados porque la gente no siempre entiende exactamente cómo funcionan o qué implica el entrenamiento". Dado que la investigación en este campo es aun relativamente poco desarrollada, explicó, "se termina con algoritmos pobremente entrenados que dan resultados sesgados".] (Traducción propia).

El resultado de lo dicho es que el algoritmo produzca resultados que discriminen injustamente. Por ejemplo, sistemas predictivos de abuso infantil que asocian pobreza con abusos, sistemas predictivos de delitos que asocian delincuencia a racismo policial histórico, sistemas predictivos que detectan erróneamente fraudes de beneficios sociales a personas honestas de escasos recursos, entre otros casos.[258]

Y es que la discriminación algorítmica tira a la borda la creencia de que los algoritmos son objetivos y neutrales, pues como se ha visto, al existir intervención humana en su desarrollo, no están libres de sesgos y prejuicios. Existe un grado de discriminación que es justificable por parte del algoritmo, en tanto sea capaz de predecir algo y que ese resultado sea explicable. Sin embargo, existe un grado de discriminación no justificable, en tanto que el resultado puede ser perjudicial para grupos históricamente discriminados.

En su aplicación, los algoritmos incorporan visiones y valores de las personas que los desarrollan. Pensemos, por ejemplo, en los sesgos que podrían tener los algoritmos desarrollados en "*Sillicon Valley"*, en donde prevalece la visión

[258] MUÑOZ GUTIÉRREZ, C., "La discriminación en una sociedad automatizada: Contribuciones desde América Latina", en *Revista Chilena de Derecho y Tecnología*, vol. 10, no. 1, 2021, p. 273.

masculina, de gente blanca y del norte industrializado. Basta cuestionarse por qué a la fecha no ha habido ninguna mujer gurú ahí, o por qué ningún director ejecutivo de las grandes empresas tecnológicas es negro.

A decir de esta falta de neutralidad algorítmica:

> *The lack of neutrality of the algorithms is due to their dependency on big data, the databases are not neutral and present the prejudices inherent in the hardware with which they have been collected, the purpose for which they have been collected, the purpose for which they have been compiled and the unequal data landscape […]. The application of algorithms trained with these data can disseminate the prejudices present in our culture like a virus, giving rise to vicious circles and the marginalization of sector of society*. [259] [Las bases de datos no son neutrales y presentan prejuicios inherentes en el hardware con el cual ellas han sido recolectadas, el propósito para el cual han sido compiladas y el inequitativo paisaje de datos. La aplicación de algoritmos entrenados con estos datos puede diseminar los prejuicios presentes en nuestra cultura como un virus, dando lugar a círculos viciosos y la marginalización de sectores de la sociedad] (Traducción propia).

Regresando al caso de "*Sillicon Valley"*, hay constancia de que en ese lugar falta la diversidad y abunda la misoginia. Un caso relevante fue el llamado *Donglegate*, reportado por la revista *Wired* y ocurrido en el 2013 en la *Python Conference* de California. En esa ocasión Adria Richards, ex desarrolladora y ex empleada de una compañía de tecnología, experimentó de manera directa el cómo un par de hombres que se encontraban detrás de ella, hacían algunos juegos de palabras sexuales, sin importar su presencia. [260]

Siguiendo con el hilo, en "*Sillicon Valley"* la meritocracia también es un mito, pues existen barreras que impiden la presencia de mujeres y personas de color y

[259] ÁLVARO, S., Living with Smart Algorithms, en *CCBLAB*, 3 de mayo de 2017, [en línea], <https://lab.cccb.org/en/living-with-smart-algorithms/>, [consulta: 18 de septiembre de 2023].

[260] MARWICK, A., Donglegate: Why the Tech Community Hates Feminists, en *Wired,* 3 de abril de 2013, [en línea], <https://www.wired.com/2013/03/richards-affair-and-misogyny-in-tech/>, [consulta:18 de septiembre de 2023].

latinas en puestos vinculados con las tecnologías de la información.[261] De casos como éste se deriva también el riego de que los algoritmos terminen reflejando preferencias de quienes los desarrollan y de sus prejuicios políticos, religiosos o étnicos.

La respuesta a este problema actual de la IA viene de distintos frentes. Desde el ámbito técnico, la vía para evitar sesgos en los algoritmos radica en que éstos cuenten con datos lo suficientemente diversos para que puedan aprender y hacer predicciones que tomen en consideración a grupos históricamente discriminados. A esto, también debe sumársele el enfoque inclusivo con el cual deben conformarse los grupos de desarrolladores de algoritmos, la idea de esto es que los datos que se utilicen sean lo más diversos posibles.

El otro frente es el regulatorio, que implica evaluar y fortalecer los actuales marcos legales que existen en materia de igualdad y no discriminación. Esta tarea, sin embargo, tiene su complejidad, pues las normas de igualdad y no discriminación:

> [...] Se han construido bajo complejas e históricas tensiones políticas y sociales, existiendo, al día de hoy, diversas teorías e interpretaciones. Los conceptos de igualdad y discriminación tienen distintos contenidos y alcances en cada sistema jurídico, poseyendo, además, particularidades propias en cada país, siendo en definitiva conceptos elaborados en virtud de su contexto social. [262]

Así pues, la discriminación algorítmica es un tema vigente, el cual implica un desafío para la igualdad y la no discriminación, y que a la fecha es objeto de estudio, no sólo desde el derecho, sino también desde la informática y la ética.[263]

1.2. La privacidad en la era de la Inteligencia Artificial

El algoritmo es también la herramienta que posibilita la transformación de los datos,

---

[261] *Idem.*

[262] MUÑOZ GUTIÉRREZ, C., "La discriminación en una sociedad automatizada: Contribuciones desde América Latina", *op. cit*., *supra*, nota 258, p. 274.

[263] XENEDIS, R., "Tuning EU equality law to algorithmic discrimination: Three pathways to resilience, en *Maastricht Journal of European and Comparative Law*, Vol. 26, 2020, p. 737.

con ella se extraen características que, desde el campo de la privacidad, terminan por desentrañar a las personas. De este primer aspecto, el que la privacidad sea una preocupación creciente en el mundo de la IA.

Pero ¿cómo se explica la relación entre la IA y la privacidad, en particular con la protección de datos personales? Bien, por principio de cuentas ha de señalarse que la IA utiliza una cantidad importante de datos para funcionar, como señalan Christopher Kuner y Fred Cate, en el campo de la IA: "*With few expections, more data is better than less, and there is almost never enough*" [Con pocas expectativas, más datos es mejor que menos, y casi nunca hay suficientes] (Traducción propia).[264] Asimismo, dichos autores, nos dicen que la IA requiere de los datos para alcanzar su potencial y para evitar la monopolización de dicha tecnología, pero también para evitar sesgos,[265] como los ya referidos en el apartado anterior.

Hablando de datos, cabe referir que con la irrupción de Internet se dio pie a que las personas usuarias de dicha tecnología generen una ingente cantidad de datos en el espacio digital susceptible de explotación:

> Toda la información que generamos en el espacio digital, desde el correo electrónico al WhatsApp, desde las transacciones virtuales a las redes sociales, adquiere valor porque existen herramientas analíticas que hacen posible "ubicarla" sin límite. Y no se trata de un bien valioso que cada uno almacena debajo de su colchón, sino de una especie de moneda encriptada que puede alimentar dos grandes "negocios" biopolíticos de titularidad ajena: uno público, la seguridad, y otro privado, el comercio de macrodatos personales.[266]

Una explicación del por qué la IA ha ganado impulso en los últimos años es precisamente por las cantidades gigantes de datos a las que tiene acceso. Existe un término en específico para esas cantidades grandes de datos, y es el de "*Big*

---

[264] KUNER, C., F. H. Cate, *et al*., "Expanding the artificial intelligence-data protection debate", en *International Data Privacy Law*, Vol. 8, No. 4, 2018, p. 289.

[265] *Idem.*

[266] CAMPIONE, R., "Recopilar y vigilar: algunas consideraciones filosófico-jurídicas sobre inteligencia artificial", en *Sociología y Tecnociencia*, no. 11, 2021, p. 128.

Data", los cuales no pueden ser analizados por los métodos tradicionales de almacenamiento. Para lograr su explotación se requiere de la implementación de algoritmos y estadísticas mediante las cuales se pueden obtener resultados.[267]

Y es en la explotación del *"Big Data"* en los que es posible desentrañar a la persona. Entre los resultados que pueden obtenerse están:

> […] el comportamiento de las personas, sus gustos, la toma de decisiones, el reconocimiento de la voz, la identificación de objetos, el diagnóstico de enfermedades, el ahorro de energía, la elaboración de perfiles para analizar o predecir aspectos relativos al rendimiento profesional, la situación económica, la salud, los intereses, la fiabilidad, el comportamiento o la ubicación, datos que se utilizan con diversos fines y se encuentran al alcance de empresas.[268]

El *"Big Data"* y su explotación por algoritmos encuentran una simbiosis en casi todos los aspectos de la vida humana. El del marketing es posiblemente el lugar idóneo de encuentro, pues es un campo fértil para probar y corregir las nuevas herramientas matemáticas y tecnológicas, debido a que tiene pocos riesgos y muchos beneficios.[269] Pero no así en otros campos, el bancario, el asegurador y el sanitario son casos en los que genera preocupaciones:

> En éstos, surgen serias dudas sobre cuándo es realmente necesaria la intervención humana para supervisar los resultados alcanzados por algoritmos. Una de las tendencias es hacer que el humano siga formando parte del proceso de toma de decisiones; pero otros opinan que esto sería contraproducente. Así por ejemplo, en el sector bancario numerosas empresas están acudiendo al análisis de los macrodatos, siguiendo el principio básico de la banca de 'conozca a su cliente'. El análisis de datos permite conocer a los prestatarios mejor que nunca y predecir si

[267] MARTÍNEZ DEVIA, A., "La Inteligencia Artificial, el Big Data y la era digital: ¿una amenaza para los datos personales", en *Revista la Propiedad Inmaterial*, no. 27, 2019, p. 8.
[268] *Idem.*
[269] GIL, E., *Big Data, Privacidad y Protección de Datos*, Agencia Española de Protección de Datos, 2016, p. 43.

> devolverán sus préstamos de forma más certera que si únicamente se estudia su historial crediticio. Este sistema depende de algoritmos que analizan de forma compleja y automatizada […].[270]

En efecto, el *"Big Data"* y los algoritmos contribuyen a la pérdida gradual de la privacidad, algo que es característico de nuestra era. Además, dice Byung Chul-Han, que el *"Big Data"* da lugar a una sociedad de clases digital, algo estrechamente vinculado con la discriminación que de la cual se habló en el apartado anterior, en donde los individuos se agrupan en categorías, las cuales a su vez se ofertan como mercancías:

> Aquellos con un valor económico escaso se les denomina *waste* […]. Los consumidores con un valor de mercado superior se encuentran en el grupo *Shooting Star*. Son dinámicos, de 36 a 45 años, se levantan temprano para hacer *footing,* no tienen hijos, están casados, les gusta viajar y la serie de televisión *Seinfeld*.[271]

Frente al "*Big Data*" hay conocimiento absoluto, en donde todo es mensurable y cuantificable:

> Las cosas delatan sus correlaciones secretas que hasta ahora habían permanecido ocultas. Igual de predecible debe ser el comportamiento humano. Se anuncia una nueva era del conocimiento. Las correlaciones sustituyen a las causalidades. El ello es así sustituye al por qué. La cuantificación de lo real en búsqueda de datos expulsa al espíritu del conocimiento.[272]

El espíritu, comprendido como una totalidad en la que las partes son integradas con sentido, se atrofia frente al "*Big Data*": *"El conocimiento total de datos es un*

[270] *Idem*.
[271] HAN, B., *La Sociedad de la Transparencia*, España, Herder, 2013, p. 99.
[272] *Ibidem*, p. 102.

desconocimiento absoluto en el grado cero del espíritu".[273] A lo anterior ha de sumarse que el "*Big Data*" es ciego al acontecimiento, "*no a lo estadísticamente probable, sino a lo improbable, lo singular, el acontecimiento determinará la historia, el futuro humano. Así pues, el Big Data es ciego ante el futuro*". [274]

El "*Big Data*" y su explotación por el algoritmo desentraña a la persona y tiene el potencial de reducir a la condición de cosa a la persona. Como indica Yuval Noah Harari, en relación con el avance inmenso en inteligencia informática: "*los humanos corren peligro de perder su valor porque la inteligencia se está desconectando de la conciencia*". [275] Hasta este punto la explicación de la relación entre IA y la privacidad.

Es en dicha relación de donde también emanan las preocupaciones desde el campo de lo jurídico. Por principio de cuentas, la IA se hace presente en el Derecho al hablar de privacidad, especialmente cuando se analiza el desafío que dicha tecnología supone para las normas de protección de datos personales.

Muchas de las leyes en materia de datos personales alrededor del mundo reflejan los principios establecidos en 1980 por la OCDE, en sus Directrices sobre protección de la privacidad y flujos transfronterizos de datos personales[276] (principio de limitación de recogida, de la calidad de los datos, de especificación del propósito, de limitación de uso, de salvaguardia de la seguridad, de transparencia, de participación individual y de responsabilidad). El reto, nos dice Christopher Kuner y Fred Cate, es el ajustar dichos principios al contexto de la IA cuando los datos pueden potencialmente alcanzar resultados imprevistos por medio de algoritmos avanzados de los que no siempre se comprende su resultado por su programador, o de los cuales incluso no sean creados por humanos, sino por computadoras.[277] A esto ha de sumarse lo siguiente:

Implicit in the OECD Guidelines, and made explicit in the EU's General

[273] *Ibidem*, p. 105.
[274] *Ibidem*, p. 113.
[275] HARARI, Y. N., "*Homo Deus, Breve historia del mañana*", Debate, 2016, p. 341.
[276] KUNER, C., F. H. Cate, *et al*., "Expanding the artificial intelligence-data protection debate", en *International Data Privacy Law*, *op. cit.*, *supra*, nota 262, p. 290.
[277] *Idem.*

> *Data Protection Regulation (GDPR) and other modern data protection laws, in another widely shared principle: data minimization, ie that Personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which those data are processed'. However, with data, in the words of the Singapore Personal Data Commission, constituting the 'basic building block of the digital economy', the concept of data minimization stands in tension with developing AI technologies*.[278] [Implícito en las Directrices de la OCDE, y señalado también en el Reglamento General de Protección de Datos (RGPD) de la UE y en otras leyes modernas de protección de datos, en otro principio ampliamente compartido: la minimización de datos, es decir, que los datos personales sean adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que se procesan esos datos. Sin embargo, dado que los datos, en palabras de la Comisión de Datos Personales de Singapur, constituyen el bloque básico de la economía digital, el concepto de minimización de datos entra en tensión con el desarrollo de las tecnologías de lA] (Traducción propia).

Es en esa tensión que se vuelve crucial que cualquier esfuerzo por proteger la privacidad encuentre un equilibrio con el desarrollo de la IA:

> We need to consider to what extent that transformation should extend to data protection law itself and the tools that give it meaning. Otherwise, we run the risk of leaving data privacy inadequately protected -or the benefits of AI inadequately developed- in our rapidly transforming world.[279] [Debemos considerar hasta qué punto esa transformación debe extenderse a la propia ley de protección de datos y a las herramientas que le dan sentido. De lo contrario, corremos el riesgo de dejar la privacidad de los datos inadecuadamente protegida –o los beneficios de la IA inadecuadamente desarrollados– en nuestro mundo en rápida transformación] (traducción propia).

[278] *Idem.*
[279] *Ibidem*, p. 292.

Algunos de los puntos en los que se hace patente la protección de los datos personales, con relación al desarrollo de la IA son los siguientes:

I) Con relación a la obtención de datos personales: en particular pensando a la IA en el contexto del Internet de las Cosas (el proceso que permite a elementos físicos cotidianos al Internet como puede ser un foco inteligente, hasta dispositivos médicos, abarcando también prendas y accesorios personales inteligentes e incluso los sistemas de las ciudades inteligentes),[280] y en donde la regla es emplear los datos generados por la persona para ofrecer servicios que pueden contribuir al bienestar de la persona. Sin embargo, también existe el riesgo de que la información personal que se obtenga sea tratada para fines distintos y de los cuales desconoce el titular de los datos personales.

II) Con relación a la explotación de los datos y la generación de conocimiento a través de ellos: pensando en las herramientas analíticas utilizadas para extraer valor de los datos, y el poder con el que estas cuentan para perfilar a la persona, extrayendo de ésta su comportamiento, por ejemplo. Un caso ocurre con las páginas que emplean "*cookies*" (las cuales permiten que las páginas web puedan identificar la computadora por la cual navegamos en Internet, y consultar la actividad previa del navegador utilizado, y que son clave para la publicidad en internet, especialmente en lo que respecta a los avisos publicitarios personalizados),[281] y mediante las cuales es posible identificar las preferencias personales que las personas consumidoras tienen. En ocasiones, los sitios web que emplean las "cookies" lo hacen sin contar con una política de privacidad clara que tenga en cuenta el consentimiento de la persona, obteniendo de manera abusiva los datos personales de los usuarios.

En fin, puede afirmarse que es todo un reto actual el encontrar el equilibrio entre la privacidad y el desarrollo de la IA. De hecho, se afirma que la privacidad es un

---

[280] RED HAT, *¿Qué es el Internet de las Cosas (IoT)?*, 8 de enero de 2019, [en línea], <https://www.redhat.com/es/topics/internet-of-things/what-is-iot>, [consulta: 18 de septiembre de 2023].

[281] BBC NEWS*, Qué ocurre cuando aceptas las cookies y por qué es conveniente borrarlas del navegador de vez en cuando*, 29 de junio de 2017, en línea], <https://www.bbc.com/mundo/noticias-40443519>, [consulta: 18 de septiembre de 2023].

ejemplo paradigmático de la desconexión entre validez y eficacia de la ley, algo que es causa de la creciente aceleración de las aplicaciones vinculadas a la IA.[282]

1.3. El trabajo frente a la Inteligencia Artificial

Se dice que, por primera vez en la historia, las máquinas realizan tareas que anteriormente eran exclusivas del cerebro humano, y que está tendencia irá al alza con el desarrollo de las tecnologías de la información.[283] Esto, como es de suponerse, ha encendido las inquietudes de distintos sectores de la sociedad, especialmente del obrero, pues se infiere que, con el desarrollo de la IA, esta tecnología los suplirá.

En la historia, la irrupción de la tecnología ha terminado con industrias y sectores de empleo. Por ejemplo, en los Estados Unidos, la mecanización de la agricultura acabó con millones de trabajos, y empujó la migración de trabajadores, de los campos agrícolas a las fábricas.[284] Otro caso fue el de la Primera Revolución Industrial, iniciada alrededor de 1770, la cual trajo consigo grandes impactos de cara a los empleos que en ese momento histórico existían. A decir de esto último, un buen caso es el de la máquina de tejer que, al momento de ser incorporada por la industria textil, generó disturbios en Gran Bretaña y otros países de Europa. Incluso, dos siglos antes, en 1589, el temor de que este invento trajera consigo efectos perjudiciales para trabajadores, fue la causa de que la Reina Isabel I rechazara su patente, argumentando que ese invento sería la ruina de sus pobres súbditos al privarlos de empleo y convertirlos en mendigos.[285]

Así pues, no es sorpresa que en esta era, las alarmas de que la IA transforme radicalmente al mundo laboral se hayan encendido. El argumento parece el mismo que en otras etapas históricas se ha utilizado de cara a la irrupción tecnológica: se perderán millones de empleos, provocando la desgracia económica de los trabajadores que serán suplidos por la tecnología. Este argumento, sin embargo,

---

[282] CAMPIONE, R., "Recopilar y vigilar: algunas consideraciones filosófico-jurídicas sobre inteligencia artificial", en *Sociología y Tecnociencia*, no. 11, 2021, p. 128.

[283] FORD, M., "Could Artificial Intelligence Create an Unemployment Crisis?", en *Communications of the ACM*, vol. 56, no. 7, julio 2013, p. 38.

[284] *Idem*.

[285] MANZANO, F., "¿Avance tecnológico o desempleo?", *Reporte Técnico*, 2019, p. 2.

también tiene sus detractores, aquellos que dicen que no es posible afirmar tal cosa, debido a que en la actualidad no se tiene ni siquiera claridad respecto al concepto de IA, Automatización y Robótica, por ejemplo.

En este breve apartado, se busca identificar tres cosas: 1) el impacto presente y futuro de la IA en el mundo laboral; 2) algunos contraargumentos a la idea de que la IA suplirá al humano en el ámbito laboral y 3) las posibles soluciones que, desde el Derecho, pero también desde otras áreas, se vislumbran para hacer frente a un probable escenario en el que la IA arrase con sectores de empleo.

1.3.1. La Inteligencia Artificial y su impacto en el mundo laboral

De acuerdo con una proyección hecha por McKinsey & Company se estima que, con la irrupción de la IA, el desempleo tecnológico llegué a los 800 millones de trabajadores para 2030. En Brasil, por ejemplo, se estima que esta tendencia impacte en el 15% de sus empleos.[286] Esto puede correlacionarse con otro dato, el cual sirve para comprender por qué países como Estados Unidos y China apuestan por el desarrollo de la IA, y que indica que dicha tecnología permitirá la creación de una riqueza sin precedentes en la historia:

> PricewaterhouseCoopers (PwC) calcula que su adopción generalizada aumentará en alrededor de 15.7 billones de dólares el PIB mundial en 2030, es decir, un poco más de diez años. Este aumento continuará su trayectoria exponencial hasta 2050. La incorporación de la IA puede generar enormes beneficios empresariales, pero la creación de la riqueza no será uniforme. Como he indicado en mi libro "IA Superpowers: China, Sillicon Valley, and the New World Order", las ganancias generadas por las primeras innovaciones en el sector de la IA presentan un escenario prácticamente monopolístico, puesto que dos colosos económicos - Estados Unidos y China- ya están sentando la pauta al albergar a todos los grandes gigantes corporativos del sector. Según los pronósticos de

[286] VARELA DE ALBURQUERQUE DALPRÁ, J., "La protección del trabajo digno mediante los impactos de las nuevas formas de robótica laboral, inteligencia artificial y nuevas tecnologías", en *Revista Internacional y Comparada de Relaciones Laborales y Derecho del Empleo*, vol. 8, no. 3, 2020, p. 214.

> PwC ya mencionados, se cree que el crecimiento más relevante se producirá en China, en gran medida por su enorme población, que representa casi un quinto del total mundial. [287]

En efecto, es válido pensar que la gran apuesta en el sector empresarial será el desarrollo de la IA, con efectos previsibles, como el hecho de que muchas de las soluciones desarrolladas impacten directamente al mundo laboral. En la actualidad, es posible identificar casos en los que la IA impacta directamente en el trabajo. Por ejemplo, existen algoritmos que miden con precisión la productividad humana:

> En la industria, empresas como Coca-Cola, Johnson & Johnson y Shell utilizan Ignition, un software de supervisión, control y adquisición de datos tipo SCADA, que permite gestionar procesos industriales de forma remota "cerrando la brecha entre producción y tecnologías de la información". Estas plataformas administran los procesos industriales de punta a punta, y por supuesto, incluyen la posibilidad de calendar cada microtarea que deben desempeñar los operarios, quienes a la vez están integrados a brazos mecánicos conectados a Internet, rampas que deslizan al ritmo de la inteligencia artificial o motores que emiten la exacta cantidad de gas que les prescribe el algoritmo. [288]

Otro caso es el del reclutamiento digital, el cual es simple y menos costoso que los métodos comunes. Sobre este punto también se abordan cuestiones que preocupan con relación a los derechos fundamentales de los postulantes:

> Por ejemplo, podría la computadora detectar que, usualmente, los postulantes que tienen hijos no marcan el casillero del formulario que indica "el trabajador está dispuesto a ser trasladado al extranjero". En esa línea, podría concluir que, para puestos de trabajo en los que se requiera

[287] LEE, K., "La inteligencia artificial y el futuro del trabajo: una perspectiva china", en *El Trabajo en la Era de los Datos*, OpenMind BBVA, 2019, p. 4.
[288] OTTAVIANO, J. M., "La amenaza fantasma: Inteligencia Artificial y derechos laborales", en *Nueva Sociedad*, no. 294, 2021, p. 90.

trasladados al extranjero, ningún postulante con hijos podría acceder. [289]

Una idea que ha ganado impulso en los últimos años es que la automatización terminará por desplazar a los trabajadores de sus puestos. Y es en este punto en donde se allana el camino para pensar en el futuro. Se dice pues, que la automatización permite que muchas ocupaciones o puestos de trabajo puedan dejar de realizarse por el trabajo humano y pasen a ser realizados por las nuevas tecnologías:

> La automatización reemplazará a quienes desempeñan tareas rutinarias, repetitivas, de baja complejidad o simples (en movimientos y/o cálculos), estandarizadas o fácilmente estandarizables (que puedan ser traducidas a algoritmos informáticos programados en aplicaciones de software y en acciones de robots con sus 'actuadores') mediante mecanismos que transfieren las tareas realizadas por el ser humano a los nuevos dispositivos tecnológicos. [290]

La postura de que la automatización acabará con el trabajo humano está vinculada con el pesimismo o fatalismo tecnológico. Dicha postura también apunta a que la Cuarta Revolución Industrial no vendrá acompañada de la creación de múltiples sectores productivos y de servicios que remplazarán las ocupaciones destruidas -algo que sí aconteció en la Primera Revolución Industrial-. Se dice pues, que:

> […] los empleos descualificados basados en tareas simples y rutinarias serán realizados por la inteligencia artificial digital y robotizada y sólo se crearán, mucha menor medida, empleos de alta cualificación, basados en la adquisición educativa de competencia y conocimientos matemáticos e

[289] TOYAMA MIYAGUSUKU, J. y A. Rodríguez León, "Algoritmos labores: big data e inteligencia artificial", en *Themis-Revista de Derecho*, no. 75, 2019, p. 258.

[290] LAHERA SÁNCHEZ, A., "El debate sobre la digitalización y la robotización del trabajo (humano) del futuro: automatización de sustitución, pragmatismo tecnológico, automatización de integración y heteromatización", en *Revista Española de Sociología*, no. 30, 2021, p. 3.

> informáticos, que no estarán al alcance de la mayoría de la ciudadanía (competencias STEM: credenciales educativas en ciencia, tecnología, ingeniería, matemáticas…).[291]

La postura pesimista también se alimenta de la siguiente idea, la cual es parte del informe periodístico sobre el "*Brain Trust Global*", organizado por Mijaíl Gorbachov en noviembre de 1995, y en donde se señala que en el próximo siglo (este siglo) el 20% de la población activa bastará para mantener en marcha la economía global.[292] Lo preocupante de esta idea es que como señala Susan George, para que sobreviva ese 20% de la población basta con bajar la producción para autosustentarse y dejar que el otro 80% perezca mediante sutiles técnicas de pandemias y guerras regionales.[293]

Frente al pesimismo tecnológico existen posturas moderadas, que permiten hacer análisis más críticos del impacto que la IA tiene en el mundo laboral.

1.3.2. Falta tiempo para la sustitución (o destrucción) del empleo provocada por la Inteligencia Artificial

Parte del contraargumento para decir que la sustitución del empleo por la IA está ya a la vuelta de la esquina, tiene como base un hecho: las innovaciones en materia de IA están todavía en un nivel de desarrollo que las acerca más a ser prototipos que artefactos tecnológicos directamente aplicables y rentables a corto plazo.[294]

---

[291] *Ibidem*, p. 4. Sólo como ejemplo, algunos ámbitos susceptibles de ser robotizados, en el caso del sector público, son los siguientes: transportes, empleados de correos: carteros y manipuladores; trabajos de carácter administrativo y auxiliar; trabajos de carácter burocrático de elevado nivel encargados de tareas de tramitación de expedientes, gestión económica y de personal; cuerpos de seguridad; personal penitenciario; fuerzas armadas; trabajadores sanitarios, y trabajadores en los servicios sociales. En el mismo tenor, las nuevas profesiones derivadas de esta revolución son los siguientes: analistas y programadores de Internet de las cosas; arquitecto de nuevas realidades; profesionales de big data; diseñador de órganos; profesionales de la robótica; diseñador de redes neuronales; terapeuta de empatía artificial; impresor 3-D; protésico robóticas; ingeniero de nanobots médicos; abogados especializados en drones y ciberseguridad. RAMIÓ MATAS, C., "El impacto de la inteligencia artificial y la robótica en el empleo público", en *GIGAPP Estudios Working Papers*, no. 98, 2018, pp. 406-409.

[292] PARODI, G. F., "La renta básica ciudadana, opción frente al desempleo tecnológico y la corrupción", en *ACADEMO Revista de Investigación en Ciencias Sociales y Humanidades*, vol. 2, no. 1, 2012, p. 3.

[293] *Idem.*

[294] LAHERA SÁNCHEZ, A., "El debate sobre la digitalización y la robotización del trabajo (humano)

Con esto, hay razones para pensar que la sustitución del trabajo humano será más lenta de lo pronosticado y que, por tal motivo, el efecto global pueda ser menor que el anticipado por el enfoque de la automatización de la sustitución [295]: "*No hay duda de que los robots del futuro serán capaces de ejecutar tareas extraordinarias, pero actualmente, sus aplicaciones comerciales han sido limitadas. El futuro está todavía probablemente algo lejos en el tiempo y llegará gradualmente*".[296]

A esta idea, debemos sumarle otra, el hecho de las imprecisiones conceptuales que impiden predecir el impacto de la IA en el sector del trabajo humano. Por ejemplo, no es lo mismo decir IA que robots industriales o digitalización:

> Esto deriva en cierta dificultad para poder articular y comparar los análisis realizados, en tanto que la bibliografía no siempre se especifica bien a cuál de estos fenómenos se estaría haciendo referencia. Más allá de esta imprecisión conceptual, de todos modos, se puede sostener que los pronósticos más catastróficos respecto del desempleo tecnológico masivo son infundados, tanto por las limitaciones metodológicas en los estudios que sostienen estas tesis como por los condicionamientos que todavía evidencia la modernización digital. [297]

En efecto, asumir que un gran cambio está por ocurrir en el sector laboral producto de la IA resulta ser mera especulación, debido a lo difícil que se vuelve comparar categorías (IA, Robótica, Automatización) cuando éstas son utilizadas indistintamente al momento de analizar un fenómeno como el de la sustitución laboral. Esto, sin embargo, no quiere decir que se niegue los efectos de las transformaciones tecnológicas, especialmente de la IA, en el mercado laboral: "*El mayor impacto seguramente se verá más en la estructura del empleo, estancando*

---

del futuro: automatización de sustitución, pragmatismo tecnológico, automatización de integración y heteromatización", en *Revista Española de Sociología*, no. 30, 2021, p. 6

[295] *Idem.*

[296] EUROFOUND, *The future of manufacturing in Europe*, Luxemburgo: Publications Office of the European Union, 2019, p. 42 *apud* LAHERA SÁNCHEZ, A., "El debate sobre la digitalización y la robotización del trabajo (humano) del futuro…", *op. cit*., *supra*, nota 294, p. 6.

[297] NAVA, A. y F. Naspleda, "Inteligencia artificial, automatización, restructuración capitalista y futuro del trabajo: un estado de cuestión", en *CEC*, no. 12, 2020, p. 110.

los salarios, polarizando el mercado laboral e implementando un mayor control en el proceso productivo que generando un fenómeno de desempleo masivo".[298]

Así pues, la idea de que la sustitución o destrucción laboral está próxima a ocurrir, parece no tener la suficiente fuerza desde nuestro presente. De nueva cuenta, no se trata de negar los cambios producidos por la tecnología en el campo laboral. Existe evidencia de cómo la globalización y la robótica en particular, han facilitado la telemigración laboral, la cual consiste en que compañías de países ricos contraten trabajadores de países de salarios bajos para que realicen tareas específicas en plataformas en línea.[299] En el mismo tenor, también existen aplicaciones de IA que se asoman como una forma de suplir a periodistas deportivos o de negocio, o de software que es capaz de hacer labores de abogados, por ejemplo, revisando contratos y demás documentos necesarios para armar casos judiciales.[300]

Sin embargo, también puede decirse que en la actualidad la IA está lejos de replicar a la persona humana en aspectos que son muy propios de la mente humana, por ejemplo, trabajos en los que se requiere una alta dosis de empatía, creatividad, inteligencia social y negociación.[301] Entonces, el inferir que en el corto o mediano plazo existirá una destrucción total del empleo humano parece muy aventurado, en razón de que en el presente la IA sólo es capaz de hacer una única actividad a la vez, por ejemplo, si una red neuronal fue programada para detectar expresiones faciales, sólo será capaz de hacer eso, y nunca de detectar imágenes de gatos o perros.

Ahora bien, también es necesario en pensar en las soluciones que se han puesto sobre la mesa, de cara a una probable sustitución o destrucción del empleo por parte de la IA.

---

[298] *Idem.*

[299] HUMPRIES, J. y B. Schneider, "El trabajo en el siglo XXI", en *El Trimestre Económico*, vol. 87, no. 346, 2020, p. 555.

[300] FORD, M., "Could Artificial Intelligence Create an Unemployment Crisis?", *op. cit.*, *supra*, nota 281, p. 38.

[301] RAMIÓ MATAS, C., "El impacto de la inteligencia artificial y la robótica en el empleo público", en *GIGAPP Estudios Working Papers*, no. 98, 2018, p. 406.

1.3.3. Soluciones desde el Derecho para la sustitución del empleo humano provocado por la Inteligencia Artificial

¿Qué hay entonces para la gente que pueda ser sustituida de sus trabajos por la IA? Esta pregunta se vuelve crucial suponiendo que la sustitución o la destrucción del empleo por parte de la IA ocurrirá. Una posible respuesta va de la mano con las políticas públicas, específicamente para cartografiar qué empleos, ocupaciones, tareas, sectores, perfiles profesionales y personas se estarán convirtiendo en los menos favorecidos de cara la digitalización, la robotización y la IA.[302]

Otra vía de solución, que también deviene de políticas públicas, es la de la llamada renta básica o salario ciudadano, una propuesta que tiene las siguientes características: 1) el pago se hace en efectivo y de manera regular; 2) lo paga el Estado; 3) es para todos los ciudadanos y ciudadanas de un país determinado, y 4) es incondicional, es decir, que no está atada a determinados comportamientos, actividades o actitudes.[303] La Renta Básica Universal, como también se le conoce, se presentaría entonces como una solución viable para el problema del desempleo causado por la IA:

> Se trata de una solución ética que no implica ningún cambio ideológico profundo, facilitando también la transparencia de los Planes Sociales de los gobiernos. Desde el punto de vista de las personas, cumple con el deber Republicano de libertad, ya que toda persona será libre de elegir su vida sin necesidad de vivir del permiso de los demás.[304]

Ahora bien, desde el ámbito jurídico se han propuesto soluciones a la problemática del desempleo causado por la IA. Por ejemplo, a fines de abril de 2021, la Comisión Europea envió un proyecto al Parlamento de la Unión en el que se considera el uso de la IA en la administración del trabajo como un sistema de alto riesgo, ordenando a los proveedores a corregir los modelos algorítmicos que no se

---

[302] LAHERA SÁNCHEZ, A., "El debate sobre la digitalización y la robotización del trabajo (humano) del futuro…", *op. cit*., *supra*, nota 294, p. 6.

[303] PARODI, G. F., "La renta básica ciudadana, opción frente al desempleo tecnológico y la corrupción", en *ACADEMO Revista de Investigación en Ciencias Sociales y Humanidades*, vol. 2, no. 1, 2012, pp. 6-7.

[304] *Ibidem*, p. 11.

adecúen a los protocolos de respeto por el Estado de Derecho (en el sentido de conjunto de normas y derechos de un sistema jurídico). Por ejemplo, el gobierno español, poco después, también aprobaría una ley que ordena a las empresas a informar a los sindicatos sobre la composición de los algoritmos o sistemas de IA que impactan en las condiciones de trabajo.[305]

Existen voces que también apuestan a que se garanticen los derechos sociales de los trabajadores impactados por la IA. Esta vía implica que los trabajadores reciban capacitaciones para adquirir competencias y habilidades para la inserción en el nuevo patrón de trabajo. La idea con esto es que el trabajador se reinvente, vuelva a calificarse y adaptarse al nuevo mercado [306]:

> Que el desarrollo de políticas y programas de inversiones, capacitaciones, subsidios financieros […], junto con la concientización de la clase trabajadora a través de su unión y búsqueda colectiva de sus derechos, con la asistencia básica del derecho laboral y las normas de protección interpretadas y aplicadas de acuerdo con las formas actuales de trabajo y sus relaciones, pueda convertirse en prioridades en diversos países, como una forma de apoyar esta fase de transición y construir un futuro que en lugar de enterrar los derechos y la dignidad de los trabajadores, puedan actuar juntos. [307]

En efecto, el rol del Estado es clave de cara a una probable sustitución o destrucción del empleo causada por la IA, ya que se requiere garantizar el derecho al trabajo por distintas vías, especialmente la vía de la política pública y la vía jurídica. Si bien parece lejano —incluso de ciencia ficción— el escenario caótico en donde las máquinas suplan al ser humano en el trabajo, no está de más pensar desde ahora en otras soluciones que contribuyan a proteger al trabajo humano.

---

[305] OTTAVIANO, J., "La amenaza fantasma: Inteligencia Artificial y derechos laborales", en *Nueva Sociedad*, no. 294, 2021, pp. 93 – 94.

[306] VARELA DE ALBURQUERQUE DALPRÁ, J., "La protección del trabajo digno mediante los impactos de las nuevas formas de robótica laboral, inteligencia artificial y nuevas tecnologías", en *Revista Internacional y Comparada de Relaciones Laborales y Derecho del Empleo*, vol. 8, no. 3, 2020, p. 225.

[307] *Ibidem*, pp. 230 - 231.

1.4. La desinformación derivada del uso de la Inteligencia Artificial

Otro problema relevante, en el cual la IA tiene un rol activo, es el de los "*Deep Fakes*". Por "*Deep Fake*" se entiende a los videos hiperrealísticos utilizando intercambio de caras y que dejan poco rastro de que dichos videos hayan sido manipulados.[308] "*Deep Fake"* es un término de reciente acuñación, se hizo popular en el 2017 para describir a las falsificaciones realistas de fotos, audio y videos generadas con IA.[309]

Existen ejemplos de "*Deep Fakes*" que han sido objeto de preocupaciones. Uno de los más conocidos es de un video pornográfico en el que el rostro de Gal Gadot (famosa actriz de Hollywood) aparecía en lugar del rostro de la actriz pornográfica del video. El video era tan realista que incluso era posible identificar la sincronización entre los labios del rostro de Gal Gadot con el diálogo del video. Dicho video fue producto de una red neuronal artificial entrenada con videos pornográficos y con cientos de imágenes del rostro de Gal Gadot disponibles en Internet. La red lentamente aprendería a resolver el "problema" de mapear dinámicamente el rostro de Gal Gadot para incorporarlo al cuerpo de la actriz pornográfica.[310]

El riesgo de los "*Deep Fakes*" es que estos pueden suponer un riesgo para las sociedades insertas en la era de la posverdad. Por ejemplo, videos falsificados de líderes políticos que puedan llevar a la "manipulación diplomática", capaces de generar reacciones no sólo en la sociedad sino también en el sector militar.[311] El riesgo se potencia en tanto que al público se le dificulta creer lo que sus ojos u oídos perciben, incluso cuando se trata de información verídica.[312]

Los "*Deep Fakes*", además, implican un riesgo en tanto que producen el

---

[308] WESTERLUND, M., "The Emergence of Deepfake Technology: A Review", en *Technology Innovation Management Review*, vol. 9, no. 11, 2019, p. 39.

[309] SALYER, K. M., "Deep fakes and National Security", en *Congressional Research Service*, 2020, p. 1.

[310] FLETCHER, J., "Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance", en *Theatre Journal*, vol. 70, no. 4, 2018, p. 461.

[311] *Ibidem*, p. 465.

[312] *Idem*.

"dividendo de la mentira", esto es, la idea de que los individuos pueden negar fácilmente la veracidad de lo que ven o escuchan, incluso tratándose de un contenido que es real: [313]

> The Liar's Dividend could become more powerful as deep fake technology proliferates and public knowledge of the technology grows. Some reports indicate that such tactics have already been used for political purposes. For example, political opponents of Gabon President Ali Bongo asserted that a video intended to demonstrate his good health and mental competency was a deep fake, later citing it as part of the justification for an attempted coup. Outside experts were unable to determine the video's authenticity, but one expert noted, "in some ways it doesn't matter if [the video is] a fake … It can be used to just undermine credibility and cast doubt".[314] [El dividendo de la mentira podría ser más poderoso a medida que prolifere la tecnología de falsificación profunda y crezca el conocimiento público de la misma. Algunos informes indican que estas tácticas ya se han utilizado con fines políticos. Por ejemplo, los oponentes políticos del presidente de Gabón, Alí Bongo, afirmaron que un video que pretendía demostrar su buena salud y competencia mental era una falsificación profunda, y posteriormente lo citaron como parte de la justificación de un intento de golpe de estado. Los expertos externos no pudieron determinar la autenticidad del video, pero uno de ellos señaló que "en cierto modo no importa si el video es falso… Puede utilizarse simplemente para socavar la credibilidad y sembrar la duda".] (Traducción propia).

El problema se agrava si se considera que en la actualidad la tecnología utilizada para hacer "*Deep Fakes*" es asequible para las personas, por ejemplo, en Reddit (una famosa red social de marcadores sociales y agregador de noticias) se puso a disposición de los usuarios una aplicación llamada "*FakeApp*", la cual, incluso con poco conocimiento en materia de Aprendizaje Automatizado y programación, podría

[313] SALYER, K. M., "Deep fakes and National Security", *op. cit*., *supra*, nota 309, p. 1.
[314] *Idem.*

crear imágenes y videos falsos.[315]

Así pues, la IA sube a un nivel más arriba a la desinformación, permitiendo la falsificación de videos capaces de generar confusión en las personas. Retomando una frase de Yuval Noah Harari, con relación a la desinformación provocada por las noticias falsas: "*The more people believe in free will… the easier it is to manipulate them, because they won't think that their feelings are being produced and manipulated by some external system"*.[316] [Cuanto más crea la gente el libre albedrío… más fácil será manipularla, porque no pensará que sus sentimientos están siendo producidos y manipulados por algún sistema externo] (Traducción propia). En efecto, puede inferirse que los "*Deep Fakes*" pueden provocar efectos perjudiciales en las sociedades democráticas que abrazan la libertad como principio.

Se tiene conocimiento de que existen al menos cuatro grandes productores de "*Deep Fakes"* en el mundo: 1) las comunidades que por afición hacen "*Deep Fakes*"; 2) actores políticos como pueden ser los gobiernos que buscan influir en decisiones políticas de otros países; 3) defraudadores, y 4) actores que actúan de buena fe, por ejemplo, cadenas televisivas.[317] A decir de los defraudadores y criminales que utilizan los "*Deep Fakes*":

> Deepfakes are also increasingly being deployed by fraudsters for the purpose of conducting market and stock manipulation, and various other financial crimes. Criminals have already used AI-generated fake audios to impersonate an executive on the phone asking for an urgent cash transfer. In the future, video calls will also be able to be faked in real-time. Visual material required to produce impersonations of executives are often available on the Internet. Deepfake technology can make use of visual an

[315] ALBAHAR, M. y J. Almalki, "Deepfakes: Threats and Countermeasures Systematic Review", en *Journal of Theoretical and Applied Information Technology*, vol. 97, no. 22, 2019, p. 3243.

[316] ANTONHY, A., "Yuval Noah Harari: The Idea of Free Information is Extremely Dangerous", en The Guardian, [5 de agosto de 2018, en línea], *apud* MANHEIM, K. y L. Kaplan, "Artificial Intelligence: Risk to Privacy and Democracy", en *The Yale Journal of Law & Technology*, vol. 21, no. 106, 2019, p. 145.

[317] WESTERLUND, M., "The Emergence of Deepfake Technology: A Review", *op. cit*., *supra,* nota 308, p. 41.

> audio impersonations of executives, for example, TED Talk videos available on YouTube. [318] [Los estafadores también utilizan cada vez más las falsificaciones para manipular los mercados y las acciones, así como para cometer otros delitos financieros. Los delincuentes ya han utilizado audios falsos generados por la IA para hacerse pasar por un ejecutivo al teléfono pidiendo una transferencia urgente de dinero. En el futuro, las videollamadas también podrán ser falsificadas en tiempo real. El material visual necesario para producir suplantaciones de ejecutivos suele estar disponible en Internet. La tecnología Deepfake puede utilizar limitaciones visuales y sonoras de ejecutivos, por ejemplo, videos de Ted Talks disponibles en YouTube] (Traducción Propia).

Pensando en actores políticos que quieran sacar provecho de los "*Deep Fakes*", podemos encontrar el siguiente escenario: *"Imagine a future where … a fake video of a president incites a riot or fells the market"*.[319] [Imagina un futuro donde … un video falso del presidente incite a una revuelta o haga caer al mercado] (Traducción propia). En efecto, un escenario así seguramente mostraría la vulnerabilidad del humano frente a la distorsión de la realidad, y un recordatorio de lo frágil que pueden ser las democracias frente a las mentiras replicadas a gran escala.

Ahora bien, al pensar en soluciones para hacer frente a los "*Deep Fakes*" se puede decir que estas provienen de tres vías: 1) la regulación; 2) la educación del receptor del contenido y 3) la tecnología anti "*Deep Fake*" de que se disponga.[320]

A decir de la primera vía, esta es compleja puesto que es muy delgada la línea entre el problema de los "*Deep Fakes*" que agravian libertades y derechos, y la línea de la libertad de expresión.[321] En el mismo tenor, pensar en prohibir la tecnología que hace posible a los "*Deep Fakes*", implica un riesgo, y es el de inhibir el desarrollo tecnológico de la IA, algo que puede ser perjudicial si lo que se busca es sacar provecho de los dividendos que ésta trae consigo.

Con relación a la solución de generar capacidades en la persona para discernir

[318] *Ibidem*, p. 42.
[319] *Idem*.
[320] *Ibidem*, p. 44.
[321] *Idem.*

un video falso de otro que no lo es, puede decirse que dicha solución resulta clave. Para esto se requiere concientizar a las personas sobre los potenciales usos perjudiciales de la IA, [322] dando la pauta para que éstas se cuestionen si el video que tienen frente a ellos es real o si se trata de una manipulación lograda a través de dicha tecnología.

Por último, la tecnología anti "*Deep Fake*" contribuye a detectar los "*Deep Fakes*", autenticar el contenido si es el caso, y evitar que los contenidos se utilicen para producir "*Deep Fakes*". [323]

2. EL FUTURO

2.1. Transhumanismo e Inteligencia Artificial

Si bien es cierto que actualmente la IA no ha sido capaz de exceder la inteligencia humana, también lo es que a futuro plantea una serie de escenarios sobre los cuales es importante comenzar a reflexionar. Uno de estos escenarios es el de la simbiosis humano-máquina, un punto en el que la tecnología elimina las barreras que impiden al ser humano trascender sus límites biológicos.

A decir de Elon Musk, con el paso del tiempo probablemente veremos una fusión más y más clara de la inteligencia biológica y la digital, un hecho que mantendrá la relevancia del ser humano en la era de la IA.[324] En efecto, bien vale ir reflexionando sobre las cuestiones que desde el ámbito del Derecho podrían suscitarse con motivo de dicha fusión. Así, este apartado busca, además de exponer lo qué es el transhumanismo, esbozar algunos puntos de reflexión jurídica para el futuro.

El concepto de transhumanismo no es nuevo. Hay quienes afirman que fue utilizado por primera vez en 1957, por el biólogo humanista Julian Hurxley, en su texto "*New bottles for a new wine*", y cuya idea era que el ser humano debe mejorarse a sí mismo, a través de la ciencia y la tecnología, ya sea desde el punto

[322] *Ibidem*, p. 45.
[323] *Idem.*
[324] PASTOR, J., "Elon Musk: o los humanos se fusionan con las máquinas, o la inteligencia artificial nos hará irrelevantes", en *Xataka*, [23 de abril de 2019, en línea], <https://www.xataka.com/robotica-e-ia/elon-musk-o-los-humanos-se-fusionan-con-las-maquinas-o-la-inteligencia-artificial-nos-hara-irrelevantes> [consulta: 12 de septiembre de 2023].

de vista genético o desde el punto de vista ambiental y social.[325] Desde un panorama actual las mejoras comprenden el tratamiento médico de las enfermedades, el incremento del potencial humano natural, por ejemplo, aumentar el coeficiente intelectual de una persona, y las mejoras suprahumanas, como puede ser el proporcionar un sonar o un coeficiente intelectual por encima de los 200 puntos.[326]

Se entiende al transhumanismo como una escuela del pensamiento que se niega a aceptar las limitaciones humanas tradicionales como la muerte, la enfermedad y otras debilidades tecnológicas.[327] De la mano con la anterior noción, el transhumanismo propone el mejoramiento humano haciendo uso de las nano-bio-tecnologías para hacer realidad el deseo humano de trascender sus limitaciones. Desde esta postura, se trata de avanzar hacía lo posthumano, acelerando así la evolución de la vida inteligente.[328] Se dice también que rescata al deseo de trascendencia presente en las narraciones míticas de antiguas culturas, y que también recurre al imaginario de la ciencia ficción para explorar nuevos límites.[329]

Una forma de entender el transhumanismo, a decir de Antonio Diéguez, sería como la convicción de que el ser humano se encuentra contenido en un soporte inadecuado -su cuerpo- y que es posible perfeccionarlo mediante la tecnología.[330] De igual manera, el autor sitúa al transhumanismo en continuidad con los intentos de mejora del ser humano a lo largo de la historia y, al mismo tiempo, lo distingue de los medios tradicionales de perfeccionamiento; en ese tenor, el transhumanismo propone ciertas intervenciones directas por medio de la tecnología, conocidas como biomejoras, que van desde medicamentos, ingeniería genética y hasta la unión con

---

[325] VILLAROEL, R., "Consideraciones Bioéticas y Biopolíticas acerca del transhumanismo. El debate en torno a una posible experiencia posthumana", en *Revista de Filosofía*, vol. 71, 2015, p. 179.

[326] *Idem.*

[327] PIEDRA ALEGRÍA, J., "Transhumanismo: hacia un nuevo cuerpo", en *Daimon, Revista Internacional de Filosofía*, 2016, p. 490.

[328] GAYOZZO HUAMANCHUMO, P. A., "Singularidad tecnológica y transhumanismo", en *Teknokultura, Revista de Cultura Digital y Movimientos Sociales*, no. 18, 2021, p. 196.

[329] *Idem.*

[330] QUER, M. "Reseña de Cuerpos inadecuados. El desafío transhumanista a la filosofía, Diéguez, Antonio (2021), Herder, Barcelona. 2014", en *Revista de Filosofía Open Insight*, vol. XIII, núm. 27, 2022. [s.f., en línea], <https://www.redalyc.org/journal/4216/421669968008/html/> [consulta: 12 de septiembre de 2023].

la máquina.[331]

Nick Bostrom, uno de los principales impulsores del movimiento transhumanista, indica lo siguiente:

> Transhumanism is a loosely defined movement that has developed gradually over the past two decades. It promotes an interdisciplinary approach to understanding and evaluation the opportunities for enhancing the human condition and the human organism opened up by the advancement of technology. Attention is given to both present technologies, like genetic engineering and information technology, and anticipated futures ones, such as molecular nanotechnology and artificial intelligence. The enhancement options being discussed include radical extension of human health-span, eradication of disease, elimination of unnecessary suffering, and augmentation of human intellectual and emotional capacities. Other transhumanist themes include space colonization and the possibility of creating superintelligent machines, along with other potential developments that could profoundly alter the human condition.[332] [El transhumanismo es un movimiento poco definido que se ha desarrollado gradualmente en las dos últimas décadas. Promueve un enfoque interdisciplinario para comprender y evaluar las oportunidades de mejora de la condición humana y del organismo humano que ofrece el avance de la tecnología. Se presta atención tanto a las tecnologías actuales, como la ingeniería genética y la tecnología de la información, como a las futuras, como la nanotecnología molecular y la inteligencia artificial. Las opciones de mejora que se discuten incluyen la ampliación radical de la duración de la salud humana, la erradicación de las enfermedades, la eliminación del sufrimiento innecesario y el aumento de las capacidades intelectuales y emocionales del ser humano. Otros temas transhumanistas incluyen la colonización del espacio y la posibilidad de crear máquinas superinteligentes, junto con otros desarrollos potenciales que podrían alterar profundamente la condición

[331] *Idem*.

[332] BOSTROM, N., “Transhumanist Values”, en *Journal of Philosophical Research*, Vol. 30, 2005, p. 3.

humana] (Traducción propia).

En efecto, el transhumanismo apuesta por el mejoramiento de la condición humana a través de las tecnologías. La intención es alterar la condición natural del humano, despojándolo de sus límites biológicos. Se dice también que el transhumanismo abraza los ideales de progreso y pretende aliviar el sufrimiento, evitar desastres y superar las enfermedades mediante los recursos tecnológicos.[333]

Ahora bien, ¿qué relación tiene la IA con el transhumanismo?, por principio de cuentas, se sabe que la IA es uno de los temas de mayor importancia para el transhumanismo, lo anterior se explica mejor con la siguiente cita:

> Como bien indica Bostrom, el desarrollo de inteligencia artificial es también uno de los temas de mayor importancia para el transhumanismo. No solo porque puede ser usada como una herramienta para el mejoramiento cognitivo, para la transferencia de una conciencia humana a dispositivos computacionales (mind uploading) o para la fusión humano-máquina, sino porque es posible extrapolar algunos problemas sobre el posible estatus moral de un transhumano con los del estatus moral de una superinteligencia artificial.[334]

La IA se vincula con el transhumanismo en tanto que con ella es posible el mejoramiento cognitivo, pero también debido al vínculo que existen entre las reflexiones sobre el estatus moral del agente transhumano y aquellas del agente superinteligente. Y es que si se piensa en lo ciborg —un fenómeno en el que la fusión humano-máquina ocurre, por ejemplo, mediante prótesis computacionales de alto nivel y de todo tipo que pueden acoplarse e interaccionar directamente con nuestro tejido neuronal,[335] o en el que lo artificial suple a lo orgánico—, surgen muchas interrogantes, ¿es humano aquel agente que sólo conserva su consciencia,

[333] GAYOZZO HUAMANCHUMO, P. A., “Singularidad tecnológica y transhumanismo”, *op. cit*., *supra*, nota 328, p. 196.
[334] *Idem.*
[335] NÚÑEZ PARTIDO, J. P., “Hombres y máquinas. Futuro y límites del transhumanismo”, en *Razón y Fe*, no. 1434, 2018, p. 76.

pero ya no su cuerpo biológico, pues este ha sido cambiado por prótesis tecnológicas?

Como mera anécdota, existe una escena eliminada de la película "*RoboCop*", dirigida por Paul Verhoeven, en donde Bob Morton —el ejecutivo de Security Concepts y líder del proyecto "*RoboCop*"— protagoniza la siguiente entrevista:

> Periodista: 'Sr. Morton, ¿es este un hombre o es una máquina?
> Morton: 'Bueno, técnicamente es un ciborg integrado. Sería reacio a categorizarlo como a ti, para decirte la verdad' […]. [336]

¿Qué es un ciborg? ¿humano o máquina?, la pregunta persiste. ¿El ente superinteligente, es más próximo a una máquina o a un ser humano? Además de las reflexiones filosóficas que lo transhumano y la IA comparten, otro punto de convergencia entre ambos temas son los conceptos referenciales que son clave para la comprensión científica de los mecanismos que fundamentan el pensamiento y el comportamiento humano inteligente y su incorporación en las máquinas:[337]

> Dichos mecanismos en los que se basa la Inteligencia Artificial, permiten identificar un punto de correlación tecnocientífico frente a lo denominado 'transhumanismo', pues ambos convergen [lo transhumano y la Inteligencia Artificial] en conceptos referenciales tales como bio (vida), info (información), cogno (conocimiento) y nano (simplicidad), equivalentes a la biotecnológica, la información tecnológica, la ciencia cognitiva y la nanotecnología —elementos básicos de la convergencia NBIC [Tecnologías convergentes]—, cuyo propósito busca "reconocer el grado de correlación y amplitud de las máquinas en un contexto específicamente alternativo al servicio del desarrollo humano".[338]

---

[336] RESPUESTAS AQUÍ, *¿Sabía el público en general que RoboCop era un Cyborg?*, [s.f., en línea], <https://respuestas.me/q/sabia-el-publico-en-general-que-robo-cop-era-un-cyborg-61527303970> [consulta: 01 de octubre de 2023].

[337] VILLALBA GÓMEZ, J. A., "Problemas bioéticos emergentes en la inteligencia artificial", en *Diversitas: Perspectivas en Psicología*, Universidad Santo Tomás, Colombia, vol. 12, no. 1, 2016, p. 139.

[338] *Idem*.

De cara al mejoramiento de las capacidades humanas por medio de la creación y despliegue de las tecnologías convergentes, aparece un concepto que ayuda comprender más la unión entre el transhumanismo y la IA, la singularidad:

> In physics, a singularity is a point in space or time, such as the center of a black hole or the instant of the Big Bang, where mathematics breaks down and our capacity for comprehension along with it. By analogy, a singularity in human history would occur if exponential technological progress brought about such dramatic change that human affairs as we understand them today came to an end. […] Our very understanding of what it means to be human — to be an individual, to be alive, to be conscious, to be part of the social order— all this would be thrown into question, not by detached philosophical reflection, but through force of circumstances, real and present.[339] [En física, una singularidad es un punto en el espacio o en el tiempo, como el centro de un agujero negro o el instante del Big Bang, en el que las matemáticas se rompen y con ellas nuestra capacidad de comprensión. Por analogía, una singularidad en la historia de la humanidad se produciría si el progreso tecnológico exponencial provocara un cambio tan drástico que los asuntos humanos, tal y como los entendemos hoy, llegarán a su fin. […] Nuestra propia comprensión de lo que significa ser humano —ser un individuo, estar vivo, ser consciente, formar parte del orden social— todo esto se pondría en cuestión, no por una reflexión filosófica aislada, sino por la fuerza de las circunstancias reales y presentes] (Traducción propia).

La singularidad es el momento en el que lo humano se fusiona con la tecnología. La IA se presenta aquí como un potenciador de la inteligencia humana, como bien lo refiere Ray Kurzweil:

> Un hardware necesario para emular la inteligencia humana con

[339] SHANAHAN, M., *The Technological Singularity*, The MIT Press Essential Knowledge Series, Massachusetts Institute of Technology, 2015, p. XV.

> superordenadores, (…) modelos de software de la inteligencia humana (…) lo cual indicará una inteligencia indistinguible a la de humanos biológicos (…) cuando se alcance este nivel de desarrollo, los ordenadores podrán combinar los tradicionalmente puntos fuertes de la inteligencia humana, con los puntos fuertes de la inteligencia de las máquinas.[340]

Así pues, el transhumanismo como movimiento también echa mano de la IA en tanto que mejora la inteligencia humana. Ahora bien, regresando al objetivo de este apartado, ¿qué desafíos jurídicos trae aparejado el transhumanismo? Para responder a esta pregunta, es importante referir a los retos sociales que devendrían de la irrupción ciborg.

De acuerdo con Juan Pedro Núñez, algunos de los retos son los siguientes:

- El incremento de las relaciones virtuales, incluyendo la interacción con sistemas artificiales de tipo androide que se ajusten a nuestros gustos y necesidades. Esto traerá como efecto que, como personas humanas, al reducir nuestras relaciones, también se reducirá nuestro desarrollo social: "*En el futuro no tendremos que afrontar el aburrimiento ni la soledad, ni la dificultad de la gestión de los conflictos en persona y será más fácil eludir la propia responsabilidad gracias a la distancia o el anonimato* […]"[341].
- Se estará frente a una revisión drástica de nuestro sistema de valores. Algunas de los dilemas son los siguientes: ¿es infiel el cónyuge que mantiene relaciones sexuales en un espacio virtual con un ciborg?, ¿existe la maldad en la crueldad ejercida en el mundo virtual o sobre seres artificiales? A la par de estos dilemas, también deberá de ser revisitada la responsabilidad jurídica de las personas, por lo difícil que podría ser diferenciar entre un fallo de las aplicaciones neurales, o un acto deliberado en el sentido clásico.[342]

---

[340] KURZWEIL, R., *The singularity is near: When humans transcend biology*, Lola Books, 2005, p. 148 *apud* VILLALBA GÓMEZ, J. A., "Problemas bioéticos emergentes en la inteligencia artificial", *op. cit*., *supra*, nota 335, p. 143.

[341] NÚÑEZ PARTIDO, J. P., "Hombres y máquinas. Futuro y límites del transhumanismo", *op. cit*, *supra*, nota 335, pp. 81-82.

[342] *Idem.*

- También se experimentará un crecimiento en la desigualdad social, entre clases, diferencias insalvables en niveles funcionales y ejecutivos de todo tipo.[343]
- Las corporaciones que controlen los sistemas más vendidos en el mercado podrían acumular un poder difícilmente controlable por los gobiernos y sociedades que dependen de ellos. [344]

Con base en estos retos, es posible identificar algunos problemas jurídicos a futuro, ¿será posible responsabilizar jurídicamente al ciborg cuyas aplicaciones neurales mejoradas fallaron, provocando que matara a una persona?; ¿pensando en el alza de las interacciones virtuales, podrían encontrarse razones jurídicas para validar un matrimonio ocurrido en el mundo virtual?; ¿qué tiene por hacer el Derecho para evitar que se generen y amplíen brechas de desigualdad, entre una clase social que tiene el privilegio de acceder a las mejoras tecnológicas y genéticas y entre una clase social que no tendrá esa oportunidad?, y ¿cómo pueden actuar los gobiernos para evitar que los monopolios tecnológicos del futuro se conviertan de hecho en quienes controlen a la sociedad y en quienes anulen la capacidad coercitiva del Estado?

Si bien es cierto que las preguntas planteadas parecieran partir de una especulación futurista, lo cierto es que en el presente se han planteado cuestiones jurídicas a partir de hechos acontecidos: ¿una persona, tiene derecho a escoger los órganos que quiere? Neil Harbisson, el primer ciborg reconocido oficialmente por un gobierno ha defendido públicamente el derecho de los seres humanos a incorporar tecnología en su cuerpo y a diseñarse como especie: "*cada uno debe ser libre de escoger qué órganos quiere como especie*".[345] Frente a esta demanda, ¿podrían los gobiernos oponerse, por ejemplo, frente a la decisión de una persona por incorporar a su cuerpo componentes mecánicos para asemejarse a un robot (pensemos en un robot del tipo "*Terminator*")?

---

[343] *Idem.*

[344] *Idem*.

[345] LA VANGUARDIA., "*El primer cyborg del mundo defiende 'el derecho a diseñarse como especie'*", [4 de octubre de 2019, en línea], < https://www.lavanguardia.com/vida/20191004/47798233444/el-primer-ciborg-del-mundo-defiende-el-derecho-a-disenarse-como-especie.html> [consulta: 2 de octubre de 2023].

Las anteriores son preguntas que pueden ayudar a reflexionar sobre los escenarios a los que habrá de enfrentarse el Derecho. Valdría entonces la pena que, desde los distintos ámbitos, tanto profesionales como académicos, comiencen a analizarse estas cuestiones, con la finalidad de que cuando el futuro sea nuestro presente, podamos ganar algo del tiempo al ya de por sí vertiginoso desarrollo tecnológico.

2.2. Derechos de los Robots

La reflexión de este apartado conlleva un desafío, la de establecer bases que permitan en un futuro justificar la decisión de adscribir o no derechos a los sistemas dotados de IA. Aquí vale hacer una primera delimitación, no todos los sistemas dotados de IA merecen ser objeto de esta reflexión. Pensar que un asistente virtual, como puede serlo Alexa, Siri o Google Home, merezca derechos es por demás ocioso. Sin embargo, si nos referimos a un agente moral artificial, supongamos un robot social capaz de demostrar afecto hacia nosotros o incluso demostrar sufrimiento, entonces sí vale la pena analizar la idea de adscribirle derechos, o al menos, algún tipo de consideración moral.

Y ¿qué hay con relación a los ciborgs? Pero no ciborgs del presente, como es el caso de Neil Harbisson, sino ciborgs futuros. Suponiendo que "*RoboCop*" fuera real, ¿tendría derechos? En efecto, nuestra primera delimitación está establecida: no todos los sistemas de IA merecen ser objeto de esta reflexión, en todo caso, el análisis tiene una proyección a futuro, con sistemas de IA que repliquen o superen la mente humana o con ciborgs cuya composición sea más de componentes artificiales que de componentes biológicos.

Dicho esto, ¿por qué reflexionar sobre los derechos de los robots? Por principio de cuentas, nos encontramos ante una nueva posibilidad derivada de los avances tecnológicos en temas de aprendizaje automatizado y profundo que son capaces de trastocar las relaciones con otros sujetos de derecho.[346] Díez Spelz, refiriendo a

[346] DÍEZ SPELZ, J. F., "¿Robots con derechos?: la frontera entre lo humano y lo no-humano. Reflexiones desde la teoría de los derechos humanos", en *IUS, Revista del Instituto de Ciencias Jurídicas de Puebla*, vol. 15, núm. 48, 2021, pp. 261-262.

Yuval Noah Harari, nos dice que esta nueva posibilidad es uno de los principales retos para nuestro siglo, pues, en algún punto, podría ser incluso la causa de que nuestro concepto de humanidad cambie.[347]

Otra razón para justificar la presente reflexión se cimenta en el presente. Sophia, un robot humanoide desarrollada por la compañía *Hanson Robotics*, se convirtió en el 2017 en la primer robot con ciudadanía de un país, en su caso, la saudí. Este hecho suscito una serie de controversias, ¿qué implicaciones tendría dicho acontecimiento para el derecho? ¿Sophia debía ser protegida igual que el resto las personas? ¿podía Sophia participar en las elecciones de Arabía Saudita? Si bien es cierto que el reconocimiento de ciudadanía fue más un movimiento de marketing,[348] también lo es que dio la pauta para avivar las discusiones en torno a los derechos (y obligaciones) de los robots. En efecto, pienso que vale la pena analizar el tema de los derechos de los robots con base en una razón presente, justo para disipar confusiones o minimizar el impacto de ideas de la cultura popular que pueden encaminar al Derecho para un camino equivocado.

La última razón para reflexionar sobre los derechos de los robots es el futuro. No hay certeza de cuándo será el momento en el que una máquina con superinteligencia, o al menos capaz de replicar a la mente humana, aparezca; no obstante, se vuelve necesario analizar desde ahora si estas máquinas pueden ser sujetos de derechos. El desafío de analizar con base en consideraciones futuras es el de evitar caer en la perogrullada de que, cuando las máquinas sean conscientes, en un futuro próximo o lejano, serán sujetos de derechos. Probablemente se caiga en ella, pero de dicho ejercicio será posible pensar en soluciones presentes, más aterrizadas a la realidad jurídica de nuestro tiempo, que permitan abordar problemas éticos asociados a los sistemas con IA.

Teniendo en cuenta la anterior justificación, resulta importante hacerse otra pregunta, ¿qué antecedentes de discusiones en foros jurídicos existen con relación al estatus jurídico de los robots? Quizá el más relevante es el de *Robo-Law*,

---

[347] *Idem.*

[348] SCALITER, J., “Sophia, el primer robot con ciudadanía”, en *La Razón*, [29 de octubre de 2017, en línea], <https://www.larazon.es/tecnologia/sophia-el-primer-robot-con-ciudadania-OE16746549/> [consulta: 2 de octubre de 2023].

impulsado y financiado por la Unión Europea entre el 1 de marzo de 2012 y el 22 de septiembre de 2014. Dicho proyecto:

> […] perfiló las líneas más importantes a tener en cuenta en una futura normativa sobre robótica, y trata de dar respuestas a tres cuestiones básicas: las herramientas legales que mejor se adaptan al objetivo de regular la tecnología robótica y la inteligencia artificial, el tipo de análisis ético que se debe realizar sobre la innovación tecnológica y la metodología a aplicar, respetando la política de la Unión Europea sobre Investigación e Innovación Responsable (RRI) y los contenidos de la regulación, con especial atención a la necesidad de proteger los derechos fundamentales de los ciudadanos europeos.[349]

El 3 de septiembre de 2013, la Comisión Europea junto con un consorcio de diez empresas europeas puso en marcha el proyecto PETROBOT, el cual estaba orientado a fabricar robots que pudieran sustituir a los seres humanos en tareas de inspecciones de los recipientes a presión y los tanques de almacenamiento utilizados en la industria petroquímica.[350]

Frente al escenario que supone la sustitución laboral, Marc Tarabella, miembro del Parlamento Europeo, solicitó a la cámara pronunciarse acerca del estatuto jurídico de los robots y de su posición sobre un eventual reconocimiento de su personalidad jurídica.[351]

---

[349] GONZÁLEZ GRANADO, J., *De la persona algorítmica, A propósito de la personalidad jurídica de la inteligencia artificial,* Colección de Bioética, Edicions de la Universitat de Barcelona, España, 2020, p. 61.

[350] *Idem*.

[351] *Idem*. Marc Tarabella lanzó las siguientes interrogantes con relación a los derechos de los robots: "*Des avocats planchent sur la création d'une «personnalité robot» et l'attribution d'un numéro de sécurite sociale; le ministére du redressement productif travaille sur un project de charte éthique non contraignante, la Commission européenne envisage de leur conférer la personnalité morale… de maniére protéiforme, le droit des robots semble émerger. Des robots remplaçant des hommes, c'est ce que commence à faire la Commission européenne avec son projet Petrobot. Associée à un consortium de dix entreprises européennes dirigé par le pétrolier Shell, la Commission souhaite élaborer des robots pouvant se substituer aux êtres humains pour «l'inspection des cuves à pression et des réservoirs de stockage largement utilisés dans l'industrie du pétrole, du gaz et de la pétrochimie». La Commission précise que le fait de «conférer un statut légal aux robots et aux systèmes intelligents est une option, mais c'est*

La respuesta de la Comisión Europea, del 15 de noviembre de 2013, fue la siguiente:

> […] la tecnología aún no está lista para ofrecer el grado de autonomía necesario para otorgar personalidad o estatus legal a los robots. Sin embargo, la comunidad de la robótica y de los sistemas cognitivos artificiales tiene muy en cuenta las cuestiones éticas, jurídicas y sociales relacionadas con los sistemas futuros, dotados de mayor inteligencia y autonomía.[352]

De igual manera, afirmaba que existían tendencias mundiales que dejaban entrever que en el futuro podrían ser una realidad los sistemas completamente autónomos: *"[…] el desarrollo de opciones políticas y la comprensión de las consecuencias jurídicas de los sistemas plenamente autónomos es una forma de preparar el futuro y que la concesión de personalidad jurídica a los robots es actualmente un debate académico".* [353]

Posteriormente, el Parlamento Europeo en una resolución del 16 de febrero de 2017, sugeriría que a futuro sería conveniente considerar a algunos robots como personas electrónicas, sobre todo para efectos de la responsabilidad civil por daños

---

seulement une option». 1. Qu'en est-il vraiment?; 2. Quel est le but?, 3. Quel est le budget autour de cette thématique?". [Los juristas trabajan en la creación de una "personalidad robótica" y en la atribución de un número de seguridad social; el Ministerio de Recuperación Productiva trabaja en un proyecto de carta ética no vinculante; la Comisión Europea estudia la posibilidad de darles personalidad jurídica... de manera repentina, el derecho de los robots parece estar surgiendo. Que los robots sustituyan a los humanos es lo que está empezando a hacer la Comisión Europea con su proyecto Petrobot. Junto con un consorcio de diez empresas europeas lideradas por la petrolera Shell, la Comisión quiere desarrollar robots que puedan sustituir a los humanos para "la inspección de recipientes a presión y tanques de almacenamiento ampliamente utilizados en las industrias del petróleo, el gas y la petroquímica". La Comisión afirma que "dar a los robots y a los sistemas inteligentes un estatus legal es una opción, pero es sólo una opción". 1. 1. ¿Qué es realmente? 2. ¿Cuál es el objetivo? 3. ¿Cuál es el presupuesto que lo rodea?] (Traducción propia). PARLAMENTO EUROPEO, *Questions parlamentaires, Question avec demande de réponse écrite E-011289-13 à la Commission, Article 117 du reglament, Marc Tarabella* (S&D), [3 de octubre de 2013, en línea], <https://www.europarl.europa.eu/doceo/document/E-7-2013-011289_FR.html> [consulta: 2 de octubre de 2023].

[352] GONZÁLEZ GRANADO, J., De la persona algorítmica, A propósito de la personalidad jurídica de la inteligencia artificial, *op. cit*., *supra*, nota 349, p. 61.

[353] *Ibidem*, pp. 61-62.

que puedan cometer. De esta resolución, también se desprende que los robots no podrían beneficiarse de los derechos humanos, en todo caso lo que se pretende es dotarlos de un estatus similar al de las corporaciones (persona moral con capacidad jurídica) para temas de responsabilidad.[354]

A esto debemos sumarle un par de acotaciones. La primera acotación la hace el Grupo Europeo sobre Ética de Ciencia y las Nuevas Tecnologías, al referir que, aunque algunos sistemas de IA puedan presentar capacidades intelectuales o cognitivas muy avanzadas, el concepto de "autonomía" debe vincularse únicamente a los seres humanos y a su dignidad. La segunda acotación es de la Carta abierta firmada por más de 150 expertos en temas de IA y robótica, dirigida a la Comisión Europea. Díez Spelz aludiendo a la Carta, parafrasea lo siguiente:

> […] sostienen (los 150 expertos) que la atribución de un estatuto jurídico de personas a máquinas, robot o sistemas de Inteligencia Artificial no debería fundarse ni en el modelo de la persona física, ni en el de la persona moral o jurídica. En el primero de los casos porque equiparar a estos sistemas a las personas físicas supondría atribuirles derechos humanos, lo que es esencialmente problemático; en el segundo caso porque supondría reconocer que detrás de la acción de los robots habría intereses humanos, lo cual podría contradecirse con aquellos sistemas con capacidad de "autonomía" o "autoaprendizaje".[355]

En efecto, otorgar derechos a los robots o algún tipo de personalidad implicaría sobrevalorar las capacidades reales de la IA en la actualidad, con un imprevisible impacto económico, legal, social y ético.[356] Además también se dice que: "*la personalidad legal que se les quiere otorgar ahora a los robots no tiene nada que ver con regular sus derechos, sino que tiene una motivación económica y busca*

[354] DÍEZ SPELZ, J.F., "¿Robots con derechos?: la frontera entre lo humano y lo no-humano, *op. cit*., supra, nota, 346, p. 267.
[355] *Ibidem*, p. 268.
[356] CHÁVEZ VALDIVIA, A. K., "Hacia otra dimensión jurídica: el derecho de los robots", en *IUS, Revista del Instituto de Ciencias Jurídicas de Puebla*, vol. 15, 2021, p. 333.

eximir a los fabricantes de responsabilidad en los actos de los robots"[357]

Con referencia al impacto jurídico, el reconocer una personalidad jurídica a los robots implicaría: 1) que puedan ser sujetos de derecho; 2) que tengan la posibilidad de ejercer derechos y deberes, y 3) el que puedan disfrutar de derechos implica que los robots tienen conciencia y libertad.[358] La personalidad electrónica, si bien no es lo mismo a la personalidad jurídica, se trata de un esfuerzo para considerar como responsables de reparar los daños causados a los robots, en particular los que gozan de mayor autonomía:

> […] la personalidad electrónica —que no persona electrónica— puede ser reputada como un enfoque plausible al problema, tanto para los robots dotados de un cuerpo como para los robots software que exhiben un cierto grado de autonomía e interactúan con las personas, en cuanto que tendrían la posibilidad de ser titulares de relaciones jurídicas con sus correspondientes derechos y obligaciones, y tener un cierto reconocimiento jurídico de sus subjetividad, fundamentalmente en derechos de naturaleza patrimonial, pero no los constitucionales ni los de la personalidad, absolutamente consustanciales a la dignidad de los seres humanos tal y como recuerdan textos internacionales recientes como la Declaración Universal de la UNESCO sobre bioética y derechos humanos de 2005 o la Carta de los Derechos Fundamentales de la Unión Europea de 2000. [359]

Si se piensa en que la personalidad jurídica conlleva a una posibilidad de que se cuente con derechos, pero también con obligaciones, ¿qué obligaciones tendría un robot hacía el ser humano, por ejemplo? La literatura universal nos una idea de cuáles podrían ser. En su obra, "*Yo, Robot*", Isaac Asimov señala Tres Leyes de la Robótica, a saber: 1) Un robot no puede dañar a un ser humano o, por su inacción,

---

[357] *Ibidem*, p. 334.

[358] DÍEZ SPELZ, J. F., "¿Robots con derechos?: la frontera entre lo humano y lo no-humano", *op. cit*., *supra*, nota 346, p. 269.

[359] BARRIO ANDRÉS, M., "Hacia una personalidad electrónica para los robots", en *Revista de Derecho Privado*, no. 2, marzo 2018, p. 104 *apud* DÍEZ SPELZ, J. F., "¿Robots con derechos?: la frontera entre lo humano y lo no-humano", *op. cit*., *supra*, nota 346, p. 276.

dejar que un ser humano sufra daño; 2) Un robot debe obedecer las órdenes que le son dadas por un ser humano, excepto cuando estas órdenes se oponen a la Primera Ley y 3) Un robot debe proteger su propia existencia hasta donde esa protección no entre en conflicto con la Primera o Segunda Ley.

En esencia, dichas reglas buscan la supervivencia y el no daño de la humanidad como obligaciones de los robots. Ahora bien, tomando como referencia el documento "*European civil law rules in robotics*", realizado por la Parlamento Europeo, es posible identificar principios a salvaguardar de cara a la interacción humana con robots.

Dichos principios son cimentados desde la ética y también podríamos entender, del principio uno al sexto, como obligaciones para los robots: 1) proteger a los seres humanos de los daños causados por robots; 2) respetar la decisión de no ser atendido por un robot; 3) proteger la libertad humana frente a los robots; 4) proteger la humanidad contra las violaciones de la privacidad cometidas por un robot; 5) gestionar responsablemente los datos personales procesados por robots; 6) proteger a la humanidad contra el riesgo de manipulación por los robots; 7) evitar la disolución de los vínculos sociales; 8) igualdad de acceso al progreso en la robótica, y 9) restringir el acceso humano a las tecnologías de mejora.[360]

Regresando a la cuestión de los derechos de los robots, puede decirse que no es un tema trivial. Los esfuerzos por intentar responder a problemáticas asociadas a la responsabilidad por daños provocados por sistemas artificiales con una autonomía amplia son de por sí problemáticos; por ejemplo, el dotar de una personalidad electrónica a un robot desata una serie de cuestiones asociadas al propio concepto de personalidad. Se dice que la personalidad, en el ámbito del derecho, implica la posibilidad de la persona para actuar (capacidad) como sujeto activo o pasivo en las relaciones jurídicas,[361]en otras palabras, la personalidad

[360] PARLAMENTO EUROPEO, *European Civil Law Rules in Robotics, Study for the JURI Committee*, 2016. [octubre 2016, en línea], <https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf > [consulta: 2 de octubre de 2023]. Los principios séptimo y octavo están orientados a evitar la generación de brechas de desigualdad producto del acceso a las tecnologías vinculadas al uso de la IA y la mejora biológica a través de la tecnología.

[361] LÓPEZ HERNÁNDEZ, E., "La personalidad en el juicio", en *Revista Mexicana de Derecho*, no. 2,

refiere a la "*idoneidad para ser sujeto de derechos y obligaciones*".[362] En efecto, si reconocemos personalidad a un sistema con IA, también estaríamos reconociendo que cuenta con derechos y con deberes.

Aunque la justificación de dotar de personalidad "*electrónica*" a los robots sea la de proteger al fabricante frente a daños causados por un robot que cuente con un grado alto de autonomía, esta no resulta satisfactoria frente a cuestionamientos como el de ¿si algo o alguien con personalidad, también tendría que contar con derechos? Y en caso de que cuente con ellos, ¿de qué tipo de derechos estaríamos hablando? Y otra pregunta relevante ¿ante quién tendría que reclamar sus derechos y por qué razones lo haría? [363]

Y de nueva cuenta, aquí saldría a luz el concepto de la Agencia Moral para determinar la capacidad y personalidad jurídica. Atribuir derechos también nos habla de que ese ente es un agente moral. Si es así, entonces supondría, como ya se señaló en el tercer capítulo, que la IA cuenta con autonomía —en el sentido filosófico de la palabra—, intencionalidad y que es capaz de experimentar el dolor y el placer.

Y también como ya se dijo, el agente moral artificial no es posible en la actualidad. En este tenor, es posible ir vislumbrando la siguiente conclusión: en la actualidad no es viable que un robot ejerza derechos ni que tenga deberes. Quien tiene derechos y guarda deberes es el fabricante, como persona física, incluyendo a todas las personas que participan en el desarrollo y despliegue de la IA. Pero esto no supone que la reflexión deba terminar aquí. En todo caso, resulta conveniente reflexionar con miras al futuro, a fin de establecer bases que contribuyan a justificar la decisión de por qué un robot tendría que contar con derechos y en qué casos existirían límites a esos derechos, similar con lo que ocurre a las restricciones

---

Colegio de Notarios del Distrito Federal, [en línea], <http://historico.juridicas.unam.mx/publica/librev/rev/mexder/cont/2/cnt/cnt2.pdf> [consulta: 2 de octubre de 2023].

[362] DE PINA, R. y R. de Pina Vara, *Diccionario de Derecho*, ed. Porrúa, México, 2010, p. 405.

[363] Estas cuestiones van de la mano con la fórmula básica de un derecho humano como derecho subjetivo presentada por A. Gewirth y expuesta por J.F. Díez Spelz, donde A tiene un derecho a X exigible a B, por virtud de Y. DÍEZ SPELZ, J. F., "¿Robots con derechos?: la frontera entre lo humano y lo no-humano. Reflexiones desde la teoría de los derechos humanos", *op. cit, supra*, nota 346, p. 271.

constitucionales de los derechos humanos de las personas en nuestro sistema jurídico. Así pues, a continuación, se señalan algunas de estas bases, las cuales son tomadas del libro "*Robot Rules, Regulating Artificial Intelligence*", de Jacob Turner.

2.2.1. El argumento del sufrimiento

Para este argumento, la consciencia juega un rol importante. De acuerdo con Jacob Turner, si la IA adquiere dicha cualidad, entonces podría calificar para tener hacía ella algún tipo de consideración moral.[364] En este tenor, un ente consciente tendría que ser capaz de 1) percibir estímulos; 2) percibir sensaciones y, 3) tener sentido de sí mismo, esto es, una concepción de su propia existencia en el espacio y el tiempo.[365]

La pregunta aquí es si la IA es capaz de lo anterior. En particular, si es capaz de sufrir. Al respecto, Jacob Turner escribe:

> If and when an AI becomes conscious, the final question is whether the conscious AI can suffer. AI technology today appears capable of achieving this result—even if it is not yet conscious. Reinforcement learning involves programs analyzing data, making decisions and then being informed by a feedback mechanism whether these decisions were more or less correct. The feedback mechanism qualifies results by assigning them a score according to how desirable the decision was. Each time this process takes place, the computer learns more about its tasks and its environment, gradually honing and perfecting its abilities. Most human children discover at some point that if they touch a sharp object, it can hurt. If pain is just a signal which encourages an entity to avoid something undesirable, then it is not difficult to acknowledge that robots can experience it. In 2016, German researchers published a paper indicating that they had created a robot which could "feel" physical pain when its skin was pricked with a

[364] TURNER, J., *Robot Rules, Regulating Artificial Intelligence*, Palgrave Macmillan, Reino Unido, 2019, p. 146.
[365] *Ibidem*, p. 147.

> pin.[366] [Si una IA llega a ser consciente, la cuestión final es si la IA consciente puede sufrir. La tecnología de IA actual parece capaz de lograr este resultado, aunque todavía no sea consciente. El aprendizaje por refuerzo consiste en que los programas analicen datos, tomen decisiones y luego sean informados por un mecanismo de retroalimentación sobre si esas decisiones fueron más o menos correctas. El mecanismo de retroalimentación califica los resultados asignándoles una puntuación en función de lo deseable que haya sido la decisión. Cada vez que se produce este proceso, el ordenador aprende más sobre sus tareas y su entorno, afinando y perfeccionando gradualmente sus habilidades. La mayoría de los niños descubren en algún momento que, si tocan un objeto afilado, puede doler. Si el dolor es solo una señal que anima a una entidad a evitar algo indeseable, no es difícil reconocer que los robots pueden experimentarlo. En 2016, investigadores alemanes publicaron un artículo en el que indicaban que habían creado un robot que podía sentir el dolor físico cuando se le picaba la piel con un alfiler] (Traducción propia).

Si es posible identificar que un sistema dotado con IA sufre, entonces podemos decir que cuenta con consciencia. Es verdad que es complejo argumentar que el entrenamiento de algoritmos, un proceso de aprendizaje para el sistema de IA representa un indicio de consciencia; no obstante, y como refiere Jacob Turner, la consciencia puede existir en grados, no se trata de algo binario — o se tiene o no se tiene—. Jacob indica que la consciencia puede variar en al menos tres niveles.

El primer nivel, que ocurre en los organismos vivos, se da cuando pasan de un estado de sueño profundo —un estado mínimamente consciente— hasta estar completamente despierto. El segundo nivel ocurre también en los organismos vivos, en especial en los mamíferos que continúan desarrollándose significativamente después del nacimiento, pues un recién nacido parece tener menos consciencia que un adulto completamente desarrollado. El tercer nivel, es el hecho de que la consciencia puede diferir entre las especies.

A decir de este último nivel, Turner señala que no hay ninguna razón lógica por

[366] *Ibidem*, pp. 151-152

la que los seres humanos en un estado de vigilia normal deban ocupar el pináculo de dicha experiencia consciente:

> […] we know that certain animals possess the ability to sense phenomena outside the bounds of human senses or even comprehension. It is generally accepted that humans are limited to five senses through which the world can be experienced: taste, sight, touch, smell and sound. Bats experience the world through sonar. Other animals can sense and act on the basis of electromagnetic waves. Humans can observe such forces via other media, such as visual representations on a computer screen, but we cannot know or even accurately imagine what it is to experience them directly. Owing to these additional senses, some animals are arguably more conscious than us, or at the very least conscious in different ways.[367]
> [Sabemos que ciertos animales poseen la capacidad de percibir fenómenos que escapan a los sentidos humanos o incluso a su comprensión. En general, se acepta que los humanos están limitados a cinco sentidos a través de los cuales se puede experimentar el mundo: el gusto, la vista, el tacto, el olfato y el sonido. Los murciélagos experimentan el mundo a través del sonar. Otros animales pueden percibir y actuar en función de las ondas electromagnéticas. Los humanos podemos observar esas fuerzas a través de otros medios, como las representaciones visuales en la pantalla de una computadora, pero no podemos saber, ni siquiera imaginar con exactitud, lo que es experimentarlas directamente. Gracias a estos sentidos adicionales, algunos animales son posiblemente más conscientes que nosotros, o al menos tienen una consciencia diferente] (Traducción propia).

Suponer entonces que la persona humana es el único ente capaz de vivir una experiencia consciente completa reduce el espectro para siquiera iniciar una reflexión en torno a la consciencia de la IA. Es verdad que en la actualidad no existe evidencia de que la IA sea consciente, sin embargo, no debemos descartar en el futuro de que lo sea:

[367] *Ibidem*, p. 153.

> Susan Schneider of the University of Connecticut considers that AI might bypass what we consider to be consciousness and develop an entirely different form of operating. First, she notes that in humans, consciousness is correlated with novel learning tasks that require concentration, and when a thought is under the spotlight of attention, it is processed in a slow, sequential manner … A superintelligence would surpass expert-level knowledge in every domain, with rapid-fire computations ranging over vast databases that could encompass the entire internet. It may not need the mental faculties that are associated with conscious experience in humans.[368] [Susan Schneider, de la Universidad de Connecticut, considera que la IA podría eludir lo que consideramos la consciencia y desarrollar una forma de funcionamiento totalmente diferente. En primer lugar, señala que en los seres humanos, la consciencia está relacionada con tareas de aprendizaje novedosas que requieren concentración, y cuando un pensamiento está en el punto de mira de la atención, se procesa de forma lenta y secuencial … Una superinteligencia superaría los conocimientos de los expertos en todos los ámbitos, con cálculos rápidos que abarcarían vastas bases de datos que podrían abarcar todo Internet. Es posible que no necesite las mismas facultades mentales que asocian a la experiencia de la consciencia en los seres humanos] (Traducción propia).

Así pues, suponer que la consciencia sólo es posible en los organismos biológicos, y además poner en el punto más alto de la experiencia consciente al ser humano, no contribuye al debate relacionado a si un sistema artificial autónomo puede en algún punto de su existencia percibir estímulos y sensaciones, como el dolor, por ejemplo. Si bien es cierto que es difícil pensar en una IA consciente en la actualidad, también lo es que conforme irrumpan nuevos sistemas, que demuestren — o al menos nos hagan creer— qué sienten, estaremos reflexionando más sobre la posibilidad de que la IA sea consciente en algún grado.

[368] *Ibidem*, p. 154.

2.2.2. El argumento de la compasión

Turner nos dice aquí que nosotros protegemos ciertas cosas porque tenemos una reacción emocional en el caso de que resulten dañadas. Los derechos humanos son este ejemplo, se protegen porque al final se "siente" mal el ver como otra persona sufre. En efecto, la empatía se convierte una razón para crear derechos.[369]

La empatía provoca una respuesta emocional, que permite la colaboración con otros miembros de nuestra especie porque podemos imaginar cómo es ser ellos. El éxito de la raza humana tiene como uno de sus factores la empatía.[370]

Con relación a la IA, el debate en este punto es el momento en el que como seres humanos intervendríamos frente actos que puedan dañar o destruir a las máquinas con IA. El ejemplo al que refiere es de los robots sexuales. Turner, refiriendo al título "*Love and Sex with Robot: The Evolution of Human-robot Relationships*" de David Levy, menciona que llegará el momento en el que los robots alcancen un grado alto de sofisticación, al punto de que sean capaces de satisfacer necesidades psicológicas del ser humano. Pero qué ocurrirá cuando la persona humana, vinculada a un robot sexual, realice con éste actividades que son prohibidas con los humanos: "*It is wrong to allow a human to play out a rape fantasy using a robot as a victim? Would anything change if the robot was aware that it was being used for activities which, if carried out on a human, would be consider moral depraved?*".[371] [¿Está mal permitir que un humano represente una fantasía de violación utilizando un robot como víctima? ¿Cambiaría algo si el robot fuera consciente de que está siendo utilizado para actividades que, si se llevaran a cabo en un humano, se considerarían moralmente depravadas?] (Traducción propia).

Es en este punto en el que el argumento de la compasión entraría en juego: no se quiere dañar al robot, porque al hacerlo se estimula la inmoralidad o ilegalidad de actos, y con ello se perjudica a la sociedad. Si bien esto pudiera traducirse en una prohibición para él o la propietaria del robot, también pudiera dar pie a pensar en maneras de que exista un trato igualitario —similar al que tenemos en las sociedades contemporáneas— para los sistemas con IA:

---

[369] *Ibidem*, pp. 155-156.
[370] *Ibidem*, p. 157.
[371] *Ibidem*, p. 158.

> Would you rather live in, say, a Westworld universe filled with humans who feel free to rape and maim the park's mechanical inhabitants, or on the deck of Star Trek: The Next Generation, where advanced robots are treated as equals? The humans of one world seem a lot more welcoming that the other, don't they?. [372] [¿Preferirías vivir, por ejemplo, en un universo de Westworld lleno de humanos que se sienten libres de violar y mutilar a sus habitantes mecánicos del parque, o en la cubierta de Star Trek: The Next Generation, donde los robots avanzados son tratados como iguales? Los humanos de un mundo parecen mucho más acogedores que los del otro, ¿no?] (Traducción propia).

En efecto, la compasión puede convertirse en una razón sobre la cual fundar la idea de dotar de derechos, al menos el de no ser objeto de tratos degradantes, a los robots.

2.2.3. El argumento desde el valor que los robots tienen para la humanidad

Este argumento consta de dos ideas. La primera de ellas tiene que ver con la reciprocidad en el respeto hacía la IA. El argumento se construye de la siguiente forma: suponiendo que la IA sea racional y busque preservarse a sí misma y a sus intereses, entonces, adoptar una actitud de coexistencia mutua provocará una actitud similar por parte de la IA hacía con los humanos. [373]

La segunda idea parte del valor inherente de la IA. Al igual que a lo largo de la historia el ser humano ha dado un valor a objetos por razones culturales e históricas, podría llegar el momento en el que sea necesario darle un valor a la IA por su importancia para el ser humano:

> For instance, it may be necessary for legal forensic purposes to preserve

[372] DASHEVSKY, E., "Do Robots and AI Deserve Rights", en *PC Magazine*, [16 de febrero de 2017, en línea], <http://uk.pcmag.com/robotics-automation-products/87871/feature/do-robots-and-ai-deserve-rights> [consulta: 2 de octubre de 2023] *apud* TURNER, J., *Robot Rules*, Regulating Artificial Intelligence, *op. cit*., *supra*, nota 364, p. 159.

[373] *Ibidem*, p. 164

> a version of an AI system at the time a relevant event took place so as to be able to make enquiries as to its functioning and thought process. Similarly, if an update or patch leads to unforeseen problems, it may well be necessary to "roll back" a program to its previous version in order to rectify the issue. Both of these motivations underline the importance to humanity of preserving types of AI in some way.[374] [Por ejemplo, puede ser necesario, a efectos forenses, conservar una versión de un sistema de IA en el momento en que se produjo un hecho relevante para poder hacer averiguaciones sobre su funcionamiento y proceso de pensamiento. Del mismo modo, si una actualización o un parche provocan problemas imprevistos, puede ser necesario revertir un programa a su versión anterior para rectificar el problema. Ambas motivaciones subrayan la importancia que tiene para la humanidad preservar de algún modo los tipos de IA] (Traducción propia).

Probablemente, este argumento sea el más débil al momento de justificar la posibilidad de que la IA cuente con derechos. No basta con que un objeto sea valioso para que este cuente con derechos, pues como ya se ha dejado entrever, es necesario que ese ente puede ser calificado como un agente moral. Sin embargo, este argumento sirve en vías de construir razones para tener consideración moral hacía las IA, pues no se querrá destruir o dañar algo que, por su valor histórico o cultural, es de valor para la humanidad.

2.2.4. El argumento desde lo posthumano

Por último, encontramos el argumento que toma como base el transhumanismo. Como ya se vio, el transhumanismo apuesta por la mejora humana a través de las tecnologías. Turner invita a la reflexión al plantearse qué ocurriría si los humanos que han sido mejorados por la IA, ¿deberíamos dejar de protegerlos? El ejemplo que da es el siguiente: "*We would not deny someone their human rights if they were 1% augmented by AI. What about if 20%, 50% or 80% of their mental functioning*

[374] *Ibidem*, p. 166

was the result of computer processing powers?"[375]. [No negaríamos a alguien sus derechos humanos si tuviera un 1% de aumento por la IA. ¿Y si el 20%, el 50% o el 80% de su funcionamiento mental fuera el resultado de los poderes de procesamiento informático?] (Traducción propia).

La respuesta inicial es que no debería perder derechos sólo por haber aumentado su funcionamiento mental; sin embargo, y en consonancia con la idea de John Searle, podría llegar el momento en que la sustitución elimine gradualmente a la experiencia consciente.[376]

Turner también refiere que el remplazo o aumento de funciones físicas humanas con artificiales no implica que alguien merezca más o menos derechos:

> Someone who loses an arm and has it replaced with a mechanical version is not consider less human. The same argument might be made in the future, for instance if someone suffers a brain injury causing persisting amnesia and undergoes surgery to fit a processor replacing this mental function.[377] [Alguien que pierde un brazo y lo sustituye por una versión mecánica no se considera menos humano. El mismo argumento podría esgrimirse en el futuro, por ejemplo, si alguien sufre una lesión cerebral que le provoque una amnesia persistente y se somete a una operación quirúrgica para que le coloquen un procesador que sustituya esta función mental] (Traducción propia).

Así pues, la idea de lo transhumano sirve también como un buen parámetro al momento de reflexionar sobre los derechos de los que pudiera ser objeto la IA. Un humano biológicamente aumentado por la IA, y la tecnología en general, es merecedor de derechos, básicamente porque se trata de un agente moral.

A manera de conclusión de este apartado, pensar que en el presente un sistema dotado de IA es merecedor de derechos resulta una idea poco fundamentada. Afirmar que en la actualidad les debemos consideración moral también resulta

[375] *Idem.*
[376] *Ibidem*, p. 167
[377] *Ibidem*, p. 168

controversial. No obstante, pienso que conforme la IA se vuelva más sofisticada, siendo capaz de tomar formas humanas y de expresar sentimientos humanos, será posible encontrar justificaciones para pensar que un robot debería tener derechos, con las limitaciones que el propio sistema normativo determine necesarias.

Antes de que esto ocurra sería necesario reflexionar también sobre las implicaciones que conlleva el reconocer derechos a la IA. Y más importante aún, en el caso de que les sean reconocidos derechos, ¿también la IA tendrá dignidad?, recordemos que este es un atributo de los seres humanos, y que poco a poco ha ganado terreno con los derechos de los animales —la muestra de que la idea de derechos no es inmutable, sino más bien cambiante, atendiendo las necesidades de las sociedades, particularmente las occidentales, contemporáneas—.

Y así pueden enlistarse otras implicaciones. Para el derecho, el reconocimiento de derechos a la IA iría de la mano con reconocer que los robots tienen obligaciones hacía nosotros, igual como nosotros las tendríamos para ellos. Esto es ya de por sí problemático, ¿qué ocurriría si la IA decide no actuar conforme a su obligación de no dañar a una persona humana? ¿si tienen derechos, son agentes morales artificiales, entonces serían completamente responsables de sus actos?

En fin, vale la pena ir reflexionando sobre este tema, pero también vale la pena cerrar este apartado con la siguiente idea: si pensamos que un robot será merecedor en el futuro de derechos, entonces podríamos decir que, en el futuro, fallamos como desarrolladores de esta tecnología. Pienso que la IA, por más sofisticada que sea, no merece derechos, pues se tratan de objetos que fueron desplegados por la creación humana para el servicio de la humanidad y, en su caso, para la protección del entorno en el que habita. Sin embargo, pienso que la IA con un alto grado de sofisticación sí tendría que ser objeto de consideración moral. En efecto, no tendría que justificarse de ninguna manera tratos inmorales, antiéticos o ilegales hacía una IA avanzada que sea capaz de sentir.

3. REFLEXIONES FINALES

En este capítulo se identificaron problemas actuales y futuros asociados al desarrollo de la IA. En el presente, se enmarca a la discriminación algorítmica, la

preocupación en torno a la privacidad en la era de la IA, el impacto de la IA en el mundo laboral y la desinformación provocada por dicha tecnología. A decir de estos cuatro problemas enunciados, todos parecen tener soluciones con base en la regulación, el desarrollo de estándares técnicos y en la implementación de políticas públicas.

En el caso de la discriminación algorítmica, una parte de la solución involucra que los algoritmos cuenten con datos lo suficientemente diversos para que puedan aprender y hacer predicciones que tomen en consideración a grupos históricamente discriminados; esto se complementa con el enfoque regulatorio, el cual implica evaluar y fortalecer los actuales marcos existentes en materia de igualdad y no discriminación.

La preocupación en torno a la privacidad debe mirarse desde un enfoque regulatorio, pero también técnico-regulatorio porque implica ajustar los principios sobre los cuales giran las leyes en materia de datos personales a la realidad impuesta por la IA y desde el técnico porque los desarrollos tecnológicos deben adoptar desde su diseño estándares orientados a la protección de la privacidad de sus usuarios.

Con relación al impacto de la IA en el campo laboral, y frente a una posible sustitución del empleo humano provocado por la IA, es necesario identificar en primer término qué empleos, ocupaciones, tareas, sectores, perfiles profesionales y personales serán los menos favorecidos ante la IA. Adicionalmente, existen propuestas como la de la Renta Básica Universal para hacer frente a escenarios en los que el desempleo provocado por la IA sea una constante; de igual manera, también estás las propuestas que apuestan por garantizar los derechos sociales de los trabajadores impactados por la IA, priorizando que éstos cuenten con competencias y habilidades para la inserción en el nuevo patrón de trabajo.

La desinformación generada por los videos hiperrealísticos llamados “*Deep Fakes*”, implica un serio reto, pues pensar en prohibirlos supone abrir el debate entre la delgada línea que separa los “*Deep Fakes*” que atentan contra las libertades y derechos, de la libertad de expresión. La idea para contrarrestar los efectos perjudiciales que pueden traer consigo, es la de generar capacidades en las

personas, con la finalidad de discernir videos falsos de reales; adicionalmente, también se requiere contar con tecnología que permita detectar fácilmente los "*Deep Fakes*".

Ahora bien, en lo que respecta a los problemas futuros asociados al desarrollo de la IA, ha de decirse que estos traen consigo serios desafíos éticos y jurídicos. Hablar, por ejemplo, de transhumanismo implica cuestionarse qué ocurriría en el caso de que un ciborg cuyas aplicaciones neuronales mejoradas fallen mate a una persona, ¿será posible responsabilizarlo jurídicamente?, en el supuesto de que sea posible utilizar a la IA como potenciadora de la inteligencia humana, ¿Qué debe hacerse desde el campo jurídico y de las políticas públicas para evitar que se generen y amplíen brechas de desigualdad entre aquellos que tienen el privilegio de acceder a dicha mejora y entre una clase social que no tendrá esa oportunidad?

Otro problema que se asoma a la luz del desarrollo de la IA es el de si es posible dotar a los robots de derechos. Este debate debe analizarse desde distintas aristas, a saber: 1) si el robot tiene consciencia con la cual pueda experimentar dolor, entonces puede justificarse que tengan derechos; 2) si el dañar a un robot supone una inmoralidad o ilegalidad, o si con ello se perjudica un valor social, entonces también podría justificarse que este tenga derechos, en particular para no ser objeto de tratos degradantes; 3) si la IA tiene un valor por su importancia para el ser humano o en caso de que desarrolle un sentido de preservación hacía ella misma, entonces cabe la posibilidad de dotarle de derechos, y 4) las implicaciones que lo posthumano tiene frente al tema, particularmente la utilidad de los problemas asociados a ello para reflexionar sobre los derechos de los que pudiera ser objeto la IA.

CAPÍTULO V
LOS COMITÉS DE ÉTICA EN INTELIGENCIA ARTIFICIAL Y LA AGENCIA MORAL ARTIFICIAL

Ya se ha mencionado en este trabajo que la aparición del Agente Moral Artificial es improbable en nuestra actualidad. Sin embargo, suponiendo que en un futuro no muy lejano el poder de la IA supere las expectativas que hoy día se tienen sobre ella, siendo capaz de traer consigo a sistemas con superinteligencia, los cuales a su vez puedan crear Agentes Morales Artificiales, seres con la capacidad de sentir placer y dolor, ¿qué reglas y regulaciones tendría que haber producido el Derecho para evitar que los agentes morales artificiales se conviertan en una amenaza incluso para la propia existencia humana?

Seguramente llegados a ese momento, existirán reglas y regulaciones que establezcan responsabilidades y sanciones tanto a los desarrolladores, dueños e incluso a los propios agentes morales artificiales; sin embargo, ¿cuál tendría que ser un primer paso que permita zanjar el terreno a todo el marco regulatorio que para ese entonces exista? A esta pregunta, podría decirse que los Comités de Ética podrían ser esa primera acción necesaria de cara a los Agentes Morales Artificiales.

En ese sentido, este apartado analiza a los Comités de Ética en IA como una forma de regulación frente al desarrollo y uso de la IA. Cabe mencionar, que se optó por el análisis de dichos Comités, en virtud de que son un complemento a formas de regulación dura, como pueden ser las normas vinculantes, pero particularmente por el enfoque que tienen en los aspectos éticos relacionados con el desarrollo y uso de dicha tecnología.

En un primer momento, se describe cómo la Ética sirve reflexionar críticamente frente al desarrollo tecnológico, ubicando a la Ética de la IA, una ética práctica, como aquella que tiene como tarea orientar, mediante valores y principios, el desarrollo de la IA.

Posteriormente, se analiza el rol de los Comités de Ética, mencionando los tipos de Comité existentes, haciendo una acotación particular en los Comités de Bioética y en los principios éticos básicos que emplean y que en su conjunto sirven como método argumentativo para deliberar frente a dilemas éticos.

Asimismo, se describen a los Comités de Ética en IA, describiendo su composición, núcleo de principios y el rol de éstos de cara a los Agentes Morales Artificial. Se concluye con la propuesta de que la Universidad Nacional Autónoma de México cuente con un Comité de Ética en IA, a fin de convertirse en un referente a nivel nacional en el tema de Ética de la IA.

1. LA ÉTICA COMO UNA REFLEXIÓN CRÍTICA FRENTE AL DESARROLLO TECNOLÓGICO

La era actual no ha sido la única que ha despertado inquietudes con relación al desarrollo tecnológico y sus consecuencias, de hecho, si se mira al pasado, particularmente a la Revolución Industrial del siglo XIX, es posible identificar una oposición a la industrialización, por el temor de que los trabajadores manuales fuesen desplazados por las máquinas.[378] Algo parecido ocurrió también en los años 50 y 70 del siglo XX, con la llamada Revolución Verde, la cual permitió un auge descomunal en la agricultura gracias a los descubrimientos de la época en el campo de la química y la genética, pero con efectos perjudiciales en el ecosistema.[379]

En el contexto de estos debates, la Ética desempeña un papel central al abordar problemas derivados del desarrollo tecnológico, más aún cuando se trata de tomar decisiones complejas. En la actualidad, por ejemplo, algunas preguntas asociadas al consumo de información que podrían responderse desde el análisis ético podrían

[378] *Vid.* CHAVES PALACIOS, J., "Desarrollo tecnológico en la primera revolución industrial", en Norba. *Revista de Historia*, vol. 17, 2004, p. 98-99.

[379] *Vid.* CECCON, E., "La revolución verde en dos actos", en *Ciencias*, Universidad Nacional Autónoma de México, vol. 1, no. 91, 2008, pp. 21-22.

ser las siguientes: 1) ¿cómo deben los usuarios de Internet comportarse frente a las inmensas cantidades de información que existen?, y 2) ¿por qué se vuelve relevante validar las noticias que se consumen directamente de redes sociales? Otras preguntas, asociadas al auge de las criptomonedas, son las siguientes: 1) ¿es la minería un mecanismo sostenible y justo?, y 2) ¿cómo garantizar la redistribución ética de los recursos y dividendos en un ecosistema descentralizado?

En síntesis, se puede decir que la Ética se posiciona como una reflexión crítica frente al desarrollo tecnológico, específicamente frente a sus nuevos problemas. Sin embargo, esta afirmación requiere de un sustento teórico que nos permita ir más adelante, particularmente si lo que busca defenderse es que la IA requiere de una ética aplicada que permita su desarrollo conforme a principios establecidos.

Así, pues, en este apartado se identificará lo qué es la Ética, su diferencia con la moral, de igual manera se distinguirá a la ética general de la ética práctica, en la cual situaremos a la Ética de la IA.

1.1. Aproximación al concepto de Ética

Se entiende a la Ética como la disciplina filosófica que analiza el lenguaje moral y que ha elaborado diferentes teorías y maneras de justificar o de fundamentar y de revisar críticamente las pretensiones de validez de los enunciados morales.[380] Aquí vale la pena hacer la siguiente acotación, la Ética es distinta a la moral, pues esta última se refiere a "*un conjunto de principios, mandatos, prohibiciones, permisos, patrones de conducta, valores e ideales de la vida buena que en su conjunto conforman un sistema más o menos coherente, propio de un colectivo humano concreto en una determinada época histórica*."[381]

En efecto, no debemos confundir a la Ética con la moral. La Ética, como refiere Adolfo Sánchez Vázquez, se encuentra con una serie de morales ya dadas, y partiendo de ellas trata de establecer la esencia de la moral, su origen, las condiciones objetivas y subjetivas del acto moral, las fuentes de valoración de la moral, la naturaleza y función de los juicios morales, los criterios de justificación de

[380] DE ZAN, J., *La ética, los derechos y la justicia*, Fundación Konrad-Adenauer, Uruguay, 2004, p. 19.
[381] CORTINA, A. y E. Martínez, *Ética*, Editorial Akal, España, 3ª edición, 2001, p. 14.

dichos juicios y el principio que rige el cambio y sucesión de diferentes sistemas morales.[382]

Se dice también que la Ética es una reflexión sobre las cuestiones morales, pues con ella se pretende desplegar conceptos y argumentos que permitan comprender la dimensión de la persona humana en cuento a tal dimensión moral.[383] Así pues, la Ética se sitúa como la disciplina filosófica que tiene por objeto de estudio a la moral.

A la Ética le interesa analizar lo que acontece en el mundo de la moral, y en ese sentido, cumplir con una triple función: 1) aclarar qué es lo moral, cuáles son sus rasgos específicos; 2) fundamentar la moralidad, es decir, tratar de averiguar cuáles son las razones por las que tiene sentido que los seres humanos se esfuercen en vivir moralmente, y 3) aplicar a los distintos ámbitos de la vida social los resultados obtenidos en las dos primeras funciones, de manera que se adopte en esos ámbitos sociales una moral crítica, es decir, racionalmente fundamentada, en lugar de un código moral dogmáticamente impuesto o de la ausencia de referentes morales.[384]

No es la intención de este apartado hacer un recorrido histórico sobre la Ética, en cambio, lo que se pretende es entenderla y ubicarla como una disciplina filosófica que estudia la moral, y a partir de esto, distinguirla de las éticas aplicadas que en las últimas décadas se ha desarrollado en el campo de la Ética, pues será ahí en donde habrá de ubicarse a la Ética de la IA.

En efecto, a la Ética como disciplina filosófica que estudia la moral habremos de situarla dentro de una categoría que estableceremos aquí como ética general, dentro de la cual se ubican las grandes teorías éticas, por ejemplo, las Éticas occidentales, como pueden ser la Ética de la virtud, de Aristóteles, la Ética sentimentalista de David Hume, la Ética del deber de Immanuel Kant o la Ética utilitarista, de Jeremy Bentham, por mencionar algunas de ellas.

Por su parte, las éticas aplicadas aluden a algunas nuevas especialidades que destacan la resolución práctica: "*La ética aplicada debe ser vista como una actividad interdisciplinaria en la que se procura resolver racionalmente problemas profesionales que se plantean en situaciones complejas, en las que intervienen*

[382] SÁNCHEZ VÁZQUEZ, A., Ética, Debolsillo, México, 2009, p. 22.
[383] CORTINA, A. y E. Martínez, Ética, *op. cit*., *supra*, nota 381, p. 9.
[384] *Ibidem*, p. 23.

distintas ciencias".[385] Algunos ejemplos de éticas aplicadas son la Ética de los Negocios, la Ética del Deporte y la Bioética.

El contexto es relevante para la ética aplicada, pues al momento de la resolución de casos se da un valor al análisis de consecuencias y a la toma de decisiones. La Bioética, por ejemplo, se caracteriza por la casuística asociada a ella, pues cada caso que enfrenta es distinto, así como también por ser dilemática, esto es que existe argumentos a favor y en contra, lo cual a su vez implica que, para ofrecer una solución al problema al cual se hace frente, es necesario comprender ambas caras de la moneda. Se buscan consensos, pero también se asumen los disensos que se pudieran generar.

La ética teórica y la ética aplica no son ajenas entre sí, ya que la última hace uso de los principios desarrollados desde alguna teoría ética. Por ejemplo, el utilitarismo —teoría que busca que se alcance un alto número de consecuencias positivas para el máximo número de personas[386]— se hizo presente en la Bioética cuando se decidió vacunar contra el covid-19 primero al personal médico antes que a las juventudes, el principal argumento es que con esa decisión se desea preservar la estabilidad social en medio de la pandemia, pues al proteger a los médicos se garantizaba su disponibilidad para atender a los contagiados. [387]

Teniendo en cuenta la distinción entre la ética teórica y la ética aplicada, es adecuado hablar de la Ética de la IA. La justificación de abordar esta ética aplicada es contar con una mejor comprensión de la necesidad actual de reflexionar críticamente en torno a los problemas que se suscitan con el desarrollo de la IA, preparando además el terreno para analizar el rol que los Comités de Ética en IA tendrían que desempeñar en el presente, pero también en el futuro, específicamente con relación al desarrollo de los agentes morales artificiales.

---

[385] DE ZAN, J, La ética, los derechos y la justicia, *op. cit*., *supra*, nota 380, pp. 38 - 39.

[386] *Vid.* ARNAU, H., J. M. Gutiérrez y G. Navarro, ¿Qué es el utilitarismo?, *Colección Universitas-39*, PPU, 1992, pp. 10-11.

[387] *Vid.* REUTERS*, Primero médicos y ancianos, México se prepara para aplicar vacuna contra COVID-19,* 8 de diciembre de 2020*,* [en línea], <https://www.reuters.com/article/salud-coronavirus-mexico-vacunas-idLTAKBN28I1YG>, [consulta: 03 de octubre de 2023].

2. LA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

En el cuarto capítulo de este trabajo se habló de algunas de las repercusiones actuales de la IA. Asuntos como la discriminación algorítmica y el desplazamiento laboral provocado por dicha tecnología, además de levantar una serie de preocupaciones en la sociedad, también han provocado un profundo cuestionamiento ético. Aludiendo a Adela Cortina:

> Y sigue siendo cierto que el exponencial progreso de lo que hoy llamamos ya 'tecnociencias' plantea una gran cantidad de preguntas éticas para las que es necesario ir encontrando respuestas. Precisamente porque lo moral no consiste en mapas de carreteras, ya cerrados, sino en una brújula que señala el norte, es posible y necesario encontrar mejores caminos ante los nuevos descubrimientos tecnocientíficos.[388]

Adela Cortina trae a colación el caso de Michihito Matsuda, un robot androide con rasgos femeninos que se presentó como uno de los candidatos para las elecciones municipales en un distrito de Tokio llamado Tama New Town. Sorprendentemente, Michihito logró quedar en tercer puesto en la segunda vuelta con 4,013 votos, y entre sus promesas estaba el de acabar con la corrupción y ofrecer oportunidades justas y equilibradas para todos. [389]

Michihito funcionaba gracias a todo un equipo de personas humanas, que apostaban al uso de un algoritmo capaz de sustituir las debilidades emocionales de los seres humanos, causa de malas decisiones políticas, corrupción, nepotismo y conflictos, por un análisis objetivo de los datos generados acerca de las opiniones, expectativas, preferencias y costumbres de la ciudadanía.[390]

Con esta historia, Adela pone sobre la mesa una cuestión ética, la cual sirve como preámbulo para abordar a la Ética de la IA:

> Que Matsumoto [el humano que desarrolló a Michihito] se presente a las

[388] CORTINA, A., "Ética de la Inteligencia Artificial", en *Anales de la Real Academia de Ciencias Morales y Políticas*, no. 96, 2019, p. 379.
[389] *Ibidem*, p. 380.
[390] *Idem*.

> elecciones y, si gobierna, se sirva de Michihito como ayuda para tomar decisiones, no es lo mismo que poner el gobierno en manos de Michihito. Con los grandes avances en temas de IA, ¿se trata de que los seres humanos utilicen los sistemas inteligentes como instrumentos o de que estos sustituyan a los seres humanos? Y aquí surge la gran pregunta: ¿una ética de la inteligencia artificial es la que deben practicar los sistemas inteligentes desde sus propios valores, o es la que los seres humanos deberíamos adoptar para servirnos de los sistemas inteligentes?[391]

Ambas cuestiones reflejan la preocupación de que el desarrollo de la IA ocurra sin antes haber establecido las bases éticas que faciliten su convivencia con los seres humanos. Pero ¿cómo establecer esas bases? Es aquí donde adquiere sentido pensar en una ética aplicada, como lo es la Ética de la IA.

Por Ética de la IA habremos de entender a una ética aplicada que trata de orientar, a través de valores y principios, el desarrollo de la IA. Esta Ética también se ocupa del cambio tecnológico y su impacto en la vida de los individuos, pero también de las transformaciones en la sociedad y en la economía.[392]

En la actualidad existen muchas iniciativas que buscan poner en marcha a la Ética de la IA. Por ejemplo, el Parlamento Europeo publicó en el 2018 el primer borrador de la Guía Ética para el uso responsable de la IA. Esta Guía enuncia estándares morales dirigidos a los desarrolladores de IA, tales como: asegurar que la IA esté centrada en el ser humano; prestar atención a los grupos vulnerables; respetar los derechos fundamentales y la regulación aplicable; ser técnicamente robusta y fiable; funcionar con transparencia y no restringir la libertad humana.[393]

También existen iniciativas hechas desde la comunidad técnica. Por ejemplo, los estándares propuestos por la IEEE, los principios sobre IA hechos por el *Future of Life Institute*, o la declaración para el desarrollo responsable de la IA, elaborado por

[391] *Ibidem*, p. 381.
[392] COECKELBERGH, M., *AI Ethics*, The MIT Press, Estados Unidos de América, 2020, p. 7.
[393] VILLALBA, J., "Algor-ética: la ética en la inteligencia artificial", en *Revista Anales de la Facultad de Ciencias Jurídicas y Sociales*, Universidad Nacional de la Plata, no. 50, 2020, p. 684.

el *Forum of the Socially Responsable Development of Artificial Intelligence*.[394]

Sin embargo, pese a los esfuerzos hechos, sigue sin ser muy claro qué tendría que hacerse, o qué camino en particular debería emprenderse para hacer frente a los problemas éticos generados por el desarrollo de la IA:

> Today there are a lot of initiatives in this area and this should be applauded. However, it is not clear what should be done, what precise course of action should be taken. For example, it is not so clear how to deal with transparency or bias, given the technologies as there are, existing bias in society, and divergent views on justice and fairness. There are also many possible measures to choose from: policy can mean regulation by means of laws and directives, say, legal regulation, but there are also other strategies that may or may not be connected to legal regulation, such as technological measures, codes of ethics, and education. And within regulation there are not only laws but also standars such as ISO norms. Moreover, other sorts of questions also need to be answered in policy proposals: not only what should be done, but also why it should be done, when it should be done, how much should be done, by whom it should be done, and what the nature, extent, and urgency of the problem are.[395] [Hoy en día hay muchas iniciativas en este ámbito y hay que aplaudirlas. Sin embargo, no está claro qué hay que hacer, qué curso de acción precisa hay que seguir. Por ejemplo, no está tan claro cómo lidiar con temas como el de la transparencia o el de los sesgos, dadas las tecnologías tal y como son, los prejuicios existentes en la sociedad y las opiniones divergentes sobre justicia y equidad. También hay muchas medidas posibles entre las cuales elegir: políticas de actuación puede significar regulación a través de leyes y directivas, en otras palabras, regulación jurídica, pero también hay otras estrategias que pueden o no pueden estar relacionadas con la regulación jurídica, como medidas tecnológicas, códigos éticos, y educación. Y dentro de la regulación no solo hay leyes, también normas, cómo la ISO. Además, en las propuestas

[394] *Ibidem*, p. 688.
[395] COECKELBERGH, M., *AI Ethics*, *op. cit.*, *supra*, nota 392, pp. 145-146.

> de política de actuación también hay que responder a otro tipo de preguntas: no sólo qué debe hacerse, sino también por qué debe hacerse, cuándo debe hacerse, quién debe hacerlo y cuál es la naturaleza, el alcance y la urgencia del problema] (Traducción propia).

De manera general, las distintas iniciativas que proponen una ruta ética a los problemas asociados al despliegue y desarrollo de la IA no están necesariamente orientadas a prohibir cosas, ni tampoco se encuentran restringidas a un tipo de actor en particular. Ya anteriormente se daba cuenta de iniciativas de instancias gubernamentales y también de la comunidad técnica, a estas hemos de incluir también al sector privado y a la comunidad académica.

Referente al sector privado, encontramos la alianza en IA, por parte de compañías como DeepMind, IBM, Intel, Amazon, Apple, Sony y Facebook. En lo que respecta a Google, esta empresa ha publicado sus principios éticos para el desarrollo de la IA, los cuales son: ofrecer beneficios sociales, evitar o crear sesgos injustos, reforzar la seguridad, promover la excelencia científica y limitar aplicaciones potencialmente dañinas o abusivas, como pueden serlo armas o tecnologías que violen los principios de la ley internacional y los derechos humanos.[396]

Otro caso es el de Microsoft, con su iniciativa "*IA para el bien*", la cual propone principios como el de la justicia, fiabilidad e inocuidad, privacidad y seguridad, inclusividad, transparencia y responsabilidad. Accenture, por su parte, apuesta a principios como el de la privacidad, la inclusión y la transparencia.[397]

Aquí vale la pena señalar que las empresas también reconocen como algo necesario a la regulación externa, y no solamente apostar por la autorregulación. Por ejemplo, Tim Cook, CEO de Apple, ha señalado que la regulación tecnológica es inevitable en aras de preservar la privacidad, debido a que el libre mercado no ha cumplido bien con esta función.[398]

En lo que respecta a la academia, encontramos ejemplos como el de la

[396] *Ibidem*, p. 159.
[397] *Idem.*
[398] *Idem.*

Declaración de IA responsable de Montreal, propuesta por la Universidad de Montreal, y cuyo punto central está en el afirmar que el desarrollo de la IA debe promover el bienestar de todas las criaturas sintientes y la autonomía de los seres humanos, eliminar todo tipo de discriminación, respetar la privacidad personal, protegernos de la propaganda y la manipulación, promover el debate democrático y responsabilizar a las distintas partes interesadas a la hora de trabajar contra los riesgos de la IA. También existen voces académicas, como es el caso de Luciano Floridi, que han propuesto principios como el de beneficencia, no maleficencia, autonomía, justicia y explicabilidad. Por su parte, universidades como la de Cambridge y Standford trabajan en el campo de la Ética de la IA, y la Universidad de Santa Clara, a través de su Centro Markkula para la Ética Aplicada, ofrece una serie de teorías éticas como base para la práctica tecnológica y de ingeniería, muy útiles para la ética de la IA.[399]

Aquí una primera acotación: la Ética de la IA implica hablar de una ética aplicada que echa mano de principios y valores para hacer frente a los desafíos vinculados al desarrollo de la IA. Relacionado con lo anterior, actualmente existen distintas iniciativas, propuestas por distintos actores interesados, que ofrecen bases éticas para el abordaje de los dilemas éticos aparejados al desarrollo de la IA.

De cara a esta primera acotación, se debe enfatizar lo identificado por Mark Coeckelbergh, cuando señala que en la actualidad las iniciativas relacionadas con las políticas de actuación, las cuales incluyen una Ética de la IA, son numerosas; pero que aún sigue sin estar claro qué es lo que debería hacerse ni cuáles son las líneas de actuación más apropiadas.[400]

En efecto, la Ética de la IA es una ética aplicada en desarrollo. Es también una muestra de que la ética no es algo estático, sino más bien algo dinámico, que evoluciona. José Ignacio Latorre muestra con claridad este aspecto de la ética:

> La ética evoluciona, de la misma manera que las leyes que nos rigen se han adaptado a la realidad social que viene condicionada por todo gran

[399] *Ibidem*, p. 157.
[400] *Ibidem*, p. 145.

> cambio tecnológico. Sociedades primitivas podían usar el precepto del ojo por ojo, hoy no es posible. Si alguien choca con su coche por detrás del nuestro, no zanjamos el asunto dándole otro golpe por detrás a su coche. Si alguien espía nuestro correo electrónico, necesitamos nuevas leyes que no existían en el siglo V. La evolución del sistema jurídica es lenta pero necesaria. La evolución de la ética también es lenta, esquiva, a veces retrógrada, pero inevitable.[401]

Así pues, frente a problemas de una sociedad avanzada que, apuesta por la integración de la IA en sus distintos ámbitos, se vuelve necesario reflexionar críticamente sobre los dilemas éticos que se asocian a esa tecnología. De ahí la importancia y la necesidad de impulsar a la Ética de la IA.

La Ética de la IA está en auge, y seguramente lo estará durante un lago tiempo. En un futuro próximo, refiere José Ignacio Latorre, las discusiones referentes a la Ética de la IA — o Roboética, como la llama dicho autor— estarán dominadas por tres ideas, a saber: "*I) Los robots operan de forma autónoma. La ausencia de supervisión puede ser crítica; II) Los robots restan puestos de trabajo, y III) Los robots pueden llevar armas, pueden matar a humanos*".[402]

Como complemento de estas tres ideas, Ignacio Latorre indica lo siguiente:

> Cada vez que un robot cometa un error, las leyes deberán saber cómo y a quién juzgar. Si un robot sustituye a un humano, aparece un problema económico y ético. Todo robot asesino conlleva decadencia moral. Dejar que la creación de robos evolucione libremente es un gran error. De ahí que instituciones, empresas y gobiernos se estén apresurando a establecer códigos internos, manifiestos e incluso leyes que controlen los robots.[403]

Así pues, la Ética de la IA se muestra como una solución frente a los desafíos que la IA ha traído y seguirá trayendo consigo. Esta Ética se refleja en iniciativas,

[401] LATORRE SENTÍS, J.I., *Ética para Máquinas*, Editorial Planeta, España, 2019, p. 157.
[402] *Ibidem,* p. 193.
[403] *Idem.*

tanto técnicas como no técnicas. En este último rubro hemos de situar a las guías y códigos de ética, así como las distintas regulaciones existentes en la materia.

Con relación a las iniciativas técnicas, hay algunas que vale la pena identificar en este trabajo. La primera de ellas refiere a la "Ética por diseño", un método técnico que ha recibido mucha atención en los últimos años. La "Ética por diseño" implica que desde el diseño previo de los algoritmos que controlan a los sistemas inteligentes, se pueda garantizar el comportamiento ético de estos. A su vez, dentro de la "Ética por diseño", es posible encontrar distintos planteamientos, como lo es "*security-by-desing*" o "*privacy-by-design*". Las maneras a través de las cuales es posible lograr esto son varias:

> En primer lugar, se podría hacer que los sistemas y robots provistos de IA adquieran patrones éticos de conducta 'observando' el comportamiento humano en situaciones concretas. Un segundo planteamiento consistiría en establecer una serie de normas que el sistema siempre debería seguir o enumerar los comportamientos y acciones que siempre debería evitar. Por último, los sistemas inteligentes podrían adquirir los principios éticos para orientar su conducta según la situación en la que se encontrasen. En algunas situaciones, prevalecerían los principios de carácter universal, mientras que, en otras, se daría más importancia al comportamiento que a las circunstancias específicas demandan.[404]

Otra solución técnica propuesta es la de la IA explicable, la cual también ha adquirido relevancia en los últimos años, la intención de la IA explicable es contar con sistemas que sean transparentes, esto es, que dispongan de mecanismos para mostrar de forma clara su funcionamiento y su razonamiento interno, la intención de esto es disipar la opacidad de muchos de los actuales sistemas de IA, particularmente aquellos dentro del campo del Deep Learning. [405]

Un método más es el de "*Prueba y validación del producto*", el cual busca

[404] MARÍN GARCÍA, S., *Ética e Inteligencia Artificial*, Cuadernos de la Cátedra CaixaBank de Responsabilidad Social Corporativa, no. 42, septiembre de 2019, Universidad de Navarra, España, p. 20.
[405] *Idem*.

garantizar la seguridad de los dispositivos y asegurar que su diseño y programación responda de manera adecuada a lo planeado, sometiéndolos a mecanismos de examen y validación:

> Este tipo de pruebas permiten comprobar que los diversos componentes funcionen correctamente y sirven para detectar de forma temprana posibles fallas o consecuencias imprevistas. Estas pruebas ya han desempeñado un papel clave en el desarrollo de algunas aplicaciones de Inteligencia Artificial. Por ejemplo, los vehículos automáticos desarrollados por varias compañías han sido sometidos durante los últimos años a distintos exámenes y comprobaciones. Los fallos y accidentes ocurridos durante estas pruebas han permitido a los ingenieros desarrollar sistemas de navegación más precisos y seguros.[406]

Como es posible identificar, todas estas soluciones técnicas implican la intervención humana, particularmente en los patrones de conducta ética que debe aprender la IA, los cuales deben ser enseñados por seres humanos, al menos en la actualidad.

En ese sentido, en el presente, la tarea de los desarrolladores de IA es definir esa "*función error*"[407] que dictará lo que aprenda una red neuronal artificial, mediante el entrenamiento supervisado, entrenamiento no supervisado o entrenamiento por refuerzo. La Ética de la IA también se codifica en la "*función error*" que deseamos minimizar, y aquí lo valioso sería cuestionarse quién ha de decidir la Ética que se

---

[406] *Ibidem*, pp. 20-21.

[407] José Ignacio Latorre señala que por cada error cometido por una red neuronal artificial, nos es posible imponerle una penalización: "*Si la red acierta cuando nos da un resultado le damos un cero, si se equivoca le damos un uno. Cuanto mayor puntación reciba, peor es su rendimiento. Podemos enseñar un millón de fotos a una red neuronal que pueden ser gatos o perros. Por cada error sumamos un uno. Si al final la red tiene una puntuación de 500,000, la verdad es que es una red muy poco útil. Ahora vamos modificando los pesos de la red. El error va bajando, falla 200,000 veces. Con muchísimo entrenamiento logramos una red que tiene una penalización de, por ejemplo, 23. Del total de un millón de fotos, solo se ha equivocado en 23. Minimizar el error es la forma matemática de aprender una tarea específica*". LATORRE SENTÍS, J. I., *Ética para Máquinas*, *op. cit.*, *supra*, nota 401, p. 173.

En ese sentido, José Ignacio Latorre explica que el error no es más que una función, denominada "función error", y que nuestra tarea es definir la función error que dictará lo que aprende una red neuronal.

programará en las máquinas.

Este último cuestionamiento nos sirve para avanzar en el análisis de este trabajo: la Ética de la IA necesita de un espacio en el que se reflexione críticamente los dilemas éticos propuestos por el desarrollo de la IA, incluyendo quién y cómo debería programar a las máquinas, pero, sobre todo, si un espacio así puede servir de cara a un Agente Moral Artificial con la suficiente capacidad de tomar decisiones con repercusiones morales y jurídicas por sí mismo.

3. LOS COMITÉS DE ÉTICA

Previo al desarrollo de este apartado, es conveniente aclarar que cuando hablemos de Comités de Ética, será para referirnos exclusivamente a aquellos relacionados con el quehacer clínico, y particularmente con los Comités de Bioética. La justificación por la cual se opta por estos Comités es debido a la estrecha relación que puede identificarse entre la Bioética y la Ética de la IA.

Esta relación se traduce en que los principios éticos básicos utilizados como método argumentativo de la Bioética, también pueden utilizarse en la Ética de la IA, con reservas, sin ignorar que los dilemas éticos que acontecen en el campo de la IA son diferentes a los enfrentados en el de la Bioética. De igual manera, la relación entre la Bioética y la Ética de la IA se ubica también en el acento multidisciplinario que debe caracterizar a los Comités de Bioética y los de IA. En fin, para este apartado, la Bioética y sus Comités han de ser nuestros referentes de cara a la propuesta, que se desarrolla en un apartado posterior, de instituir Comités de Ética en IA con la intención de hacer frente a la idea de una posible Agencia Moral Artificial.

Ahora bien, abordando el tema de este apartado, el origen de los Comités de Ética se ubica en la última mitad del siglo pasado, centrándose en el problema de la tecnologización de la medicina.[408] El primer intento de creación se ubica en los años sesenta, en los Estados Unidos, con relación a la selección de pacientes para diálisis y frente al problema de que no sería posible atender a todos los que

[408] VIESCA, C., “Perspectiva histórica de los Comités de Ética”, en *Revista CONAMED*, Año 4, no. 12, 1999, p. 27

requirieran de ese tratamiento.[409]

En palabras de Carlos Viesca: "*Una gran tecnificación de la medicina produce una deshumanización de la atención médica, una sensación de vacío. Por un lado, de los pacientes y por el lado de los médicos nos damos cuenta de que hace falta algo*".[410]

En la actualidad, el desarrollo y expansión de estos Comités suponen que se hayan hecho cargo de los problemas alrededor de temas como la muerte, la reproducción, de la situación de incapacidad, de la investigación en seres humanos y de la agresividad de los tratamientos, entre otros temas en torno a las relaciones clínicas.[411]

En ese tenor, los Comités de Ética también encuentran su justificación en el hecho de que, con la tecnologización de la medicina, los problemas éticos en la relación clínica se han complejizado, sin que éstos se puedan resolver negándolos o simplificándolos. Además, estos problemas difícilmente pueden resolverse atendiendo sólo a regulaciones jurídicas, debido a que éstas suelen ser menos dinámicas y flexibles que la realidad:

> Las decisiones éticas más delicadas y difíciles sobre la vida y la muerte se deben tomar no por los jueces sino por los médicos, los pacientes y las familias dentro del ámbito de la clínica. El tribunal no es el foro adecuado para los muchos problemas éticos que surgen hoy en día en la clínica. Algunos jueces pidieron la formación de Comités de Ética no para hacer de ellos una especie de falsos jueces, sino para que sirvieran de foro intermedio en el que todos los datos adecuados de un caso pudieran analizarse desde tantas perspectivas como personas haya directamente afectadas. [412]

Efectivamente, los Comités de Ética ofrecen soluciones a dilemas éticos

[409] DOLORES RUIZ, M. M., *UD3: Los comités de ética*, Cuaderno para Curso: Bioética, 2009, p. 3
[410] VIESCA, C., "Perspectiva histórica de los Comités de Ética", *op. cit*., *supra*, nota 408, p. 27.
[411] DOLORES RUIZ, M. M., *UD3: Los comités de ética*, *op. cit., supra,* nota 409*,* p. 3.
[412] DRANE, J., La ética médica entra en el Hospital, en Jano, no. 781, 1987, p. 81 *apud* DOLORES RUIZ, M.M, *op. cit., supra,* nota 409*,* p.2.

vinculados con la tecnologización de la medicina. Además, estos dilemas suelen estar aparejados a un alto grado de complejidad en cuanto a su solución, lo cual se traduce en que los marcos habituales para resolver problemas, como bien pudieran ser las regulaciones jurídicas, no resulten tan efectivos al momento de querer aplicarlos. En este punto, podemos identificar una característica común entre la Bioética y la Ética de la IA: que ambos suelen ser más efectivos para solucionar problemas que demandan dinamismo para su solución, como pueden ser las decisiones involucradas con cuestiones clínicas o con cuestiones tecnológicas respectivamente, que suponen dilemas éticos.

En tiempos recientes, los Comités de Ética se han visto robustecidos gracias a la Bioética, la cual es la rama de la ética normativa que se ocupa de problemas éticos surgidos en el contexto de la medicina y de las ciencias biomédicas[413]. Algunos de los temas relevantes para esta rama de la ética normativa son el aborto, la eutanasia, los trasplantes de órganos, la manipulación genética, la clonación, la relación médico-paciente, la investigación biomédica con sujetos humanos, los derechos de los animales y la justicia en la distribución de los servicios de salud.[414]

Cabe referir que la Bioética, al tratarse también de una ética práctica, no es puramente conceptual ni puramente descriptiva, sino que prescribe qué acciones, normas de conducta e instituciones deberían respaldarse o rechazarse en el contexto de la biomedicina.[415]

En ese sentido, es posible ubicar hitos relevantes que nos permiten comprender en la actualidad mejor a los Comités de Ética, y de forma más precisa, a los Comités de Ética de la Investigación y Bioética. Estos últimos Comités se originan por medio del acuerdo del 19 de octubre de 2004, en el que la UNESCO adoptó la Declaración Universal de Bioética y Derechos Humanos, a través del cual se sentaron parámetros universales rectores de toda investigación, a fin de proteger la dignidad del ser humano y fomentando el respeto por los seres humanos, animales y el medio

[413] RIVERA LÓPEZ, E. Derecho y Bioética, en Fabra Zamora, J.L y E. Spector (Coords.), "*Enciclopedia de filosofía y teoría del derecho*", Volumen, III, Universidad Nacional Autónoma de México, Instituto de Investigaciones Jurídicas, México, 2015, p. 2735.
[414] *Idem.*
[415] *Idem.*

ambiente.[416]

Los Comités de Bioética y Ética de la Investigación están orientados a enfrentar dilemas bioéticos y erradicar los abusos por parte de los investigadores. Además, son clave para el desarrollo de la ciencia e investigación en virtud de que implican órganos de evaluación, aprobación y seguimiento para dar mayor desarrollo de conocimiento científico.[417]

Dichos Comités se tratan de grupos interdisciplinarios que buscan clarificar y resolver los conflictos de valores que pueden surgir en la práctica clínica o en la investigación de la salud. La razón por la cual se constituyen es para la deliberación y argumentación racional sobre los dilemas de orden moral que se presentan en el campo de las ciencias de la vida y la salud, con el horizonte puesto en proteger la dignidad y los derechos de las personas.[418]

A decir de la UNESCO, existen cuatro tipos de Comités de Bioética, a saber: 1) Comité normativo y/o consultivo; 2) los comités de asociaciones de profesionales de la salud; 3) Comité de ética asistencial/hospitalaria, y 4) Comité de ética de la investigación. Estos cuatro tipos de Comités se basan en la premisa de que todas las personas con capacidad mental o sus representantes son agentes morales y que los presidentes y miembros de dichos comités, a su vez agentes morales, están obligados a intervenir en controversias sobre Bioética y en particular, en las polémicas y dilemas planteados durante las actividades cotidianas de los comités normativos y consultivos gubernamentales de ámbito nacional, en asociaciones de profesionales de la salud, hospitales y otras instituciones de asistencia médica, así como en centros de investigación clínica.[419]

Los dos primeros comités no suelen crear confusión. Por una parte, los comités normativos y/o consultivos están orientados a asesorar a funcionarios públicos y a influir en la adopción de políticas científicas y de salud de ámbito nacional. Por la

[416] MEDINA ARELLANO, M. J. y M. I. Castañeda Avalos, “Comité Universitario de Bioética en la Universidad Autónoma de Nayarit”, en *Revista Bio Ciencias*, 2013, p. 5.
[417] *Idem.*
[418] ÁLVAREZ, M. Y., *et al*., “Comités de bioética en las instituciones del sistema de salud”, en *Red Sociales, Revista del Departamento de Ciencias Sociales*, vol. 08, no. 2, pp. 71-72.
[419] UNESCO, *Funcionamiento de los comités de bioética: procedimientos y políticas*, Guía no. 2, Francia, 2006, p. 7.

otra, los comités de asociaciones de profesionales de la salud se centran en cuestiones de Bioética planteadas por profesionales de la asistencia sanitaria.[420]

Los Comités de Ética asistencial tienen tres funciones: 1) la ayuda en la resolución de conflictos, 2) la elaboración de protocolos y 3) la función pedagógica. Además, cuentan con otras funciones de orden interno, como pueden ser la elaboración de memorias anuales o el establecimiento de un reglamento interno que regule su actividad.[421]

Por último, los Comités de Ética de la Investigación son definidos como: "*un cuerpo independiente y multidisciplinario encargado de la revisión de investigación que involucran la participación de humanos, para asegurar que su integridad, derechos y bienestar estén protegidos*".[422]

Estos comités evalúan las investigaciones empleando el principio de la justicia:

> Este principio no se expresa solamente en la 'justicia' al momento de 'seleccionar' los sujetos de investigación. Elementos relacionados con el uso de los recursos y la justificación de desarrollar o no una investigación en razón de su calidad, son algunos factores que debe considerar un comité en la evaluación de un proyecto de este tipo.[423]

Los Comités de Ética de la investigación abordan las cualidades científicas, el riesgo al que se exponen los participantes y los posibles beneficios para futuros pacientes que no participan como sujetos en ensayos clínicos.[424] El origen de este tipo de comités es comprensible a luz de la historia, particularmente después del interés internacional que se suscitó por cuestiones de bioética relativas a la investigación en seres humanos a raíz de los juicios de Nuremberg, en 1946. En ese momento, se centró la atención en buscar un equilibrio entre la libertad de

---

[420] *Ibidem*, pp. 7-8.
[421] DOLORES RUIZ, M. M., *UD3: Los comités de ética*, *op. cit., supra,* nota 408*,* pp. 4-5.
[422] ARANGO-BAYER, G. L., "Los Comités de Ética de la Investigación, Objetivos, funcionamiento y principios que buscan proteger", en *Investigación en Enfermería: Imagen y Desarrollo*, vol, 10, no. 1, Bogotá, 2008, p. 11.
[423] *Idem.*
[424] UNESCO, *Funcionamiento de los comités de bioética: procedimientos y políticas*, Guía no. 2, Francia, 2006, pp.7-8.

científicos y médicos para efectuar investigación clínica y la protección de los individuos que habían consentido en participar como sujetos en experimentos, ensayos o estudios clínicos.[425]

De cara a problemas como el anteriormente señalado, los comités de ética juegan un rol decisivo:

> Al supervisar la investigación, deben facilitar los ensayos, instrumentos indispensables para el progreso de la medicina clínica. No obstante, también deben proteger los intereses de los participantes, en especial cuando carecen de facultades plenas para protegerse a sí mismos. Los comités de ética de la investigación han hecho suyo ese deber fundamental, no sólo en beneficio de los participantes, sino en el de la sociedad en general, pues si la investigación cayese en descrédito quedaría mermada y se privaría a la humanidad de muchos de sus maravillosos frutos. Numerosos Estados Miembros [de la UNESCO] han creado comités de ética de la investigación, precisamente debido a los rápidos adelantos de la biotecnología y la medicina biomédica y conductual. [426]

Los miembros de estos Comités de Ética de la investigación se caracterizan por su deseo y capacidad de diálogo, pues este es la base que facilita el intercambio de información técnica, social y moral entre “expertos y profanos”. Estos Comités también se caracterizan por la interdisciplinariedad de sus miembros:

> Entre los miembros del comité hay personas conocedoras de las diferentes dimensiones de la información pertinente, pero, sólo en la profundidad de una de ellas; requiriendo el saber de los otros miembros para completar su conocimiento o apreciación de cada caso. Esta es la dinámica de interdisciplinariedad que caracteriza el trabajo del CEI [Comité de Ética de la Investigación] y se estructura como su método

[425] *Ibidem*, p. 13.
[426] *Ibidem*, p. 17.

específico.[427]

Cabe mencionar que, de cara a las investigaciones clínicas y en concreto a la experimentación con seres humanos, se han identificado principios éticos básicos, que en su conjunto conforman al principialismo,[428] uno de los métodos de la argumentación en Bioética. A continuación, refiero a dichos principios, con la intención de contar con una visión de 360 grados en torno al tema de los Comités de Bioética:

- **Principio de autonomía**: el cual refiere a la capacidad racional con la que se cuenta para elegir lo más conveniente: *"la autonomía juega un papel sustantivo en la auto-definición y el auto-determinamiento como personas libres que, más allá de las presiones externas, vislumbran la importancia de actuar, pensar y decidir con libertad*".[429]

  El respeto a este principio exige el reconocimiento a lo siguiente: 1) el derecho a tener sus propios puntos de vista; 2) el derecho a tomar sus propias opciones y, 3) el derecho a actuar en conformidad con su escala de valores. [430] En efecto, la autonomía de la persona se respeta cuando se le reconoce el derecho a mantener puntos de vista, a hacer elecciones y a realizar acciones basadas en valores y creencias personales. Este principio obliga a los profesionales a revelar información, a asegurar la comprensión y la voluntariedad y a potenciar la participación del paciente en la toma de decisiones.[431]
- **Principio de no maleficencia**: con este principio se hace referencia a la

[427] RUEDA CASTRO, L., "Interdisciplinariedad y Comités de Ética", en *Revista Latinoamericana de Bioética*, vol. 12, no. 2, 2012, p. 74.

[428] El principialismo es una postura jurídica caracterizada por el papel preponderante que concede a los principios, a su aplicación y ponderación. Desde dicha postura, es habitual identificar la moral con el derecho, con la intención de tener garantías de buena argumentación y buen discernimiento a la hora de deliberar sobre determinados asuntos. MEDINA ARELLANO, M. J. y J. Hincapié Sánchez, *Bioética: teorías y principios*, Serie Libros Digitales, núm. 1, Instituto de Investigaciones Jurídicas, UNAM, México, 2019, pp. 14-15.

[429] *Ibidem*, p. 15.

[430] *Idem.*

[431] SIURANA APARISI, J. C., Los principios de la bioética y el surgimiento de una bioética intercultural, en "*Veritas"*, no. 22, 2010, p. 124.

obligación de no dañar de forma intencionada a ningún ser vivo, ya sea por acciones relacionadas con nuestra profesión o por las cotidianas. Son cuatro las obligaciones generales aparejadas a este principio: 1) no se debe hacer mal o daño; 2) se debe prevenir el mal o daño; 3) se debe eliminar el mal o daño y 4) se debe hacer o promover el bien.[432]

Las reglas típicas vinculadas al principio de No maleficencia son las siguientes: 1) no mate; 2) no cause dolor o sufrimiento a otros; 3) no incapacite a otros; 4) no ofenda a otros y 5) no prive a otros de aquello que aprecian en la vida.[433]

- **Principio de beneficencia**: este principio refiere al deber de proporcionar bienestar a la sociedad. Se trata de la obligación moral de actuar de forma objetiva en beneficio de los demás, con lo cual se va mucho más allá de la simple benevolencia como una mera actitud o disposición de querer el bien para los demás.[434]

  La benevolencia no es lo mismo que la beneficencia:

> En el lenguaje habitual, la beneficencia hace referencia a actos de buena voluntad, amabilidad, caridad, altruismo, amor o humanidad. La beneficencia puede entenderse, de manera más general, como todo tipo de acción que tiene por finalidad el bien de otros. Si la benevolencia se refiere a la voluntad de hacer el bien, con independencia de que se cumpla o no la voluntad, la beneficencia, en cambio, es un acto realizado por el bien de otros.[435]

Así, el principio de beneficencia implica sólo a aquellos actos que son una exigencia ética en el ámbito de la medicina. Algunos ejemplos

[432] MEDINA ARELLANO, M. J. y J. Hincapié Sánchez, *Bioética: teorías y principios, op. cit., supra,* nota 428*,* p. 16.

[433] SIURANA APARISI, J. C., "Los principios de la bioética y el surgimiento de una bioética intercultural", *op. cit*., supra, nota 431, p. 125.

[434] MEDINA ARELLANO, M. J. y J. Hincapié Sánchez, *Bioética: teorías y principios, op. cit., supra*, nota 428, p. 17.

[435] SIURANA APARISI, J. C., "Los principios de la bioética y el surgimiento de una bioética intercultural", *op. cit*., *supra*, nota 431, pp. 125-126.

de reglas de beneficencia son estos: 1) protege y defiende los derechos de otros; 2) previene el daño que pueda ocurrir a otros; 3) quita las condiciones que causarán daño a otros; 4) ayuda a personas con discapacidades y, 5) rescata a personas en peligro.[436]

- **Principio de justicia:** este principio está relacionado con lo que es debido a las personas, con aquello que les pertenece o les corresponde de alguna manera. Desde el mirador médico, la justicia que es de interés es la distributiva, la cual refiere a la distribución equitativa de los derechos, beneficios, responsabilidades y cargas en la sociedad.[437]

  Juan Carlos Siurana, al respecto, señala:

> El término relevante en este contexto es el de justicia distributiva que, según estos autores [Beauchamp y Childress] se refiere la 'distribución imparcial, equitativa y apropiada en la sociedad, determinada por normas justificadas que estructuran los términos de la cooperación social'. Sus aspectos incluyen las políticas que asignan beneficios diversos y cargas tales como propiedad, recursos, privilegios y oportunidades. Son varias las instituciones públicas y privadas implicadas, incluyendo al Gobierno y al sistema sanitario.[438]

Los criterios empleados para recibir un trato igualitario son: 1) a cada persona una porción igual; 2) a cada persona según sus necesidades; 3) a cada persona según sus esfuerzos; 4) a cada persona según su aportación; 4) a cada persona según su mérito y, 5) a cada persona según las reglas de intercambio en un mercado libre.[439]

Recapitulando, los Comités de Ética y Bioética son espacios interdisciplinarios

---

[436] *Ibidem*, p. 126.
[437] MEDINA ARELLANO, M. J. y J. Hincapié Sánchez, *Bioética: teorías y principios, op. cit., supra*, nota 428, p. 19.
[438] SIURANA APARISI, J. C., "Los principios de la bioética y el surgimiento de una bioética intercultural", *op. cit.*, *supra*, nota 431, p.127.
[439] MEDINA ARELLANO, M. J. y J. Hincapié Sánchez, *Bioética: teorías y principios, op. cit., supra*, nota 428, p. 19.

en los que es posible resolver conflictos de valores surgidos en la práctica clínica o en la investigación de la salud. Existen cuatro tipos de comités de bioética: 1) Comité normativo y/o consultivo; 2) los comités de asociaciones de profesionales de la salud; 3) Comité de ética asistencial/hospitalaria, y 4) Comité de ética de la investigación.

Quienes se dan cita en dichos comités, se sirven de un método argumentativo que consta de los siguientes principios: 1) autonomía; 2) No maleficencia; 3) Benevolencia y, 4) Justicia. Estos principios, en términos generales, están orientados a proteger la dignidad y los derechos de las personas. Es importante tomar conciencia de los Comités de Ética y Bioética, pues con ellos se instrumentaliza la reflexión ética con base en los mencionados principios, siendo también una responsabilidad de todo profesional o especialista que participa de la investigación científica o de la práctica médica el participar en ellos, en virtud de que con ello defiende y pretende garantizar el respeto, la paz y la dignidad de nuestra especie.[440]

Así pues, dicho lo anterior, es posible tomar como referencia a los Comités de Ética y Bioética como base para pensar en el establecimiento de Comités de Ética en IA, que además de deliberar sobre dilemas éticos actuales, también sirvan para reflexionar sobre aquellos dilemas que comienzan a ser visibles y que se vinculan estrechamente con los agentes morales artificiales, por las siguientes razones: 1) tanto en la Bioética como en la IA se presentan dilemas éticos que requieren de espacios en los que se pueda deliberar y proponer soluciones con el fin de preservar las consideraciones morales que existen hacía los seres humanos, pero también hacía los demás seres vivos que puedan verse impactados por el desarrollo de la medicina y la IA, respectivamente; 2) las reflexiones críticas en torno al desarrollo e impacto de los desarrollos tecnológicos en el campo de la medicina, así como de la IA requiere de la interdisciplinariedad; los Comités de Ética —tanto en el campo de la Bioética como en el de la IA— sirven para satisfacer este interés; 3) el método

[440] MEDINA ARELLANO, M. J., M. E. Cancino Marentes, A. Gascón Cervantes y A. de Lara Ramírez, Comités *de ética y bioética*, Serie Libros Digitales, núm. 3, Instituto de Investigaciones Jurídicas, UNAM, México, 2019, pp. 48-49.

argumentativo empleado en la Bioética también podría funcionar para la Ética de la IA, sin ignorar las adecuaciones necesarias para abordar los dilemas éticos presentados en el campo de la IA.

Considerando lo expuesto , podemos avanzar en el desarrollo del siguiente apartado, el cual tendrá como finalidad explicar la manera en la que un Comité de Ética en IA podría resultar clave frente a la existencia de los agentes morales artificiales.

4. LOS COMITÉS DE ÉTICA EN INTELIGENCIA ARTIFICIAL Y LA AGENCIA MORAL ARTIFICIAL

Cabe destacar que la creación de Comités de Ética en IA no es una idea nueva. Empresas líderes en tecnología como Google, Amazon, Meta, IBM y Microsoft han establecido alianzas para el desarrollo de estándares que garanticen que las investigaciones en este campo se realicen de manera segura, ética y transparente.[441] En el caso de Google, su Comité de Ética en IA resonó mucho en el año de 2019, cuando fue desmantelado apenas una semana después de su creación, debido al nombramiento como miembro de dicho comité de un personaje que era tildado de transfóbico. [442]

La idea de que los Comités de Ética en IA son necesarios para reflexionar de forma crítica y coadyuvar a que el desarrollo de posibles Agentes Morales Artificiales esté acorde a un núcleo de principios establecidos con anterioridad a su creación, por su parte, aporta un valor agregado a las discusiones que buscan contribuir sobre cómo convivir con la IA.

Un Comité de Ética en IA tiene como función principal identificar de forma sistemática y exhaustiva y ayudar a mitigar los riesgos éticos de los productos con

[441] OBSERVATORIO DE INTELIGENCIA ARTIFICIAL, *Google, Facebook, Amazon, IBM y Microsoft forman una alianza en materia de Inteligencia Artificial*, 30 de septiembre de 2016, [en línea], <https://observatorio-ia.com/alianza-en-materia-de-inteligencia-artificial>, [consulta: 02 de octubre de 2023].

[442] BUSINESS INSIDER, *Google cancela su Comité para regular la Ética de la Inteligencia Artificial tan solo una semana después de su creación*, 5 de abril de 2019. [en línea], <https://www.businessinsider.es/google-cancela-comite-regular-etica-inteligencia-artificial-400609>, [consulta: 02 de octubre de 2023].

IA que se desarrollan de forma interna o se compran a terceros.[443]

Asimismo, los Comités de Ética en IA también son útiles para tomar decisiones sobre los riesgos para el bienestar humano, los animales, el medio ambiente y la infraestructura crítica que pueden derivarse de la investigación en IA, lo cual comprende al aprendizaje automático, el Big Data y la investigación en redes neuronales artificiales. [444]

Un Comité de Ética en IA no busca llevar a cabo un análisis de viabilidad tecnológica ni criticar los métodos de investigación, sino responder a preguntas concretas sobre el riesgo de daño y los tipos de daño que surgen de proyectos específicos. Tampoco busca ser parte en la adjudicación de disputas sobre el valor general de la investigación en IA o tratar de supervisar la investigación básica que sólo está relacionada provisionalmente con la IA en un futuro hipotético.[445]

Desde dicho mirador, un Comité de Ética en IA resulta útil frente al Agente Moral Artificial, específicamente de cara al análisis de los dilemas éticos que éste pudiera llegar a proponer. Ya en el segundo capítulo de este trabajo se dio cuenta de algunos de los tipos de IA que, de ser realidad, supondrían un desafío para nuestro sistema actual de valores: la IA Fuerte y la Superinteligencia. Estos dos tipos de IA son, a opinión, los que podrían en su momento dar pie al Agente Moral Artificial.

En términos generales, el Agente Moral Artificial, al contar con condiciones internas como la autonomía, la intencionalidad y la capacidad de experimentar dolor y placer, será capaz de tomar decisiones por sí mismo, ¿deberíamos permitir eso o es preferible que exista en todo momento un tipo de supervisión humana? La otra gran pregunta sería: ¿a un Agente Moral Artificial se le deberá consideración moral o también consideración jurídica?

Un Comité de Ética en IA resulta de utilidad para el análisis de ambas preguntas, principalmente cuando lo que se pretende es contar con una posible respuesta cimentada en un núcleo de valores compartidos por quienes en él participan.

---

[443] BLACKMAN, R., "Why You Need an AI Ethics Committee", en *Harvard Business Review*, Julio-Agosto 2022, [en línea], <https://hbr.org/2022/07/why-you-need-an-ai-ethics-committee>, [consulta: 02 de octubre de 2023].

[444] JORDAN, S. R., *Designing an Artificial Intelligence Research Review Committee*, Future of Privacy Forum, Estados Unidos de América, 2019, p. 14.

[445] *Idem.*

También vale la pena referir que un Comité de Ética en IA resulta idóneo como complemento frente a las regulaciones tradicionales, las cuales suelen ser incapaces de adaptarse a las realidades producidas por el avance de la tecnología en general. De esta forma, los Comités de Ética en IA proporcionan una alternativa ágil para enfrentar las cuestiones asociadas al desarrollo de dicha tecnología.

En el mismo contexto de los agentes morales artificiales, tanto la Ética de la IA y los Comités de Ética en IA podrían ayudar a determinar, de ser el caso, cuándo y cómo otorgar derechos y responsabilidades a estos agentes, así como evaluar y controlar los posibles riesgos y beneficios asociados con su creación y uso. Estas cuestiones pueden ayudar a que los agentes morales artificiales, si llegan a desarrollarse, no se conviertan en una amenaza para la humanidad, sino en aliados valiosos que contribuyan al bienestar y el progreso social.

Aquí una advertencia, lo que se pretende en este apartado no es discurrir sobre propuestas de respuestas a ambas preguntas, sino identificar lo qué es un Comité de Ética en IA, su funcionamiento y estructura.

4.1. El núcleo de principios del Comité de Ética en Inteligencia Artificial

Un Comité de Ética en IA requiere de un núcleo de principios que proteger y promover. En términos generales, el núcleo de principios que el Comité de Ética en IA decida abanderar, describirán lo que éticamente debe hacerse para protegerlos y promoverlos. El articular claramente estos principios es clave, pues con ellos se guiará el trabajo del comité de ética, un ejemplo de esto es lo siguiente:

> For example, if informed consent and maintaining anonymity are core principles or requirements for sharing user data with other firms or organizations, then the role of the ethics committee is to ensure that any proposal to share data — any program, technical design, partnership, or initiative that would do this — meets the standards of informed consent and anonymity. As another example, if avoiding adverse impacts on socially and economically disadvantaged groups is a core principle for using AI or machine learning in decision-making, the role of the committee is to ensure the design and implementation adequately considers,

> monitors, audits, and adjust to ensure this.[446] [Por ejemplo, si el consentimiento informado y el mantenimiento del anonimato son principios o requisitos fundamentales para compartir los datos de los usuarios con otras empresas u organizaciones, entonces la función del Comité de Ética es garantizar que cualquier propuesta para compartir datos -cualquier programa, diseño técnico, asociación o iniciativa realizada- cumpla con las normas de consentimiento informado y anonimato. Otro ejemplo es el siguiente, si evitar impactos adversos en los grupos sociales y económicamente desfavorecidos es un principio básico para el uso de la IA o el aprendizaje automático en la toma de decisiones, la función del Comité es garantizar que el diseño y la implementación consideren, supervisen, auditen y ajusten adecuadamente para garantizar esto.] (Traducción propia).

En efecto, el núcleo de principios es primordial para el funcionamiento del Comité de Ética en IA. A consideración de este trabajo, un núcleo de principios propicio para abordar cuestiones éticas en las que pudieran involucrarse los agentes morales artificiales es el brindado por la UNESCO, en su Recomendación sobre la Ética de la IA. Si bien es cierto que los principios y valores a los que refiere dicho documento buscan ser respetados por todos los actores durante el ciclo de vida[447] de los sistemas de IA, también lo es que desempeñan una importante función como ideales que motivan la orientación de las medidas de políticas y normas jurídicas de conformidad con el derecho internacional, y como tales, también los principios y valores son deseables por sí mismos.

En este punto vale la pena distinguir entre principios y valores de acuerdo con la Recomendación de la UNESCO, pues a decir de ese documento:

---

[446] SANDLER, R. y J. Basl, *Building Data and AI Ethics Committees*, Accenture y Ethics Institute at Northeastern University, 2019, p. 11. [en línea] <https://ethics.harvard.edu/news/how-build-data-and-ai-ethics-committees > [consulta: 02 de octubre de 2023].

[447] La Recomendación de la UNESCO señala que el ciclo de vida relativo a los sistemas de Inteligencia Artificial va desde la investigación, la concepción y el desarrollo y hasta el despliegue y la utilización, pasando por el mantenimiento, el funcionamiento, la comercialización, la financiación, el seguimiento y la evaluación, la validación, el fin de la utilización, el desmontaje y la terminación. UNESCO, *Informe de la Comisión de Ciencias Sociales y Humanas (SHS)*, 41 C/73, 22 de noviembre de 2021, p.17.

> Mientras que el conjunto de valores que se enuncia a continuación inspira, por tanto, un comportamiento deseable y representa los fundamentos de los principios, los principios, por su parte, revelan los valores subyacentes de manera más concreta, de modo que estos últimos puedan aplicarse más fácilmente en las declaraciones de políticas y las acciones.448

Así pues, los principios se muestran como bases que facilitan la consecución de los objetivos específicos enmarcados en el multicitado documento. Cabe señalar que la Recomendación sobre la Ética de la IA tiene, como algunos de sus objetivos, el proteger, promover y respetar los derechos humanos y las libertades fundamentales, la dignidad humana y la igualdad, incluida la de género; salvaguardar los intereses de las generaciones presentes y futuras; preservar el medio ambiente, la biodiversidad y los ecosistemas; y respetar la diversidad cultural en todas las etapas del ciclo de vida de los sistemas de IA.[449]

En efecto, el que la Recomendación de la UNESCO tenga como objetivo el salvaguardar los intereses de las generaciones presentes y futuras, y que este conexa a un marco universal de valores y principios que orienten la toma de decisiones en el ámbito político y jurídico, es una justificación para concluir que, en específico, los principios enunciados por la UNESCO también contribuyen a abordar cuestiones en las que pudieran involucrarse los agentes morales artificiales.

Los principios a los que refiere la Recomendación sobre la Ética de la IA son los siguientes: 1) Proporcionalidad e inocuidad; 2) Seguridad y protección; 3) Equidad y no discriminación; 4) Sostenibilidad; 5) Derecho a la intimidad y protección de datos; 6) Supervisión y decisión humanas; 7) Transparencia y explicabilidad; 8) Responsabilidad y rendición de cuentas; 9) Sensibilización y educación, y 10) Gobernanza y colaboración adaptativas y de múltiples partes interesadas. A continuación, se describe a cada uno de estos principios.

---

448 *Ibidem*, p.19.
449 *Ibidem*, p. 18.

4.1.1. Proporcionalidad e inocuidad

La proporcionalidad implica que los procesos relacionados con el ciclo de vida de los sistemas de IA no pueden ir más allá de lo necesario para lograr propósitos u objetivos legítimos, y esos procesos deberían ser adecuados al contexto. Este principio hace patente la necesidad de garantizar la aplicación de procedimientos de evaluación de riesgos y la adopción de medidas para impedir que se produzcan daño para los seres humanos, los derechos humanos y las libertades fundamentales, las comunidades y la sociedad en general, o para el medio ambiento y los ecosistemas.

Además, este principio señala que la decisión de utilizar sistemas de IA y la elección del método de IA deben justificarse de las siguientes maneras: a) el método de IA elegido debe ser adecuado y proporcional para lograr un objetivo legítimo determinado; b) el método de IA elegido no debería vulnerar los valores fundamentales enunciados en la Recomendación sobre la Ética de la IA, y c) el método de IA elegido debería ser adecuado al contexto y basarse en fundamentos científicos rigurosos. Señala también que en los casos en que se entienda que las decisiones tienen un impacto irreversible o difícil de revertir o que puedan implicar decisiones de vida o muerte, la decisión final debería ser adoptada por un ser humano. Igualmente, refiere que los sistemas de IA no deberían utilizarse con fines de calificación social o vigilancia masiva. [450]

4.1.2. Seguridad y protección

Este principio señala que los daños no deseados (riesgos de seguridad) y las vulnerabilidades a los ataques (riesgos de protección) deberían ser evitados y deberían tenerse en cuenta, prevenirse y eliminarse a lo largo del ciclo de vida de los sistemas de IA para garantizar la seguridad y la protección de los seres humanos, del medio ambiente y de los ecosistemas. Nos dice también que la seguridad y la protección de la IA se propiciarán mediante el desarrollo de marcos de acceso a los datos que sean sostenibles, respeten la privacidad y fomenten un mejor entrenamiento y validación de los modelos de IA que utilicen datos de

[450] *Ibidem*, p. 21.

calidad.[451]

4.1.3. Equidad y no discriminación

Este principio señala que los actores de la IA deben promover la justicia social, salvaguardad la equidad y luchar contra todo tipo de discriminación, de conformidad con el derecho internacional. Con esto, se busca garantizar que los beneficios de la IA estén disponibles y sean accesibles para todos, teniendo en cuenta las necesidades específicas de los diferentes grupos de edad, los sistemas culturales, los diferentes grupos lingüísticos, las personas con discapacidad, las niñas y las mujeres y las personas desfavorecidas, marginadas y vulnerables o en situación de vulnerabilidad.

De igual manera, este principio implica que los actores de la IA deberían hacer todo lo razonablemente posible para reducir al mínimo y evitar reforzar o perpetuar aplicaciones y resultados discriminatorios o sesgados a lo largo del ciclo de vida de los sistemas de IA, con la intención de garantizar la equidad de dichos sistemas.[452]

4.1.4. Sostenibilidad

Este principio refiere que el desarrollo de sociedad sostenible depende del logro de un complejo de objetivos relacionados con distintas dimensiones humanas, sociales, culturales, económicas y ambientales, y que la llegada de la IA puede beneficiar a los objetivos de sostenibilidad o dificultar su consecución. En efecto, este principio implica que se realicen evaluaciones continuas de los efectos humanos, sociales, culturales, económicos y ambientales de las tecnologías de la IA. [453]

4.1.5. Derecho a la intimidad y protección de datos

Este principio apuesta que los datos para los sistemas de IA se recopilen, utilicen, compartan, archiven y supriman de forma coherente con el derecho internacional y acorde con los valores y principios que enuncia la Recomendación,

[451] *Ibidem,* p. 22.
[452] *Idem.*
[453] *Ibidem*, pp. 22-23.

respetando al mismo tiempo los marcos jurídicos nacionales, regionales e internacionales pertinentes.

Señala también que los algoritmos requieren de evaluaciones adecuadas del impacto en la privacidad, que incluyan también consideraciones sociales y éticas de su utilización y un empleo innovador del enfoque de privacidad desde la etapa de concepción; asimismo, implica que los actores de la IA deben asumir la responsabilidad de la concepción y la aplicación de los sistemas de IA manera que se garantice la protección de la información personal durante todo el ciclo de vida del sistema de IA.[454]

4.1.6. Supervisión y decisión humanas

Este principio implica el que los Estados Miembros velen para que siempre sea posible atribuir la responsabilidad ética y jurídica, en cualquier etapa del ciclo de vida de los sistemas de IA, así como los casos de recurso relacionados con sistemas de IA, a personas físicas o a entidades jurídicas existentes.

Con este principio se hace énfasis a no ceder el control a los sistemas de IA en contextos limitados, ya que éstos no podrán reemplazar la responsabilidad final de los seres humanos y su obligación de rendir cuentas; por regla general, nos dice este principio, las decisiones de vida o muerte no deberían cederse a los sistemas de IA.[455]

4.1.7. Transparencia y explicabilidad

Con el principio de transparencia se busca que las personas debieran estar informadas cuando una decisión se basa en algoritmos de IA o se toma a partir de ellos, en particular cuando afecta a su seguridad o a sus derechos humanos. En efecto, las personas deberían tener la oportunidad de solicitar explicaciones e información al actor de la IA o a las instituciones del sector público correspondientes.

De igual manera, las personas deberían poder conocer los motivos por los que ha tomado una decisión que afecta a sus derechos y libertades y tener la posibilidad de presentar alegaciones a un miembro del personal de la empresa del sector

[454] *Ibidem*, p. 23.
[455] *Idem.*

privado o de la institución del sector público habilitado para revisar y enmendar la decisión.

Por su parte, la explicabilidad supone hacer inteligibles los resultados de los sistemas de IA y facilitar información sobre ellos. La explicabilidad de los sistemas de IA también se refiere a la inteligibilidad de la entrada, salida y funcionamiento de cada componente algorítmico y la forma en que contribuye a los resultados de los sistemas. [456]

4.1.8. Responsabilidad y rendición de cuentas

Este principio está orientado a que los actores de la IA y los Estados Miembros respeten, protejan y promuevan los derechos humanos y las libertades fundamentales, y fomenten también la protección del medio ambiente y los ecosistemas, siempre asumiendo su responsabilidad ética y jurídica respectiva, de conformidad con el derecho nacional e internacional, en particular las obligaciones de los Estados Miembros en materia de derechos humanos, y con las directrices éticas establecidas durante todo el ciclo de vida de los sistemas de IA. [457]

4.1.9. Sensibilización y educación

Este principio busca la promoción de la sensibilización y la comprensión del público respecto de las tecnologías de la IA y el valor de los datos, siempre mediante una educación abierta y accesible, la participación cívica, las competencias digitales y la capacitación en materia de la IA, la alfabetización mediática e informacional y la capacitación dirigida conjuntamente por los gobiernos, las organizaciones intergubernamentales, la sociedad civil, las universidades, los medios de comunicación, los dirigentes comunitarios y el sector privado, partiendo de la diversidad lingüística, social y cultural existente, a fin de garantizar una participación pública efectiva, de modo que todos los miembros de la sociedad puedan adoptar decisiones informadas sobre su utilización de los sistemas de IA y estén protegidos de influencias indebidas. [458]

---

[456] *Ibidem*, pp. 23-24.
[457] *Ibidem*, pp. 24-25.
[458] *Ibidem*, p. 25.

4.1.10. Gobernanza y colaboración adaptativas y de múltiples partes interesadas

Para este principio es crucial la participación de las diferentes partes interesadas a lo largo del ciclo de vida de los sistemas de IA, a fin de garantizar enfoques inclusivos de la gobernanza de la IA, de modo que los beneficios puedan ser compartidos por todos, y para contribuir al desarrollo sostenible. Entre las partes interesadas están los gobiernos, las organizaciones intergubernamentales, la comunidad técnica, la sociedad civil, los investigadores y los círculos universitarios, los medios de comunicación, los responsables de educación, los encargados de formular políticas, las empresas del sector privado, las instituciones de derechos humanos y los organismos de fomento de la igualdad, los órganos de vigilancia de la lucha contra la discriminación y los grupos de jóvenes y de la niñez. [459]

En efecto, lo que se pretende es que estos diez principios constituyan el núcleo de principios del Comité de Ética en IA, a fin de que los proyectos de investigación y desarrollo en materia de dicha tecnología que se sometan a su análisis cumplan con estándares éticos que contribuyan a que la irrupción del Agente Moral Artificial sea en el entendido de dichos principios y siempre considerando el impacto de los Agentes Morales Artificiales en la sociedad.

4.2. Sobre la estructura del Comité de Ética en Inteligencia Artificial

Un Comité de Ética en IA debe integrarse por un grupo interdisciplinario de expertos. Dicho requerimiento puede justificarse por varias razones. Una de ellas es que la IA, al ser una tecnología compleja y con implicaciones directas e indirectas en una variedad de campos, requiere de expertos capaces de abordar dichas implicaciones desde distintas perspectivas.

Otra razón es que el enfoque interdisciplinario facilita una mejor comprensión de los problemas asociados a la IA. Por ejemplo, quienes son expertos en la tecnología de la IA podría proporcionar información sobre las capacidades y limitaciones reales de dicha tecnología, a fin de evitar especulaciones respecto al funcionamiento de

[459] *Idem.*

esta.

Una razón adicional es que un grupo interdisciplinario de expertos, que cuenten con diferentes experiencias y conocimientos, puede ayudar a identificar problemas potenciales y ofrecer posibles soluciones al problema presentado. Como ejemplo, un grupo interdisciplinario en el que participen expertos en el desarrollo de la IA, expertos en Ética y expertos en Derecho podrían ayudar a identificar las implicaciones de un sistema de IA que es dado a discriminar a ciertos grupos de personas, además de señalar aspectos que den luz sobre cómo entrenar mejor a ese sistema, las implicaciones éticas que conlleva que éste discrimine y la información sobre las leyes y regulaciones que estarán dejándose al margen en caso de que el problema continúe.

En ese sentido, es posible identificar los siguientes perfiles profesionales como parte de un Comité de Ética en IA:

4.2.1. Personas expertas en Ética

Podrían ser personas con doctorados en filosofía, especializados en Ética, o personas con maestrías en Ética. Blackman refiere que estos perfiles están en el Comité porque tienen la capacitación, el conocimiento y la experiencia necesario para comprender y detectar una gran variedad de riesgos éticos, además de que se encuentran familiarizados con conceptos y distinciones que ayudan a las deliberaciones Éticas con mayor facilidad y son hábiles en ayudar a los grupos a evaluar objetivamente los problemas éticos. [460]

4.2.2. Personas expertas en Derecho

Blackman señala que, de cara al problema de los sesgos, las y los abogados cuentan con el conocimiento para determinar si una métrica particular para la equidad empleada por un técnico experto en IA podría ser considerado discriminatoria frente a una determinada ley. De igual forma, las y los abogados pueden ayudar a identificar si el uso de herramientas técnicas para lograr la equidad en los sistemas de IA es legal, o en su caso prohibido.[461] Adicionalmente, los

[460] BLACKMAN, R., "Why You Need an AI Ethics Committee", *op. cit.*, *supra,* nota 443.
[461] *Idem.*

expertos en Derecho son cruciales para garantizar el cumplimiento legal, así como para identificar áreas en las cuales las normas jurídicas resultan ambiguas o inadecuadas. [462]

4.2.3. Personas estrategas empresariales

este perfil profesional resulta útil si el Comité de Ética en IA tiene lugar al interior de una empresa. Al respecto, Reid Blackman señala que los estrategas empresariales son útiles cuándo se trata de identificar tácticas de mitigación de riesgos empresariales que devienen de los riesgos éticos asociados a la IA. [463]

4.2.4. Personas expertas en tecnología

Un experto en tecnología puede ayudar a entender los fundamentos técnicos de los modelos de IA, la probabilidad de éxito de diversas estrategias de mitigación de riesgos y la viabilidad de éstas. Este perfil profesional es necesario si lo que se busca en que los demás integrantes del Comité cuenten con una adecuada comprensión de los aspectos técnicos asociados a la IA.[464]

4.2.5. Profesionales expertas en disciplinas sobre las cuales la Inteligencia Artificial tiene una repercusión directa

Por ejemplo, si la IA tiene implicaciones con la prestación de servicios médicos, entonces lo deseable es que los profesionales clínicos sean los expertos requeridos en el Comité. En general, los expertos en disciplinas sobre las cuales la IA puede afectar son útiles para identificar como dicha tecnología podría repercutir en las prácticas actuales asociadas a su campo de trabajo.[465]

4.2.6. Participación ciudadana

Por último, la participación ciudadana representa las perspectivas y preocupaciones públicas.[466] Más aún, si tomamos en cuenta que son los principales usuarios de la IA, y que cuentan puntos de vista valiosos que ayudan a entender mejor la

[462] SANDLER, R. y J. Basl, *Building Data and AI Ethics Committees*, *op. cit., supra,* nota 446.
[463] BLACKMAN, R., "Why You Need an AI Ethics Committee", *op. cit.*, *supra*, nota 443.
[464] SANDLER, R. y J. Basl, *Building Data and AI Ethics Committees*, *op. cit., supra*, nota 446.
[465] BLACKMAN, R., "Why You Need an AI Ethics Committee", *op. cit.*, *supra*, nota 443.
[466] SANDLER, R. y J. Basl, *Building Data and AI Ethics Committees*, *op. cit. supra,* nota 446.

afectación de la IA en la sociedad en general.

Un Comité de Ética en IA debe nutrirse desde un enfoque interdisciplinario con el cual se pueda abordar adecuadamente las cuestiones éticas relacionadas con el desarrollo de la IA. De esta manera, también se asegura que se consideren todos los aspectos posibles asociados a un problema en concreto, desde los impactos a corto y largo plazo, hasta las implicaciones sociales y éticas. Además, un Comité de Ética en IA en el cual participen expertos de distintas áreas, ayuda a asegurar que la IA se desarrolle de manera responsable y en sintonía con los principios que se consideren base en toda deliberación en la cual tenga lugar dicha tecnología.

5. PROPUESTA: EL COMITÉ DE ÉTICA EN INTELIGENCIA ARTIFICIAL DE LA UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Llegados a este punto, es conveniente pensar en cuál tendría que ser el rol de la UNAM de cara a los agentes morales artificiales. La posible respuesta está vinculada directamente con la creación de un Comité de Ética en IA en el seno de nuestra Universidad, y bajo la supervisión del Comité Universitario de Ética.[467] La justificación más próxima para su creación tiene que ver con el abanico de dilemas éticos asociados a la IA que requieren análisis desde enfoques interdisciplinarios,

[467] Creado el 29 de agosto de 2019, el Comité Universitario de Ética es el órgano colegiado central, permanente, de carácter honorífico, técnico y especializado, que se constituye para supervisar y fomentar prácticas éticas universitarias. Sus decisiones y dictámenes son vinculantes para todos los Comités de Ética de la Universidad. CUÉTICA, Quiénes somos, [s.f., en línea], <https://cuetica.unam.mx/quienes-somos> [consulta: 17 de marzo de 2024]. Planteado desde el anterior mirador, los Comités de Ética Universitaria funcionan como la instancia mediante la cual se deliberan y debaten prácticas de integridad académica y científica. Es relevante destacar que, en línea con este enfoque, la División de Estudios de Posgrado de la Facultad de Derecho llevó a cabo en 2023 la elaboración del documento "Estándares para la integración de los trabajos de investigación. Ética académica y aparato crítico en el posgrado en Derecho", el cual tiene como objetivo establecer y proporcionar una metodología y criterios para la integración del formato y aparato crítico de los trabajos académicos, con especial énfasis en la integridad y honestidad académica para evitar el plagio en alguna de sus modalidades; así como para sumarse a las fuentes de consulta en metodología y técnicas de investigación. *Vid*. HEREDIA GARCÍA, Ana Eloísa *et al*., *Estándares para la integración de los trabajos de investigación ética académica y aparato crítico en el posgrado en Derecho*, División de Estudios de Posgrado de la Facultad de Derecho, UNAM, México, 2023 [en línea], <EstandaresParaLaIntegracionDeLosTrabajosDeInvestigacion.EticaAcademicaYaparatoCriticoEnElPosgradoEnDerecho2023_07_06.pdf (unam.mx)> [consulta: 17 de marzo de 2024].

particularmente los derivados de los sistemas desarrollados dentro de las instalaciones de la UNAM.

Una justificación un poco más alejada del presente es que contar con un Comité de Ética que exclusivamente delibere sobre cuestiones referentes a la IA en la UNAM, servirá también para que un futuro para que desde la Universidad se maduren conocimientos y experiencias que contribuyan a la investigación y desarrollo de agentes morales artificiales alineados a principios éticos y legales, tomando en cuenta los posibles impactos de éstos en la sociedad.

En ese tenor, ¿qué aspectos deberían considerarse para que la UNAM cuente con un Comité de Ética en IA?, un buen punto de partida para responder a esta pregunta se encuentra en el Acuerdo por el que se establecen los Lineamientos para la Integración, Conformación y Registro de los Comités de Ética en la UNAM, publicado en Gaceta UNAM el 20 de agosto de 2019. [468]

Dichos Lineamientos tienen por objeto regular la integración, registro y funcionamiento de los Comités de Ética en Investigación y Docencia, en sus distintas modalidades, y del Comité Universitario de Ética. Cabe señalar que este último Comité, de conformidad con el punto 29 del mencionado Acuerdo, es el órgano colegiado central, permanente, de carácter honorífico, técnico y especializado, que se constituye para supervisar y fomentar las prácticas éticas universitarias.

Los Lineamientos también establecen, en su punto 9, que los Comités de Ética en Investigación y Docencia podrán formarse en cualquier coordinación, entidad académica o dependencia universitaria que así lo requiera. Por lo que respecta a los tipos de Comité de Ética de Investigación y Docencia que pueden integrarse, el punto 11 de los Lineamientos señala los siguientes: I) Comité de Bioética; II) Comité de Bioseguridad; III) Comité de Integridad Académica y Científica; IV) Comité de Ética en Investigación Pública; V) Comité de Ética de la Investigación; VI) Comité

[468] UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO, *Acuerdo por el que se establecen los Lineamientos para la Integración, Conformación y Registro de los Comités de Ética en la Universidad Nacional Autónoma de México*, Publicado en Gaceta UNAM el 29 de agosto de 2019. [en línea], <https://www.abogadogeneral.unam.mx/sites/default/files/archivos/LegUniv/701-AcuerdoEstablecenLineamientosComitesEtica_290819.pdf>, [consulta: 03 de octubre de 2023].

de Cuidado y Uso de Animales para la Enseñanza e Investigación; VII) Comité de Ética Ambiental, y VII) Comité de Ética Administrativa.

Partiendo del hecho de que la IA tiene implicaciones directas en el ser humano, por ejemplo, en el caso de los sesgos producidos por los algoritmos, resulta conveniente pensar en medios a través que puedan servir como contención de cara a investigaciones realizadas en la UNAM en las que se utiliza a la IA, con la finalidad de que cumplan con estándares éticos apropiados.

El objetivo, a final de cuentas, sería el de evitar consecuencias graves en ámbitos cruciales para los seres humanos como la salud, la educación y la seguridad, donde los algoritmos son utilizados para tomar decisiones importantes. En efecto, la integración de un Comité de Ética de la Investigación resultaría la modalidad idónea para el Comité de Ética en IA de la UNAM. Cabe referir que, de acuerdo con los Lineamientos, el Comité de Ética de la Investigación se conforma cuando se llevan a cabo actividades de investigación o decencia en seres humanos que participan como sujetos de investigación en el marco de las disciplinas de las ciencias exactas, naturales, biológicas, humanidades y sociales, con el objetivo de salvaguardar sus derechos.

Los Lineamientos también dan luz sobre los principios aplicables en los casos en los que participan los seres humanos como sujetos de investigación. Estos principios son los siguientes:

- **Principio de autonomía:** implica el reconocimiento de la capacidad del sujeto de la investigación para la toma decisiones que se plasma en el consentimiento informado.

  Principio de no maleficencia: se entiende como la responsabilidad del personal académico o alumnado de minimizar los daños y riesgos reales o potenciales de quienes participan como sujetos de investigación.
- **Principio de beneficencia:** implica maximizar los beneficios para los sujetos participantes en la investigación, además de garantizar que el riesgo de las investigaciones sólo puede tomarse cuando no exista una alternativa.
- **Distribución justa**: es un principio donde la distribución de cargas y

beneficios es equilibrada o razonable atendiendo a las circunstancias de los sujetos de investigación.

Dicho esto, el primer aspecto a considerar es que el Comité de Ética en IA de la UNAM debe ser integrado como un Comité de Ética de la Investigación, en los términos señalados por los ya referidos Lineamientos.

Adicionalmente, tendríamos que considerar en qué coordinación, entidad académica o dependencia universitaria debería tener sede el Comité de Ética en IA. La opción principal es la de la Coordinación de Investigación Científica de la UNAM, la cual agrupa a 24 institutos y 6 centros, agrupados en tres grandes áreas del conocimiento: Ciencias Químico-Biológicas y de la Salud, Ciencias Físico Matemáticas y Ciencias de la Tierra e Ingenieras. La razón por la cual tendría que ser la opción principal es que del área de Ciencias Físico Matemáticas se desprenden investigaciones y actividades en materia de IA. [469]

Como una segunda opción se asoma el Instituto de Investigaciones en Matemáticas Aplicas y en Sistemas. Esta opción se justificaría en virtud de actividades recientes que implican el uso de IA, por ejemplo, el 4 de agosto de 2022 se inauguró un Laboratorio en IA y Alta Tecnología. Dicha noticia se dio en el marco de la firma de intención de la Alianza para Promover el Desarrollo de Capacidades Digitales en México, cuyo objetivo es apoyar e incentivar proyectos de innovación tecnológica, con énfasis en la IA y que contribuya a la solución de problemáticas sociales. [470]

En breve, es posible que la UNAM cuente con un Comité de Ética en IA, más aún si se consideran las investigaciones realizadas en el seno de la UNAM que implica el uso de IA. El contar con un Comité como el señalado podría resultar

[469] *Vid*. LÓPEZ, P., "Centurión, robot creado por empresa incubada en la UNAM", en *Gaceta UNAM*, 23 de enero de 2020. [en línea]<https://www.gaceta.unam.mx/empresa-incubada-en-la-unam-crea-robot-capacitado-para-ser-vendedor-estrella/>, [consulta: 03 de octubre de 2023] y SAAVEDRA, D., "Pumas crean algoritmo y ganan premio en Francia", en *Gaceta UNAM*, 13 de enero de 2020. [en línea]<https://www.gaceta.unam.mx/pumas-crean-algoritmo-y-ganan-premio-en-francia/>,[consulta: 03 de octubre de 2023].

[470] HERNÁNDEZ, M., "Inaugura el IIMAS Laboratorio de Inteligencia Artificial y Alta Tecnología", en *Gaceta UNAM*, 4 de agosto de 2022. [en línea]< https://www.gaceta.unam.mx/inaugura-el-iimas-laboratorio-de-inteligencia-artificial-y-alta-tecnologia/>, [consulta: 03 de octubre de 2023].

beneficioso por muchos sentidos, pero también redituable para la propia UNAM, pues podría convertirse en un referente a nivel nacional en cuanto a la Ética de la IA.

La propuesta enunciada también se nutre del interés reciente por parte de otras áreas del conocimiento, como lo es el Derecho, al interior de la UNAM. En ese sentido, no pueden dejarse de lado esfuerzos recientes en el tema, por ejemplo, desde el IIJ de la UNAM se cuenta con una Línea de Investigación en Derecho e IA, el cual ha servido como escaparate a temas contemporáneos, tales como el del uso de la IA en respuesta al Covid-19, la IA y la defensa de los Derechos Humanos, así como los dilemas y consecuencias para el quehacer académico del uso de procesadores de lenguaje natural. [471]

6. REFLEXIONES FINALES

El presente capítulo tuvo como objetivos los siguientes: situar a la Ética, como una forma de reflexión crítica de frente a cuestiones morales; referir a la Ética de la IA como una ética aplicada para hacer frente a los desafíos vinculados al desarrollo de la IA, e identificar a los Comités de Ética en IA como una acción necesaria para el presente y con miras a un hipotético escenario en el que el desarrollo de la IA traiga consigo a los agentes morales artificiales.

Con relación a este último objetivo, y ante el escenario descrito, los Comités de Ética podrían contribuir en el análisis de los dilemas éticos que el Agente Moral Artificial trajera consigo. Considerando que un Comité de Ética en IA tiene como uno de sus principales objetivos la identificación sistemática y exhaustiva de riesgos éticos de productos de IA que sean desarrollados de manera interna o comprados a terceros, es posible suponer que algunos de los cuestionamientos más relevantes a los cuales se enfrentarán ante la irrupción de los agentes morales artificiales, es si estos agentes serán sujetos de derechos y responsabilidades similares a los de los agentes morales humanos, y si tendrán que estar sujetos a leyes y regulaciones

[471] *Vid*. INSTITUTO DE INVESTIGACIONES JURÍDICAS, *Informe de Gestión 2022-2023*, UNAM, [en línea] <https://www.juridicas.unam.mx/informe-2022-2023/detalle/47>, [consulta:03 de octubre de 2023].

específicas, pero fundamentalmente, responder a la pregunta de si es éticamente justificable crear agentes morales.

En este capítulo también se enunció el núcleo de diez principios del Comité de Ética en IA, los cuales son fundamentales para su funcionamiento, pues con ellos se pretende que los proyectos de investigación y desarrollo en materia de IA cumplan con estándares éticos que contribuyan a la protección de derechos y sirvan como mecanismos para garantizar que la IA no cause perjuicios en los individuos y las comunidades.

En ese mismo sentido, también se hizo mención de que un Comité de Ética en IA requiere necesariamente de un grupo interdisciplinario de personas expertas en razón de que los dilemas éticos vinculados a la IA necesitan de un enfoque interdisciplinario para su mejor comprensión; en ese orden de ideas, el Comité de Ética en IA debería contar con expertos en Ética, en Derecho, en Tecnología, y en aquellas disciplinas sobre las cuales la IA pueda tener una repercusión directa; además, es necesaria la participación ciudadana, y de ser el caso, de estrategas empresariales.

Finalmente, se hizo alusión a una propuesta que pudiera colocar a la UNAM como pionera en el ámbito de la Ética de la IA en México, apostando por la creación de un Comité de Ética de la IA derivado del Comité de Ética Universitario a nivel central en la UNAM, cuyas funciones sean las de un Comité de Ética en Investigación y Docencia, y con sede en la Coordinación de Investigación Científica de la UNAM, en virtud de las investigaciones y actividades en materia de IA con las que cuenta.

CONCLUSIONES
EL COMITÉ DE ÉTICA EN INTELIGENCIA ARTIFICIAL COMO FACILITADOR DE ENFOQUES ÉTICOS FRENTE A LA AGENCIA MORAL ARTIFICIAL

La IA es una disciplina en constante evolución, la cual busca dotar a las computadoras de autonomía y adaptabilidad para realizar tareas de manera eficiente sin intervención humana. A lo largo de este trabajo, se exploraron los campos relacionados con la IA, como el aprendizaje automático, el aprendizaje profundo y la ciencia de datos, y se han discutido sus aplicaciones, beneficios, limitaciones y desafíos éticos y jurídicos. También, se examinó los conceptos de Agencia Moral y Agencia Moral Artificial, así como los desafíos socio-jurídicos y futuros en el desarrollo de la IA y el rol que juegan los comités de ética en IA frente a ellos, y en particular frente a la irrupción del agente moral artificial.

En el primer capítulo, se analizaron las tres categorías de la IA, incluyendo la IA aplicada, la IA Fuerte y la Superinteligencia. También se abordaron aplicaciones prácticas de la IA, como los vehículos autónomos, las armas autónomas y el Chat GPT, y se discutieron implicaciones éticas y jurídicas derivadas de su implementación.

En el segundo capítulo, se abordó el concepto de la Agencia Moral, y se identificaron las tres condiciones necesarias para su existencia: intencionalidad, autonomía y capacidad de actuar. Este concepto es crucial para comprender si es posible atribuir una Agencia Moral a los sistemas artificiales autónomos.

El tercer capítulo analizó las posturas más relevantes en torno a la idea de la Agencia Moral Artificial, concluyendo que en la actualidad no es posible atribuir Agencia Moral a los sistemas artificiales autónomos debido a las condiciones necesarias para atribuir una Agencia Moral. Sin embargo, no se descartó que en un futuro la posibilidad de que una superinteligencia pueda cubrir esas condiciones. En este tenor, la posibilidad de una Agencia Moral Artificial se ha convertido en temas centrales en el debate sobre la IA. En ese tenor, se enuncian brevemente alguna de las posturas exploradas:

- Desde la teoría de Luciano Floridi, el agente moral artificial es posible, y es identificado como un sistema de transición, interactivo, autónomo y adaptativo que puede ejecutar acciones susceptibles de calificación moral.
- Del debate entre el grupo de los "*Computational Modelers*" y el de "*Computers-in-Society*" se destaca la defensa que hace este último grupo del vínculo que debe prevalecer entre la tecnología con quienes la desarrollan, despliegan y usan, pues es en el ser humano en quien debe recaer la responsabilidad de los actos de los sistemas artificiales autónomos, desde esta postura la Agencia Moral Artificial no es posible.

Al explorar los tipos de agencia que pudieran desempeñar los sistemas artificiales autónomos, se concluye que estos pueden presentar un tipo de eficacia causal o una que actúe en nombre de la agencia humana, pero nunca de tipo autónoma, la cual se reserva sólo a los seres humanos.

En el cuarto capítulo se discutieron los desafíos socio-jurídicos presentes y futuros asociados al desarrollo de la IA. Los problemas actuales, como la discriminación algorítmica, la privacidad, el impacto en el empleo y la desinformación requieren de enfoques regulatorios, técnicos y de políticas públicas. En tanto que los problemas futuros como el transhumanismo y la posibilidad de otorgar derechos a los robots plantean desafíos éticos y jurídicos significativos, sobre los cuales vale la pena ir reflexionando desde ahora.

El quinto capítulo se centró en los Comités de Ética en IA y el rol que desempeñan frente los dilemas éticos relacionados con la Agencia Moral Artificial. Estos Comités pueden contribuir al análisis de estos dilemas y garantizar que los

proyectos de investigación y desarrollo en IA cumplan con estándares éticos para proteger los derechos y evitar daños a individuos y comunidades.

Entre los puntos más relevantes del quinto capítulo se encuentran los siguientes:

- La Ética se sitúa como una forma de reflexión crítica frente a cuestiones morales. En ese sentido, la Ética de la IA se trata de una ética aplicada que contribuye a hacer frente a los desafíos éticos vinculados al desarrollo de la IA.
- En el presente, los Comités de Ética en IA son necesarios para contribuir en el análisis de riesgos éticos de productos de IA; en el futuro, pueden contribuir en el análisis de los dilemas éticos que el Agente Moral Artificial traiga consigo. En ese tenor, es posible vislumbrar que los cuestionamientos más relevantes a los cuales se enfrentarán a la irrupción de dichos agentes es si estos serán sujetos de derechos y responsabilidades similares a los de los agentes morales humanos, y si tendrán que estar sujetos a leyes y regulaciones específicas, pero fundamentalmente, responder a la pregunta de si es éticamente justificable crear agentes morales artificiales.
- Los proyectos de investigación y desarrollo en materia de IA deben cumplir con estándares éticos, para ello, se enunció el núcleo de diez principios del Comité de Ética en IA. Este núcleo puede contribuir a la protección de derechos y como mecanismo para garantizar que la IA no cause perjuicios en los individuos y comunidades.
- El Comité de Ética en IA requiere también de un enfoque interdisciplinario, contando con personas expertas en Ética, Derecho, Tecnología y en aquellas disciplinas en las que la IA pueda tener una repercusión directa; adicionalmente, también es requerida la participación ciudadana y, en los casos requeridos, de estrategas empresariales.
- Como propuesta, la UNAM puede posicionarse como pionera en el ámbito de la Ética de la IA en México, con la creación de un Comité de Ética en Investigación y Docencia enfocado en temas de IA desde el aparato central existente.

Con las conclusiones señaladas, es posible señalar que a medida que la IA evoluciona y se desarrollan sistemas más avanzadas y autónomos, resulta crucial que las personas dedicadas a la investigación científica, aquellas en el ámbito legislativo, además de profesionales en todas las disciplinas y la sociedad en general estén preparados para enfrentar los desafíos éticos, jurídicos y sociales asociados a la IA, y en especial, para garantizar que la IA se utilice de manera responsable y ética.

Las preocupaciones relacionadas a la IA tienen cada vez mayor eco, por ejemplo, distintos expertos han llamado ya a una moratoria para detener durante seis meses el desarrollo y pruebas de sistemas de IA más poderosos que el Chat GPT, pues existe la preocupación latente de que este último modelo es ya capaz de competir con los humanos en un creciente número de tareas, destruyendo empleos y difundiendo desinformación. Cabe señalar que Elon Musk es una de las personas que rubricaron dicha petición, junto con el historiador Yuval Noah Harari, el cofundado de Apple Steve Wozniak, entre otros expertos.[472]

Si bien es cierto que una moratoria no parece ser la solución, y que hacer eso podría resultar una solución no efectiva para abordar preocupaciones éticas, sociales y tecnológicas asociadas con el avance de la IA, también lo es que sirve como un parteaguas para llamar la atención no sólo de los expertos, sino también de los gobiernos y la sociedad en general, con relación a la importancia de debatir y tener muy en cuenta las implicaciones sociales que la IA trae consigo.

Por lo tanto, resulta relevante identificar de una vez a los Comités de Ética en IA como guardianes y facilitares de un enfoque ético, asegurando de que los desarrollos en este campo estén alineados con los valores humanos y el bienestar de la sociedad, pero sobre todo a futuro, tratándose de evaluar posibles riesgos asociados a la idea de atribuir derechos y responsabilidades a los agentes morales artificiales.

[472] PASCUAL, M., “Expertos en inteligencia artificial reclaman frenar seis meses la ‘carrera sin control’ de los ChatGPT”, en *El País*, 29 de marzo de 2023, [en línea], <https://elpais.com/tecnologia/2023-03-29/expertos-en-inteligencia-artificial-reclaman-frenar-seis-meses-la-carrera-sin-control-de-los-chatgpt.html>, [consulta: 05 de octubre de 2023].

No sabemos con certeza qué depare el futuro próximo, pues siempre se asoma como un océano incierto, ante el cual siempre es mejor estar preparados con un barco resistente que nos pueda llevar a un puerto seguro. Ese puerto seguro se vislumbra como un Comité de Ética en IA que agrupe expertos capaces de hacer frente a la idea de un Agente Moral Artificial.

FUENTES CONSULTADAS

1. BIBLIOHEMEROGRÁFICAS

AGGARWAL, C. C., Neural Networks and Deep Learning, A Textbook, Springer, Estados Unidos de América, 2018. [ISBN: 978-3-319-94462-6].

ALBAHAR, M. y J. Almalki, “Deepfakes: Threats and Countermeasures Systematic Review”, en *Journal of Theoretical and Applied Information Technology*, vol. 97, no. 22, 2019. [ISSN: 1992-8645].

ALLEN, C., G. Varner, y J. Zinser, “Prolegomena to Any Future Artificial Moral Agent”, en *Machine Ethics and Robot Ethics*, núm. 12, 2020, pp. 53–63. [https://doi.org/10.4324/9781003074991-5].

ALLEN, R. y D. Masters, “Artificial Intelligence: the right to protection from discrimination caused by algorithms, machine learning and automated decision-making”, en *ERA Forum 20*, Springer, 2 de octubre de 2019. [https://doi.org/10.4324/9781003074991-5].

ÁLVAREZ, M. Y., *et al*., “Comités de bioética en las instituciones del sistema de salud”, en *Red Sociales, Revista del Departamento de Ciencias Sociales*, vol. 08, no. 2. [ISSN: 2362-4434].

AMIR, E., “Reasoning, and decision making”, en *The Cambridge Handbook of Artificial Intelligence,* Cambridge University Press, Reino Unido, 2014. [ISBN: 978-0-521-87142-6].

ARANA, J., “Acción Física y Acción Mental”, en *Thémata, Revista de Filosofía,* núm. 30, 2003, pp. 55-70. [ISSN: 0212-8365].

ARANGO-BAYER, G. L., “Los Comités de Ética de la Investigación, Objetivos, funcionamiento y principios que buscan proteger”, en *Investigación en Enfermería: Imagen y Desarrollo*, vol, 10, no. 1, Bogotá, 2008. [ISSN: 0124-2059]

ARKOUDAS, K. y S. Bringsjord, “Philosophical foundations”, en FRANKISH, K. y W. M. Ramsey, eds.*, The Cambridge Handbook of Artificial Intelligence,* Cambridge University Press, Reino Unido, 2014. [ISBN: 978-0-521-87142-6].

ARNAU, H., J. M. Gutiérrez y G. Navarro, ¿Qué es el utilitarismo?, *Colección Universitas-39*, PPU, 1992. [ISBN: 84-477-0056-9].

AUDI, R., *Diccionario Akal de Filosofía,* España, Akal, 2004. [ISBN 978-84-460-0956-6].

BARRIO ANDRÉS, M., "Consideraciones jurídicas acerca del coche autónomo", en *Actualidad Jurídica Uría Menéndez*, no. 52, 2019. [ISSN 1578-956X].

BARTNECK, C., C. Lütge, A. Wagner y S. Welsh, *An Introduction to Ethics in Robotics and AI*, Springer, Suiza, 2021. [ISBN: 978-3-030-51109-8] [https://doi.org/10.1007/978-3-030-51110-4].

BEHDADI, D. y C. Munthe, "A Normative Approach to Artificial Moral Agency", en *Minds and Machines*, núm. 30, 2020, pp. 195-218. [https://doi.org/10.1007/s11023-020-09525-8].

BLACKMORE, S., *Consciousness, An Introduction*, Hodder Education, Reino Unido, 2010. [ISBN: 978 1 444 104 875].

BOSTROM, N. y E. Yudkowsky, "The ethics of artificial intelligence", en Keith F. y M. R. William, eds.*, The Cambridge Handbook of Artificial Intelligence, op. cit., supra*, nota 2. [ISBN: 978-0-521-87142-6].

BOSTROM, N., "Ethichal Issues in Advanced Artificial Intelligence", en S. Schneider, ed., *Science Fiction and Philosophy, From Time Travel to Superintelligence*, Reino Unido, Wiley-Blackwell, 2009. [ISBN: 978-1-405-14906-8].

----------., "Transhumanist Values", en *Journal of Philosophical Research*, Vol. 30, 2005. . [ISBN: 1053-8364] [https://doi.org/10.5840/jpr_2005_26].

BROZEK, B. y B. Janik, "Can artificial intelligences be moral agents", en *New Ideas in Psychology*, núm. 54, 2019. [https://doi.org/10.1016/j.newideapsych.2018.12.002].

CABEZAS HERNÁNDEZ, M., "Autonomía y emocionalidad en el agente moral", en *Factótum, Revista de Filosofía,* núm. 7, 2010. [ISSN 1989-9092].

CALVO PÉREZ, J. L., "Debate Internacional en torno a los sistemas de armas autónomos letales. *Consideraciones tecnológicas, jurídicas y éticas", en Revista General de Marina*, Vol. 278, no. 4, 2020. [ISSN 0034-9569].

CAMPBELL, L., “Kant, autonomy and bioethics”, en *Ethics, Medicine and Public Health*, vol. 3 núm. 3, 2017, pp. 1-12. [DOI: 10.1016/j.jemep.2017.05.008].

CAMPIONE, R., "Recopilar y vigilar: algunas consideraciones filosófico-jurídicas sobre inteligencia artificial”, en *Sociología y Tecnociencia*, no. 11, 2021. [ISSN-e 1989-8487].

CASTRO MANZANO, J. M., *Revisión de Intenciones desde el Aprendizaje BDI,* Tesis de maestro, México, Universidad Veracruzana, Departamento de Inteligencia Artificial, 2008, p. 28.

CECCON, E., “La revolución verde en dos actos”, en *Ciencias*, Universidad Nacional Autónoma de México, vol. 1, no. 91, 2008. [ISSN: 0187-6376].

CHAVES PALACIOS, J., “Desarrollo tecnológico en la primera revolución industrial”, en Norba. *Revista de Historia*, vol. 17, 2004. [ISSN-e: 0213-375X].

CHÁVEZ VALDIVIA, A. K., “Hacia otra dimensión jurídica: el derecho de los robots”, en *IUS, Revista del Instituto de Ciencias Jurídicas de Puebla*, vol. 15, 2021. [ISSN: 1870-2147] [https://doi.org/10.35487/rius.v15i48.2021.681].

CHRISTEN, M. y M. Alfano, “Outlining the Field - A Research Program for Empirically Informed Ethics”, en M. Christen *et al.*, ed., *Empirically Informed Ethics: Morality between Facts and Norms*, Springer, Suiza, 2014, pp. 3-27. [ISBN 978-3-319-01369-5].

COECKELBERGH, M., *AI Ethics*, The MIT Press, Estados Unidos de América, 2020. [ISBN: 9780262538190].

CORRAL CUARTAS, Á., “En busca de la dimensión intencional de las emociones y los estados de ánimo”, en *Ideas y Valores*, vol. 66 núm. 3, 2017, [en línea], <http://www.scielo.org.co/pdf/idval/v66s3/0120-0062-idval-66-s3-00047.pdf>, [consulta: 11 de septiembre de 2023]. [ISSN 2011-3668] [DOI: 10.15446/ideasyvalores.v66n3Supl.65636].

CORTINA, A. y E. Martínez, *Ética*, Editorial Akal, España, 3ª edición, 2001. [ISBN: 84-460-0674-X].

CORTINA, A., “Ética de la Inteligencia Artificial”, en *Anales de la Real Academia de Ciencias Morales y Políticas*, no. 96, 2019. [ISSN: 0210-4121].

CRAWFORD, K., *Atlas of AI, Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, Reino Unido, 2021. [ISBN: 0300252390].

DANÓN, L., "Explicaciones intencionales y explicaciones teleológicas de la conducta animal", en *Revista Argentina de Ciencias del Comportamiento*, vol. 3 núm. 1, 2011, pp. 54-63 [ISSN-e 1852-4206]

DE ASÍS, R., *Inteligencia Artificial y Derechos Humanos*, Materiales de Filosofía del Derecho, Universidad Carlos III de Madrid, No. 2020. [ISSN: 2531-0240].

DE PINA, R. y R. de Pina Vara, *Diccionario de Derecho*, ed. Porrúa, México, 2010. [ISBN: 9789700769097].

DE ZAN, J., *La ética, los derechos y la justicia*, Fundación Konrad-Adenauer, Uruguay, 2004. [ISBN: 9974-7868-2-7].

ĐERIĆ, M., "What is Autonomy Anyway?, en M. Kühler y V. L. MItrovic, eds., *Theories of the Self and Autonomy* in Medical Ethics, vol. 83, The International Library of Bioethics, Springer, Suiza, 2020, p. 17.

DEVITT, M., "Realismo Moral: Una Perspectiva Naturalista", en *ARETÉ Revista de Filosofía,* vol. XVI núm. 2, 2004, pp. 185-206. [ISSN 1016-913X].

DÍEZ SPELZ, J. F., "¿Robots con derechos?: la frontera entre lo humano y lo no-humano. Reflexiones desde la teoría de los derechos humanos", en *IUS, Revista del Instituto de Ciencias Jurídicas de Puebla*, vol. 15, núm. 48, 2021. [https://doi.org/10.35487/rius.v15i48.2021.742].

DOLORES RUIZ, M. M., *UD3: Los comités de ética*, Cuaderno para Curso: Bioética, 2009. [ISSN: no disponible].

DURANTE, M., *Ethics, Law and the Politics of Information, A Guide to the Philosophy of Luciano Floridi,* vol. 18, The International Library of Ethics, Law and Technology, Springer, Holanda, 2017. [ISBN: 978-94-024-1150-8] [https://doi.org/10.1007/978-94-024-1150-8].

DWORKIN, G., *The Theory and Practice of Autonomy,* EEUU, Cambridge University Press, 1988. [ISBN: 9780521344524]

FERRATER MORA, J., *Diccionario de Filosofía*, Tomo I, A-K, Argentina,

Montecasino, 1964.

FLETCHER, J., “Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance”, en *Theatre Journal*, vol. 70, no. 4, 2018. [ISSN: 1086-332X].

FLORIDI, L. y J. W. Sanders, “On the Morality of Artificial Agents”, en *Minds and Machines*, núm. 14, 2004. [https://doi.org/10.1023/B:MIND.0000035461.63578.9d].

FLORIDI, L., “Ethics in the Infosphere”, *versión revisada y presentada en la 161va reunión del Consejo Ejecutivo de la UNESCO “Las nuevas tecnologías de la información y la comunicación para el desarrollo de la educación*, 31 de mayo de 2001.

----------., “Ética de la Información: su naturaleza y alcance”, en *Isegoria*, núm. 34, trad. de Roberto Feltrero y Paula Olmos, 2006. [https://doi.org/10.3989/isegoria.2006.i34.2].

FORD, M., “Could Artificial Intelligence Create an Unemployment Crisis?”, en *Communications of the ACM*, vol. 56, no. 7, julio 2013. [https://doi.org/10.1145/2483852.2483865].

FRANA, P. L., y J.K. Michael, eds., Voz “Nick Bostrom”, en *Encyclopedia of Artificial Intelligence, The Past, Present and Future of AI,* ABC-CLIO, Estados Unidos de América, 2021. [ISBN: 978-1-4408-5327-2].

FRANKISH, K. y W. M. Ramsey, *The Cambridge Handbook of Artificial Intelligence, The Cambridge Handbook of Artificial Intelligence,* Cambridge University Press, Reino Unido, 2014. [ISBN: 978-0-521-87142-6].

FRANKLIN, S, “History, motivations, and core themes”, en K. Frankish y W. M. Ramsey, eds., *The Cambridge Handbook of Artificial Intelligence, The Cambridge Handbook of Artificial Intelligence,* Cambridge University Press, Reino Unido, 2014. [ISBN: 978-0-521-87142-6].

GARCÍA, J. *et al*., *Ciencia de Datos, técnicas analíticas y aprendizaje estadístico*, editorial Alfaomega, Colombia, 2018. [ISBN: 978-607-538-252-4].

GAYOZZO HUAMANCHUMO, P. A., “Singularidad tecnológica y transhumanismo”, en *Teknokultura, Revista de Cultura Digital y Movimientos Sociales*, no. 18,

2021. [https://doi.org/10.5209/tekn.74056].

GIL, E., *Big Data, Privacidad y Protección de Datos*, Agencia Española de Protección de Datos, 2016. [ISBN: 978-84-340-2309-3].

GONZÁLEZ GRANADO, J., *De la persona algorítmica, A propósito de la personalidad jurídica de la inteligencia artificial,* Colección de Bioética, Edicions de la Universitat de Barcelona, España, 2020. [ISBN: 978-84-9168-429-9].

GUERRA ESPINOSA, R. A. y J. Tisné Niemann, "Vehículos autónomos y estado de necesidad: Análisis desde la perspectiva del peatón sujeto a una situación de peligro", en *Revista Chilena de Derecho y Tecnología*, vol. 10, núm. 2, 2021. [ISSN: 0719-2584] [http://dx.doi.org/10.5354/0719-2584.2021.57874].

GUNKEL, D. J, "A Vindication of the Rights of Machines" en *Machine Ethics and Robot Ethics*, núm. 511–30, 2020. [https://doi.org/10.4324/9781003074991-41].

HAN, B., *La Sociedad de la Transparencia*, España, Herder, 2013. [https://doi.org/10.5209/CIYC.52976].

HARARI, Y. N., "*Homo Deus, Breve historia del mañana*", Debate, 2016. [ISBN: 978-6073149891].

HEREDIA GARCÍA, Ana Eloísa *et al*., *Estándares para la integración de los trabajos de investigación ética académica y aparato crítico en el posgrado en Derecho*, División de Estudios de Posgrado de la Facultad de Derecho, UNAM, México, 2023 [en línea], <EstandaresParaLaIntegracionDeLosTrabajosDeInvestigacion.EticaAcademicaYaparatoCriticoEnElPosgradoEnDerecho2023_07_06.pdf (unam.mx)> [consulta: 17 de marzo de 2024]. [ISBN 978-607-30-7630-2].

HUMPRIES, J. y B. Schneider, "El trabajo en el siglo XXI", en *El Trimestre Económico*, vol. 87, no. 346, 2020. [ISSN 2448-718X] [https://doi.org/10.20430/ete.v87i346.1070].

JOHANSSON, L., "Robots and Moral Agency", en *Journal of Technoethics*. 2011. [en línea] <http://www.diva-

portal.org/smash/get/diva2:410512/FULLTEXT02.pdf>, [consulta: 03 de octubre de 2023].

JOHNSON, D. y K. W. Miller, “Un-making artificial moral agents”, en *Ethics and Information Technology*, núm. 10, 2008, pp. 123-133. [https://doi.org/10.1007/s10676-008-9174-6].

JORBA GRAU, M., “La intencionalidad: entre Husserl y la filosofía de la mente contemporánea”, en *Investigaciones Fenomenológicas, Revista de la Sociedad Española de Fenomenología,* núm. 8, 2011. [e-ISSN: 1885-1088].

JORDAN, S. R., *Designing an Artificial Intelligence Research Review Committee*, Future of Privacy Forum, Estados Unidos de América, 2019. [ISSN: no disponible].

KAPLAN, J., *Artificial Intelligence, What Everyone Needs to Know*, Oxford University Press, Estados Unidos de América, 2016. [ISBN: 978-0-19-060239-0] [https://doi.org/10.1177/0170840618792173].

KELLEHER, J. D., *Deep Learning*, MIT Press, Estados Unidos de América, 2019. [ISBN: 0262354896].

KELLEHER, J. D. y B. Tierney, *Data Science*, The MIT Press, Estados Unidos de América, 2018. [ISBN: 9780262535434].

KLEINBERG, J., J. Ludgwig, S. Mullainathan y C. R. Sunstein, “Discrimination in the age of algorithms”, en *Journal of Legal Analysis*, Vol. 10, 2018. [https://doi.org/10.1093/jla/laz001].

KORSGAARD, C., *Las Fuentes de la Normatividad*, UNAM, Instituto de Investigaciones Filosóficas, México, 2000. [ISBN:968–36–8389–4]

KRAMER, Z., “What is Chat GPT and how does it work?”, en *Fresered.com*, febrero 2023. [en línea] <https://www.freshered.com/what-is-chat-gpt-and-how-does-it-work/> [fecha de consulta: 06 de septiembre de 2023].

KREUTZER, R. y M. Sirrenberg, *Understanding Artificial Intelligence, Fundamentals, Use Cases and Methods for a Corporate AI Journey*, Springer, Suiza, 2020. [ISBN: 978-3-030-25271-7] [https://doi.org/10.1007/978-3-030-25271-7].

KUNER, C., F. H. Cate *et al.,* “Expanding the artificial intelligence-data protection

debate", en *International Data Privacy Law*, Vol. 8, No. 4, 2018. [https://doi.org/10.1093/idpl/ipy024].

LAHERA SÁNCHEZ, A., "El debate sobre la digitalización y la robotización del trabajo (humano) del futuro: automatización de sustitución, pragmatismo tecnológico, automatización de integración y heteromatización", en *Revista Española de Sociología*, no. 30, 2021. [https://doi.org/10.22325/fes/res.2021.66].

LANE FOX, R., *The Classical World, An Epic History of Greece and Rome*, Estados Unidos de América, Penguin, 2006. [ISBN 10: 0141021411].

LATORRE SENTÍS, J.I., *Ética para Máquinas*, Editorial Planeta, España, 2019. [ISBN: 9789584280053].

LEE, K., "La inteligencia artificial y el futuro del trabajo: una perspectiva china", en *El Trabajo en la Era de los Datos*, OpenMind BBVA, 2019. [ISSN: no disponible].

LÓPEZ HERNÁNDEZ, E., "La personalidad en el juicio", en *Revista Mexicana de Derecho*, no. 2, Colegio de Notarios del Distrito Federal, [en línea], <http://historico.juridicas.unam.mx/publica/librev/rev/mexder/cont/2/cnt/cnt2.pdf> [consulta: 2 de octubre de 2023].

MABASO, B., "Artificial Moral Agents Within an Ethos of AI4SG", en *Philosophy and Technology*, 2020. [https://doi.org/10.1007/s13347-020-00400-z].

MAINZER, K., *Artificial Intelligence — When do machines take over?,* Springer, Suiza, 2020. [ISBN: 978-3-662-59717-0] [https://doi.org/10.1007/978-3-662-59717-0].

MALLE, B. F., "Moral Competence in Robots?" en *Frontiers in Artificial Intelligence and Applications,* núm. 273, 2014. [https://doi.org/10.3233/978-1-61499-480-0-189].

MANHEIM, Karl y L. Kaplan, "Artificial Intelligence: Risk to Privacy and Democracy", en *The Yale Journal of Law & Technology*, vol. 21, no. 106, 2019. [ISSN: No disponible]

MANZANO, F., "¿Avance tecnológico o desempleo?", *Reporte Técnico*, 2019. [http://dx.doi.org/10.13140/RG.2.2.22300.44164].

MARÍN GARCÍA, S., *Ética e Inteligencia Artificial*, Cuadernos de la Cátedra CaixaBank de Responsabilidad Social Corporativa, no. 42, septiembre de 2019, Universidad de Navarra, España. [https://dx.doi.org/10.15581/018.ST-522].

MARQUISIO AGUIRRE, R., "El ideal de la autonomía moral", en *Revista de la Facultad de Derecho*, núm. 74 jul-dic. 2017, Universidad de la República, Uruguay. [*ISSN:* 2301-0665] [https://doi.org/10.22187/rfd2017n2a4].

MARTÍNEZ DEVIA, A., "La Inteligencia Artificial, el Big Data y la era digital: ¿una amenaza para los datos personales", en *Revista la Propiedad Inmaterial*, no. 27, 2019. [https://doi.org/10.18601/16571959.n27.01].

MARTÍNEZ MERCADAL, J. J., "Vehículos Autónomos y Derecho de Daños", en *Revista de la Facultad de Ciencias Económicas*, Número 20, 2018. [http://dx.doi.org/10.30972/rfce.0203267].

MAZO ÁLVAREZ, H.M., "La Autonomía: Principio Ético Contemporáneo", en *Revista Colombiana de Ciencias Sociales*, vol.. 3 núm. 1, 2012, pp. 115-132. [ISSN: 2216-1201].

MCCARTHY, J., *Ascribing Mental Qualities to Machines,* Computer Science Department, EEUU, Universidad de Standford, 1979.

MEDINA ARELLANO, M. J. y J. Hincapié Sánchez, *Bioética: teorías y principios*, Serie Libros Digitales, núm. 1, Instituto de Investigaciones Jurídicas, UNAM, México, 2019. [ISBN electrónico: 978-607-30-2492-1].

MEDINA ARELLANO, M. J., M. E. Cancino Marentes, A. Gascón Cervantes y A. de Lara Ramírez, Comités *de ética y bioética*, Serie Libros Digitales, núm. 3, Instituto de Investigaciones Jurídicas, UNAM, México, 2019, pp. 48-49. [ISBN electrónico: 978-607-30-2480-8]

MEDINA ARELLANO, M. J. y M. I. Castañeda Avalos, "Comité Universitario de Bioética en la Universidad Autónoma de Nayarit", en *Revista Bio Ciencias*, 2013. [ISSN 2007-3380].

MEZA RIVAS, M., "Los sistemas de armas completamente autónomos: un desafío para la comunidad internacional en el seno de las Naciones Unidas", en *Instituto Español de Estudios Estratégicos*, Gobierno de España, 18 de

agosto de 2016. [ISSN-e: 2530-125X].

MONASTERIO ASTOBIZA, A., “Ética algorítmica: implicaciones éticas de una sociedad cada vez más gobernada por algoritmos”, en *Revista Dilemata,* núm. 24 año 9, 2017. [ISSN-e: 1989-7022].

MORALES ZÚÑIGA, H., “Estatus moral y el concepto de persona”, en F. Vergara Ceballos, ed., *Problemas Actuales de la Filosofía Jurídica*, Chile, Librotecnia, 2015. [ISBN: 9789563271300].

MORENO, A, *et al., Aprendizaje automático,* Edicions de la Universitat Politecnica de Catalunya, España, 1994. [ISBN: 9788483019962].

MUÑOZ GUTIÉRREZ, C., “La discriminación en una sociedad automatizada: Contribuciones desde América Latina”, en *Revista Chilena de Derecho y Tecnología*, vol. 10, no. 1, 2021. [https://doi.org/10.5354/0719-2584.2021.58793].

NATH, R. y V. Sahu, “The problem of machine ethics in artificial intelligence”, en *AI & Society*, núm. 35, 2017. [https://doi.org/10.1007/s00146-017-0768-6].

NAVA, A. y Naspleda, F., “Inteligencia artificial, automatización, restructuración capitalista y futuro del trabajo: un estado de cuestión”, en *CEC*, no. 12, 2020. [ISSN: 2408-400X].

NÚÑEZ PARTIDO, J. P., “Hombres y máquinas. Futuro y límites del transhumanismo”, en *Razón y Fe*, no. 1434, 2018. [ISSN: 0034-0235].

NYHOLM, S., *Humans and Robots, Ethics, Agency and Anthropomorphism,* Rowman & Littlefield, Reino Unido, 2020.

OCHOA CHEHAB, X., W. Fierro Saltos y J. Cárdenas Benavides, “La Ciencia de los datos”, en ILLESCAS ESPINOZA, W., *et al*, coords., *Las Ciudades Inteligentes*, Ediciones UTMACH, Ecuador, 2018, pp. 23-24.

OTTAVIANO, J. M., “La amenaza fantasma: Inteligencia Artificial y derechos laborales”, en *Nueva Sociedad*, no. 294, 2021. [ISSN 0251-3552].

PAPACCHINI, A., “El porvenir de la ética: La autonomía moral, un valor imprescindible para nuestro tiempo”, en *Revista de Estudios Sociales, Universidad de los Andes*, núm. 5, 2000, s.p. [en línea], <https://journals.openedition.org/revestudsoc/30167#quotation>, [consulta:

12 de septiembre de 2023]. [ISSN: 0123-885X].

PARLAMENTO EUROPEO, *European Civil Law Rules in Robotics, Study for the JURI Committee*, 2016. [octubre 2016, en línea], <https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf > [consulta: 2 de octubre de 2023].

PARODI, G. F., "La renta básica ciudadana, opción frente al desempleo tecnológico y la corrupción", en *ACADEMO Revista de Investigación en Ciencias Sociales y Humanidades*, vol. 2, no. 1, 2012. [ISSN-e: 2414-8938].

PIEDRA ALEGRÍA, J., "Transhumanismo: hacia un nuevo cuerpo", en *Daimon, Revista Internacional de Filosofía*, 2016. [https://doi.org/10.6018/daimon/270011].

QUER, M. "Reseña de Cuerpos inadecuados. El desafío transhumanista a la filosofía, Diéguez, Antonio (2021), Herder, Barcelona. 2014", en *Revista de Filosofía Open Insight*, vol. XIII, núm. 27, 2022. [s.f., en línea], <https://www.redalyc.org/journal/4216/421669968008/html/> [consulta: 01 de octubre de 2022].

RAMIÓ MATAS, C., "El impacto de la inteligencia artificial y la robótica en el empleo público", en *GIGAPP Estudios Working Papers*, no. 98, 2018. [ISSN: 2174-9515].

RICHART, A., "El origen evolutivo de la agencia moral y sus implicaciones para la ética", en *Pensamiento*, *Revista de Investigación e Información Filosófica*, vol. 72 núm. 273, 2016. [ISSN 0031-4749] [DOI: 10.14422/pen.v72.i273.y2016.005]

RIVERA LÓPEZ, E. Derecho y Bioética, en J. L. Fabra Zamora y E. Spector (Coords.), "*Enciclopedia de filosofía y teoría del derecho*", Volumen, III, Universidad Nacional Autónoma de México, Instituto de Investigaciones Jurídicas, México, 2015. [ISBN 978-607-02-6617-1].

RODRÍGUEZ CORONEL, M. M., "A vueltas con agencia moral: una perspectiva crítica", en *Quaderns de filosofia i ciència*, núm. 42, 2012, pp.127-138. [ISSN-e 0213-5965].

RÖNNEGARD, D., *The Fallacy of Corporate Moral Agency*, vol. 44, Issues in

Business Ethics, Springer, Holanda, 2015. [ISBN 978-94-017-9756-6].

RUEDA CASTRO, L., “Interdisciplinariedad y Comités de Ética”, en *Revista Latinoamericana de Bioética*, vol. 12, no. 2, 2012. [ISSN: 1657-4702].

RUSELL, S. J. y P. Norvig, *Artificial Intelligence, A Modern Approach*, 3ª. edición, Pearson, Estados Unidos de América. [ISBN-13: 978-0-13-604259-4].

SALYER, K. M., “Deep fakes and National Security”, en *Congressional Research Service*, 2020, p. 1.

SÁNCHEZ VÁZQUEZ, A., Ética, Debolsillo, México, 2009, p. 22.

SÁNCHEZ-PESCADOR, J., “En torno a la intencionalidad”, en *Revista de Filosofía*, Universidad Complutense Madrid, vol. 3 núm. 14, 1995, pp. 30-44. [DOI: 10.5209/RESF].

SANDLER, R. y J. Basl, *Building Data and AI Ethics Committees*, Accenture y Ethics Institute at Northeastern University, 2019, p. 11. [en línea] <https://ethics.harvard.edu/news/how-build-data-and-ai-ethics-committees > [consulta: 02 de octubre de 2023]. [ISSN: no disponible].

SCHMIDT, N. y B. Stephens, “An Introduction to Artificial Intelligence And Solutions to the Problems of Algorithmic Discrimination”, en *Quarterly Report*, Vol. 73, No. 2, 2019. [https://doi.org/10.48550/arXiv.1911.05755].

SEARLE, J., “Minds, brain, and programs”, en *The Behavioral and Brain Sciences*, núm. 3, 1980. [https://doi.org/10.1017/S0140525X00005756].

----------., *Minds, Brain and Sciece*, Harvard University Press, Estados Unidos de América, 2003. [ISBN: 9780674576339].

SHANAHAN, M., *The Technological Singularity*, The MIT Press Essential Knowledge Series, Massachusetts Institute of Technology, 2015. . [ISBN: 9780262527804].

SIE, M., “Moral Agency, Conscious Control, and Deliberative Awareness”, en *Inquiry: An Interdisciplinary Journal of Philosophy*, vol. 52, núm. 5, 2009, pp. 516-531. [DOI: 10.1080/00201740903302642].

SINGH, M. P., *Multiagent Systems: A theoretical framework for intentions, know how, and communication,* núm. 799, Lecture Notes in Computer Sciences, Springer, Holanda, 1995. [eBook ISBN 978-3-540-48418-9] [DOI:

10.1007/BFb0030531].

SIURANA APARISI, J. C., Los principios de la bioética y el surgimiento de una bioética intercultural, en "*Veritas"*, no. 22, 2010. [ISSN: 0718-9273].

SUGARMAN, J., "Persons and Moral Agency" en *Theory & Psychology*, vol. 15, núm. 6, 2005. [DOI: 10.1177/0959354305059333].

THOMAS, M., Voz "Driverless Cars and Trucks", en P. L. Frana y M. J. Klein, eds., *Encyclopedia of Artificial Intelligence, The Past, Present and Future of AI*, ABC-CLIO, Estados Unidos de América, 2021. [ISBN: 978-1-4408-5327-2].

THORNE LUPKER, J. A., Voz "Deep Learning", en P. L. Frana y M. J. Klein, eds., *Encyclpedia of Artificial Intelligence, The past, Present, and Future of AI,* ABC-CLIO, Estados Unidos de América, 2021. [ISBN: 978-1-4408-5327-2].

TOLLON, F., "Moral Agents or Mindless Machines? A Critical Appraisal of Agency in Artificial Systems", en *Hungarian Philosophical Review*, vol. 4 núm. 63, 2019, pp. 9-23. [ISSN: no disponible].

TOYAMA MIYAGUSUKU, J. y A. Rodríguez León, "Algoritmos labores: big data e inteligencia artificial", en *Themis-Revista de Derecho*, no. 75, 2019. [https://doi.org/10.18800/themis.201901.018].

TURNER, J., *Robot Rules, Regulating Artificial Intelligence*, Palgrave Macmillan, Reino Unido, 2019. [ISBN: 978-3-319-96235-1].

UNESCO, *Funcionamiento de los comités de bioética: procedimientos y políticas*, Guía no. 2, Francia, 2006. [ISSN: no disponible].

VARELA DE ALBURQUERQUE DALPRÁ, J., "La protección del trabajo digno mediante los impactos de las nuevas formas de robótica laboral, inteligencia artificial y nuevas tecnologías", en *Revista Internacional y Comparada de Relaciones Laborales y Derecho del Empleo*, vol. 8, no. 3, 2020.

VÉLIZ, C., "Moral zombies: why algorithms are not moral agents", *AI & Society*, 2021. [en línea], <https://link.springer.com/article/10.1007/s00146-021-01189-x>, [consulta: 12 de septiembre de 2023]. [https://doi.org/10.1007/s00146-021-01189-x.].

VIESCA, C., "Perspectiva histórica de los Comités de Ética", en *Revista CONAMED*, Año 4, no. 12, 1999.

VIGEVANO, M., “Inteligencia artificial aplicable a los conflictos armados: límites jurídicos y éticos, en *Arbor Ciencia, Pensamiento y Cultura*, no. 197, abril-junio 2021. [https://doi.org/10.3989/arbor.2021.800002].

VILLALBA GÓMEZ, J. A., “Problemas bioéticos emergentes en la inteligencia artificial”, en *Diversitas: Perspectivas en Psicología*, Universidad Santo Tomás, Colombia, vol. 12, no. 1, 2016, p. 139.

VILLALBA, J., “Algor-ética: la ética en la inteligencia artificial”, en *Revista Anales de la Facultad de Ciencias Jurídicas y Sociales*, Universidad Nacional de la Plata, no. 50, 2020. [ISSN: 1794-9998].

VILLAROEL, R., “Consideraciones Bioéticas y Biopolíticas acerca del transhumanismo. El debate en torno a una posible experiencia posthumana”, en *Revista de Filosofía*, vol. 71, 2015. [ISSN: 0718-4360].

VIVEROS ÁLVAREZ, J., *Sistemas de Armas Autónomas: el dilema de la rendición de cuentas*, Instituto de Investigaciones Jurídicas, Universidad Nacional Autónoma de México, 2021. [ISBN: 978-607-30-1256-0].

VON MISES, L., *Human Action, A Treatise on Economics*, Estados Unidos de América, Fox & Wilkes, 1996. [ISBN 0-930073-18-5].

WANG, J., Y. Shao, Y. Ge y R. Yu, “A Survey of Vehicle to Everything (V2X) Testing”, en *Sensors*, no. 19, 2019. [https://doi.org/10.3390/s19020334].

WARREN, M. A., “Moral Status”, en R.G. Frey y C. Heath Wellman, coords., *A Companion To Applied Ethics*, Blackwell Publishing, Reino Unido, 2002. [ISBN 9781557865946] [DOI: 10.1002/9780470996621].

WESTERLUND, M., “The Emergence of Deepfake Technology: A Review”, en *Technology Innovation Management Review*, vol. 9, no. 11, 2019. [ISSN: No disponible].

XENEDIS, R. y L. Senden, “EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination”, en Ulf, B. et al (eds.), *General Principles of EU law and the EU Digital Order*, Kluwer Law International, 2020. [ISBN: 9789403511658].

XENEDIS, R., “Tuning EU equality law to algorithmic discrimination: Three pathways to resilience, en *Maastricht Journal of European and Comparative Law*, Vol.

26, 2020. [https://doi.org/10.1177/1023263X20982173].

ZUIDERVEEN, F. J., “Strengthening legal protection against discrimination by algorithms and artificial intelligence”, en *The International Journal of Human Rights*, Vol. 24, no. 10, 2020. [https://doi.org/10.1080/13642987.2020.1743976].

2. LEGISLACIÓN

UNESCO, *Informe de la Comisión de Ciencias Sociales y Humanas (SHS)*, 41 C/73, 22 de noviembre de 2021. [ISSN: no disponible].

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO, *Acuerdo por el que se establecen los Lineamientos para la Integración, Conformación y Registro de los Comités de Ética en la Universidad Nacional Autónoma de México*, Publicado en Gaceta UNAM el 29 de agosto de 2019. [en línea], <http://www.abogadogeneral.unam.mx:6060/acuerdos/view/701>, [consulta: 03 de octubre de 2023].

3. PÁGINAS WEB

AGÜERA MARTÍNEZ, M.B., “¿Qué es ChatGPT y cómo puede revolucionar la Inteligencia Artificial?, en *Lisa News*, 4 de enero de 2023, [en línea], <https://www.lisanews.org/tecnologia/que-es-chatgpt-y-como-puede-revolucionar-la-inteligencia-artificial/>, [consulta: 10 de septiembre de 2023].

ALLENDE LÓPEZ, M., “¿Cómo funciona la computación cuántica?”, en *Abierto al público, Banco Interamericano de Desarrollo*, 31 de mayo de 2019. [en línea] <https://blogs.iadb.org/conocimiento-abierto/es/como-funciona-la-computacion-cuantica/> [fecha de consulta: 08 de septiembre de 2023].

ÁLVARO, S., Living with Smart Algorithms, en *CCBLAB*, 3 de mayo de 2017, [en línea], <https://lab.cccb.org/en/living-with-smart-algorithms/>, [consulta: 18 de septiembre de 2023].

ANAYA HUERTAS, A., “¿Animales ante los tribunales”, en *Nexos*, el juego de la

Suprema Corte, 22 de junio de 2010. [en línea], <https://eljuegodelacorte.nexos.com.mx/%C2%BFanimales-ante-los-tribunales/>, [consulta: 12 de septiembre de 2023].

BBC NEWS*, Qué ocurre cuando aceptas las cookies y por qué es conveniente borrarlas del navegador de vez en cuando*, 29 de junio de 2017, en línea], <https://www.bbc.com/mundo/noticias-40443519>, [consulta: 18 de septiembre de 2023].

BLACKMAN, R., "Why You Need an AI Ethics Committee", en *Harvard Business Review*, Julio-Agosto 2022, [en línea], <https://hbr.org/2022/07/why-you-need-an-ai-ethics-committee>, [consulta: 02 de octubre de 2023].

BUSINESS INSIDER, *Google cancela su Comité para regular la Ética de la Inteligencia Artificial tan solo una semana después de su creación*, 5 de abril de 2019. [en línea], <https://www.businessinsider.es/google-cancela-comite-regular-etica-inteligencia-artificial-400609>, [consulta: 02 de octubre de 2023].

CHAKRABORTY, P., "Applied AI is Neither Weak Nor Narrow: Isn't It Time To Change the AI Nomenclature?", en *Analytics India Magazine,* 2017. [en línea] <https://analyticsindiamag.com/applied-ai-neither-weak-narrow-isnt-time-change-ai-nomenclature/> [Fecha de consulta: 07 de septiembre de 2023].

COPELAND, B.J., "Artificial Intelligence", en *Encyclopedia Britannica*, 11 de agosto de 2020. [en línea] <https://www.britannica.com/technology/artificial-intelligence> [fecha de consulta: 06 de septiembre de 2023].

CUÉTICA, Quiénes somos, [s.f., en línea], <https://cuetica.unam.mx/quienes-somos> [consulta: 17 de marzo de 2024].

GONZÁLEZ, I., "GPT3: ¿Que 'es y por qué todo el mundo habla de ello"?, en *IEBS Business School*, 14 de diciembre de 2022, [en línea], <https://www.iebschool.com/blog/que-es-chat-gpt3-openai-tecnologia/>, [consulta: 10 de septiembre de 2023].

HAKSAR, V., "Moral agents", en *Routledge Encyclopedia of Philosophy*, [en línea], <https://www.rep.routledge.com/articles/thematic/moral-agents/v-

1/sections/understanding#>, [consulta: 11 de septiembre de 2023].

HERNÁNDEZ, M., “Inaugura el IIMAS Laboratorio de Inteligencia Artificial y Alta Tecnología”, en *Gaceta UNAM*, 4 de agosto de 2022. [en línea]< https://www.gaceta.unam.mx/inaugura-el-iimas-laboratorio-de-inteligencia-artificial-y-alta-tecnologia/>, [consulta: 03 de octubre de 2023].

INSTITUTO DE INVESTIGACIONES JURÍDICAS, *Informe de Gestión 2022-2023*, UNAM, [en línea] <https://www.juridicas.unam.mx/informe-2022-2023/detalle/47>, [consulta:03 de octubre de 2023].

KNIGHT, W., “La inminente y poderosa simbiosis de los ordenadores cuánticos y la IA”, en *MIT Technology Review*, 04 de abril de 2019. [en línea]<https://www.technologyreview.es/s/11028/la-inminente-y-poderosa-simbiosis-de-los-ordenadores-cuanticos-y-la-ia> [fecha de consulta: 08 de septiembre de 2023].

LA VANGUARDIA., “*El primer cyborg del mundo defiende ‘el derecho a diseñarse como especie’*”, [4 de octubre de 2019, en línea], < https://www.lavanguardia.com/vida/20191004/47798233444/el-primer-ciborg-del-mundo-defiende-el-derecho-a-disenarse-como-especie.html> [consulta: 2 de octubre de 2023].

LÓPEZ, P., “Centurión, robot creado por empresa incubada en la UNAM”, en *Gaceta UNAM*, 23 de enero de 2020. [en línea]<https://www.gaceta.unam.mx/empresa-incubada-en-la-unam-crea-robot-capacitado-para-ser-vendedor-estrella/>, [consulta: 03 de octubre de 2023]

MARWICK, A., Donglegate: Why the Tech Community Hates Feminists, en *Wired,* 3 de abril de 2013, [en línea], <https://www.wired.com/2013/03/richards-affair-and-misogyny-in-tech/>, [consulta:18 de septiembre de 2023].

NATIONAL INSTITUTE OF BIOMEDICAL IMAGING AND BIOENGINEERING, *Modelado Computacional*, s.f. [en línea], <https://www.nibib.nih.gov/espanol/temas-cientificos/modelado-computacional#pid-2186>, [consulta: 12 de septiembre de 2023].

OBSERVATORIO DE INTELIGENCIA ARTIFICIAL, *Google, Facebook, Amazon,*

IBM y Microsoft forman una alianza en materia de Inteligencia Artificial, 30 de septiembre de 2016, [en línea], <https://observatorio-ia.com/alianza-en-materia-de-inteligencia-artificial>, [consulta: 02 de octubre de 2023].

OXFORD, *Diccionario Lexico*, voz: acción. [en línea], <https://web.archive.org/web/20220823141312/https://www.lexico.com/es/definicion/accion>, [consulta: 12 de septiembre de 2023]

PARLAMENTO EUROPEO, *Questions parlamentaires, Question avec demande de réponse écrite E-011289-13 à la Commission, Article 117 du reglament, Marc Tarabella* (S&D), [3 de octubre de 2013, en línea], <https://www.europarl.europa.eu/doceo/document/E-7-2013-011289_FR.html> [consulta: 2 de octubre de 2023].

PASCUAL, M., "Expertos en inteligencia artificial reclaman frenar seis meses la 'carrera sin control' de los ChatGPT", en *El País*, 29 de marzo de 2023, [en línea], <https://elpais.com/tecnologia/2023-03-29/expertos-en-inteligencia-artificial-reclaman-frenar-seis-meses-la-carrera-sin-control-de-los-chatgpt.html>, [consulta: 05 de octubre de 2023].

PASTOR, J., "Elon Musk: o los humanos se fusionan con las máquinas, o la inteligencia artificial nos hará irrelevantes", en *Xataka*, [23 de abril de 2019, en línea], <https://www.xataka.com/robotica-e-ia/elon-musk-o-los-humanos-se-fusionan-con-las-maquinas-o-la-inteligencia-artificial-nos-hara-irrelevantes> [consulta: 12 de septiembre de 2023].

PÉREZ COLOMÉ, J., "'Funciona muy bien, pero no es magia': así es ChatGPT, la nueva inteligencia artificial que supera límites", en *El País*, 07 de diciembre de 2022, [en línea], <https://elpais.com/tecnologia/2022-12-07/funciona-muy-bien-pero-no-es-magia-asi-es-chatgpt-la-nueva-inteligencia-artificial-que-supera-limites.html>, [consulta: 10 de septiembre de 2023].

RED HAT, *¿Qué es el Internet de las Cosas (IoT)?*, 8 de enero de 2019, [en línea], <https://www.redhat.com/es/topics/internet-of-things/what-is-iot>, [consulta: 18 de septiembre de 2023].

RESPUESTAS AQUÍ, *¿Sabía el público en general que RoboCop era un Cyborg?*, [s.f., en línea], <https://respuestas.me/q/sabia-el-publico-en-general-que-

robo-cop-era-un-cyborg-61527303970> [consulta: 01 de octubre de 2023].

REUTERS*, Primero médicos y ancianos, México se prepara para aplicar vacuna contra COVID-19,* 8 de diciembre de 2020*,* [en línea], <https://www.reuters.com/article/salud-coronavirus-mexico-vacunas-idLTAKBN28I1YG>, [consulta: 03 de octubre de 2023].

SAAVEDRA, D., "Pumas crean algoritmo y ganan premio en Francia", en *Gaceta UNAM*, 13 de enero de 2020. [en línea]<https://www.gaceta.unam.mx/pumas-crean-algoritmo-y-ganan-premio-en-francia/>,[consulta: 03 de octubre de 2023].

SCALITER, J., "Sophia, el primer robot con ciudadanía", en *La Razón*, [29 de octubre de 2017, en línea], <https://www.larazon.es/tecnologia/sophia-el-primer-robot-con-ciudadania-OE16746549/> [consulta: 2 de octubre de 2023].

SOCIETY OF AUTOMOTIVE ENGINEERS, *SEA Levels of Driving Automation, Refined for Clarity and International Audience*, 21 de mayo de 2021, [en línea], <https://www.sae.org/blog/sae-j3016-update>, [consulta: 10 de septiembre de 2023].

STOP KILLER ROBOTS, Less autonomy, more humanity, 2021. [en línea], <https://www.stopkillerrobots.org/>, [consulta: 10 de septiembre de 2023].

THE TEHCNOLAWGIST, *GPT-3 en el sector legal: ¿asistente o enemigo?*, 8 de febrero de 2023, [en línea], <https://www.thetechnolawgist.com/2023/02/08/gpt-3-en-el-sector-legal-asistente-o-enemigo/>, [consulta: 10 de septiembre de 2023].

UNIVERSITY OF HELSINKI, Elements of AI, Part 1: *Introduction to AI,* 2019. [en línea] <https://www.elementsofai.com/> [fecha de consulta: 06 de septiembre de 2023].