

Sobre la necesidad de unificación de las iniciativas para un uso ético de la Inteligencia Artificial

On the need to unify initiatives for an ethical use of Artificial Intelligence

Carlos Lázaro Nodals García¹  

Cómo citar

Nodals García¹, C.L. (2021). Sobre la necesidad de unificación de las iniciativas para un uso ético de la Inteligencia Artificial. *Socialium*, 5(2), 318-334. <https://doi.org/10.26490/uncp.sl.2021.5.2.880>

¹ Magister en Estudios Socioculturales, Licenciado en Derecho.
Universidad de Cienfuegos
Carlos Rafael Rodríguez,
Cienfuegos, Cuba.
clnodals@ucf.edu.cu

RESUMEN

El trabajo tiene como objetivo revisar, exponer y comprender la necesidad de unificación de las iniciativas para un uso ético de la Inteligencia Artificial (IA). Expone la relación entre ciencia, ética, sociedad y comportamiento del profesional de la ciencia. A partir de los abordajes básicos de esta tecnología, se describen áreas de beneficios probados. En la metodología, consultando las bases de datos Dialnet y Scielo, se hizo una revisión sistemática de libros, artículos científicos, informes y declaraciones de parlamentos, agencias gubernamentales y foros internacionales, así como artículos de especialistas en el tema. Se revisaron los resúmenes y artículos completos, en español e inglés, considerando los que trataban la unificación viable de todas las iniciativas tendientes al uso ético de la IA. Los resultados muestran los riesgos de la implementación acrítica y poco ética de la misma, las implicaciones del desarrollo de aplicaciones que incorporan Inteligencia Artificial, para la sociedad y las personas en su vida privada. Se habla de la toma de conciencia por los gobiernos y los ciudadanos, de los riesgos de esta tecnología, y de la urgencia de unificar en un solo modelo, su uso ético, seguro y transparente. Las conclusiones arrojan que prevalece en la literatura consultada un consenso sobre la necesidad de unificar las iniciativas para su uso ético y verificable, lo que demuestra una conciencia de los riesgos potenciales de su desarrollo sin un control consensuado. Pero prima el entusiasmo por los beneficios de esta tecnología, lo cual es un obstáculo para lograrlo.

Palabras clave: *inteligencia artificial; aprendizaje automático; ética; ética de la ciencia; deontología.*

ABSTRACT

The objective of the work is to review, expose and understand the need to unify initiatives for an ethical use of Artificial Intelligence (AI). It exposes the relationship between science, ethics, society and the behavior of the science professional. Starting from the basic approaches of this technology, areas of proven benefits are described. In the methodology, consulting the databases Dialnet and Scielo, a systematic review was made of books, scientific articles, reports and declarations of parliaments, government agencies and international forums, as well as articles by specialists on the subject. The abstracts and full articles, in Spanish and English, were reviewed, considering those that dealt with the viable unification of all initiatives aimed at the ethical use of AI. The results show the risks of its uncritical and unethical implementation, the implications of the development of applications that incorporate Artificial Intelligence, for society and people in their private lives. There is talk of awareness by governments and citizens of the risks of this technology, and the urgency of unifying in a single model, its ethical, safe and transparent use. The conclusions show that a consensus prevails in the consulted literature on the need to unify the initiatives for their ethical and verifiable use, which shows an awareness of the potential risks of their development without a consensual control. But enthusiasm for the benefits of this technology prevails, which is an obstacle to achieving it.

Keywords: *artificial intelligence; machine learning; ethics; ethics of science; deontology.*

Arbitrado por pares ciegos

Recibido: 09/03/2021

Aceptado: 29/06/2021

Introducción

Con el fin de la llamada Guerra Fría y la llegada del siglo XXI, las relaciones internacionales se vieron ante nuevos desafíos en el ámbito político, económico y científico. Los dilemas éticos en las investigaciones médicas y genéticas se convierten en temas de conocimiento y discusión pública, debido a su tratamiento mediático. Otras cuestiones como la eutanasia y el aborto adquieren rango de debate nacional en algunos países. Las representaciones sociales sobre la ciencia y su relación con la sociedad cambian drásticamente. La sociedad civil reclama tener incidencia en las decisiones de los gobiernos en relación a estos problemas.

Se discute si es posible constituir una ética en la ciencia, si es alcanzable una ética profesional en el ámbito científico. Los científicos pueden, o no, tener principios que los guíen como profesionales en el ejercicio de su actividad. Pero hay que tener en cuenta que la moral social se manifiesta en el comportamiento del científico. Es cada día más evidente que los rápidos y cambiantes progresos de la ciencia y la tecnología actuales, reclaman un enfoque ético y un abordaje humanistas del campo científico. Existe una cuestión medular en esta problemática: ¿la ciencia es éticamente neutral?

La cantidad de dilemas éticos que hoy presenta el avance científico tecnológico a las sociedades, generan una diversidad de problemas éticos, morales, filosóficos, jurídicos, entre otros; que justifican la necesidad y pertinencia de al menos, una responsabilidad del científico. Añádase a esto, que las tecnologías de la información y la comunicación han moldeado una opinión pública cada vez más concedora de los aspectos relativos a la ciencia, y que reclama una cuota de influencia en las decisiones que se adopten en el abordaje de determinados desafíos, cuyos resultados pueden impactar en la naturaleza humana y extrahumana.

Hay líneas de investigación y proyectos de alto impacto, de tal envergadura, que pueden revolucionar la naturaleza humana y la existencia misma del homo sapiens. Entre ellas se encuentra el desarrollo de la Inteligencia Artificial (IA). De ahí que es necesario que el rigor científico y las potencialidades de una investigación estén en concordancia con valores como el humanismo y la honestidad. Pero ante la eventual aplicación irresponsable de un resultado científico, la principal virtud del profesional de la ciencia tiene que ser el coraje. Todo lo anterior puede subsumirse en el sentido de la justicia.

La ciencia no puede trabajar como una secta. “Los llamados posmodernos se limitan a hacer afirmaciones mientras más herméticas y menos fundamentadas mejor”. (Bunge, 2000, p. 50). Por ende, apoyamos la existencia de un vínculo entre ciencia, humanismo y moral. Todo conocimiento que emane de ella incide en el destino del hombre y debe ser verificable. Concebimos la ciencia como una actividad humana

especial, un sistema lógicamente estructurado de obtención de conocimientos verificables, cuya esencia es el descubrimiento de las leyes de los fenómenos de la realidad objetiva y su correspondiente explicación.

Apartado teórico

Para entender cómo se puede desarrollar la IA, existen cuatro abordajes básicos.

Los sistemas que piensan como humanos, aquellos que modelan las propiedades del procesamiento cognitivo de los humanos. Los sistemas que actúan como humanos, que pueden realizar cosas específicas que hacen los humanos, lo que incluye el procesamiento del lenguaje natural, razonamiento automático, representación del conocimiento, aprendizaje automático, visión computacional y robótica. Los sistemas que piensan racionalmente, con un pensamiento estructurado, tales como silogismos y lógica formal. Y por último, los sistemas que actúan racionalmente, o sea, que hacen cosas racionales. (Russell y Norvig, 2009, p. 5)

De aquí emanan cinco perspectivas fundamentales. Cada una prevé los resultados esperados del avance e implementación de los logros de la IA. ¿Qué puede depararnos la misma en el mediano y largo plazos? ¿Que los humanos y las máquinas luchen por la hegemonía y estas últimas salgan vencedoras? ¿Que la presencia de las máquinas en todos los ámbitos de la vida social humana traiga un progreso y una riqueza sin parangón? ¿O que las máquinas sean la causa de una sociedad escindida entre las personas súper especializadas y con competencias actualizadas, y aquellas que malviven de profesiones arcaicas y/o su esfuerzo físico?

Hay muchas definiciones de IA. Con diferentes enfoques, pero en torno a una idea fundamental: crear y dar forma a programas de ordenador y/o máquinas que sean capaces de desarrollar conductas que si las realizara un humano, serían consideradas inteligentes. Cualquier sistema de IA debe poseer algunas características básicas. La capacidad de aprender, lo que en el campo computacional se denomina *machine learning*. Otra, es manejar la incertidumbre y la información probable, así como formar conceptos a partir de representaciones combinatorias que se usan en el razonamiento lógico e intuitivo, y poseer esos caracteres básicos.

El núcleo teórico en el campo de la IA está en determinar si una máquina puede pensar. Probar la habilidad de una máquina y comparar el conocimiento de esta con el de un ser humano y verificar si ese conocimiento es cualitativamente similar o superior. Existe una constante reflexión filosófica sobre las posibilidades pensantes de una máquina. Ello ha traído un fuerte interés en el desarrollo de la IA en dos direcciones: una que tiene que ver con una teoría de la información que sea más sólida para el aprendizaje artificial. Y otra hacia el aspecto práctico, de resolución de problemas concretos y de

ámbitos específicos. A pesar de todo, no existe una postura unánime en torno al futuro de esta tecnología.

Muchos intelectuales cuando abordan el campo de la IA, centran sus reflexiones en la proyección de la misma hacia el futuro. Unos hablan de súper inteligencia y otros, de singularidad. Lo innegable es que nos encontramos ante un desarrollo indetenible en la robótica y la IA, que tiene que ser pensada desde la filosofía, la juridicidad y la ética, pues ese desarrollo trae aparejado numerosos e importantes desafíos. Dilemas ecológicos y políticos globales que necesitan un enfrentamiento y resolución solidarios. La relación del hombre con las máquinas ha modificado el mundo. Las aplicaciones y productos de esta tecnología son cada vez más autónomos. Su control ético es una necesidad intrínseca.

La ética tradicional no puede dar una respuesta a este escenario donde los efectos adversos para la biosfera provienen de variadas fuentes, la mayoría tecnológicas. Ahora se trata de preservar la existencia de la humanidad en términos colectivos. El desarrollo de la IA tiene que ser considerado desde la responsabilidad, no abusar de la capacidad que puede llegar a tener. Al respecto, Kurzweil (2017) plantea que cuando arranque la década de 2030, se producirá un volumen de cálculo no biológico casi igual a toda la inteligencia humana biológica viviente. Que esto no representará la Singularidad, pero en la década siguiente, la inteligencia anual creada será aproximadamente mil millones de veces más poderosa que toda la inteligencia humana actual. Por eso se fija la Singularidad para el 2045, caracterizándola como la transformación más profunda y disruptiva en la capacidad humana.

Con frecuencia se tiende a negar la entrada de la IA en la vida cotidiana por parte de las personas. Esto ocurre por la idea prefabricada que proviene del cine de ficción, que ha representado básicamente los efectos adversos de esa tecnología. Pero la IA es un factor de progreso y humanización de las condiciones de existencia del ser humano en su desarrollo bio-psico-social, si se implementa según parámetros y regulaciones adecuadas. La lista de aplicaciones que pueden mejorar la vida de la sociedad va desde optimizar el uso de fondos públicos, las políticas de igualdad, la salud, la educación y la distribución de la energía. También el control fiscal y la seguridad ciudadana. Es una de las ramas de la tecnología de más rápida expansión y desarrollo.

La IA hace posible que las máquinas aprendan de la experiencia, se ajusten a nuevas aportaciones y realicen tareas como seres humanos. La mayoría de los ejemplos de los que se oye hablar en la actualidad, recurren básicamente al aprendizaje profundo y al procesamiento del lenguaje natural. Con estas tecnologías, las computadoras son entrenadas para realizar tareas procesando grandes cantidades de datos e identificando patrones en ese cúmulo de datos. No es lo mismo que la automatización de robots, esto último se hace basado en hardware y su objetivo es fundamentalmente sustituir al hombre en tareas

manuales o industriales. La IA agrega inteligencia a productos y aplicaciones existentes, mejora esos productos y tecnologías.

Tiene la capacidad de adaptación mediante algoritmos de aprendizaje progresivo, para que los datos realicen programación. Encuentra estructuras, regularidades y patrones en los datos. El algoritmo adquiere una habilidad, clasifica, empatrona. Si se proveen nuevos datos, los algoritmos se adaptan y ajustan. Analiza datos profundos y complejos utilizando redes neuronales de varias capas. Cada día crece el poder de cómputo gracias al desarrollo de la tecnología, con el *big data*, por ejemplo. Esto ha permitido entrenar modelos de aprendizaje profundo que aprenden directamente de los datos. La calidad o importancia de los datos redundan en competitividad y éxito, y en última instancia, en poder: se han vuelto estratégicos.

Los datos imponen que cualquier sistema deberá ser capaz de aprender continuamente. Incorporar un número ilimitado de conocimientos mientras se encuentre operativo.

En definitiva, es imprescindible diseñar sistemas que integren percepción, representación, razonamiento, acción y aprendizaje. Este es un problema muy importante en IA, ya que todavía no sabemos cómo integrar todos estos componentes de la inteligencia. Necesitamos arquitecturas cognitivas que integren estos componentes de forma adecuada. Los sistemas integrados son un paso previo fundamental para conseguir algún día inteligencias artificiales de tipo general. (Forbus, 2012, p. 55)

Esa es la meta que busca alcanzar el desarrollo de esta tecnología. Los principales proyectos de investigación están enfocados en el salto hacia determinados niveles de inteligencia general que traigan aparejados un salto cualitativo en la relación humano-máquina.

Apartado contextual-histórico

En los inicios de la IA en los años 50 del siglo XX, se esbozaron las posibles aplicaciones y áreas de desarrollo de la misma. En la década siguiente comenzaron los esfuerzos por aplicarla en la resolución de problemas prácticos y en operaciones de predicción.

Los algoritmos genéticos – denominados originalmente planes reproductivos genéticos – fueron desarrollados por John H. Holland a principios de la década de 1960, quien interesado en estudiar los procesos lógicos que se daban en la adaptación, e inspirado por los estudios realizados en esa época con autómatas celulares y redes neuronales, se percató que el uso de reglas simples podía conllevar a comportamientos flexibles visualizando así, la posibilidad de estudiar la evolución en sistemas complejos. (Ponce *et al.*, 2014, p. 117)

Desde el comienzo del siglo XXI, la IA se ha transformado en una manera cualitativamente más compleja de transformación de datos. La conexión de la sociedad a internet, proceso que se acrecienta, ha resultado en la digitalización de toda la producción cultural humana. Esa enorme cantidad de información hace de los datos un activo básico para cualquier organización, empresa o país. Es el resultado de un sistemático avance en la informática. La competición entre la inteligencia humana y la inteligencia artificial está servida para las próximas décadas. Como humanidad, no hemos logrado comprender todo lo que está en juego, las implicaciones que tiene el desarrollo de la IA, y el riesgo que supone.

Desde los años 80 del siglo pasado los sistemas expertos se convirtieron en la tecnología estándar para actividades industriales, financieras, médicas y militares, entre otras. Si repasamos los logros industriales y productivos de la época se constata que la IA se había aplicado en la economía de punta en varios países. También, desde ese momento, surgió una diversidad de visiones sobre el futuro de esa tecnología. Esta situación se mantiene hasta la actualidad.

La opinión de los expertos sobre el futuro de la IA varía enormemente. No hay un acuerdo sobre la sucesión temporal de los acontecimientos ni sobre qué formas podría llegar a adoptar la IA. Las predicciones sobre el futuro desarrollo de la inteligencia artificial, señaló un estudio reciente, son tan firmes como diversas. (Bostrom, 2016, p.16)

El período entre 1990 y mediados de la primera década del 2000, puede catalogarse como de salto histórico. Comienza el desarrollo del aprendizaje automático (*machine learning*).

Es el subcampo de las ciencias de la computación y una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan. Se dice que un agente aprende cuando su desempeño mejora con la experiencia; es decir, cuando la habilidad no estaba presente en su genotipo o rasgos de nacimiento. (Russell y Norvig, 2009, p. 229)

Esto abrió la puerta a las aplicaciones que incorporan IA, con las cuales interactuamos en la cotidianidad: búsqueda de canciones, películas y videoclips, fotos, entre otros. Se considera que fue la irrupción de los teléfonos inteligentes los que la colocaron en el accionar diario de la ciudadanía.

Esta tecnología pasó de la especulación científica y la ciencia ficción a la realidad. Mejora, viabiliza y beneficia la vida de las personas. Pero sus efectos adversos ya han sido experimentados por la sociedad. Dígase las armas autónomas, la utilización de datos públicos y privados para fines ilegítimos e ilegales, entre otros; que han hecho que los ciudadanos y los gobiernos tengan la percepción de que esta tecnología, sin un control ético y legal, puede traer consecuencias que afecten incluso el modo de existencia humano, tal como lo hemos concebido hasta hoy. Son preocupaciones legítimas, han puesto a la comunidad mundial ante la problemática de cómo implementar una iniciativa global, mancomunada, que garantice su uso ético y verificable.

Más allá del sensacionalismo y las elucubraciones teóricas de algunos medios de comunicación, personas y organizaciones poco serias; lo cierto es que estamos ante la impostergabilidad de dotarnos de un marco de actuación en la IA, que sea reconocido por todos los actores. Que garantice la responsabilidad de las empresas en su desarrollo y expansión presente y futuro, la seguridad para los usuarios que utilizan productos y aplicaciones que incorporan la misma. El respeto a los derechos humanos fundamentales, el equilibrio existencial de los seres humanos, que no puede ser alterado o destruido por ninguna tecnología. En nuestra opinión, esto solo puede venir de una regulación gubernamental conjunta, única, con un basamento y reconocimiento internacional.

La vida social contemporánea ha producido eventos en lo político, científico, tecnológico, y otras áreas de la práctica humana; que han provocado una falta de credibilidad en lo que dicen los expertos. El exceso de entusiasmo, los vaticinios ingenuos, las decisiones apresuradas e inconsultas, las actuaciones irresponsables y unilaterales; han reforzado esta percepción. La IA no está fuera de este debate, es un tema central en los temores ciudadanos. La respuesta a estos desafíos requiere cooperación internacional, abierta y franca, el establecimiento de reglas de actuación válidas para todos. Es la única forma en que esta tecnología traerá un estadio humano superior, en lo existencial y lo social.

Método

Tipo de estudio. Se utilizó el método de revisión de materiales bibliográficos dedicados a las iniciativas para implementar un uso ético de la Inteligencia Artificial, tanto en formato electrónico como físico: libros, artículos científicos, páginas webs, informes y declaraciones de parlamentos, agencias gubernamentales y foros internacionales. Asimismo, artículos de especialistas en el tema, publicados en revistas y prensa especializada. También publicaciones secundarias como reseñas, referencias y citas de revistas, presentando los antecedentes y teorías que amplían el tema.

Población y muestra. Se llevó a cabo una búsqueda de documentos publicados en español e inglés, en el contexto internacional y en Cuba; sobre las estrategias de unificación para el uso ético y verificable de la IA. Posteriormente, se realizó una búsqueda de revisiones sistemáticas de la literatura científica en las bases de datos Dialnet y Scielo. No se estableció por año de publicación pero se puso como límite que el idioma fuera inglés o español. Se analizaron además las referencias bibliográficas de los artículos seleccionados con el fin de rescatar otros estudios potencialmente incluíbles para la revisión. En la búsqueda de literatura gris se incluyó todo tipo de documentos aportados por órganos legislativos, agencias gubernamentales, corporaciones y foros internacionales, con recomendaciones, pronunciamientos y posicionamientos sobre las iniciativas que coexisten para un uso ético de la IA.

Respecto a los libros y artículos científicos se aplicó como criterio de inclusión que los mismos incorporaran abordajes y recomendaciones para la unificación viable de todas las iniciativas tendientes al uso ético de la IA. El principal criterio de exclusión recayó en cualquier literatura revisada que no incluyese información sobre los beneficios y peligros de la IA y la necesidad de su regulación ética verificable.

Instrumentos de recolección de datos. El material impreso en formato físico se obtuvo del Centro de Información de la Sede Central de la Universidad de Cienfuegos. Los documentos electrónicos fueron localizados a través de Google Scholar. Mediante los descriptores del Tesoro UNESCO se introdujeron los términos de búsqueda objeto del estudio en los buscadores Google y Bing. La discriminación de la información se realizó según el tipo y calidad de la fuente. Se utilizó una ficha de análisis con núcleos conceptuales.

Procedimientos de la recolección de datos. La información analizada se estructuró en dos apartados: uno dedicado a la Inteligencia Artificial como tecnología, beneficios y peligros de su implementación. Y otro dedicado a las iniciativas para su uso ético y verificable, así como a los esfuerzos para la unificación de las mismas. De los libros consultados se extrajeron los elementos teóricos e históricos de la IA, su evolución y fundamentos. De los informes, se obtuvo el estado de las propuestas para la unificación de las iniciativas, la visión jurídica y el posicionamiento político administrativo en torno al tema. De los artículos científicos de revistas indexadas y de las publicaciones de expertos en la prensa especializada, se obtuvieron elementos de actualidad y de desarrollo del tema abordado. Tras la búsqueda inicial se localizaron 102 documentos, entre libros impresos y artículos científicos en revistas, de los cuales se excluyeron 36 que no fueron relevantes para el objetivo de esta revisión. Finalmente se seleccionaron 7 libros, 13 informes gubernamentales, 12 artículos científicos y 34 materiales en publicaciones especializadas. Para proceder a la selección se revisaron los resúmenes y en caso necesario los artículos completos, para decidir si la información que contenían estaba o no relacionada con el objetivo del trabajo y la relevancia de la misma.

Aspectos éticos. Se cumplió con la pauta ética de tomar información de documentos contenidos en fuentes especializadas en el tema, reconocidas y confiables. Se evitó la literatura especulativa, sesgada o con planteamientos contrarios al método científico, la neutralidad y la responsabilidad social.

Análisis de datos. Se aplicaron métodos y técnicas de investigación cualitativa como el análisis de contenido, la lectura crítica de teorías y conceptos, y la interpretación de textos. A través de los métodos aplicados, se obtuvieron los resultados.

Resultado

En el contexto de la pandemia de COVID-19, la IA mostró sus potencialidades de aplicación en beneficio de la humanidad. Algunos ejemplos sentaron precedentes. Jaimovich (2020) menciona los robots que atienden comensales en Holanda, monitorean el distanciamiento social en Singapur, en otros países limpian establecimientos, hacen entregas, atienden pacientes por medio de telemedicina y entregan diplomas de graduación. El *big data* se aplicó en los mapas desarrollados por la Universidad Johns Hopkins, para conocer el avance de la pandemia en el mundo. Y los mapas de Google y Apple permitieron conocer la forma de moverse de los usuarios durante las cuarentenas.

A esto se suman las aplicaciones de rastreo de contacto que permitieron identificar y frenar posibles cadenas de contagios, por medio de celulares. Para monitorear y evaluar el ámbito social en lugares públicos y privados, surgieron otras iniciativas que implican la integración de IA en cámaras de seguridad. El enfrentamiento al coronavirus SarsCov-2, mostró nuevas opciones o utilidades para los múltiples desarrollos tecnológicos basados en el Aprendizaje Automático y el *big data*.

La gran feria de tecnología CES 2021, mostró productos futuristas que constituyen la base de lo que será en próximas décadas el hogar inteligente. La Corporación sudcoreana Samsung reportó en su sitio digital sus últimas creaciones. Samsung web (2021) describe las novedades de las heladeras inteligentes que permiten al usuario saber qué ingredientes tiene con la pantalla *View Inside*. También heladeras que ofrecen recetas. En la pantalla de inicio del *display* hay un menú que muestra fotos, videos, notas o el calendario. Reproduce la pantalla del celular o del *smart TV*. La tecnología *SmartThings Cooking* ofrece un planificador automático de comidas basado en las necesidades y preferencias del usuario, gracias a la Inteligencia Artificial. Se pueden comprar ingredientes, elegir o enviar recetas para dispositivos conectados, desde un mismo espacio.

Si analizamos esta línea de productos, se comprende que la IA en su aplicación doméstica, busca reconocer y comprender los comportamientos del usuario. El objetivo principal es convertirse en un compañero dentro del hogar y ofrecer acompañamiento personalizado. Fue Samsung precisamente, la que presentó lavarropas que sugieren ciclos de lavado. Permite controlar, de forma conjunta, la lavadora y la secadora, y calcula la cantidad óptima de agua y detergente. Una aspiradora con tecnología de reconocimiento de objetos que identifica elementos y decide cómo moverse sin tropiezos. Remueve la suciedad incluso de los rincones y funciona como cámara de seguridad, ya que el usuario puede supervisar qué sucede en el hogar cuando no está.

Pero no es la única. Muchas corporaciones de tecnología de punta en el mundo, están desarrollando una línea de robots para el hogar. Algunos modelos, ya operativos, reconocen y recogen objetos. Potenciados por IA, pueden diferenciar entre un vaso de vidrio y un plato de cerámica, y responde al usuario cuando

lo consulte al respecto. Otros, guardan alimentos y ponen la mesa. Las últimas versiones de estos robots, están mejorando las capacidades, para que puedan aprender de los comportamientos del usuario, de su rutina diaria y horarios, para convertirse en un compañero que ofrece ayuda, compañía y soporte psicológico.

Sin embargo, ya se han dado escenarios donde la IA ha sido utilizada con objetivos malsanos. Esto ha disparado las alertas a nivel mundial. La Guardia Revolucionaria de Irán confirmó su empleo en el asesinato su científico Mohsen Fajrizadé en Teherán. Cubatedebate (2021) plantea que contaba con una escolta de once efectivos y la explosión del vehículo desde el que se hicieron los disparos por control remoto, tenía como objetivo eliminar su protección. Los disparos se efectuaron por satélite desde una camioneta conducida por control remoto. El arma enfocaba sólo a Fajrizadé. Su esposa no sufrió disparos, a pesar de estar a unos centímetros de distancia. El jefe del equipo de protección también recibió cuatro porque se arrojó sobre él. Nadie estaba en el lugar del asesinato, la ametralladora estaba equipada con Inteligencia Artificial.

La venta de juguetes sexuales es otra actividad donde está presente esta tecnología, y que crece cada año. La categoría de estos dispositivos conectados se llama *teledildonics*. Están ganando popularidad como parte del concepto sexnología: una combinación de sexo y tecnología. Los juguetes sexuales incorporan chats grupales, mensajes multimedia, videoconferencias, transferencia de archivos, sincronización con listas de canciones o audiolibros, la capacidad de conectarse con Alexa y otros asistentes inteligentes y la posibilidad de ceder el control del dispositivo a usuarios remotos. Las compañías especializadas en ciberseguridad concuerdan en que, con la expansión de ese mercado, crecerán los ataques a los dispositivos sexuales.

Los juguetes sexuales inteligentes cuentan con una amplia gama de funciones. Permiten otorgar el control remoto de su dispositivo a otras personas desde una *app* móvil o accederse remotamente a través de internet, forman parte del fenómeno Internet de las cosas (*IoT*) por sus siglas en inglés. Dada la amplia gama de funcionalidades que ofrecen estos productos la superficie de ataque es bastante grande. Sobre el particular, Pastorino y Giusto (2021) refieren que hay ciertas características de diseño que los atacantes pueden explotar como la conexión Bluetooth que algunas veces está desprotegida, vulnerabilidades en el servidor o en las aplicaciones y conexiones Wi-Fi inseguras.

La información procesada por los juguetes sexuales y las aplicaciones es sensible. Nombre, contactos, parejas sexuales, fotos y videos íntimos. Su uso revela las preferencias sexuales, los patrones y horarios preferidos. En caso de robo, esto podría usarse contra la víctima, exponiendo su intimidad o hacerlo objeto de chantaje. También se pueden explotar vulnerabilidades en los protocolos que se utilizan para recopilar información o conectarse al dispositivo burlando mecanismos de autenticación deficientes. Si se

toma el control de un dispositivo sexual sin la autorización del usuario mientras lo está usando, e incluso envía diferentes comandos al dispositivo. ¿Cómo contempla esta acción la legislación actual? ¿Configura una agresión sexual? ¿La legislación tipifica esta acción como un delito y prevé la sanción correspondiente?

La era de los juguetes sexuales inteligentes está comenzando. Los últimos avances en la industria incluyen modelos con capacidades de realidad virtual y robots sexuales impulsados por IA que incluyen cámaras, micrófonos y capacidades de análisis de voz basadas en esa misma tecnología. Determinar el responsable del daño causado por uno de estos aparatos es un desafío. Otro ejemplo, en un accidente en el que esté implicado un coche de conducción autónoma, ¿quién cubre los daños? ¿El propietario, el fabricante o el programador? Liberar de responsabilidad al fabricante puede desincentivar el logro de calidad en el producto. Las regulaciones estrictas desmotivan la innovación. Y en ambas variantes, merma la confianza de las personas hacia esa tecnología. Todo un dilema tecnológico.

Uno de los aspectos que más preocupan son las implicaciones políticas. Los resultados de la IA dependen de su uso y los datos utilizados. Existe la posibilidad de sesgar, intencional o involuntariamente, tanto el diseño como los datos. Algunos aspectos importantes de un asunto podrían no estar programados en el algoritmo. O podrían programarse para reflejar y replicar sesgos estructurales. Un uso discriminatorio sería utilizar la IA en la toma de decisiones influenciadas por la etnia, el sexo o la edad incluidos en los datos, al contratar o despedir de un empleo u ofrecer préstamos en el Banco. También acarrea riesgos a la privacidad y la protección de datos al utilizarse en equipos de reconocimiento facial o seguimiento en línea, y la creación de perfiles de personas.

Los sistemas de IA también pueden usarse para crear vídeos, audios o imágenes falsas pero realistas, conocidos como *deepfakes*, en inglés. Este contenido puede implicar riesgos financieros, daños a la reputación y dignidad de las personas y problemas en la toma de decisiones. Puede conducir a la separación y polarización en la esfera pública, la manipulación de elecciones, y de la selección de la persona idónea para un cargo de dirección o plaza laboral. La libertad de reunión y manifestación está también amenazada por la IA. Con ésta se puede rastrear y controlar a las personas vinculadas a ciertas creencias, acciones, grupos sociales, o militancias políticas.

La acumulación de información puede distorsionar la competencia y hacerla desleal. Los actores con más información tendrían una ventaja sobre los competidores. Pueden aprovechar los desequilibrios en el acceso a la información. Un proveedor en línea puede usar la IA para predecir cuánto está dispuesto a pagar un consumidor al observar su comportamiento en línea o una campaña de marketing puede adaptar su mensaje. Las aplicaciones que están dentro del cuerpo humano, pueden presentar riesgos de seguridad si están mal diseñadas, o son hackeadas. Se habla de ciber asesinato, si se hackea un

marcapasos, por ejemplo. El uso no regulado de la IA en las armas puede provocar la pérdida del control humano sobre las armas de exterminio masivo.

Algunos timos y falsificaciones en el sector de las finanzas y la administración pública, se han ejecutado utilizando la IA. A pesar del esfuerzo de las empresas por evitarlo y del endurecimiento de las puniciones, se siguen reportando con tendencia al aumento. Prang y Tweh (2019) narran en el *Wall Street Journal* la historia de un caso que ocurrió en marzo de ese año. En el mismo, los criminales utilizaron un software basado en IA para imitar la voz del máximo responsable de la gestión y dirección administrativa (CEO) de una empresa de energía alemana. Así llevaron adelante un engaño que permitió a los estafadores robar 220.000 euros.

A pesar de estos hechos, se afianza el enfoque mercantilista. Sobre todo en las áreas de desarrollo más complejas y que pueden tener implicaciones de dominación geopolítica. Rosa (2019) refiere que a partir del 2014, el mercado mundial del Aprendizaje Profundo muestra un crecimiento continuo. Asimismo, que en el 2016, *Grand View Research (GVR)* calculó que ascendía a 272 millones de dólares, de los cuales el 20% pertenecía a la industria aeroespacial y a la de defensa. Y agrega que el último informe de *GVR* indica que este mercado alcanzará el valor de 10.200 millones de dólares a finales de 2025.

Discusión

Vivir con la IA en la cotidianidad requiere actuar en tiempo real. Evitar la elaboración legislativa después del daño o efecto adverso. Hacer análisis transdisciplinarios, desde la filosofía, la ética aplicada, el derecho y el método científico. Fortalecer la confianza de la sociedad con una IA transparente en todos los sectores de la vida pública y privada. Hay preocupaciones legítimas y desconfianza en las aplicaciones que la incorporan. El Parlamento Europeo (2020) plantea la urgencia de adoptar sistemas transparentes que generen confianza de los desarrolladores y del público en general. Se basa en un estudio de la Universidad de Oxford y la Comisión Europea que arroja que el 84% de los estadounidenses y el 88% de los europeos creen que la IA debe ser manejada con cuidado, debido a su potencial para producir resultados sociales dañinos.

Las iniciativas que se implementen deben tener como objetivo el despliegue seguro de la tecnología y el visto bueno de gobiernos, empresas, instituciones y personas. Hay al menos 175 de ellas para que los sistemas de IA sean confiables. Son de empresas privadas, otras gubernamentales y de la ciudadanía. Todas ellas surgen después de que se hayan producido varias controversias como los sistemas de reconocimiento facial, las *deep fake* o la creación de *bots* con voces de personas fallecidas. Y es que

junto con las posibilidades y beneficios, se acrecientan las dudas sobre su alcance, usos y empleos éticos. Se presentan potenciales dilemas que tienen un impacto multidimensional: jurídico, filosófico y ético.

Herranz (2021) analiza la convivencia no coordinada de variados enfoques en el control de estos hechos. Y exalta el papel mediador que puede jugar la *Global AI Action Alliance (GAIA)*, lanzada por el Foro Económico Mundial para resolver la dispersión de iniciativas para un uso ético de la IA. Con la misma se busca la unificación y materialización de todos estos proyectos para eliminar barreras de conocimiento y aprendizaje, y que se fomente una colaboración mundial en ese campo. Lograr que las prácticas y herramientas que han demostrado utilidad se universalicen. Empresas como Accenture, BBVA, Dell, Deloitte, EY, Facebook, HPE, Huawei, Microsoft, Netflix, Palantir, Salesforce, VMware, entre otras, forman parte de GAIA.

Están comprometidas en el diseño conjunto de los marcos de referencia y protocolos de gobernanza de datos, para acelerar los beneficios y minimizar los riesgos de la IA. Marcos de referencia que serán probados en proyectos piloto de bajo riesgo para validar su impacto. La IA está aprendiendo más sobre cómo trabajar con y sobre los humanos. Puede aprender a identificar vulnerabilidades en los hábitos y comportamientos humanos, y usarlos para influir en la toma de decisiones. El Aprendizaje Automático pronto podrá guiar a las personas hacia acciones específicas y aprender a dirigir la toma de decisiones humanas a través de sus interacciones con ellos.

Sobre este aspecto, la relación humano-máquina; orbitan las principales preocupaciones. Whittle (2021) dice que puede parecer un cliché el decir que la IA está transformando las formas de vivir y trabajar, pero es cierto. Que aunque no posee inteligencias y emociones similares a las humanas, está funcionando en campos como el desarrollo de vacunas, la gestión ambiental y la administración de oficinas, sus capacidades son poderosas y se desarrollan rápidamente (...) no hay necesidad de preocuparse por la toma de posesión de una máquina por el momento, pero este descubrimiento reciente destaca el poder de la tecnología y afianza la necesidad de una gobernanza adecuada para evitar el mal uso.

Los avances de la IA en la industria han hecho que las naciones líderes en esta tecnología reevalúen sus planes de desarrollo de la misma. A la par de los logros, aparecen áreas de potenciales conflictos éticos. Estados Unidos tiene un plan estratégico nacional de investigación y desarrollo de la IA. Este país realizó la actualización del mismo, contenía siete estrategias y adicionó una. El Comité Selecto de Inteligencia Artificial (2019) amplió las asociaciones público-privadas para acelerar los avances en Inteligencia Artificial, con el objetivo de vincular empresas, academia, al público y socios internacionales; a fin de acelerar los desarrollos en esta tecnología. Del mismo modo, esta nueva versión señala diferentes desafíos para la adopción de la inteligencia artificial, entre los cuales se pueden resaltar la creación de

estándares técnicos, seguridad y protección y creación de confianza entre otros; cada uno de estos ligados a una o más estrategias anteriormente mencionadas.

El desarrollo y la investigación crecen con rapidez. Países, empresas y organizaciones destinan grandes presupuestos para desarrollar esta tecnología. Y asistimos aquí al principal obstáculo en el logro de una iniciativa unificada para su control y uso ético: la tecnología evoluciona con rapidez, su incorporación es expedita en la sociedad, pero esta rapidez provoca vacíos jurídicos. Las instituciones y el derecho no evolucionan a la par de la ciencia y los avances tecnológicos, generalmente van a la zaga. Derecho y Superestructura se desfasan con el desarrollo y su acople requiere voluntad política. Esto conlleva a la preocupación sobre lo que puede desencadenar un mal uso unido a la ausencia de marco jurídico que tipifique los hechos. De ahí el debate sobre la creación de normas, marcos regulatorios e iniciativas internacionales unificadas.

En el año 2017 varias de estas paradojas fueron expuestas en un reporte de *World Commission on the Ethics of Scientific Knowledge and Technology*, en el cual se enfatizaba en la importancia de la responsabilidad compartida en el área de la robótica. Ni los diseñadores, ingenieros, programadores, fabricantes, inversionistas, vendedores y usuarios tienen responsabilidad completa sobre el uso e impacto de estos, especialmente cuando se trata de robots que tienen un nivel de autonomía cuyas acciones pueden ser impredecibles. “Este problema no solo aborda el impacto de los robots, sino que por el contrario abre la posibilidad de doble uso, es decir, la reprogramación o cambios al propósito de robots diferente al original”. (World Commission on the Ethics of Scientific Knowledge and Technology, 2017).

Este es un ejemplo, pero que puede hacerse extensivo a otras áreas de aplicación. En todas existe la posibilidad del mal uso de la IA, en el desarrollo de sistemas de armas autónomas y el posible uso de la información que recoge esta tecnología. Lo cual da paso a pensar en la regulación de estas amenazas para crear políticas que puedan prevenir o mitigar su impacto negativo. Especial preocupación generan la protección de datos, la creación de parámetros éticos, la creación de robots autónomos, el transporte autónomo, los sistemas de inteligencia artificial y el desarrollo de sistemas de armas autónomas letales (LAWS), por sus siglas en inglés, algunas de ellas ya operativas.

Como cualquier tecnología, la Inteligencia Artificial se puede usar para el beneficio de la humanidad o para causar daño. La gobernanza adecuada y consensuada es fundamental para garantizar que se implemente de manera responsable y verificable. Toda organización, agencia, compañía, etc.; que la utilice y desarrolle, tiene que dominar lo que esta tecnología puede y no puede hacer. Así como los *know how* para evitar la violación accidental o intencionada de las reglamentaciones que se adopten con respecto a su uso ético. Tener conciencia de los posibles riesgos y beneficios, es el primer eslabón en la

asunción de una ética en la ciencia que investiga, produce y aplica la Inteligencia Artificial. Es un desafío paradigmático del siglo XXI.

Conclusión

La postura que prevalece en la literatura que aborda los problemas del manejo de la Inteligencia Artificial muestra un consenso en cuanto a la necesidad de la unificación de las iniciativas para su uso ético y verificable. Esto evidencia que existe una conciencia de los riesgos potenciales que tiene su desarrollo sin un control consensuado.

Sin embargo, se aprecia asimismo un énfasis en los beneficios de la inteligencia artificial en ámbitos como la educación, la salud, las finanzas y el internet de las cosas, entre otros, que pueden retrasar la adopción de una política mancomunada a nivel global para la implementación de su uso ético y controlado.

Si bien es cierto que los dilemas que hoy plantea la aplicación de la inteligencia artificial y los hechos dañinos que ya ha producido están identificados, descritos y pronosticados en la literatura, prima el entusiasmo por los beneficios de esta tecnología, lo cual, es un obstáculo para la adopción de un consenso internacional que produzca un equilibrio entre potenciación y limitación del uso, cumpliendo con la responsabilidad ética de la ciencia.

Referencias

- Bostrom, N. (2016). Why I Want to be a Posthuman When I Grow Up. En Gordijn, B. y Chadwick, B. (Eds.). *Medical Enhancement and Posthumanity* (pp. 107-137). Springer.
- Bunge, M. (2002). *Crisis y reconstrucción de la filosofía*. Gedisa.
- Comité Selecto de Inteligencia Artificial. (2019). *Plan estratégico nacional de investigación y desarrollo de la inteligencia artificial: actualización 2019*. Consejo Nacional de Ciencia y Tecnología.
- Cubadebate. (2020, 12 de junio). Irán confirma uso de inteligencia artificial para asesinar a científico nuclear. *Cubadebate*. <http://www.cubadebate.cu/noticias/2020/12/06/iran-confirma-uso-de-inteligencia-artificial-para-asesinar-a-cientifico-nuclear/>
- Forbus, K. D. (2012). How minds will be built. *Advances in Cognitive Systems*, 1(3), pp. 47-58.
- Herranz, A. (2021, 2 de febrero). *El Foro Económico Mundial quiere unificar todas las iniciativas éticas de la Inteligencia Artificial en Global AI Action Alliance (GAIA)*. Xataka.

- [https://www.xataka.com/temas/pro/digitalizacion/el-foro-económico-mundial-quiere-unificar-todas-las-iniciativas-éticas-de-la-inteligencia-artificial-en-global-ai-action-alliance-\(gaia\)/](https://www.xataka.com/temas/pro/digitalizacion/el-foro-económico-mundial-quiere-unificar-todas-las-iniciativas-éticas-de-la-inteligencia-artificial-en-global-ai-action-alliance-(gaia)/)
- Jaimovich, D. (2020, junio 1). El uso de robots y otras formas de inteligencia artificial en tiempos de coronavirus. *Infobae*. <https://www.infobae.com/america/tecno/2020/06/01/el-uso-de-robots-y-otras-formas-de-inteligencia-artificial-en-tiempos-de-coronavirus/>
- Kurzweil, R. (2017). *The Singularity is near*. Penguin Group.
- Parlamento Europeo. (2020). A comprehensive European industrial policy on Artificial Intelligence and robotics [Archivo PDF]. Document A8-0019/2019. https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html
- Pastorino, C. y Giusto, D. (2021, febrero 18). Teledildonics. Boom de los juguetes sexuales inteligentes: advierten por el riesgo de ciberataques. *Clarín*. https://www.clarin.com/tecnologia/boom-juguetes-sexuales-inteligentes-advierten-riesgo-ciberataques_0_gy3MxGXTJ.html
- Ponce, J.C., Torres, A., Quezada, F.S., Silva, A., Martínez, E.U., Casali, A., Scheihing, E., Túpac, Y.J., Torres, M.D., Ornelas, F.J., Hernández, J.A., Zavala, C., Vakhnia, N. y Pedreño, O. (2014). *Inteligencia Artificial*. Latin.
- Prang, A., Tweh, B. (2019, agosto 30). CEO Jack Dorsey's Twitter account was hacked, Company says. *Wall Street Journal*. <https://www.wsj.com/articles/twitter-says-ceo-jack-dorseys-twitter-account-was-compromised-11567197298>
- Rosa, A. (2019, 4 de julio). What is DL? A Look at Its Definition, Benefits, and Challenges. *UpWork*. <https://www.upwork.com/hiring/for-clients/deep-learning-definition-benefits-challenges/>
- Russell, S.J. y Norvig, P. (2009). *Artificial Intelligence. A modern approach*. Prentice hall.
- Samsung. (2021, 12 de enero). Samsung presenta sus innovaciones en el CES 2021 para una mejor normalidad. *News Samsung*. <https://news.samsung.com/latin/ces2021/samsung-presenta-sus-innovaciones-en-el-CES-2021-para-una-mejor-normalidad>
- Whittle, J. (2021, febrero 10). AI can now learn to manipulate human behaviour. *The Conversation*. <https://theconversation.com/ai-can-now-learn-to-manipulate-human-behaviour-155031>
- World Commission on the Ethics of Scientific Knowledge and Technology. (2017). *Report of COMEST on robotics ethics*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.

Fuentes de financiamiento

Autofinanciado.

Conflictos de interés

No presenta conflicto de interés.

Correspondencia:

clnodals@ucf.edu.cu