¿Pueden Discriminar las Máquinas?

Can Machines Discriminate?

Submitted: 14 April 2023
Reviewed: 20 June 2023
Revised: 10 July 2023
Accepted: 11 July 2023

Maria Karolina Urbano[*]
https://orcid.org/0000-0001-5274-3824
Lainiver Mendoza Munar[**]
https://orcid.org/0000-0003-3750-3261

Article submitted to peer blind review

DOI: https://doi.org/10.26512/lstr.v16i1.46871

Abstract

[Proposal] *To take a first step towards the demand for responsible technological development, since like any tool created by the human being, it can be used for the benefit of humanity or against it.*

[Methodology/Approach/Design] *In the initial stage of the research project titled "Discrimination and Ethics of Artificial Intelligence," a comprehensive review of literature and documents in the field of philosophy of mind and cognitive science during the latter half of the 20th century was conducted. This review aimed to establish connections with contemporary advancements in the ethics of artificial intelligence, particularly following the establishment of ethical principles by the European Community in 2019. The disciplines of philosophy of mind and cognitive science, being closely linked to neurobiology and artificial intelligence, were pioneers in recognizing the challenges associated with the development of this technology. However, at that time, artificial intelligence applications had not permeated all aspects of human activities as extensively as they do today. Consequently, the problems surrounding this technology were more speculative rather than urgent concerns. The rapid integration of artificial intelligence into various domains has presented society, governments, and companies with pressing issues that demand immediate attention. Given the recent nature of these advancements, adequate preparedness to address the associated challenges remains lacking. While ethics may lie outside the realm of scientific inquiry as it does not fall within its scope or methods, science plays a crucial role in the subject matter of ethics, as the tools created by humans can be employed for both harm and societal well-being. Once these tools*

[*]Licenciada en Filosofía, Magíster en Filosofía y profesora investigadora de la Universidad Cooperativa de Colombia, Sede Cali. Dirección: Universidad Cooperativa de Colombia, Sede Cali: Cra. 73 #2a-80, Barrio: Buenos Aires, Cali, Valle del Cauca. E-mail: maria.urbanog@campusucc.edu.co.
[**]Abogada, especialista en docencia universitaria, magister en derecho, candidata a doctorado en derecho por la Universidad Carlos III de Madrid y profesora investigadora de la Universidad Cooperativa de Colombia, Sede Cali. E-mail: lainiver.mendoza@campusucc.edu.co.

begin to generate adverse consequences for individuals and society, ethical and legal considerations become paramount. The resurgence of interest in the philosophy of mind stems from the realization that ethical dilemmas associated with artificial intelligence are rooted in long-standing ethical problems. We continue to seek improved responses to such dilemmas and grapple with questions that arise from indeterminate concepts like equality. It is important to recognize that discrimination, by definition, goes against the principle of equality, and there are instances where machines contribute to discriminatory actions or situations. It is worth noting that addressing the questions posed in this article, within the context of a research project focused on discrimination, necessitates an interdisciplinary analysis. This further justifies the reliance on the disciplines of philosophy of mind and cognitive science, which approach the challenges of artificial intelligence from scientific, technological, linguistic, and philosophical perspectives, extending their reach into the legal domain as well.

[Purpose] *This article aims to emphasize the significance of conducting responsible technological advancements, particularly in the realm of artificial intelligence. It underscores the necessity for AI-based technologies to be designed with careful consideration of ethical and moral principles, ensuring that their impact aligns with societal norms and upholds the recognition of human rights.*

[Practical Implications] *As Artificial Intelligence continues to permeate numerous facets of society, it is crucial to acknowledge the inherent risks associated with its usage, particularly in relation to discrimination. Consequently, it becomes imperative to implement measures that effectively mitigate these risks and reduce the potential for discriminatory outcomes.*

Keywords: *Artificial intelligence. Discrimination. Information technologies. Ethical framework.*

Resumen

[Propuesta] La importancia de este documento radica en que constituye un primer paso hacia la exigencia de un desarrollo tecnológico responsable, pues como toda herramienta creada por el ser humano, puede ser usada en beneficio de la humanidad o en contra de ella.

[Metodología/Enfoque/Diseño] En el desarrollo de este primer avance del proyecto de investigación "Discriminación y ética de la inteligencia artificial"[1], se hizo una revisión bibliográfica y documental de la filosofía de la mente y la ciencia cognitiva de la segunda mitad del siglo XX, para conectar con los avances que en materia de la ética de la inteligencia artificial se tienen hoy, especialmente a partir de la creación de los principios éticos establecidos en 2019 por la Comunidad Europea. Como ya se mencionó, la filosofía de la mente y la ciencia cognitiva, en tanto ramas interdisciplinarias y

[1]El presente artículo es resultado del desarrollo del proyecto de investigación titulado: " DISCRIMINACIÓN Y ÉTICA DE LA INTELIGENCIA ARTIFICIAL", investigadora principal: Maria Karolina Urbano, Coinvestigadora: Lainiver Mendoza Munar. Universidad Cooperativa de Colombia, Sede Cali.

directamente relacionadas con la neurobiología y la inteligencia artificial, fueron pioneras en la visualización de los problemas alrededor del desarrollo de este tipo de tecnología. Pero en ese entonces, los productos con inteligencia artificial no estaban incorporados a todas las actividades del ser humano como vemos ahora, en donde los problemas no son especulaciones, sino aspectos urgentes a resolver, dado que por ser tan recientes ni los gobiernos, ni las empresas, ni la sociedad están preparados para atenderlos con suficiencia. Desconocemos gran parte del impacto que este tipo de tecnología pueda provocar en el mundo y en la sociedad. Si bien la ética puede ser un tema ajeno al mundo de la ciencia, porque no hace parte de su objeto de estudio y mucho menos de su método, el uso de la ciencia sí está en el objeto de estudio de la ética, puesto que las herramientas creadas por el hombre pueden ser usadas para hacer el daño o para procurar bienestar en la sociedad. Una vez la herramienta empieza a generar consecuencias negativas en los seres humanos y la sociedad, el tema ético y jurídico se ponen a la orden del día. La necesidad de retomar la filosofía de la mente se debe a que, tal como menciona Juan Ignacio del Valle (2019), los problemas de la inteligencia artificial heredan viejos problemas éticos, para los cuales se siguen buscando mejores respuestas, como es el caso de los dilemas, pero también de aquellas cuestiones que parten de conceptos indeterminados como el de la igualdad. Recordemos que la discriminación es por definición ausencia del principio de igualdad y existen máquinas que provocan acciones o situaciones discriminatorias. Es importante anotar, que las preguntas que este artículo intenta resolver como parte de un proyecto investigativo alrededor del problema de la discriminación, exige un análisis interdisciplinario, una razón más para apoyarnos en la filosofía de la mente y la ciencia cognitiva como disciplinas que abordan el problema de la inteligencia artificial desde la visión científica, tecnológica, lingüística, filosófica, que hacemos extensiva al ámbito jurídico.

[Finalidad] Este artículo busca resaltar la importancia de llevar a cabo desarrollos tecnológicos responsables, pues las tecnologías que cuentan con inteligencia artificial deben ser concebidas para que sus efectos se materialicen en un marco ético y moral, teniendo en cuenta el contexto de las sociedades y el reconocimiento de los derechos humanos.

[Implicaciones Prácticas] La Inteligencia Artificial, cada vez, permea diversos aspectos de las sociedades, sin embargo, se hace necesario los riesgos que implica su utilización en materia de discriminación, para implementar medidas que reduzcan este riesgo.

Palabras Clave: Inteligencia Artificial. Discriminación. Tecnologías de la Información. Marco Ético.

INTRODUCTION

En 1947, Alan Turing (2010) formuló una inquietante pregunta para la época, en el ensayo Computing Machinery and Inteligence: "¿pueden pensar las máquinas?", pregunta que inspira el título de este texto. Indagar sobre la posibilidad de pensamiento y capacidad para discriminar de las máquinas es, sin duda, inquietante, pero no mucho menos misteriosa de lo que podía ser en tiempos de Turing. Es provocadora porque el

sentido común, el cual lleva a pensar que no, dado que las máquinas son producto del ingenio humano, no seres humanos, sin embargo, se encuentra en la literatura de la ética de la inteligencia artificial o el derecho informático que algunos sistemas que utilizan esta tecnología producen situaciones que generan discriminación. Dado que las consecuencias del desarrollo vertiginoso de esta tecnología son tan recientes, el lenguaje que usamos para las máquinas es el que usamos para la conducta y naturaleza humanas: "pensar", "discriminar", "tomar decisiones", "aprender", entre otras. Por ende, es hora de pensar si esto es correcto, cómo se debe entender este tipo de lenguaje en las máquinas.

En los años 80 y 90, la filosofía de la mente y en general lo que se conoce como ciencia cognitiva, se ocuparon extensamente del tema de la inteligencia de las máquinas[2], ya que era el sueño cumplido de la neurobiología, pues, una vez se conozca exactamente cómo funciona el cerebro humano, lo que resta es replicar el funcionamiento en materiales artificiales (silicio) y así construir replicas humanas. A pesar del enorme conocimiento que se tiene del funcionamiento del cerebro humano, las réplicas humanas no resultaron tan sencillas de lograr y entre tanto surgió otra transformación tecnológica que impactó de manera significativa al mundo: el fenómeno de las redes sociales. Desde esta situación, el interés por lo que pasaba al interior del desarrollo tecnológico fue desplazado por lo que producía en la sociedad, en el fenómeno de masas que permitió la interconexión de personas a través de medios virtuales. Así se empezaron a incorporar la inteligencia artificial en nuestra vida cotidiana sin ser muy conscientes de lo que esto implicaba a nivel tecnológico.

En los últimos años, el desarrollo de la inteligencia artificial ha seguido en su crecimiento vertiginoso, lo cual ha dado origen a sistemas autónomos que van más allá del funcionamiento tradicional de un algoritmo. La idea clásica de que los computadores no pueden pensar dado que hace cálculos a partir de un conjunto de órdenes (inputs) y la emisión de respuestas (outputs), se de debe replantear con los sistemas autónomos conocidos como Maching Learning, sistemas creados a partir de la inteligencia artificial en donde las respuestas no necesariamente están en el algoritmo, dado que la máquina es creada de tal manera que pueda escalar "aprendiendo" del ambiente y "tomando decisiones" autónomas. La cercanía entre el humano y la máquina parece muy cercana y las preguntas se multiplican ¿pueden las máquinas pensar, aprender y tomar decisiones? Esto nos lleva a un asunto más complejo que el simple hecho de hacer cálculos, puesto que hay una zona gris en la responsabilidad de los cálculos que hace el sistema.

No obstante, con este artículo se pretende mostrar que aún si las máquinas pueden algún día simular completamente a seres humanos, estos seres carecerán de las capacidades fundamentales del hombre, entre ellas la autoconciencia, la capacidad de dar sentido, las experiencias en primera persona, la intencionalidad, rasgos naturales del ser

[2]Por máquinas entendemos los computadores y sistemas que utilicen este tipo de programación.

humano que dan origen a aspectos sociales de primer orden en el ámbito moral, social y jurídico: la responsabilidad y la imputabilidad.

REDUCCIONISMO Y NO REDUCCIONISMO

Debemos empezar, entonces, por lo que significa la inteligencia artificial (desde ahora IA). Si por inteligencia entendemos "hacer cálculos", entonces las máquinas sí tendrán inteligencia desde esta óptica. Sí, por el contrario, concebimos la inteligencia como la capacidad de razonar, tener experiencias conscientes y autoconscientes como fenómenos intencionales, entonces no, las máquinas no pueden pensar, ni siquiera en el caso de los sistemas autónomos, *robots o cyborgs*. Sería una concepción de inteligencia exclusiva de los seres humanos.

Esto, sin embargo, no es tan obvio, en la ciencia nada es obvio ni de sentido común. El pensamiento desarrollado en los años 80 y 90 especialmente por científicos y filósofos norteamericanos se da a través de dos posturas enfrentadas: una postura reduccionista que considera que el cerebro no es más que un programa muy sofisticado de computación y que todos los *qualia*, los fenómenos subjetivos de primera persona que experimentamos como sentimientos, pensamientos, etc., no son más que cómputos complejos procesados por el cerebro, se reducen a billones de sinapsis interconectadas que producen todos estos fenómenos en milésimas de segundos. Representantes de este reduccionismo son los filósofos Daniel Dennett (1995), Paul y Patricia Churchland (Churchland, 1992), científicos como Francis Crick y David Chalmers (Searle, El misterio de la conciencia, 2000), entre otros. De otro lado, está la postura no reduccionista, en la cual, se defiende la conciencia y la intencionalidad como rasgos de los seres humanos que no pueden reducirse porque es precisamente la experiencia interna, en primera persona lo que proporciona su característica esencial: el sentido. No experimentamos sinapsis cuando tenemos pensamientos (y solo pensamos cuando somos conscientes), experimentamos un lingüístico conjunto de ideas, de signos llenos de contenido. Algunos filósofos como John R. Searle (1996), (1997), Garret Thompson o John Kearns (Botero et al., 2000) defienden esta postura y es el soporte teórico de la pregunta que se pretende responder en este artículo. Para Searle, las máquinas responden "como si" fueran seres humanos, pero no lo son, porque sus respuestas carecen de intencionalidad, solo hacen una combinación de signos, como se puede hacer con cualquier otro sistema de signos: la base lingüística de los humanos, el sistema binario o el código Morse. La diferencia es que solo para los humanos esos resultados tienen significado porque solo ellos pueden dotarlos de sentido (Urbano, 2016). Este argumento expuesto por Searle se conoce como "la habitación china" (2000), un experimento mental con el que se pretende refutar la tesis que resulta del "*Test de Turing*", la cual defiende que el pensamiento no es más que un conjunto de cálculos. El problema, entonces, pasa de un nivel neuronal y biológico a uno, pues se trata ya no solo de los cómputos como

cálculos y conjunto de signos que tienen una estructura sintáctica, sino de la semántica de estos cómputos a la que escapa la máquina.

EL LENGUAJE CIENTÍFICO, EL LENGUAJE SOCIAL Y EL LENGUAJE DE LAS MÁQUINAS

Con lo expuesto hasta el momento podemos identificar el lenguaje de la ciencia y el de las máquinas, es importante también indicar que el lenguaje científico es aquel que utiliza la ciencia natural para explicar fenómenos naturales, como el funcionamiento del cerebro. El lenguaje de las máquinas es aquel con el que se construyen los programas de computador, por ejemplo, el sistema binario es un lenguaje o código que manipula signos como el 0 y el 1. Este sistema así presentado corresponde a lo que conocemos como sintaxis, ambos lenguajes, el de la máquina y el de la ciencia, dependen o se derivan de un lenguaje que surge en el ámbito social: el lenguaje natural, que si bien, no es natural en sentido estricto, es una construcción social producto de la capacidad lingüística humana.

Es importante hacer énfasis en que esto es lo que claramente nos distingue de las máquinas y de los otros seres de la naturaleza y es el argumento más fuerte para defender la postura no reduccionista, pues el lenguaje de la máquina carece de semántica, tal como lo señala Kearns: "Las máquinas no actúan propositivamente. En cambio, los actos lingüísticos, que son portadores primarios de significación, son intencionales y propositivos" (2000, p.152), de acuerdo con este filósofo, las relaciones entre las características semánticas y sintácticas en el lenguaje natural son convencionales y contingentes, por lo tanto, no puede haber una correspondencia entre la sintaxis y la semántica, pues una misma estructura sintáctica puede contener diferencias semánticas dependiendo del contexto.

Ahora bien, esta diferencia en el lenguaje nos lleva a otra tesis del pensamiento de Searle, la tesis desarrollada en La construcción de la realidad social (1997), si bien podemos tener intuitivamente una idea de realidad unitaria, esto es, de una sola realidad, no la podemos entender desde un único lenguaje. Searle traslada su postura antirreduccionista a todas las esferas del conocimiento. Así como no podemos reducir fenómenos mentales (la conciencia, intención, etc.) a fenómenos físicos (sinapsis, funcionamientos cerebrales), tampoco podemos reducir los fenómenos sociales a fenómenos físicos sin que estos pierdan sentido. Para Searle entender una realidad o la realidad, implica una comprensión desde al menos tres ámbitos diferentes: el físico, el mental y el social. Cada uno con un lenguaje que le es propio y que solo puede entenderse de manera adecuada desde ese lenguaje, lo cual es indispensable para entender las consecuencias de defender una postura reduccionista para los productos creados desde la IA en el ámbito social y jurídico, pues es en estos donde la agencia, la voluntad, la conciencia, la razón y la moral, entre otros, son condiciones sin las cuales pueden existir la responsabilidad y la imputabilidad. Justamente es la ausencia de ellas, la naturalización

de las acciones por enfermedad o minoría de edad, lo que hace que una persona pierda la capacidad de actuar bajo su responsabilidad y por lo tanto de ser imputado.

Desde esta perspectiva elegir entre una postura reduccionista o antirreduccionista, nos llevaría a las siguientes consecuencias: la primera, nos lleva a aceptar que las máquinas piensan y, en la medida que simulen bien a los seres humanos, pues tienen las mismas capacidades de los seres humanos, es decir, podríamos decir que, si hay un error de la máquina, parte de la responsabilidad es de la máquina. Desde la segunda postura, la responsabilidad solo puede ser humana, es decir, solo puede estar detrás de la persona que crea o la entidad a cargo de la máquina, pues la máquina, al carecer de conciencia e intención, solo debe concebirse como una herramienta.

Por ejemplo, los perros policía son animales domesticados y adiestrados para actividades especiales como la detección de droga en los aeropuertos, pues gracias a que tienen un olfato más desarrollado que el del ser humano, pueden percibir el olor de la sustancia para la que fue entrenado. Sin embargo, aun así es poco probable que suceda, en algún momento puede atacar a un ser humano sin que reciba esta instrucción, en el caso hipotético que al policía encargado del perro se le olvide ponerle bozal y el can ataque a un niño que pasa cerca de él, es importante, tener en cuenta que la responsabilidad por los daños causados por el ejemplar canino será atribuida a la institución a la que pertenece, para efectos de reparación del daño y reconocimiento de indemnizaciones si diere lugar a ello.

Así como los perros tienen algunos sentidos más desarrollados que los de los seres humanos. Por otro lado, se puede indicar que las computadoras y todos los sistemas creados con IA pueden hacer cómputos a una velocidad, complejidad y capacidad de almacenamiento inalcanzables para los seres humanos, per se. Sin embargo, no por esto, las máquinas adquieren la capacidad de tener experiencias subjetivas, ser agentes a la hora de "actuar", es decir, hacerlo con una motivación e intención o tener grados de conciencia o autoconsciencia. Penalizar o responsabilizar a una máquina por el resultado del funcionamiento de su sistema, resulta tan representativo como hacerlo para los animales no humanos.

En este sentido, desde la perspectiva social y jurídica, la importancia de la irreductibilidad de su lenguaje a uno natural (o científico) resulta ineludible para la legitimación de prácticas sociales, entre ellas, la normativa. No hacerlo, sería una buena manera de evadir responsabilidades en la creación de productos tecnológicos, un asunto del que se ocupa la ética de la inteligencia artificial.

UNA APROXIMACIÓN A LA ÉTICA DE LA IA

Con el desarrollo de la inteligencia artificial y especialmente de los sistemas autónomos o modelos basados en *machine learning*, se puede afirmar que algunas máquinas tienen la capacidad de tomar decisiones, por lo cual surge la pregunta ¿son las máquinas responsables de los resultados de estas decisiones? El problema ontológico se

convierte, entonces en un problema ético y no menos importante, puesto que algunos productos diseñados con esta tecnología generan escenarios donde es evidente la discriminación de raza, sexo o condición económica.

El impacto generado por la inteligencia artificial en la sociedad ha llevado a que los mismos creadores y diseñadores, como Tristán Harris, Sandy Parakilas o Guillaume Chaslot (Orlowski, 2020), así como la socióloga y filósofa Shoshana Zuboff, planteen la necesidad de un desarrollo tecnológico responsable y de un marco normativo que regule dicha actividad. En 2019, la Comunidad Europea establece algunos principios éticos de la inteligencia artificial entre los cuales está el principio de diversidad, no discriminación y equidad, cabe anotar que el derecho a la no discriminación es un derecho fundamental, por lo cual el discurso encaminado a la protección y defensa de los derechos humanos debe incluir ahora esta vertiente que surge del desarrollo tecnológico.

Los principios éticos para el desarrollo de la IA son propuestos en el documento *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* creado por el grupo *High-Level Expert Group on Artificial Intelligence*, el ocho de abril de 2019 en Bruselas, puesto que los países más desarrollados son los que han visto con mayor fuerza el impacto de la tecnología en la sociedad y los problemas sociales, jurídicos y de gobernanza que generan. La importancia de este documento radica en que constituye un primer paso hacia la exigencia de un desarrollo tecnológico responsable, pues como toda herramienta creada por el ser humano, puede ser usada en beneficio de la humanidad o en contra de ella.

En este documento se relacionan los siguientes principios éticos para la inteligencia artificial: 1. *Human agency and oversight*, 2. *Technical robustness and safety*, 3. *Privacy and data governance*, 4. *Transparency*, 5. *Diversity, non-discrimination and fairness*, 6. *Societal and environmental well-being* y 7. *Accountability* (Comisión Europea, 2019). A pesar de la importancia que este documento tiene, no establece definiciones o explicaciones concretas sobre cómo se deben entender dichos valores. Lo cual no es problema menor, puesto que conceptos como agencia, equidad, discriminación, diversidad, bienestar, por mencionar algunos, son conceptos indeterminados que pueden ser interpretados de muchas maneras. De hecho, en este momento no hay un documento oficial por algún organismo competente que haya realizado una traducción al español de estos principios y, dado que no hay mucha literatura al respecto en este idioma, la traducción que se ha hecho de estos principios varía de un autor a otro, de ahí que en esta presentación se prefiera una traducción quizás más literal a la que ofrece José Ignacio Latorre (2019), quien los traduce así: "1. Dignidad humana, 2. Autonomía, 3. Responsabilidad, 4. Justicia, equidad y solidaridad, 5. Democracia, 6. Estado de derecho, 7. Seguridad e integridad corporal y mental, 8. Protección de datos y privacidad, 9. Sostenibilidad" (p.303). Como se puede ver, en el cuarto principio, el autor subsume la diversidad, la no discriminación y la igualdad en los términos de "justicia, equidad y

solidaridad", lo cual es bastante cuestionable, puesto que la diversidad y la no discriminación, si bien son una extensión del principio de igualdad, merecen una mención explícita, en tanto no han sido incorporadas históricamente al principio de igualdad.

En el caso del fenómeno de la discriminación, lo que se ha notado en los modelos conocidos como *machinne learning*, es que constituyen espacios para la generación de discriminación, toda vez que refuerzan estereotipos sociales que se provienen de los sesgos de los diseñadores o creadores del modelo, por lo cual hay agencia humana y responsabilidad humana, pero una vez creados estos modelos pueden "tomar decisiones autónomas" en la medida que "aprenden" reproduciendo el sesgo que tienen incorporado en sus datos. Es importante recordar el poder que tiene la inteligencia artificial de almacenar, hacer cálculos y distribuir información en el menor tiempo posible, diluyendo la responsabilidad e invisibilizando nuevamente el problema.

Desde esta perspectiva el impacto negativo en la lucha contra la discriminación es significativo y muestra, una vez más, que la discriminación es un problema estructural, que requiere ser analizado a través del concepto de discriminación estructural (de lo cual presente texto no se ocupa).

El análisis de la relación entre el concepto de discriminación estructural y los principios éticos de la inteligencia artificial, atraviesan problemas y aspectos que requieren ser atendidos por su necesidad y urgencia, pero que representan un nuevo reto para los seres humanos, así como un campo fértil y prolífico en términos de investigación.

De acuerdo con Saiph Savage (Epic Queen, 2020), directora del laboratorio de *Human Computer Interaction* de la Universidad de West Virginia y codirectora del laboratorio de Innovación Cívica de la UNAM, es imperativo hacer máquinas que piensen de manera ética y responsable, lo que significa que sus creadores y diseñadores piensen de esta manera, pues ellos son los que agregan la información, los datos, que serán utilizados por la máquina.

La IA es una clase de tecnología que se usa para que las computadoras realicen actividades que antes solo hacían los seres humanos, y dentro de esta clase de tecnología se encuentra una subárea denominada *machine learning* que consiste en la creación de modelos con datos específicos con los cuales la máquina puede aprender y tomar decisiones. Para Tom Mitchell (1997, citado por Savage, 2020) *machine learning* consiste en una máquina que es capaz de mejorar su comportamiento, basándose en la experiencia que va adquiriendo, sin necesidad de un algoritmo que le determine qué hacer. Esta dinámica es la que genera problemas éticos.

Así lo explica Grecia Macías, abogada de R3D: Red en Defensa de los Derechos Digitales, para quien el mito de la objetividad de las máquinas que no se equivocan, lleva a la idea, igualmente errónea, de que las máquinas son neutrales. Por el contrario, las máquinas, los sistemas autónomos están llenos de sesgos: "no hay bases de datos sin sesgos" (Epic Queen, 2020), puesto que son humanos los que crean el modelo bajo el

cual la máquina aprende, y estos modelos son escenarios propicios para reforzar estereotipos sociales. Los ejemplos que ofrece la abogada mexicana son claros: una aplicación de Amazon que identificó a 23 congresistas de los E.E.U.U., como exconvictos por su color de piel y rasgos no caucásicos, el lavamanos que arroja agua y jabón cuando detecta manos, pero no reconoce manos de piel oscura. Para Olga Fernández (2020): "Grandes empresas como Google y su buscador de imágenes o Youtube, se han enfrentado a importantes polémicas. Si buscamos términos como cocina en Google, solo aparecen imágenes de mujeres, pero si buscamos CEO solo aparecen hombres". Esto sin contar las campañas de desinformación, las *fakes news* y otras prácticas, al alcance de todos, que impiden cambiar los estereotipos sociales. Dice Macías: "si metes información sesgada, solo va a salir información sesgada (…) y esto perpetúa la desigualdad estructural".

Esto significa que antes de plantearnos si las máquinas pueden llegar a una autonomía completa o no, necesitamos pensar lo que sucede con lo que ya es un hecho: los modelos autónomos son construidos por seres humanos y en estos se incorporan datos que pueden ser sesgados por la ideología de quien los crea, pero también de acuerdo con unos criterios que se le pide al programador agregar a un modelo específico. En este sentido, las críticas se enfocan en los programas que reciben y clasifican las hojas de vida en las empresas. Una empresa puede crear (es lo que se estima que hacen) perfiles de tal manera que el modelo excluya candidatos por sexo, lugar de origen o raza (podría haber otras exclusiones de acuerdo con los deseos de la empresa, como la edad, por ejemplo). En este tipo de procedimientos la exclusión y discriminación puede llevarse a cabo de la manera más sutil y soslayada, pero reforzando estereotipos que es urgente superar. Para Grecia Macías (Epic Queen, 2020), una posible acción para disminuir y evitar sesgos consiste en la exigencia de un equipo, interdisciplinario y ecléctico en la creación de modelos que acuerden y aprueben colectivamente los datos con los cuales se va a crear el modelo. Esta solución suena razonable, pero uno de los problemas que encontramos es la falta de regulación y control en la creación de productos tecnológicos, algo fascinante para las empresas a cargo porque siguen generando riquezas sin directrices legales y sin medir los peligros que esta pueda tener sobre la sociedad y el ser humano.

CÓMO ELIMINAR LOS SESGOS EN LA IA

Detrás de los sesgos en los programas de IA, entonces, están las personas que crean estos modelos y es ahí, donde, siguiendo las ideas de la Dra. Macías, se debe poner la atención, pues es ahí donde se generan las experiencias de discriminación y el reforzamiento de estereotipos que conducen a la discriminación: el estereotipo de la mujer delicada y sumisa dedicada a labores del hogar, por ejemplo, los estereotipos de que las personas afrodescendientes y de bajos recursos tienden a la delincuencia o solo tienen ciertas habilidades (deportes, música, danza o trabajos que requieran mucha fuerza física). Esto es algo que podría quedar en especulación de no ser porque una de las

programadoras de Media Lab en MIT sufrió directamente una experiencia de discriminación, dado que, por su color de piel, los softwares de reconocimiento facial no funcionaban en ella, es decir, no la reconocían. Se trata de Joy Buolamwini, una joven informática canadiense de ascendencia ghanesa, quien se ha convertido en una activista de la creación de tecnología responsable y creadora de la Liga de la justicia algorítmica, un movimiento que, por un lado, busca denunciar o poner en evidencia los sistemas que producen experiencia de discriminación y de otro, promover los principios que deben regir la creación de modelos informáticos con el fin de evitar dichos efectos.

Para Buolamwini (2017), los sesgos algorítmicos son tan injustos como los de los humanos, precisamente porque provienen de los seres humanos, de sus creadores, con la diferencia, de que a través de los algoritmos estas injusticias se pueden propagar a gran escala a un ritmo acelerado, verbigracia: hoy se toman muchas decisiones con base en la información suministrada por la IA: la elección de un candidato a un puesto de trabajo, los estudiantes que entran a una universidad. La aplicación de la ley también ha empezado a usar este tipo de tecnología, de hecho, en E.U. "algunos jueces usan puntajes de riesgo generados por máquinas para determinar cuánto tiempo un individuo permanecerá en prisión", desde esta perspectiva, para Buolamwini, si se tiene en cuenta el sesgo que lleva consigo cada algoritmo se debe pensar seriamente el impacto social que este tiene. Por ello propone tres principios que deben regir a la hora de crear algoritmos: ¿Quién lo crea?, ¿cómo lo crea?, ¿para qué lo crea? Estas preguntas deben ir encaminadas a develar si hay un producto tecnológico responsable y con sentido de equidad. Aunque ella no lo menciona, de alguna manera está recurriendo a los principios de la ética de la IA: tecnología creada para el bienestar humano, que no sea discriminatoria y con equidad.

Los sesgos en los algoritmos, entonces, es un serio problema por resolver, tal como lo plantea Buolamwini, por lo cual ella creó la Liga como un medio para identificación de sesgos en las plataformas existentes, para auditar el software existente y para crear grupos más inclusivos a la hora de crear algoritmos. En la plataforma de la liga de la justicia algorítmica "todo aquel que se preocupe por la equidad, puede ayudar a combatir la mirada codificada. En codegaze.com pueden informar sesgos, solicitar auditorias, convertirse en un betatesters y unirse a la conversación en curso a través de #codegaze".

LA RESPONSABILIDAD DE LA IA

En este apartado solo vamos a analizar el ejemplo de los carros autónomos. Esto ya sucedió en España (Epic Queen, 2020), un Uber con sistema automático va dirigiendo el carro, también va un piloto humano, el sistema no detecta a la persona que va cruzando la calle, el piloto tampoco se da cuenta con tiempo suficiente que el sistema automático ha fallado y la persona es atropellada por el vehículo. En un primer momento podemos pensar que la responsabilidad es de la empresa Uber que a su vez desplazará la responsabilidad a los creadores del algoritmo. Sin embargo, se contempla otra

posibilidad, crear un protocolo que dote al carro autónomo de responsabilidad civil dándole carta de identidad legal (Latorre, 2019), recordemos que en Arabía Saudita se le otorgó ciudadanía a robot Sophía, lo cual no solo es polémico, sino que no hay consenso en que este método sea correcto.

Los peligros de la IA empiezan a aflorar, tal como señal Latorre:

> "Al atribuir a un coche responsabilidad civil, se está propiciando que las empresas creadoras de robots se desentiendan de la actuación de sus productos. Esta figura legal, de alguna manera, da cabida a la exculpación de todo error de construcción o diseño algorítmico por parte de los fabricantes. Otra crítica razonable es que este tipo de legislación abre un mundo nuevo que hemos de analizar a fondo y con calma. No es bueno precipitarse, como tampoco lo es no avanzar en este frente." (2019, p.223)

La responsabilidad civil que pueda o no otorgarse a las máquinas debe contemplar, en primer lugar, la naturaleza de estas entidades, tal como lo hemos hecho desde la filosofía de la mente, pues no se debe perder la perspectiva de, ante todo, son herramientas, tan complejas que pueden imitar a un ser humano, pero no son seres humanos. La ciencia ficción ha creado el imaginario de que algún día seremos dominados por las máquinas. Hoy estamos convencidos de que eso es completamente posible. Pero antes de que esto se convierta en realidad es ineludible seguirnos ocupando del daño que el ser humano es capaz de generar a los miembros de su propia especie. Bajo el manto de la simulación casi perfecta de humanos, estos productos siguen siendo lo que en principio son: herramientas generadoras de servicios que a su vez producen gran cantidad de dinero, y ese dinero no llega a la máquina sino a sus creadores, aún si la máquina tuviera una retribución económica y pagara impuestos (algo que de alguna manera hace una persona jurídica), ese dinero generado solo tiene sentido para los seres humanos, especialmente para los dueños de ese producto. Desviar este sentido, puede ocasionar grandes posibilidades para la impunidad y para otras conductas injustas como el de la discriminación, tal como se ha señalado a lo largo del texto.

INTELIGENCIA ARTIFICIAL Y DISCRIMINACIÓN

Teniendo en cuenta que el concepto de inteligencia artificial fue desarrollado en el presente documento anteriormente es importante re relacionarlo con la discriminación, para revisar el concepto de discriminación es importante tener en cuenta lo que ha determinado la Comisión Interamericana de Derechos Humanos en la CONVENCIÓN INTERAMERICANA CONTRA TODA FORMA DE DISCRIMINACIÓN E

INTOLERANCIA[3], en su artículo 1, hace referencia a la discriminación de la siguiente manera:

> "Discriminación es cualquier distinción, exclusión, restricción o preferencia, en cualquier ámbito público o privado, que tenga el objetivo o el efecto de anular o limitar el reconocimiento, goce o ejercicio, en condiciones de igualdad, de uno o más derechos humanos o libertades fundamentales consagrados en los instrumentos internacionales aplicables a los Estados Partes.
>
> La discriminación puede estar basada en motivos de nacionalidad, edad, sexo, orientación sexual, identidad y expresión de género, idioma, religión, identidad cultural, opiniones políticas o de cualquier otra naturaleza, origen social, posición socioeconómica, nivel de educación, condición migratoria, de refugiado, repatriado, apátrida o desplazado interno, discapacidad, característica genética, condición de salud mental o física, incluyendo infectocontagiosa, psíquica incapacitante o cualquier otra." (Comisión Interamericana de Derechos Humanos, 2013)

En el artículo 1, antes referido la Comisión Interamericana de Derechos Humanos, indica que existen dos tipos de discriminación indirecta y múltiple y agravada, las cuales se detalla a continuación:

Discriminación Indirecta	**Discriminación Múltiple o Agravada**
"(…) se produce, en la esfera pública o privada, cuando una disposición, un criterio o una práctica, aparentemente neutro es susceptible de implicar una desventaja particular para las personas que pertenecen a un grupo específico, o los pone en desventaja, a menos que dicha disposición, criterio o práctica tenga un objetivo o justificación razonable y legítimo a la luz del derecho internacional de los derechos humanos." (Comisión Interamericana de Derechos Humanos, 2013)	"(…) cualquier preferencia, distinción, exclusión o restricción basada, de forma concomitante, en dos o más de los motivos mencionados en el artículo 1.1 u otros reconocidos en instrumentos internacionales que tenga por objetivo o efecto anular o limitar, el reconocimiento, goce o ejercicio, en condiciones de igualdad, de uno o más derechos humanos y libertades fundamentales consagrados en los instrumentos internacionales aplicables a los Estados Partes, en cualquier ámbito de la vida pública o privada." (Comisión Interamericana de Derechos Humanos, 2013)

Tabla 1 – Tipos de Discriminación[4]

[3] Adoptada en La Antigua, Guatemala, el 5 de junio de 2013 en el cuadragésimo tercer período ordinario de sesiones de la Asamblea General

Cabe anotar que además de los tipos de discriminaciones puesto la Comisión Interamericana de Derechos Humanos (2013), también indica que se pueden generar actos de intolerancia, los actos de intolerancia entendidos como manifestaciones de irrespeto rechazo o desprecio de la dignidad, características convicciones u opiniones de los seres humanos por ser diferentes o contrarias y se pueden manifestar como marginación y exclusión de participación en cualquier ámbito de la vida.

De acuerdo a lo que indica Muñoz Gutiérrez (2021), las tecnologías de la información deben gozar de neutralidad y objetividad sin embargo la inteligencia artificial no es considerada neutra, si bien al hacer uso de las tecnologías para su desarrollo goza de una aparente objetividad sin embargo los algoritmos que se plantean en ellas son desarrollados por humanos que se encuentran situados en un contexto específico atravesando un lugar histórico puntual y sus desarrollos son realizados desde una visión particular, lo cual, lleva necesariamente a que se materialice discriminación y en este sentido es importante que quienes hacen tecnologia con IA, reconozcan estos riesgos y tomen las medidas necesarias para ofrecer soluciones de IA que no solo apalanquen el crecimiento tecnológico de las sociedades, sino que también contribuyan a una mejor convivencia y puedan alinearse con los valores éticos y morales de las sociedades.

En este sentido el Banco Interamericano de Desarrollo, (2020), publicó el documento titulado: " Adopción ética y responsable de la inteligencia artificial en América Latina y el Caribe", donde reconoce que la inteligencia artificial tiene potencial de ayudar a la sociedad a superar diversos desafíos: la reducción de la pobreza avances en educación mejora en los servicios de salud erradicación de enfermedades incremento en la producción de alimentos entre otros, sin embargo, resulta muy importante la aplicación de principios dados por la Organización para la Cooperación y el Desarrollo Económicos, en adelante -OCDE- para propiciar que los desarrollos de inteligencia artificial propician la convivencia armónica de las naciones y relaciona los principios éticos de la IA dados por la OCDE, dentro de los cuales incluye: Valores centrados en el ser humano y la equidad:

> "Los actores del ecosistema de IA deben respetar el estado de derecho, los derechos humanos y los valores democráticos a lo largo de todo su ciclo de vida. Entre estos últimos sobresalen la libertad, la dignidad y la autonomía, la privacidad y la protección de los datos, la no discriminación y la igualdad, la diversidad, la equidad, la justicia social y los derechos laborales internacionalmente reconocidos. Con este fin, los actores de la IA deben implementar mecanismos y salvaguardias de protección de derechos como el de la autodeterminación de los individuos. Estos deben ajustarse al contexto y

[4] Fuente: Elaboración propia a partir de la CONVENCIÓN INTERAMERICANA CONTRA TODA FORMA DE DISCRIMINACIÓN E INTOLERANCIA de la (2013).

> ser consistentes con el estado del arte." OCDE (2019), citado por Banco Interamericano de Desarrollo, (2020)

Conforme a lo anterior, es importante que quienes desarrollan la los sistemas con inteligencia artificial, actúen reconociendo el marco ético y legal de las sociedades, teniendo en cuenta que la IA, trae grandes ventajas para las sociedades y también grandes desafíos para su regulación.

CONCLUSIONES

Para concluir esta introducción a un tema complejo, que seguramente en unos años lo será más, es necesario volver al problema desde la perspectiva de la filosofía de la mente. Desde la postura cientificista de las teorías reduccionistas en las que todas las capacidades humanas son el producto de cálculos muy complejos del cerebro humano, es legítimo conceder a las máquinas todos los atributos del ser humano, incluyendo derechos y deberes. Pero si asumimos una postura no reduccionista no tendríamos que hacer estas concesiones, pues, en cualquier caso, es solo una simulación que nos permite concebir a la máquina como lo que es desde su origen: una herramienta.

No hay que perder de vista que, por lo menos en este momento, es el ser humano el que le otorga a la máquina su identidad, lo que quiere que esta sea, por lo cual la determinación que se tome no puede estar sujeta a los intereses de las empresas que crean estos productos. La necesidad de un marco normativo y jurídico que proteja al ciudadano de lo que las empresas (dirigidas y creadas por seres humanos), quieran hacer con el desarrollo de la tecnología es de completa urgencia y necesidad y debe estar a cargo de los gobiernos y organismos competentes.

La responsabilidad frente a los problemas éticos que surgen del desarrollo de la inteligencia artificial no es nueva, es a la que siempre se ha tenido que enfrentar el ser humano ante las revoluciones tecnológicas. Por ello, sería mejor que no repitamos lo que ha sucedido en otros momentos de la historia de la humanidad: "no supimos ver venir los males de la revolución industrial, la contaminación, el daño a las clases trabajadoras, las injusticias flagrantes de premiar al dinero por encima de todo" (Latorre, 2019, p. 207). Con la reciente película sobre la vida de Marie Curie, recordamos que sus descubrimientos posibilitaron las bombas atómicas. El ser humano es tan maravilloso como terrorífico y antes de preocuparnos por la dominación de las máquinas sobre los seres humanos, debemos pensar en las acciones que cometemos entre nosotros mismos.

Otra conclusión parcial importante, tomando nuevamente el pensamiento de Searle, es la separación de lenguajes en los diferentes ámbitos de lo que llamamos realidad, a saber: la física, la mental y la social. En este sentido, es necesario profundizar en el concepto de responsabilidad, separando la explicación de lo moral como capacidad natural de los seres humanos, y la que se desprende del ámbito social en las categorías clásicas: social, moral y legal. Siguiendo la tradición que evita la falacia naturalista

(aquella que recuerda la tesis humana que impide derivar lógicamente el deber ser del ser), es preciso analizar la responsabilidad considerando los diferentes lenguajes en que se presenta.

Si hemos de responder al título de este trabajo, no sin saber lo que falta por explorar en el tema, la postura que se ha intentado defender a lo largo del escrito es la postura no reduccionista. El pensamiento de Searle nos invita a explorar la condición humana desde una visión múltiple y no atada exclusivamente al dictamen de la ciencia natural. Reducir las facultades humanas a cómputos o cálculos es seguir subvalorando (esto lo ha hecho la ciencia históricamente) las capacidades humanas que no dependen de la razón: los sentimientos, las sensaciones, la creatividad, la imaginación, la capacidad de soñar y proyectar un futuro, la posibilidad de inspirar ternura, la posibilidad de tener dolor, la necesidad de interactuar con el entorno, la construcción de una sociedad y por lo tanto, la capacidad de ser agentes morales, en tanto seres libres que condicionan su conducta en aras de una buena convivencia con los demás seres humanos. En este sentido, una primera respuesta a la pregunta acerca de si las máquinas pueden discriminar, sería negativa, las máquinas no discriminan en tanto no tienen la intención de hacerlo, no tienen una motivación o deseo de hacerlo, tampoco entienden el resultado de sus respuestas, pero el sistema con el que están hechas, los modelos o bases de datos con los que se dotan a la máquina pueden tener sesgos que dan origen a situaciones o respuestas discriminatorias, sin que la posibilidad de que la máquina intuya, sienta o se percate de que puede estar "haciendo" algo incorrecto.

He escrito algunas veces sobre este tema y me gusta terminar con la misma reflexión (Urbano, 2016): son impresionantes los avances tecnológicos y todo lo que puede construir la IA, no dudo que el mundo que veíamos como ciencia ficción se haga realidad. Pero cuando veo los tropiezos y dificultades que ha tenido la ciencia para hacer estas replicas "exactas" de los seres humanos me asombro más de la enorme complejidad y riqueza de la naturaleza humana, "es una lástima que con el alboroto que causan los avances tecnológicos veamos esto de manera invertida".

Es importante tener en cuenta que los desarrollo de inteligencia artificial, pueden incurrir en disminación multiple o agravada , indirecta y en actos de intoleracia, pues los algoritmos son un reflejo de la sociedad y en ellos se reprodu cen sus caracteristicas, en este sentido, contar con un marco ético y moral, que delimite los resultados de la IA, es cada vez más necesario en las sociedades, y debe ser un marco ético mundial, pues los procesos de desarrollo técnológico no conocen fronteras.

REFERENCIAS

Böckenförde, E.-W. (1993). *Escritos sobre derechos fundamentales.* (J. L. Menéndez, Trans.) Baden-Baden: Nomos.

Carlsson, U. (2003). The Rise and Fall of NWICO: From a Vision of International Regulation to a Reality of Multilevel Governance. *Nordicom Review, 2*, 31-68.

Erk, J. (Winter de 2004). Austria: A Federation without Federalism. *Publius, 34*(1), 1-20.

Häberle, P. (1962). *Die Wesensgehaltgarantie des Art. 19 Abs. 2 Grundgesetz.* Karlsruhe: C.F.Müller.

Humboldt, W. v. (1999). *On Language: On the Diversity of Human Language Construction and its Influence on the Mental Development of the Human Species.* (M. Losonsky, Ed., & P. Heath, Trad.) Cambridge: Cambridge University Press.

Luhmann, N. (2004). *Law as a Social System.* (K. A. Ziegert, Trad.) Oxford: Oxford University Press.

Baldwin, R., Cave, M., & Lodge, M. (Edits.). (2010). *The Oxford Handbook of Regulation.* Oxford: Oxford University Press.

Levy, B., & Spiller, P. (. (1996). *Regulations, Institutions and Commitment.* Cambridge: Cambridge University Press.

Price, M. E., & Noll, R. G. (1998). *A Communications Cornucopia: Markle Foundation Essays on Information Policy.* Washington, DC: Brookings Institution Press.

Rose-Ackerman, S., & Lindseth, P. L. (Eds.). (2010). *Comparative Administrative Law.* Cheltenham, UK: Edward Elgar.

Turing, A. (2010). *Maquinaria computacional e inteligencia.* Obtenido de http://xamanek.izt.uam.mx/map/cursos/Turing-Pensar.pdf

Del Valle, J. (2019). Inteligencia Artificial Ética. Un Enfoque Metaético a la Moralidad de Sistemas Autónomos. *Tesis de grado*. Universidad Nacional de Educación a distancia UNED. Obtenido de https://www.researchgate.net/publication/337797495_Inteligencia_Artificial_Etica_-_Un_Enfoque_Metaetico_a_la_Moralidad_de_Sistemas_Autonomos_TFG

Dennett, D. (1995). *La conciencia explicada.* Paidós.

Churchland, P. (1992). *Materia y conciencia.* Gedisa.

Searle, J. (2000). *El misterio de la conciencia.* Paidós.

Searle, J. (1996). *El redescubrimiento de la mente.* Grijalbo Mondadori.

Searle, J. (1997). *La construcción de la realidad social.* Paidós.

Urbano, C. (2016). *Corónica.* Obtenido de https://ensayo.revistacoronica.com/2016/05/robot-poeta.html

Orlowski, J. (Dirección). (2020). *El dilema de las redes sociales* [Película].

Latorre, J. (2019). *Ética para máquinas.* Ariel.

Fernández, O. (2020). *Mujeres en el sector público*. Obtenido de La Inteligencia Artificial: ¿un arma de discriminación masiva o un reflejo de la realidad? https://mujeresenelsectorpublico.com/la-inteligencia-artificial-un-arma-de-discriminacion-masiva-o-un-reflejo-de-la-realidad/

Buolamwini, J. (2017). *TED*. Obtenido de How I'm fighting bias in algorithms: https://www.youtube.com/watch?v=UG_X_7g63rY

Commission, E. (2019). *Ethics Guidelines for Trustworthy AI*. Obtenido de https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Queen, E. (2020). *Epic Queen* . Obtenido de Curso: Inteligencia Artificial y ética: https://www.youtube.com/watch?v=bhTz8bboPlI

Botero, J. R. (2000). *Mentes reales. La ciencia cognitiva y la naturalización de la mente.* Siglo del Hombre Editores. Universidad Nacional de Colombia.

Turing, A. (2011). *Inteligencia artificial.* Anagrama.

The Law, State and Telecommunications Review / Revista de Direito, Estado e Telecomunicações

Contact:
Universidade de Brasília - Faculdade de Direito - Núcleo de Direito Setorial e Regulatório
Campus Universitário de Brasília
Brasília, DF, CEP 70919-970
Caixa Postal 04413

Phone: +55(61)3107-2683/2688

E-mail: getel@unb.br

Submissions are welcome at: https://periodicos.unb.br/index.php/RDET