https://doi.org/10.24215/25916386e062

Algor-ética: la ética en la inteligencia artificial

POR **JORGE F. VILLALBA**(*)

Sumario: I. Introducción.- II. Definición de inteligencia artificial.- III. La ética en la inteligencia artificial.- IV. Ética y diversas aplicaciones en la IA.- V. Principios éticos para el diseño y desarrollo de la inteligencia artificial.- VI. IA y política.- VII. Interrogantes.- VIII. Conclusiones.- IV. Propuestas.- X. Bibliografía.

Resumen: mi actual tránsito por un doctorado en Educación, con el tema "*Inteligencia artificial aplicada a la sistematización de experiencias pedagógicas para el fortalecimiento del trabajo colectivo*", me lleva al abordaje de la cuestión para esta ocasión, pero desde otros enfoques disciplinarios, como la visión de la filosofía del derecho, la ética y la política. Si bien Argentina se encuentra aún en lo que podríamos llamar una etapa evangelizadora de estos temas, ya abundan expresiones del pensamiento de personas con prestigio intelectual que refieren a la inteligencia artificial (IA o AI) como "algo" que ya está entre nosotros, generando una gran revolución tecnológica con manifestaciones diversas, tales como la supresión de ciertas profesiones tal y como las conocemos hoy en día; cambio de métodos de trabajo, de prácticas forenses y demás praxis. Y es ese "algo" el que tenemos que intentar definir, delimitar y sobre todo medir cuantitativamente en términos de moralidad (1).

(*) Abogado y Notario, Facultad de Derecho y Ciencias Sociales, Universidad Católica de Córdoba. Doctorando en Educación, Universidad Católica de Córdoba (UCC). Escribano Público. Prof. becado Programa Intensivo de Nivel Doctoral sobre Altos Estudios de Derecho Histórico, Dogmático y Comparado, Pontificia Universidad Católica de Chile. Prof. Derecho Romano, Derecho Privado VI, Derecho Notarial I y Derecho Notarial II, Universidad Católica de Córdoba. Prof. de la Maestría en Derecho Administrativo, Universidad Libre Seccional Socorro de Colombia.

(1) Recuperado de https://dpicuantico.com/2019/02/04/inteligencia-artificial-en-la-corte-constitucional-colombiana-otra-experiencia-prometea/

Palabras claves: inteligencia artificial - AI - IA - algor-ética - blockchain

Algor-ethics: ethics in artificial intelligence

Abstract: *my current transition to a PhD in Education, with the theme: "Artificial intelligence applied to the systematization of pedagogical experiences for the strengthening of collective work"; leads me to approach the issue for this occasion, but from other disciplinary approaches, such as the vision of the Philosophy of Law, Ethics and Politics. Even though Argentina is still in what we could call an evangelizing stage of these topics, there are already many expressions of the thought of people with intellectual prestige that refer to Artificial Intelligence (AI) as "something" that is already among us, generating a great technological revolution with diverse manifestations such as the suppression of certain professions as we know them today; change of work methods; of forensic practices and other practices. And it is this "something" that we must try to define, delimit and above all measure quantitatively in terms of morality.*

Keywords: *artificial intelligence - AI - IA - algor-ethics - blockchain*

I. Introducción

> *Que la palabra y la tradición de la fe nos ayuden a interpretar los fenómenos de nuestro mundo. La algor-ética puede ser un puente para inscribir los principios de la Doctrina Social de la Iglesia en las tecnologías digitales.*
> Papa Francisco (Mutual, 2020)

Recientemente, la Organización para la Cooperación y el Desarrollo Económicos (en adelante OCDE), como la Unión Europea (en adelante UE), publicaron sendos documentos en los que establecen las reglas del "deber ser" de la inteligencia artificial (IA). En el caso de la OCDE, dictaminó los denominados "Principios sobre la Inteligencia Artificial", mientras que la UE se refirió a los siete principios éticos que deben regir la IA.

Debemos tener presente en este punto que Colombia adhiere al acuerdo sobre inteligencia artificial ante los países de la OCDE (2).

La OCDE sostiene que deben ser los gobiernos los que garanticen que el diseño de los sistemas de IA respete los valores y leyes imperantes. El desafío consiste en que las personas puedan confiar que su seguridad y privacidad se respetarán,

(2) Recuperado de https://www.mintic.gov.co/portal/inicio/Sala-de-Prensa/Noticias/100683:Colombia-se-adhiere-a-acuerdo-sobre-Inteligencia-Artificial-ante-los-paises-de-la-OCDE

minimizar la sensación de vulnerabilidad, considerando estas exigencias como algo prioritario en toda manipulación de sistemas de inteligencia artificial.

En el caso de Europa hay una marcada y clara estrategia de que el ser humano sea el centro de la cuestión. Esto quiere decir que las aplicaciones de IA no solo deben ajustarse a la ley, sino que también deben respetar los **principios éticos** y garantizar que su implementación evite daños involuntarios o colaterales.

Actualmente encontramos iniciativas en Europa, como el caso de España, donde observamos la creación de OdiseIA (3), el Observatorio del Impacto Social y Ético de la Inteligencia Artificial, el cual está conformado o integrado por empresas, universidades, instituciones y personas físicas, entre otras, con firme preocupación e interés por la ética de la IA.

Este espacio pretende servir de punto de discusión y análisis sobre dicha materia, garantizando, a la vez, un diálogo de carácter independiente y multidisciplinar.

Otra de las iniciativas es la denominada "*We The Humans* (4)", que nace por el interés en una inteligencia artificial ética, por pedido de una serie de profesionales de diversos ámbitos, que quieren cambiar la visión tecnológica por no aceptar un determinismo en el que la tecnología "es" de una cierta forma, poniendo el acento en que la potencialidad tiene una intencionalidad que debe ser dirigida hasta la humanización de la tecnología y no a la inversa.

II. Definición de inteligencia artificial

Una aproximación teórica de la cuestión nos permite construir una definición, como así también delimitar el ámbito de competencia de esta tecnología.

Es importante aclarar que no existe una definición única y consensuada de inteligencia artificial.

Cada definición existente parece asumir unos objetivos distintos para la IA o parte de concepciones distintas sobre la categoría "inteligencia". Cuando uso aquí el término categoría, lo hago en lenguaje educativo y no aristotélico, es decir, no me refiero a una categoría aristotélica sino a categoría como formulación de un concepto que requiere de una construcción social, que nace de dicha construcción y que debe tener una potencialidad humanizadora a los fines de permitir una utilización direccionada a la construcción de formas sociales justas, manteniendo

(3) Presentación del Observatorio del Impacto Social y Ético de la Inteligencia Artificial (OdiseIA) (19 de octubre de 2019). Recuperado de https://www.youtube.com/watch?v=y9IzF96125s

(4) Recuperado de http://wethehumansthinktank.com/sobre-nosotros/

la figura del ser humano en el centro de la cuestión en todo momento; porque es por y para el hombre que el derecho existe y que la tecnología se pone al servicio.

Intentar una definición *a priori* de inteligencia artificial nos lleva a postular la idea de combinación de algoritmos, con la intencionalidad de que permitan la creación de máquinas que puedan imitar, analogar o superar las propias capacidades del ser humano. Advierto la palabra "algoritmo", tal como proponemos en el título de este ensayo, en combinación con la palabra ética al hablar de la algor-ética, categoría o construcción terminológica empleada por su Santidad Papa Francisco al abordar la cuestión.

También podríamos hablar de tecnología que busca, por un lado, descomponer la inteligencia humana en sus procesos más simples y elementales para poder, más tarde, formalizarlos (expresarlos en el lenguaje formal de la lógica), recomponerlos en forma de algoritmos y estrategias de programación, y reproducir así la inteligencia humana en máquinas y programas.

La IA tiene la potencialidad de describir la inteligencia humana en sus elementos más simples para poder así hacerla computable y transferible mediante algoritmos y lenguajes de programación.

Muchos autores han tratado de aportar una definición más concisa, tratando de establecer una delimitación del objetivo principal de esta tecnología, refiriendo que la misma persigue imitar la forma de actuar o pensar de los seres humanos o construir sistemas que razonen o que actúen como ellos, pero la imitación tiene como intencionalidad el incremento de la potencialidad humana.

Russell y Norvig clasifican las distintas definiciones en cuatro categorías: sistemas que actúan como humanos; sistemas que piensan como humanos; sistemas que piensan racionalmente; sistemas que actúan racionalmente.

En definitiva, la IA se estructura de un conocimiento y no de una sabiduría, donde esta última es una virtud que le corresponde en forma exclusiva al ser humano. Un conocimiento elevado al punto de la categoría sabiduría implica experiencia, cultivar la prudencia, poder aplicar la equidad, tener sensibilidad por lo cotidiano; aspectos y virtudes que solo el hombre —por el solo hecho de ser hombre— puede desarrollar.

III. La ética en la inteligencia artificial

En el año 1985, Judith Jarvis Johnson puso de manifiesto en la revista *The Yale Law Journal* algunas preguntas que llevaban rondando desde principios del

siglo XX en el conocimiento humano (5). Este reconocido jurista expuso la complejidad que supondría para una inteligencia artificial encontrar la solución de un dilema humano complejo (6). Imaginemos que una persona conduce un coche sin frenos; ante él, se bifurca el camino en dos: en el camino de la izquierda hay cinco personas, en el de la derecha solo hay una. ¿Es lícito matar a una persona para salvar a otras cinco?

Un ser humano tendría que despejar las incógnitas de la ecuación atendiendo criterios éticos. Resulta sumamente difícil que una máquina pueda calibrar este tipo de decisiones, por eso es tan importante que la inteligencia artificial pueda estar dotada de una ética o código de valores que condicione, a nuestra imagen y semejanza, la esencia de sus actos.

El filósofo Nick Bostrom (7) sostiene que una inteligencia artificial avanzada podría tener la capacidad de provocar la extinción humana, ya que sus planes pueden no implicar tendencias motivacionales humanas. No obstante, Bostrom (2003) también plantea la situación contraria, en la que una *super inteligencia* podría ayudarnos a resolver problemas tediosos y constantes en la humanidad, tales como la pobreza, la enfermedad o la destrucción del planeta.

Es importante la paradoja de Bostrom (2003) que habla del "efecto perverso". Para explicar dicho efecto utiliza el mito del rey Midas al desear convertir todo lo que toque en oro. Toca un árbol y se convierte en oro, incluso su propia comida. Por ello, es importante que la IA pueda extrapolar las intenciones que tenemos, sin caer en el efecto perverso, más allá del efecto anti-ético.

Debemos ser capaces de saber cómo controlar la Inteligencia Artificial
(Bostrom, 2005)

Al incorporar la inteligencia artificial a los procesos de las organizaciones es importante dotar a esta nueva tecnología de valores y principios. Y, dentro de las organizaciones, son los desarrolladores de esta tecnología las personas que realmente han de trabajar siendo conscientes de las implicaciones morales y éticas que conlleva su trabajo.

"Cada aspecto de nuestras vidas será transformado [por AI]", y podría ser "el evento más grande en la historia de nuestra civilización" (Hawking, 2018).

---

(5) Recuperado de https://retina.elpais.com/retina/2019/02/25/tendencias/1551089772_654032.html

(6) Would you sacrifice one person to save five? Eleanor Nelsen. (12 de enero de 2019). Recuperado de https://www.youtube.com/watch?time_continue=3&v=yg16u_bzjPE&feature=emb_logo

(7) Nick Bostrom, el filósofo que advierte de los riesgos de la superinteligencia artificial (20 de abril de 2017). Recuperado de https://www.youtube.com/watch?v=5zzdVHfbj6w

La problemática planteada se está materializando en la *Asociación sobre Inteligencia Artificial*, creada por Elon Musk y Sam Altman, en la que se invita a los principales líderes tecnológicos para poder identificar dilemas éticos y prejuicios. Su objetivo primordial es establecer unas reglas de juego, basadas en un marco de comportamiento moral, donde la inteligencia artificial pueda desarrollarse en representación de la humanidad.

En tiempos actuales, muchas compañías están comenzando a formar a las personas (quienes a su vez están formando a las máquinas) en prácticas éticas aplicadas a los algoritmos que rigen la inteligencia artificial.

La gestión del aprendizaje de la inteligencia artificial es lo que actualmente conocemos como *machine learning* (8), el cual debe contener principios universales de respeto, libertad e igualdad.

La preocupación ética por la creación de nuevos tipos de inteligencia requiere de un exquisito criterio moral de las personas que diseñan estas nuevas formas de tecnología. El algoritmo ha de ser capaz de discernir y reconocer fallos cuando se centran en acciones sociales con repercusión, que antes realizaba un ser humano.

El Algoritmo no puede dañar a personas o empresas.
Wundersee, 2019

Fruto de esta preocupación, el Parlamento Europeo realizó un informe sobre robótica en 2017 llamado *Código Ético de Conducta*, publicando en 2018 el primer borrador de la *Guía Ética para el uso responsable de la Inteligencia Artificial*.

Son estándares morales dirigidos a humanos, a los creadores de tecnología, cuyos principios rectores son los siguientes: asegurar que la IA está centrada en el ser humano; prestar atención a los grupos vulnerables, como los menores de edad o las personas con discapacidades; respetar los derechos fundamentales y la regulación aplicable; ser técnicamente robusta y fiable; funcionar con transparencia y no restringir la libertad humana.

Las grandes organizaciones, empresas y gobiernos, están centrándose en los problemas que pueden surgir en el tema de la ética de la inteligencia artificial para trazar consideraciones, prácticas y marcos comunes de cara al futuro (9).

(8) Recuperado de https://www.bbva.com/es/machine-learning-que-es-y-como-funciona/

(9) Límites éticos para la inteligencia artificial. DW Documental (14 de agosto de 2019). Recuperado de https://www.youtube.com/watch?v=sHVwwriaT6k

Es importante alcanzar un acuerdo donde se pueda conceptualizar y, sobre todo, regular las prácticas derivadas. Al fin y al cabo, la tecnología es un paso más de nuestra evolución (10).

En el intento de que la IA logre replicar la inteligencia humana tenemos un debate abierto entre los expertos en IA, científicos y filósofos, sobre la existencia de múltiples definiciones de inteligencia, dando lugar a diversas maneras de concebir la naturaleza y, sobre todo, el propósito de la IA.

La relevancia ética del desarrollo de la IA no permanece ajena a este debate. El campo de la ética se extiende allá donde encuentra agentes dotados de autonomía e inteligencia, es decir, sujetos capaces de tomar decisiones y actuar de forma racional. En este sentido, el término "autonomía" suele emplearse en este contexto para designar la capacidad de escoger un curso de acción de forma libre, capacidad que tradicionalmente se ha identificado como un rasgo distintivo y exclusivo de los seres humanos (*European Group on Ethics in Science and New Technologies,* 2018).

Esta autonomía no ha de ser confundida con la capacidad que presentan otros seres vivos de dirigir sus actos.

La autonomía humana viene siempre ligada a la responsabilidad: las personas, al ser capaces de dar cuenta de sus propios motivos y razones, responden también por ello ante los actos y decisiones tomadas.

A la vista de estas consideraciones, emplear el término "autónomo" para referirse a los dispositivos provistos de IA puede llevar a confusión. Se suele afirmar que muchas de las aplicaciones, sistemas y máquinas dotados de IA son capaces de operar de forma "autónoma", es decir, de realizar operaciones y procesos por sí mismos. Sin embargo, puesto que hasta la fecha ningún sistema o artefacto inteligente es capaz de dar cuenta de sus propios actos y decisiones de la manera en la que las personas son capaces, resulta erróneo calificar a estos dispositivos de autónomos en el sentido ético de la palabra.

Es cierto que en el ámbito tecnológico ha proliferado el uso de dicho término para referirse simplemente a aplicaciones capaces de operar sin supervisión humana, pero, desde el punto de vista del análisis ético, el término adecuado para referirse a estos dispositivos sería "automático", no "autónomo".

(10) Recuperado de https://www.lanacion.com.ar/opinion/columnistas/inquietante-vinculo-entre-el-ganchao-y-el-temor-a-un-ataque-cuantico-nid2337839

Con todo, la existencia de aplicaciones capaces de tomar distintos cursos de acción por sí mismas no deja de plantear algunas dudas sobre la creencia de que solo los seres humanos son capaces de actuar de forma autónoma. Aclarar este punto es de una importancia capital, pues la relevancia ética de los sistemas provistos de IA será una u otra en función de quién sea verdaderamente responsable de los razonamientos y las decisiones de estos dispositivos: las personas encargadas de su diseño o los propios dispositivos.

En esta línea, algunos autores proponen distinguir entre una IA débil y una fuerte (11) (López de Mántaras, 2015) (12). Esta distinción entre IA débil e IA fuerte fue introducida inicialmente por el filósofo John Searle en el año 1980.

Según esta visión, la IA débil sería la ciencia que permitiría diseñar y programar ordenadores capaces de realizar tareas de forma inteligente, mientras que la IA fuerte sería la ciencia que permitiría replicar en máquinas la inteligencia humana, tiene conciencia, empatía, etc., siendo incluso capaces de comprender perfectamente cada uno de sus actos y sus consecuencias.

En otras palabras, la IA débil permitiría desarrollar sistemas con inteligencia especializada, ordenadores que juegan al ajedrez, diagnostican enfermedades o resuelven teoremas matemáticos; mientras que la IA fuerte permitiría desarrollar ordenadores (quasi-humanos) y máquinas dotados de inteligencia de tipo general. La presencia de inteligencia general —la que presentan los seres humanos— sería la condición suficiente para inferir que un dispositivo puede actuar de forma autónoma, no solo automática.

La relevancia ética de los dispositivos y sistemas dotados de una IA débil sería fácil de determinar, en virtud de que estrictamente no estamos hablando de una simetría con la conciencia humana. Estos dispositivos pueden operar de forma automática o autómata, pero no poseen inteligencia de tipo general y, por tanto, carecen de autonomía en sentido estricto. Se trata de aplicaciones programadas de antemano para realizar una única tarea, y en muchos casos han demostrado superar con creces la pericia humana, sobre todo para tareas de tipo burocráticas o mecánicas que consumen gran cantidad de horas de tiempo y mucho recurso humano. Pero se trata de sistemas incapaces de actuar de forma racional, por lo que su actividad no es éticamente imputable o reprochable según el caso.

(11) Qué es la Inteligencia Artificial Avanzada (Inteligencia Artificial Fuerte) (8 de enero de 2018). Recuperado de https://www.youtube.com/watch?v=JFpDp7H1mNM

(12) Entrevista a Ramón López de Mántaras, *OpenMind* (13 de junio de 2019). Recuperado de https://www.youtube.com/watch?v=qo_78YIBF7I

En los sistemas de este tipo, la responsabilidad ética recae plenamente sobre las personas encargadas de su diseño y funcionamiento. Puesto que estas aplicaciones solo tienen la posibilidad de operar de acuerdo a un único curso de acción, la reflexión ética en torno a este tipo de tecnología ha de centrarse en la programación y el diseño de estos sistemas.

La consideración ética de las aplicaciones dotadas de una IA fuerte resulta algo más compleja, en primer lugar, porque no resulta claro que sea posible crear máquinas con inteligencia de tipo general. Hoy sabemos que la complejidad del cerebro humano es prácticamente imposible de replicar (López de Mántaras, 2015) y que los intentos de desarrollar máquinas capaces de mostrar un comportamiento inteligente en varios ámbitos no han tenido mucho éxito. Pero hablamos de una asimilación plena, no obstante ello, muchas manifestaciones de utilización de inteligencia artificial actualmente presentan problemáticas éticas dignas de consideración, pese a no referirnos estrictamente a un sistema de inteligencia artificial fuerte.

Un ordenador capaz de vencer en el juego de ajedrez a cualquier ser humano es incapaz de aplicar ese conocimiento a un juego similar, como las damas, cuando cualquier persona podría inferir rápidamente las reglas y comenzar a jugar.

La práctica imposibilidad de crear un sistema dotado de inteligencia y sentido común hace igualmente dudosa la propuesta de la singularidad tecnológica (inteligencias artificiales dotadas de inteligencia general capaces de automejorarse progresivamente hasta superar con creces la inteligencia humana) y otorga un carácter distópico a las reflexiones en torno a los potenciales riesgos derivados de esta tecnología.

A pesar de que una IA fuerte parezca impracticable actualmente, es cierto que ya existen aplicaciones dotadas de IA con capacidad de tomar decisiones y escoger entre diversos cursos de acción, tal como dije anteriormente. Ilustramos este postulado con hechos vinculados a los vehículos automáticos o robots domésticos.

Aun así, es preciso afirmar que incluso en estos sistemas la capacidad de toma de decisiones es solo aparente: están programados para realizar un cálculo de probabilidades teniendo en cuenta una diversidad de factores y circunstancias, con alto contenido experimental producto de la carga de casos o minería de datos para evaluaciones de tipo estadísticos siguiendo patrones de repetición de decisiones ante determinados eventos; es decir, que el sistema otorga una respuesta correcta a priori según un escenario determinado y un conjunto de datos limitados, seleccionando el curso de acción óptimo.

Pero pese a encontrarnos en la esfera de la toma de decisiones en función de un cálculo probabilístico, se puede observar que aun así entran en consideración factores de relevancia ética.

En los sistemas de este tipo la responsabilidad ética sigue recayendo en su totalidad sobre las personas encargadas de su diseño y funcionamiento, pero, al ser mayor el grado de automatización, la reflexión ética ha de tener en cuenta principios de diseño más específicos, como la trazabilidad de los algoritmos, la explicabilidad, la rendición de cuentas o la supervisión humana. En definitiva, hablamos de resultados productos de ciertas tomas de decisiones con implicancias y riesgos éticos.

Ante el potencial que la IA exhibe, urge establecer unos principios y proponer una serie de pautas para que esta tecnología se emplee y siga desarrollándose de manera responsable y segura. La existencia de unos principios y estándares correctamente delimitados puede contribuir sobremanera a potenciar los beneficios derivados de la IA y a minimizar los riesgos que su uso implica.

IV. Ética y diversas aplicaciones de la IA

Las aplicaciones de la IA son de diverso tipo y su uso varía enormemente. Esto hace que sea prácticamente imposible establecer un único procedimiento o código de conducta específico para la IA en general.

Son muchos los organismos e instituciones que ya se han apresurado a elaborar una lista de principios y recomendaciones para asegurar un uso adecuado de la IA.

Entre los documentos disponibles merece la pena destacar los publicados por distintos órganos de la Unión Europea, que abordan también la dimensión regulatoria de los robots y sistemas autónomos; las recomendaciones del Foro Económico Mundial, en las que se proponen distintos métodos para asegurar un buen uso de la IA por parte de empresas y Gobiernos; y los informes publicados por el UNICRI, *Centre for AI and Robotics* de las Naciones Unidas y por la Unesco.

En los últimos años, también han visto la luz diversas iniciativas de la mano de empresas, asociaciones profesionales o académicos que han contribuido a formular criterios de buenas prácticas para el uso de la IA. Entre ellas destacan los estándares formulados por la IEEE (*Institute of Electrical and Electronics Engineers*); los principios Asilomar sobre IA propuestos por el *Future of Life Institute*; o la declaración para el desarrollo responsable de la IA, elaborada por el *Forum on the Socially Responsible Development of Artificial Intelligence*, celebrado en la Université de Montréal en el 2017.

En el ámbito español merece la pena destacar la Declaración de Barcelona sobre la inteligencia artificial, firmada en el 2017 por distintos expertos en IA y promovida por la obra social "la Caixa".

V. Principios éticos para el diseño y desarrollo de la inteligencia artificial

Un primer punto de partida para garantizar una IA segura y robusta es proponer una serie de principios éticos que abarquen las distintas aplicaciones y sistemas provistos de IA.

Estos principios no poseen una formulación tan específica como cabría esperar de un código de conducta o de una serie de normas. Se trata, más bien, de una serie de imperativos en los que se presenta un bien humano que debe ser respetado en toda situación y promovido mediante nuestras acciones.

En el ámbito de la ética y la moralidad es frecuente encontrar principios formulados de esta manera: por ejemplo, "no matarás"; o el conocido imperativo categórico de Immanuel Kant: "Obra de tal modo que uses a la humanidad, tanto en tu persona como en la persona de cualquier otro, siempre al mismo tiempo como fin y nunca simplemente como medio" (Kant, 1785). De manera similar, el campo de la robótica y de la IA es susceptible también de albergar formulaciones semejantes. Muchos de los principios que es posible encontrar en los documentos listados se inspiran previamente en las leyes de la robótica elaboradas por Isaac Asimov (1942):

1) Un robot no puede hacer daño a un ser humano o, por inacción, permitir que un ser humano sufra.

2) Un robot debe obedecer las órdenes dadas por los seres humanos, excepto si estas órdenes entrasen en conflicto con la primera ley.

3) Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o la segunda ley.

En vista de los riesgos que los sistemas y dispositivos inteligentes poseen, se hace necesario identificar qué bienes humanos entran en peligro en este escenario y formular, como consecuencia, una serie de principios que orienten el uso de la IA hacia su defensa y promoción.

Tras haber repasado en el apartado anterior algunos de estos riesgos, y tomando en consideración los principios propuestos por los organismos citados anteriormente, proponemos el siguiente listado:

a) Autonomía humana: los sistemas inteligentes deben respetar en todo momento la autonomía y los derechos fundamentales de las personas.

b) Transparencia: dado que el diseño de estos dispositivos contempla que tomen decisiones automáticamente con base en distintos cálculos y proyecciones, debe ser posible en todo momento trazar el razonamiento seguido por el sistema y explicar las consecuencias alcanzadas. El diseño y el empleo de una tecnología impredecible son incompatibles con la defensa de la autonomía humana. Es absolutamente necesario que la actividad de todos estos dispositivos sea de fácil comprensión y acceso.

c) Responsabilidad: asignación de responsabilidades ante los posibles daños y perjuicios que estos puedan ocasionar los sistemas de IA.

d) Integridad y seguridad: exigencia de algoritmos seguros, fiables y sólidos para resolver errores o incoherencias durante todas las fases del ciclo de vida útil de los dispositivos.

e) Justicia distributiva: garantizar un empleo justo de los datos disponibles para evitar posibles discriminaciones hacia determinados grupos o distorsiones en los precios y en el equilibrio de mercado.

Los principios antes establecidos llevan a considerar los mecanismos *value-by-design* (*security-by-design*, *privacy-by-design*, etc.) métodos que permitan materializar estos principios éticos en el diseño y la programación específica de los algoritmos.

Junto con la posibilidad de introducir parámetros éticos durante esta fase, el diseño de estos sistemas es de una importancia crucial debido a la propia naturaleza de la IA. Ha quedado mencionado cómo los sistemas dotados de IA son "autónomos" en un sentido restringido de la palabra: el razonamiento y el proceso de toma de decisiones que llevan a cabo responden siempre a una programación previa que, por muy amplia y compleja que pueda ser, queda también siempre dentro de los límites establecidos por el diseño; de ahí que la responsabilidad ética ligada a los diversos usos de la IA recaiga directamente sobre las personas, no sobre las máquinas.

Por esto mismo, las consideraciones éticas y sociales deben comparecer desde la primera fase de diseño de la tecnología (Buchholz y Rosenthal, 2002). Incorporar criterios éticos durante la fase de diseño reviste una gran importancia, pero no es suficiente para garantizar que las aplicaciones provistas de IA sean seguras y se empleen de manera responsable.

En resumidas cuentas, la IA no es un simple producto del que deban ocuparse los diseñadores: es un producto social en el que se ven involucrados distintos grupos de interés y en el que entran en juego fuerzas de origen social, político y económico.

En definitiva, las consideraciones éticas durante la fase de diseño de aplicaciones deben estar acompañadas también de una serie de mecanismos concretos que permitan a los distintos grupos de interés (empresas, investigadores, reguladores, gobiernos, consumidores, etc.) incorporar principios y normas al funcionamiento de estas nuevas

VI. IA y política

Carme Artigas dirige la secretaría de Estado de España. Es la primera que incluye en su título el término inteligencia artificial (IA), materia en la que es una reconocida experta. En un momento de recelo hacia esta tecnología, por cuestiones de seguridad y privacidad, destaca el poder que la IA puede otorgar a los ciudadanos para controlar a gobiernos y grandes empresas, y aboga por liderar en Europa el debate del humanismo tecnológico (13).

El uso de IA permite hacer mucho más eficiente la administración digital, teniendo como prioridad del gobierno español liderar el debate del humanismo tecnológico en Europa.

Para ello, hay una marcada necesidad de definir marcos normativos basados en la ética y en las recomendaciones. La regulación vendrá, pero cuando el mercado esté un poco más maduro.

Es fundamental que las empresas tecnológicas sean conscientes que si su desarrollo no tiene encaje en la estructura de valores de la ciudadanía, al final, esa industria no va a prosperar. La falta de presión del gobierno estimula a que las compañías tecnológicas desarrollen algoritmos éticos por diseño.

Tener un cierto nivel de confiabilidad sobre la falta de sesgos en los sistemas de IA; saber que se han entrenado bien los modelos, que se han aplicado de forma correcta, que haya transparencia con los algoritmos y que se expliquen bien su funcionamiento.

(13) Recuperado de https://elpais.com/tecnologia/2020/02/21/actualidad/1582290641_645896.html

La interrelación de los sistemas de IA con la política requiere encontrar consensos, porque todo el mundo entiende perfectamente lo que hace falta para avanzar en la transformación digital.

A nivel Estado es fundamental formular una estrategia de IA, aprender de los mejores en aquellos ámbitos en los que son buenos, sin perder de vista el hecho de tener, como país, una posición diferencial.

A nivel ciudadanía la discusión de la IA nos permite volvernos a hacer preguntas como cuáles son los valores que marcan nuestra sociedad, cuáles son los límites de la libertad de expresión o cuáles son los derechos que queremos preservar.

Queda claro que el desarrollo de la IA va a ser algo uniforme en todo el mundo. La IA no es una tecnología sino una infraestructura económica y social, y por tanto cada país va a decidir su modelo de adopción, con un gran potencial ofrecido a los ciudadanos para monitorizar a sus gobiernos y a las grandes corporaciones, exigiéndoles mayor transparencia.

La inteligencia artificial atraviesa varias áreas del conocimiento y una de ellas es la educación. Muchos países están haciendo un análisis conjunto con los ministerios de educación, ciencia, industria, y demás, para detectar cuáles van a ser las habilidades que se van a demandar en el futuro y cuáles debe tener la ciudadanía para elevar sus competencias. Creemos que podremos lanzar ese plan a finales de este año.

VII. Interrogantes

Luego de todo lo expuesto en el presente ensayo, estamos en condiciones de formular algunos interrogantes a la luz de ciertas disciplinas y enfoques.

Si desde la filosofía del derecho sostenemos la importancia de la prudencia dentro de los silogismos prácticos a la hora de resolver el reparto de derechos, la aplicación de justicia, el dar a cada uno lo suyo, y sabiendo que la prudencia es una virtud que le pertenece con exclusividad al ser humano:

· *¿Qué sucede con la Prudencia, la recta ratio agibilium en los mecanismos de resolución de conflictos por sistema de inteligencia artificial?*

Siguiendo el pensamiento de Santo Tomás de Aquino (14) y Aristóteles, es la prudencia la virtud que permite al ser humano aplicar los principios morales en los casos particulares, sin incurrir en error.

(14) Recuperado de https://www.dominicos.org/estudio/recurso/suma-teologica/

Decimos entonces que:

· *¿Iniciamos un tránsito en miras de la supresión de la prudencia a cambio del análisis estadístico fruto de la valoración empírica de casos cargados en la machine learning para la toma de decisiones o resoluciones de conflictos?*

Si el derecho natural existe para el hombre por el solo hecho de ser hombre:

· *¿Qué sucede con la captación y respeto por el derecho natural en una sociedad que puede tender a ser cada vez más deshumanizada y direccionada a la Inteligencia Artificial?*

· *¿Puede la inteligencia artificial captar la realidad sensorial igual que el ser humano?*

Desde un concepto escolástico, el obrar sigue al ser, lo que implica conocer el ser por su obrar; como así también distinguimos el hacer del obrar, donde el hacer perfecciona la cosa creada y el obrar perfecciona la cosa creada al mismo tiempo que al sujeto creador.

Entonces:

· *¿Podemos sostener una asimilación del ser humano con la inteligencia artificial al valorar el hacer y el obrar?*

· *¿De sostener que no, donde anclan sus diferencias estructurales, cuando la inteligencia artificial se va perfeccionando con cada toma de decisiones producto del sistema de* machine learning*?*

· *¿Mediante el obrar de la inteligencia artificial, conocemos su ser o la de su autor o programador?*

· *¿Qué pasa en los sistemas de inteligencia artificial fuerte en el caso de su viabilidad en nuestra sociedad?*

· *¿Hablamos de la inteligencia artificial fuerte como un nuevo tipo de ser?*

Por último, creo necesario plantear interrogantes a partir de la responsabilidad de los Estados:

· *¿Cuál es el rol, alcance, responsabilidad y compromiso de un Estado ante las manifestaciones de los sistemas de inteligencia artificial que puedan vulnerar derechos de las personas, o por qué no, invadir competencias públicas generando un nuevo mecanismo de gobierno?*

· ¿La ciencia, la teología, la filosofía, el derecho, la sociología, la antropología, la educación, etc., cuentan con el suficiente marco conceptual o estado de la cuestión sobre el tópico de inteligencia artificial con un verdadero estudio de prospectiva, antes que aquello que se cree lejano se manifieste en tiempo presente en estado acabado?

VIII. Conclusiones

Estamos en presencia de un nuevo enfoque, donde nuestro trabajo es generar, en primera medida, un acostumbramiento por el lenguaje digital.

Luego, debemos definir si el paradigma actual consiste en la apropiación de la tecnología en un territorio donde el software modela la sociedad, lo que implica contar con la firme necesidad de leer el escenario en prospectiva.

Seguidamente, deberemos empezar a contar con respuestas sólidas y consensuadas a los interrogantes planteados anteriormente, respuestas que nos permitirán construir axiomas, principios y reglas éticas que permitan delimitar el campo de aplicación de la inteligencia artificial, imposibilitando una potencialidad que vulnere los derechos del ser humano y convierta al nuevo dispositivo en una herramienta deshumanizadora.

IX. Propuesta

Aplicar cambios en la sociedad implica una tarea de aprendizaje previa, como así también el abordaje de una actividad evangelizadora, donde la capacitación y la formación son de suma importancia.

Como actores de la educación, que formamos personas desde nuestras distintas disciplinas, tenemos el desafío de abordar una matriz relacional sobre la perspectiva de la ética, con la intencionalidad de generar un espacio más humanizador en la era de la inteligencia artificial.

Tenemos numerosos dispositivos a nuestro alcance, pero el desafío de una matriz relacional permite relacionar lo que está y "crear en la realidad".

Hablar de matriz relacional acentúa el carácter creador, y es lo que el ser humano está haciendo hoy, a través de aportar toda su ciencia al fortalecimiento de los sistemas de inteligencia artificial, contando aún con plena autonomía.

La educación tiene una capacidad transformadora y es a partir de ella que podemos operar en la realidad, dotando a la sociedad de herramientas a partir del saber, sin que esta pierda el sesgo humanizador.

Superado el momento de la capacitación y enseñanza que debemos aplicar a la sociedad, es fundamental analizar la gravitación de algunas categorías en esta nueva era.

Sirva como propuesta el llamado que debemos hacer como sociedad, y sobre todo como pueblo latinoamericano, a debatir y analizar la verdadera gravitación de la ética en la inteligencia artificial para salvaguardar nuestras notas distintivas, nuestros derechos como personas y propender a una sociedad evolucionada pero cada vez más justa.

Los actores de las ciencias sociales tenemos este desafío por delante, y el resultado será alcanzado en la medida que generemos espacios de diálogo, investigación y divulgación de conocimiento.

El gran aporte del evento que nos convoca es que permite construir espacios de ciencia abierta, ya que solo en la medida en que podamos abrir la ciencia, los diseños, los campos de análisis, la divulgación y conclusiones de proyectos colaborativos de investigación, podremos lograr reglas, normas, y principios comunes que pregonen las prácticas éticas de la inteligencia artificial.

X. Bibliografía

AI Now Institute. (2018). *AI Now Report*. Recuperado de https://ainowinstitute.org/AI_Now_2018_Report.pdf

Asimov, I. y Barlett, R. (1942). Runaround. *Astounding Science Fiction*, 29 (1) (pp. 94-103).

Barrio, M. (2019). *La importancia de la ética en la inteligencia artificial*. Recuperado de https://retina.elpais.com/retina/2019/02/25/tendencias/1551089772_654032.html

Boo, J. (2020). *El Vaticano lanza el llamamiento «Ética en la Inteligencia Artificial»*. Recuperado de https://www.abc.es/sociedad/abci-vaticano-lanza-llamamiento-etica-inteligencia-artificial-202002281248_noticia.html?ref=https%3A%2F%2Fwww.google.com%2F

Bostrom, N. (2003). Are we living in a computer simulation? *Philosophical Quarterly*, vol. 53. Recuperado de https://doi.org/10.1111/1467-9213.00309

Bostrom, N. (2003). Human genetic enhancements: A transhumanist perspective. *Journal of Value Inquiry,* Nº 37. Recuperado de https://doi.org/10.1023/B:INQU.0000019037.67783.d5

Bostrom, N. (2003). The Transhumanist FAQ. Philosophy. Recuperado de https://www.nickbostrom.com/views/transhumanist.pdf

Bostrom, N. (2005a). A History of Transhumanist Thought. *Journal of Evolution and Technology,* 14 (1). Recuperado de https://www.nickbostrom.com/papers/history.pdf

Bostrom, N. (2005b). In defense of posthuman dignity. *Bioethics,* vol. 19. Recuperado de https://www.nickbostrom.com/ethics/dignity.html

Bostrom, N. (2005c). Nick Bostrom habla de nuestros más grandes problemas. *Ted global. Ted ideas worth spreading.* Recuperado de https://www.ted.com/talks/nick_bostrom_a_philosophical_quest_for_our_biggest_problems?language=es

Bostrom, N. (2009). The Future of Humanity. *A Companion to the Philosophy of Technology*. Recuperado de https://doi.org/10.1002/9781444310795.ch98

Bostrom, N. y Ćirković M. (ed.) (2009). Global Catastrophic Risks. *Risk Analysis*. Recuperado de https://doi.org/10.1111/j.1539-6924.2008.01162.x

Bostrom, N. (2013). Existential risk prevention as global priority. *Global Policy,* vol. 4. Recuperado de https://doi.org/10.1111/1758-5899.12002

Bostrom, N. y Sandberg, A. (2009). Cognitive enhancement: Methods, ethics, regulatory challenges. *Science and Engineering Ethics,* 15 (3). Recuperado de https://doi.org/10.1007/s11948-009-9142-5

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: University Press.

Bostrom, N. y Yudkowsky, E. (2014). The ethics of artificial intelligence. En W. Ramsey y K. Frankish, *The Cambridge Handbook of Artificial Intelligence*. Recuperado de https://doi.org/10.1017/cbo9781139046855.020

Bracamonte, F. (2019). *El mito griego de Talos, el primer robot*. Recuperado de https://www.ted.com/talks/adrienne_mayor_the_greek_myth_of_talos_the_first_robot/transcript?language=es#t-223044

Comisión europea (2018). *Artificial intelligence for Europe*, 237. final, 1. Recuperado de https://ec.europa.eu/digital-single-market/en/news/communication-artificialintelligence-europe

Cornieles, P. (2019). Las tres leyes de la inteligencia artificial. *Ia Latam.* Recuperado de https://ia-latam.com/2019/07/31/las-tres-leyes-de-la-inteligencia-artificial/

Corvalán, J. (2019). *Prometea, inteligencia artificial para hacer Justicia*. Recuperado de https://www.ambito.com/politica/justicia/prometea-inteligencia-artificial-hacer-justicia-n5061091

Diaz, E. (2018). *La Comisión Europea presenta un documento de principios éticos para el uso de IA*. *Blogthinkbig.com* Recuperado de https://blogthinkbig.com/inteligencia-artificial-comision

Fussell, S. (2020). *La vigilancia del consumidor entra en su fase de negociación*. Recuperado de https://www.theatlantic.com/technology/archive/2019/06/alexa-google-incognito-mode-not-real-privacy/590734/-promo-atlantic-06-04T11%3A15%3A05 [Fecha de consulta: marzo de 2020].

Hawking, S (2018). *Stephen Hawking at Web Summit 2017*. Recuperado de https://youtu.be/H41Zk1GrdRg Disponible a partir del 4 de septiembre de 2018.

Litterio, A. (2019). *De Talos a los robots colaborativos*. Recuperado de https://www.perfil.com/noticias/columnistas/de-talos-a-los-robots-colaborativos.phtml

Marín García, S. (2019). *Ética e inteligencia artificial*. IESE, ST-522. Recuperado de https://dx.doi.org/10.15581/018.ST-522

Martínez, A. (2018). *¿Qué diría Aristóteles sobre la inteligencia artificial?* Recuperado de https://www.bbva.com/es/diria-aristoteles-inteligencia-artificial/

Molina, M. (2017). *¿Inteligencia artificial? Primero hay que conocer bien nuestra propia mente*. Recuperado de https://www.elmundo.es/tecnologia/2017/08/01/597f5de6468aeb14378b4649.html

Mutual, G. (2020). *Papa a Pontificia Academia para la Vida: tecnologías bien utilizadas pueden dar buenos frutos*. Recuperado de https://www.vaticannews.va/es/papa/news/2020-02/papa-pontificia-academia-vida-palabra-tradicion-ayuden-interpret.html

Schmidhuber, J. (1992). Learning Complex, Extended Sequences Using the Principle of History Compression. *Neural Computation*. Recuperado de https://doi.org/10.1162/neco.1992.4.2.234

Schmidhuber, J. (1992). Learning to Control Fast-Weight Memories: An Alternative to Dynamic Recurrent Networks. *Neural Computation*. Recuperado de https://doi.org/10.1162/neco.1992.4.1.131

Schmidhuber, J. (1997). *A computer scientist's view of life, the universe, and everything*. Recuperado de https://doi.org/10.1007/bfb0052088

Schmidhuber, J. (2006). Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science*. Recuperado de https://doi.org/10.1080/09540090600768658

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*. Recuperado de https://doi.org/10.1016/j.neunet.2014.09.003

TeleSUR; AVR; MER (2020). *Papa Francisco insta a una inteligencia artificial con ética*. Recuperado de https://www.telesurtv.net/news/papa-francisco-inteligencia-artificial-etica-20200228-0049.html

Xataka TV (2008). *Las leyes de Asimov y la ética en robots e inteligencia artificial*. Recuperado de https://www.youtube.com/watch?v=jkyhV4QzGhY

Wundersee, P. (2019). La ética en la inteligencia artificial (8 de abril). Recuperado de https://www.youtube.com/watch?v=QWFUaDfg0Ks

Fecha de recepción: 30-03-2020 Fecha de aceptación: 29-06-2020