Universidad Autónoma de Querétaro

Facultad de Derecho

Consideraciones para una ética en la Inteligencia Artificial
Tesis

Que como parte de los requisitos para obtener el Grado de
Maestro en Ética Aplicada y Bioética

Presenta
Hugo Rodríguez-Reséndiz

Dirigido por:
Dra. Hilda Romero Zepeda

Co-Director:
Dr. Víctor Manuel Castaño Meneses

Querétaro, Qro. a 24 de mayo de 2022

Universidad Autónoma de Querétaro
Facultad de Derecho
Maestría en Ética Aplicada y Bioética

Consideraciones para una ética en la Inteligencia Artificial
Tesis

Que como parte de los requisitos para obtener el Grado de
Maestro en Ética Aplicada y Bioética

Presenta
Hugo Rodríguez-Reséndiz

Dirigido por:
Dra. Hilda Romero Zepeda

Co-Director:
Dr. Víctor Manuel Castaño Meneses

Dra. Hilda Romero Zepeda
Presidente

Dr. Víctor Manuel Castaño Meneses
Secretario

Dr. Saúl Tovar Arriaga
Vocal

Dr. Bernardo García Camino
Suplente

Dr. Jesús Armando Martinez Gómez
Suplente

Centro Universitario, Querétaro, Qro.
24 de mayo de 2022
México

Dedicatorias

Sau, sí podemos. A la memoria de Anita.

Agradecimientos

A la Universidad Autónoma de Querétaro que no sólo me ha dado la oportunidad de conocer el terreno académico/profesional, sino también por ser mi segunda casa de vida y por haberme regalado grandes personas.

A mis profesores que me han enseñado con paciencia tantas cosas y me han llevado de la mano, especialmente, la Dra. Hilda Romero Zepeda de quien he recibido acompañamiento y cariño, al igual que las enseñanzas de aula y vida del Dr. Víctor Manuel Castaño Meneses, Dr. Bernardo García Camino, Dr. Jesús Armando Martinez Gómez y Dr. Saúl Tovar Arriaga. Agradezco la lectura y diálogo de mi trabajo de investigación con el Dr. Ulises Cortes (Universitat Politècnica de Catalunya- Barcelona Supercomputing Center).

A mi Familia y Amigos, que siempre han estado incondicionalmente y de quienes no distingo las fronteras de la consanguinidad por el hecho del Amor que estimula la creatividad de las ideas y las letras. Nos hemos ganado mutuamente y siempre seremos ganados; estamos domesticados. Gracias a todos ustedes siento que inmerecidamente he sido y soy muy afortunado.

Prefacio

Toda Inteligencia Artificial (IA) e inteligencia humana tiene su razón de ser en los datos, es decir, en la información que le viene del mundo. En IA, los datos son una estela digital que se va construyendo por la interacción con la tecnología, y que en un primer momento se aparece como la historia de cada uno resumida en experiencias, deseos y formas de ser: un *yo* digital. Este vestigio es procesado por reglas computacionales para tomar decisiones que necesariamente modificarán el futuro del humano abordándose desde el terreno ético en el presente trabajo. Es así que aquí se analiza el fenómeno de planear, programar e implementar IA desde el horizonte de las humanidades para proponer una mirada ética de la tecnología. Se ha dedicado un gran esfuerzo a ello porque la tecnología es una de tantas cosas que marcan profundamente al *homo sapiens* día con día. Es entonces que se retoma la disertación agustiniana de "*Es posible haya alguien que ame conocer lo que ignora, pero nadie ama lo desconocido*", debido a que el reconocimiento de la ignorancia ha motivado a adentrarse, investigar y apasionarse con el tema en cuestión. Sólo se crean las circunstancias para que exista el conocimiento si primero se da el amor. El amor es pasión. Cuando uno se arroja a la pasión, sucumbe el tiempo y el espacio del ser que ama y del ser amado. Entonces, nace una dialéctica de responsabilidad entre dos seres, porque se crean lazos y "*Uno sólo comprende las cosas que domestica* […] *eres responsable para siempre de aquello que has domesticado*", según el secreto entregado por el zorro al Principito. Por lo tanto, la IA es un acto de conocimiento en el sentido de que todo resultado (predicción o decisión) del sistema está sujeto a los datos, al diálogo entre las reglas computacionales de discriminación y los *bits* provenientes del mundo cotidiano, que siempre "regresan" del mundo digital/binario para proponer una Realidad y Verdad en un entramado cuantitativo. Así pues, en cierta medida la IA nos conoce. Entonces, se asume el compromiso que de esta pasión ha devenido, al pensar el futuro como una posibilidad tecnológicamente que perturba desde el orden ético, lo cual habilita una preocupación y un análisis de la IA desde las humanidades en el suplicio "*mane nobiscum Domine*", que anhela un futuro con mayor paz para la humanidad frente a la incertidumbre tecnológica que aún no es, pero puede ser. El porvenir es algo intangible, algo que pertenece a los deseos, en este caso, al bien común: amor a la humanidad, de quien todo el tiempo se está conociendo.

Dignidad, Autonomía y decisiones

Hugo Rodríguez-Reséndiz

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

FACULTAD DE DERECHO

MEAB
Maestría en Ética Aplicada y Bioética

ÍNDICE GENERAL

RESUMEN

La Inteligencia Artificial (IA) es una tecnología que no es de reciente aparición, sin embargo, su desarrollo e implementación se ha potenciado en los últimos años dentro de la sociedad. Esta situación ha promovido diversas realidades en el ser humano que van desde temas económicos, salud, educativos y de relaciones personales, que necesariamente tocan temas éticos y morales. Por estas razones la comunidad global en recientes fechas ha iniciado diálogos en diversos foros para generar marcos éticos que estén dirigidos hacia una Inteligencia Artificial responsable y fiable. Ante ello, se plantea como **objetivo** el proponer conceptos de la filosofía por medio del análisis jurídico, científico y de ingeniería computacional que posibiliten la promoción de un diálogo hacia una ética en la IA. Y se plantea como **metodología** un enfoque mixto en la examinación de conceptos con participación en foros donde se discuten marcos éticos de IA a nivel local, nacional e internacional, así como la revisión de literatura en el tema. Al respecto, se tiene como **resultado** una taxonomía ética dentro de la IA que constituyen como **conclusiones** el postular conceptos que engloban y puedan guiar el quehacer reflexivo de una ética en la Inteligencia Artificial, al generar una epistemología y fenomenología digital, mientras la humanidad se aproxima a posibilidades que determinarán el comportamiento en la sociedad.

Palabras clave: Conciencia digital, Ética de IA, Fenomenología sintética, Moral computacional.

ABSTRACT

Artificial Intelligence (AI) is a technology that is not of recent appearance, however, its development and implementation has been enhanced in recent years within society. This situation has promoted various realities in the human being that range from economic, health, educational and personal relationship issues, which necessarily touch on ethical and moral issues. For these reasons, the global community has recently initiated dialogues in various forums to generate ethical frameworks that are directed towards responsible and reliable Artificial Intelligence. Given this, the **objective** is to propose concepts of philosophy through legal, scientific and computational engineering analysis that enable the promotion of a dialogue towards ethics in AI. And a mixed approach is proposed as a **methodology** in the examination of concepts with participation in forums where ethical frameworks of AI are discussed at the local, national and international level, as well as the review of literature on the subject. In this regard, the **result** is an ethical taxonomy within AI that constitutes as **conclusions** the postulation of concepts that encompass and can guide the reflexive work of ethics in Artificial Intelligence, by generating a digital epistemology and phenomenology, while humanity approaches to possibilities that will determine behavior in society.

Keywords: Digital Consciousness, AI Ethics, Synthetic Phenomenology, Computational Morality.

CAPÍTULO PRIMERO

1. Marco de referencia de la ética de la IA

Si el Parlamento hubiera creado el mundo, los derechos humanos universales podrían haber quedado enterrados en algún subcomité, junto con todo ese asunto de la física cuántica. Pero el Parlamento no creó el mundo, solo intenta darle sentido.

Yuval Noah Harari

1.1 Circunstancia digital

La Inteligencia Artificial (IA) es una tecnología que han llegado hasta la vida del *homo sapiens*, no sólo para quedarse sino para transformarla por completo. Durante mucho tiempo los humanos han buscado la forma de facilitar su principal actividad dentro del mundo: *ser*. Las diversas estrategias han intentado alcanzar una distinción de las demás especies vivas. Al apelar al talante aristotélico de la razón, se ha conferido gran poder en la capacidad cognitiva como método para percibir, modificar y dominar el entorno.

La tecnología ha sido un instrumento con el cual no sólo se puede facilitar grandes objetivos, sino también ha sido factor decisivo para marcar el rumbo de la historia de humano y del planeta entero. Así pues, no se puede negar que la bomba nuclear, el motor de combustión interna, o la síntesis de elementos químicos, llevaron al lugar en el que hoy nos encontramos a las singulares consecuencias de ayudar y problematizar la existencia.

Sin bien no existe un momento histórico en el cual apareció de forma abrupta la computación, sí hay toda una historia que data del siglo IV a.C. con la aparición del ábaco como herramienta matemática en la humanidad, un marcado desarrollo en los siglos XIX y XX d.C. en la automatización de procesos, y un destacado desarrollo en materia de capacidad, velocidad, procesos alternos, y un asentamiento del uso de la Inteligencia Artificial, por

mencionar algunos en el presente siglo XXI. Al respecto, se puede señalar al menos que tiene dos componentes secuenciales: el lingüístico y el material.

En ese sentido, se puede afirmar la concepción de Richard Braithwaite, quien propuso por primera vez el término computación: acción humana para realizar cálculos (Webb, A., 2020) y, por otro lado, la forma material más asemejada a la idea con la que se cuenta de un ordenador en estos días que se rescata de "*On computable numbers, with an application to the Entscheidungsproblem*" (Turing, A. M., 1936). Además, Alan Turing aportó una visión de abordar los problemas de la metáfora computacional: las máquinas se pueden acercar a la capacidad humana de procesar datos (Davis, M., 2018). Al mismo tiempo, por la influencia del vestigio del pensamiento de Leibniz, Turing sentó las bases para la Inteligencia Artificial en "*Computing machinery and intelligence*" (Turing, A. M. 2009) e indagó sobre la posibilidad de las máquinas para actuar de igual manera como lo hace un humano cuando se le adjudica la inteligencia (Castelfranchi, C., 2013).

En referencia a la historia lingüística y material subsecuente a la *Máquina de Turing*, fue a partir de McCarthy que se consolidó el término de Inteligencia Artificial como una ingeniería para diseñar máquinas por medio de algoritmos (McCarthy, J., 1987). Al ubicar a las matemáticas como una de las bases de la informática y en el terreno de la ciencia, se puede afirmar que todo proceso computacional que busca imitar la cognición humana y transforma la realidad es una Inteligencia Artificial. De ello se desprende que dicha tecnología es una *mímesis tecnocefálica* (Rodríguez-Reséndiz, 2020a), debido a que los procesos o reglas que se usan para esta rama de la computación son similares a la funcionalidad del cerebro humano en virtud de la frontera entre lo natural y lo artificial, la confusión creada, y que abre nuevos horizontes epistemológicos y éticos, motivo del presente trabajo de investigación.

Es así como la mímesis tecnocefálica llega hasta el día de hoy con una gran variedad de funcionalidades que abordan las problemáticas que acontecen, pero también con retos para la sociedad. Desde el sorprendente desempeño de *Deep Blue* (1996) y *AlphaGo* (2015) de IBM y Google, respectivamente, la IA ha contribuido a la mejora de la vida cotidiana en el área de la economía, la salud, el gobierno, la movilidad, arte, comunicación e interacción

humana (Cath, C. et al., 2018). En cuanto a ello, la IA se ha utilizado en el análisis la Bolsa de Valores de Nueva York por medio del Aprendizaje Automático (Machine Learning) en la predicción de mercados financieros para comprar acciones y disminuir los riesgos, dejando de lado las subjetividades humanas que antes se realizaban cargadas de un sesgado positivismo exitoso (Shen, S., Jiang, H., & Zhang, T., 2012).

De la misma manera, con el uso de IA y el Aprendizaje Automático, se ha logrado realizar una adecuada clasificación taxonómica de los genomas del Síndrome Respiratorio Agudo Severo Coronavirus 2 (SARS-CoV-2), así como su detección forma rápida y precisa gracias a los datos analizados por Procesamiento de Señales Digitales (DSP) y con ello se ha indagado sobre el origen de dicha enfermedad (Randhawa, G. S. et al., 2020). En Brasil, el gobierno utiliza algoritmos computacionales para disminuir los fraudes de los productos importados por medio de la clasificación adecuada que hace el sistema de entre más de 10,000 códigos diferentes (Digiampietri, L. A. et al., 2018). Gracias al Aprendizaje por Refuerzo Profundo (técnica de IA), se puede analizar datos en el trasporte terrestre para optimizar la calidad de la movilidad y de esta forma contribuir a que la vida humana sea más fácil (Haydari, A. & Yilmaz, Y., 2020). De igual manera, gracias a la mímesis tecnocefálica se ha logrado crear música totalmente nueva por medio de bases de datos de canciones realizadas por humanos gracias al entrenamiento de algoritmos para imitar la composición musical (Dhariwal, P. et al., 2020). No menos destacados son los algoritmos desarrollados para identificar noticas falsas en las redes sociales gracias a la ponderación de información y clasificación de Machine Learning (Regresión Logística) para la predicción potencial contenido maléfico en internet (Pinnaparaju, N. et al., 2020).

Los ejemplos del uso de la IA no sólo demuestran la facilidad de creación del ser humano, sino su capacidad de solucionar problemas de manera más eficiente. La cuestión que aparece en este despliegue tecnológico es la multiplicidad de problemas, entre ellos el de la ética, la información, la toma de decisiones y la aplicación de los sistemas computacionales y de IA, ya que la tecnología siempre y de algún modo estará afectando a los humanos por sus interacciones con y para otros. Es por ello que es necesario reflexionar entorno de los dilemas que van emergiendo a partir de su uso y aplicación. En todo caso, también es fundamental

apuntalar normas que estén orientadas al desarrollo y uso de una Inteligencia Artificial responsable y que pretenda la universalidad. De este modo la ingeniería computacional no sólo ha irrumpido de forma abrupta en la vida del *homo sapiens*, si no que ha creado paradigmas que conllevan ciertas discrepancias en el mar de la moral, conciencia, subjetividad y dispersión de *formas de ser*.

Al mismo tiempo, la aparición de otras tecnologías como Realidad Aumentada, Internet de las Cosas, Big Data, Redes Neuronales, robótica, Industria 4.0 o la nube, modifica de forma sustancial lo más significativo de la humanidad: la vida misma (Sarsanedas, A., 2015). En este sentido, la oportunidad que han tenido autores como Luciano Floridi (Oxford University), Raja Chatila (Pierre and Marie Curie University), Neal Parikh (NYC Mayor's Office), Jean-François Bonnefon (Toulouse School of Economics), Heng Xu (American University), Thomas D. Parsons, PhD (University of North Texas), Rafael Capurro (International Center for Information Ethics) o Thomas Metzinger (JGU Mainz), es inmejorable para poner en primer plano de la agenda de la Inteligencia Artificial a la ética, su creciente demanda y su escasa regulación. No menos significativo es el esfuerzo que organizaciones internacionales como OCDE, IEEE, Unión Europea o UNESCO han realizado para entablar un diálogo alrededor de la ética y la IA y generar un uso responsable de esta tecnología computacional (se dedicará un apartado para la revisión de este tema).

En ese sentido se sabe que, a diferencia de otras tecnologías, la Inteligencia Artificial es capaz de actuar por sí misma, es decir, no se mantiene estática sino que incluso toma decisiones derivadas de su programación, lo cual acarrea problemas delicados toda vez que la intervención humana en la toma de decisiones se ve mermada o incluso no tiene efecto alguno en la implementación de esta tecnología. De esta realidad, se desprenden dos posturas al respecto. La primera de ellas se relaciona con las opciones para mejorar la vida humana que nos devienen de la IA, mientras que en la segunda, orilla a pensar que la mímesis tecnocefálica contribuye al malestar del humano o incluso ocasiona su aniquilación. Esta situación antagónica lleva a pensar que la Inteligencia Artificial necesariamente ocupa un lugar trascendente a la altura de los tiempos actuales. Por ejemplo, el despojo de actividades humanas cotidianas que trae consigo la IA (Ahuja, A. S., 2019) abre la posibilidad de discutir

y abordar la forma de tomar decisiones. Pero ¿A qué costo es esto? ¿De qué manera la autonomía humana se ve confrontada con la de la máquina? ¿Si la programación computacional actúa en el mundo, se puede hablar de una moral computacional? ¿Cómo debería de regularse éticamente esta tecnología? ¿Debería de haber normas globales en el terreno ético para el despliegue de la mímesis tecnocefálica? Y más importante aún, ¿Cómo distinguir quién realmente decide?

La idea de conectividad en la que se encuentra la humanidad ahora mismo es el efecto de la cultura y comunicación que hemos desarrollado, lo cual se potencializa con la aparición de la tecnología y hace sentir que se vive en la localidad de la globalidad, o dicho en palabras de Marshall McLuhan, en una aldea global donde concurre espacio y tiempo que provoca creer que se vive en la mejor de las síntesis posibles dentro del mar de información (McLuhan, M. et al., 2011). Este argumento permite considerar la opción de que somos afectados por la singularidad de los avances tecnológicos, mismos que pueden alcanzar un bien. Es justo en este razonamiento donde la discrepancia subyace, porque si la IA es creada para beneficiar a alguien ¿el bien que se genera es unidireccional o se puede crear el mal postulando un bien particular? Desde el terreno ético ¿cómo se puede entender que la mímesis tecnocefálica contenga sesgos que a la vista de los desarrolladores no puedan ser detectados?

El hecho de que la IA tenga la capacidad de sujetar la atención o voluntad de humano está basada en la practicidad que habilita la tecnología. Esta interpelación ha provocado una dependencia en la vida actual del ciudadano y la sociedad en general. ¿Se puede imaginar una vida en estos días sin las herramientas que nos provee la técnica? Pues bien, esta incapacidad funcional de *ser-en-el-mundo* orienta la explicación de conflictos que pasan de lo digital a la vida real. El develar la existencia de sesgos en la Inteligencia Artificial ha encendido las alarmas cada vez que se genera algún daño a las personas en diversos sentidos, pero que en esencia provienen de un entramado ético, especialmente, el relacionado con la falta de atención de los Derechos Humanos.

En consecuencia de lo anterior, se puede afirmar que el sesgo es el cúmulo de problemas relacionados con la recopilación de datos y finalizan con decisiones prejuiciosas (Ntoutsi, E. et al., 2020). Entonces, en la opacidad de generar una perspectiva amplia del panorama, se deja de considerar cada una de las variantes en medio del proceso de la creación de mímesis tecnocefálica, lo cual debilita la dignidad humana y crea dilemas. Estos últimos se pueden analizar para mejorar la comprensión de las dificultades cuando la IA trastoca la vida humana y la perjudica.

Existen diferentes ejemplos de Inteligencia Artificial donde se ha develado que la implementación genera conflictos éticos como la discriminación, uso poco responsable, haber coartado Derechos Humanos, oprimir la libertad, privacidad, generar adicción o manipular la autonomía humana (Manheim, K. M., & Kaplan, L., 2018). Por ejemplo, con el uso de Machine Learning se ha abierto la brecha racial en los lugares donde opera, ya que la captura y procesamiento de imágenes con ayuda de algoritmos provoca que se vaya etiquetando a la gente en la vida real, lo cual a simple vista no tiene una implicación mayor a no ser que las personas de color “negro” sufran consecuencias negativas por esta distinción mientras que los de color “blanco” gocen de beneficios. Por ello, se abre la posibilidad de una polarización en la sociedad y el aumento de la segregación social, ya que se estigmatiza a la gente por su color, dejándoles la menor cantidad de recompensas y bienes (Benthall, S., & Haynes, B. D., 2019). En todo caso, el ejemplo anterior nos muestra cómo al mismo tiempo de resolver una problemática concreta sobre la distinción de personas, se provoca un inconveniente en el reforzamiento de sesgos raciales.

Por lo anterior, en la IA es necesario disminuir los perjuicios hacia los humanos que se pueden llegar a tener en su desarrollo, implementación y uso. Y surge la pregunta ¿Es la ética la única forma de atender esta problemática? Definitivamente no, porque el inconveniente que se presenta es estructural y no sólo se ocupa de voluntades humanas, sino de agentes externos que permeen en la sociedad como cultura, economía, medio ambiente o recursos naturales disponibles. Lo que sí es un hecho es que la ética ayuda a articular la forma de abordar las coyunturas de la IA, toda vez que esta disciplina promueve el diálogo razonable sobre el bien común porque no todas las cosas técnicamente factibles a veces son aceptables

y provocan un autoreconocimiento de la fragilidad de la ciencia y sus disciplinas emergentes, debido a que los límites son el bien hacia la humanidad misma y la responsabilidad (Bergoglio, M., 2018). Por tanto, los modelos matemáticos que se implementan en la IA deben alcanzar su funcionamiento en la realidad siempre que haya un fondo éticamente responsable que acerque al bien práctico para minimizar los daños en la búsqueda de patrones y clasificaciones digitales. Esta situación lleva a una cuestión medular: la formación ética de los desarrolladores, que no son los únicos que juegan un papel dentro de la trayectoria de la IA, pero sí son uno de los principales actores para que los códigos de programación sean establecidos con los menores sesgos posibles.

En ese sentido, se asume que es inevitable que la IA pueda ser neutral, ya que es ineludible que un solo código de programación sea aplicable a un único individuo, sino que es para la multiplicidad de formas de ser y pensar, lo que provoca que exista una multiplicidad de entendimiento y, en el trayecto de la predicción y clasificación de datos por algoritmos, se generen modelos que modifiquen el mundo real con un sentido negativo para algunas cuantas personas. Esta afección es inevitable por los sesgos sistematizados, máxime en las herramientas de minería de datos más allá de los beneficios puntuales en el desempeño del Aprendizaje Automático (Hajian, S. et al., 2014).

En los apartados posteriores se enmarcará una definición de ética, así como algunos ejemplos más puntuales sobre las discrepancias de la IA que dan pie a diversos dilemas.

1.2 Horizonte de investigación

1.2.1 Fundamentación teórica

El auge de la Inteligencia Artificial que se percibe hoy en día provoca que esta tecnología esté presente en diferentes ámbitos de la vida humana, colocándola como uno de los principales acontecimientos que han marcado el desarrollo de la sociedad, según la clasificación de la Asociación Estadounidense para el Avance de la Ciencia publicada a finales del 2019. Al respecto, se puede considerar el caso del *bot* denominado *Pluribus* que,

previamente a la competencia contra jugadores de póker, se autoentrenó jugando partidas contra sí mismo para reconocer patrones y posibles respuestas a las variables de un juego real. El resultado fue que ganó a sus contrincantes humanos en partidas diferentes al tomar en consideración las predicciones derivadas de los posibles movimientos de sus oponentes (Brown, N., & Sandholm, T., 2019)

La aproximación del ejemplo mostrado presume que la Inteligencia Artificial se presenta en actividades cotidianas e interacción con el ser humano, que van desde tareas repetitivas hasta resolver problemas complejos. Se ha especializado este conocimiento tecnológico para adentrarse en temas que normalmente el ser humano realizaba por sí mismo y que la vinculación ético-normativa estaba en función de una sola dirección: la responsabilidad del humano que actuaba en su entorno. Ahora, las tareas profesionales son complementadas con IA, lo cual permite delegar a esta tecnología la responsabilidad de abordar un problema y han traído nuevas formas de abordar la ética.

Para ahondar en lo anterior, se puede considerar el caso en el que para una valoración clínica, algunos médicos han usado Aprendizaje Profundo de Redes Neuronales Profundas Artificiales en epidemiología clínica y bioestadística para generar modelos predictivos que contribuyan a diagnósticos de pacientes (Park, S. H., & Han, K., 2018). Este ejemplo del área de la salud incorpora nuevas discusiones en el terreno bioético, debido a que tan sólo el compartir el diagnóstico clínico con la IA obliga a preguntarse sobre la responsabilidad del resultado ¿quién será el responsable de los errores cometidos en el proceso? ¿El médico que suscribe el diagnóstico, el desarrollador de Inteligencia Artificial por no haber creado reglas adecuadas, la compañía que comercializa el software, el gobierno que no estipula normas para esta tecnología o el paciente que firmaba la autorización del procedimiento sin entender la tecnología?

En cuanto a otro caso, para generar un control en las masas, la ingeniería computacional ha creado soluciones útiles para el reconocimiento facial, en tiempo real, de personas, que van más allá de las condiciones físicas donde se encuentra el humano captado por la cámara. Esta tecnología tiene la capacidad de identificar de forma muy precisa a sujetos por medio del uso

de algoritmos que calculan matemáticamente las semejanzas entre una base de datos y las personas que están siendo monitoreadas, dando como resultado la identificación instantánea de sujetos. Los gobiernos son las entidades que mayormente se ha visto beneficiadas con este tipo de software, ya que facilitan el control sobre las estrategias y atribuciones que le son estipuladas, especialmente las relacionadas a las fuerzas del orden público. Los problemas que subyacen a esta situación se vinculan en un primer momento con la precisión, seguridad, privacidad y el resguardo de derechos civiles (Lynch, J., 2020). Más allá de la sesión de sus datos biométricos para ser analizados, las personas deberían de preocuparse por la escasa supervisión normativa de este tipo de IA, ya que en ocasiones no existe la suficiente cantidad de pruebas antes de salir a operar en la sociedad y puede llegar a ser perjudicial en el mundo real, especialmente, en las decisiones relacionadas con la asociación de los datos biométricos de clasificaciones raciales, de estereotipos, estigmas o resoluciones legales.

En cuanto al tema de la privacidad, se ocupa en un primer momento que los individuos captados por las cámaras de vigilancia sepan o den su consentimiento para que puedan ser reconocidos y evitar una "*alineación perpetua*", debido a que es muy probable que los datos biométricos ya estén en otras bases de datos sin habernos dado cuenta de ello. Las aplicaciones de celulares y otros dispositivos rescatan esta información de manera pasiva sin que los usuarios sepan que está pasando. El caso es destacado cuando en las redes sociales y aplicaciones de entretenimiento se condiciona a usar fotos propias para poder funcionar (Garvie, C. et al., 2019). Así pues, existe preocupación por el uso del reconocimiento facial, ya que el uso poco ético de los datos por parte de funcionarios públicos, iniciativa privada y particulares puede llegar a crear maleficios con este tipo de software. Justo con este tipo de evidencias es cuando se refuerza la idea de sumar a la IA una perspectiva ética, pero en el mar de conceptos de que se presentan, ¿Qué entendemos por ética?

En un primer momento, se considera que la filosofía es una disciplina que se ha encargado, a lo largo de la historia occidental, de pensar al humano y sus problemas fundamentales como la vida, la muerte, el conocimiento, lo real, la verdad y el Ser. El cimiento de esta sabiduría alude principalmente a la Grecia antigua, donde se forjaron planteamientos que siguen acompañando a la civilización hasta los tiempos actuales. El paso de la Edad Media a la

Ilustración constituyó el reinado de la Razón por el postulado científico. Más adelante fue encabezado por el positivismo (De Asúa, M., 2018). Así pues, el rechazo de todo lo escatológico es una constante en la actualidad; guarda en sí mismo un dilema: para rechazar a lo llamado metafísico, se usa una pragmática que se basa en conceptos, es decir, en abstracciones que realiza el humano y que después da una percepción del mundo. El lenguaje es el vehículo con el cual se trastada lo conceptual a lo práctico.

Acerca de lo anterior, la ética es un discurso desprendido de la filosofía y por ello ha heredado una serie de características: racionalidad y criticidad. Al existir diversas formas de abordar a la ética, para el presente trabajo se conceptualizará como un ejercicio de reflexión del individuo entorno de la moral donde se busca generar una crítica de los paradigmas dados al sujeto. Ese "entorno" y "paradigmas" se encuentran en un lugar determinado: en la sujeción del sujeto, su aprensión (Kelly, M. G., 2013). Así, todo ejercicio ético corresponde a un marco de referencia que se le llama "realidad". A saber, en la Edad Media no se hablaba de ética de los robots, sino de correlaciones entre el acto humano y su pretensión con la salvación. Es por ello que, al estar situados un momento histórico donde predomina el conocimiento científico, conviene hablar de ética en la ciencia y todo lo que de ella se desprende, por ejemplo, sobre la IA.

Ahora bien, las consecuencias éticas del uso de la Inteligencia Artificial permiten una deliberación especial debido a que se corre el riesgo de que el individuo se aleje de su dignidad, por ejemplo, ante la pérdida de la toma de decisiones. De forma reciente, se ha creado polémica porque, en la posibilidad de predecir o sugerir de la IA, los sistemas computacionales inhiben o, incluso, manipulan al ser humano, pasando de "ayuda colaborativa" a "determinación del actuar" o, dicho en otras palabras, al direccionamiento por terceros de aquello que llamamos *voluntad*. Esto mismo supone ya un problema ontológico muy concreto: la IA efectúa sobre nosotros la posibilidad de *ser*. Un ejemplo de ello fue que, tras las elecciones presidenciales de Estados Unidos en el año 2016, se suscitó un escándalo ya que la empresa Cambridge Analytica fue acusada de conducir la conciencia de las personas para inclinar las preferencias de los electores hacia un candidato por medio del uso de algoritmos en las redes sociales (Kaiser, 2019). A pesar de que el caso ha llegado

al senado norteamericano y que no se ha determinado la culpabilidad, por su cuenta, la gente debe pensar si la IA puede ejercer una irrupción en su autonomía.

A propósito de que justo en la actualidad a nivel mundial se están realizando esfuerzos para regular la Inteligencia Artificial desde el campo ético, y que el 74% de las personas desconfía de la IA (Edelman, 2019), cabría preguntarse si puede llegar a haber un consenso global en torno del tema, más aún cuando nos enfrentamos a estandartes tan altos como es el respeto a la diversidad, es decir, la incapacidad de asumir la verdad o una ética universal como lo pretendía Hans Küng (Apel, K. O., 2017).

Mientras eso sucede, se puede destacar otro problema en el uso de Inteligencia Artificial: el de la autonomía del ser humano, ya que al usar esta tecnología se postulan sugerencias o predicciones que inciden en la realidad y en ocasiones de manera frontal. En ese contexto, desde la filosofía, el derecho y la tecnología, es válido preguntarse con la ayuda de autoridades intelectuales, sobre qué es la autonomía y cuál es el diálogo que sostiene con la IA en un marco ético para orientar las reflexiones del establecimiento regulatorio de dichos procesos computacionales, y así generar instrumentos que sean aplicables a los métodos de creación de IA, por ejemplo, en instituciones que se encarguen de promoverla.

El primer tópico de la autonomía en la IA se remonta a tiempos helénicos, ya que desde entonces existía la disputa de la libertad como un asunto de reflexión. La voluntad del ser humano también es una caracterización que se ajusta con los principios de la filosofía. La autonomía es darse a uno mismo su propia ley (*ατο*-auto y *νόμος*-nomos) (Palavecino, C. R., 2017) y es parte también del catálogo de propiedades de un individuo enmarcado bajo ciertas condiciones. Este axioma surge por las circunstancias históricas que envuelven al ser humano. Aquí y ahora, la tecnología marca muchas pautas para *ser* y *actuar* en el mundo. En lo que concierne a la Inteligencia Artificial, se ha cambiado la forma de valorar e, incluso, la moral se ve sujeta a los dictados de ese tipo de tecnología. No obstante, cuando la IA genera predicciones o delimita una sola opción para ser elegida por el ser humano, se crea un conflicto en el depositario del ejercicio de la voluntad y la fractura de la autonomía, ya que se permite que dichos sistemas computacionales decidan por las personas incluso sin darse

cuenta (Sciutti, A., & Sandini, G., 2017). Esta transgresión no sólo afecta dicho ámbito de la vida, sino que pone en riesgo otras cuestiones que tejen un entrelazado para que la dignidad se vaya fracturando. Un ejemplo de lo anterior es la aplicación de la Inteligencia Artificial en el consumo de bienes, predicción de enfermedades, análisis y planeaciones financieras, y todo aquello que con la estela de datos "orienta" hacia un lugar o hacia otro al *homo sapiens*.

Un segundo momento de reflexión en torno a la autonomía en los albores de la IA es de carácter ontológico, al encontrarse frente a la barrera del *aquí* y *ahora* de la vida (McConwell, A. K., & Currie, A., 2017). Toda IA está hecha para transformar el mundo, en este caso, del ser humano, por ello, tras el proceso computacional, se incide sobre los usuarios por medio de predicciones, recomendación o actuaciones. El problema se refleja en la posibilidad de que la IA indique el futuro del ser humano que está dotado de vida, siendo que ella es cambiante, por lo que todo análisis computacional debería de tomarse con esa misma condición: la posibilidad de *no-ser*, lo cual implicaría que, en un marco ético, tanto creadores como usuarios tengan en claro el sesgo propuesto por la IA que al final siempre serán posibilidades y herramientas para generar autonomía o no del ser humano.

Derivado de lo anterior, se considera necesario sumarse a los esfuerzos de la institucionalización de la ética en la IA. En los dos últimos años (2020-2022), no se han homologado criterios a nivel internacional, sino que en la búsqueda de lo universal, se disgregan las prácticas de "*hacer-diseñar-crear-implementar* y *usar-afectar*" la IA, por lo que se ocupa no aspirar a buscar máximas, sino aproximaciones éticas que partan desde los plexos de referencia del sujeto que está siendo interpelado determinadamente por lo que es y su entorno (*lo que está a la mano*) (Heidegger, M., 1993). De ello resultan dos perfiles para reflexionar en este tema: el de las personas que realizan la programación y aquellas que consumen los productos. Para ambas figuras, es necesario establecer un marco ético y mitigar el despojo de la autonomía y por lo tanto del fundamento de la dignidad humana desde la perspectiva aristotélica, kantiana, y de los Derechos Humanos, es decir, limitar la heteronomía o aquello que "nos venga desde afuera" por consecuencia de la IA.

La ética al ser una disciplina desprendida de la filosofía remite a examinar las posturas (marcos éticos) más destacadas del tema. En primer lugar, subyace la ética de la virtud de Aristóteles donde se pondera el bienestar, derivado de comportamientos, que producen felicidad como el fin del ser humano. Luego, tenemos la postura de Kant que establece una ética del deber ser (deontología) y centra su postura en el reconocimiento de la autonomía para aspirar a lo universal. También están las exposiciones que hacen Bentham y Mill quienes conceptualizan el consecuencialismo o utilitarismo para afirmar que lo relevante es una situación cuantitativa. De igual manera, Habermas postula una ética dialógica comunicativa, centrada en el lenguaje para la construcción del entendimiento en la dialéctica de la alteridad. Por su puesto, la ética de ideales y valores examinadas por Durkheim revelan un análisis de posturas a partir de perspectivas sociales. En tanto que autores como Kohlberg defienden más los contextos o plexos de referencia en el ámbito ético. De esos enfoques éticos, cabe destacar que los principios utilitarios, un equilibro reflexivo, el principialísmo y la casuística, sobresalen como metodologías para realizar deliberación moral.

Para no menoscabar estas pretensiones y dar solidez a la propuesta, es necesario considerar que la Inteligencia Artificial opera con algunas puntualizaciones. La primera de ellas es que la IA actual está clasificada como Inteligencia Artificial Estrecha, ya que realiza tareas concretas y puede llegar a tener interacción con humanos (Fjelland, R., 2020). De forma similar, hay quienes clasifican a la IA en simbólica (sistemas expertos) y no simbólica (Machine Learning), donde la primera es postulada-escrita por humanos al describir actividades de trabajo y permitir una secuencia condicional; mientras que la segunda está orientada a una menor intervención humana, al delegar el aprendizaje mayoritariamente a las máquinas, las cuales usan una gran cantidad de datos para después, por medio de reglas, realizar predicciones. Las redes neuronales y el Aprendizaje Profundo se encuentran dentro de esta categoría.

Para los fines del presente trabajo, se ocupa distinguir tres formas de Aprendizaje Automático, puesto que de ellas se desprenderán algunas reflexiones: Aprendizaje No Supervisado que es todo algoritmo que examina patrones y coincidencias de un conjunto de datos determinado; Aprendizaje Supervisado que descubre las relaciones entre datos de

entrada y salida asignados por un humano quien previamente conoce dichos datos y los etiqueta para generar una directriz adecuada; y, Aprendizaje Por Refuerzo es el algoritmo que lleva a cabo una tarea basada en las recompensas/premios y que ocupa estar constantemente probando las acciones tomadas (Mahesh, B., 2020). Otros métodos destacados en la construcción de la IA es el Procesamiento de Lenguaje Natural consistente en la aplicación del lenguaje humano a las computadoras para que estas generen un entendimiento; y la Visión por computadora que ocupa un método informático para procesar y entender imágenes del mundo sensible (Wiriyathammabhum, P. et al., 2016).

1.2.2 Hipótesis

El análisis jurídico, científico y de ingeniería computacional a partir de la comprensión de conceptos filosóficos y uso de una ética aplicada, posibilita un diálogo armónico en la postulación de una ética de la IA.

1.2.3 Objetivos

A. Objetivo general

Analizar conceptos entorno a la ética de la Inteligencia Artificial a través de la deconstrucción de ideas de dicha tecnología para posibilitar canales de comunicación en el trayecto ético y regulatorio.

B. Objetivos particulares

Delimitar los conceptos epistemológicos, jurídicos, científicos y de ingeniería computacional para generar puntos de acuerdo con la regulación de la IA desde la ética.

Proponer una taxonomía en la ética de la IA para generar un diálogo unidireccional cuando se hable de tema.

Establecer la finalidad de una ética de la IA para delimitar alcances cuando se busque la regulación de dicha tecnología.

Analizar las correlaciones entre ciencia, universalidad y metafísica para indicar la posibilidad de una ética de la IA global.

Puntualizar y describir los conceptos a considerar en una ética de la IA global para generar marcos regulatorios.

1.2.4 Metodología

Para llevar a cabo el presente proyecto, se realizó un enfoque mixto de investigación de tal forma que se pudieron examinar conceptos divergentes para adaptar necesidades, contextos y recursos que permitieron una sistematización conceptual y empírica en la obtención y análisis de datos cualitativos para una mejor compresión de la ética en la IA.

De esta forma, tras haber realizado la planeación y validación institucional del presente proyecto de investigación, se ha revisado el estado del arte correspondiente y el planteamiento del problema. Esto ha permitido focalizar grupos de interés que discuten el mercado regulatorio ético de la Inteligencia Artificial a nivel global. A continuación, se incorporó al diálogo internacional por medio de asistencias a foros, congresos, debates, mesas de diálogo donde se ha discutido el tema en cuestión. En este paso, se dictaron conferencias regionales, nacionales e internacionales en foros académicos, gubernamentales, de iniciativa ciudadana y académica para contrastar las ideas propuestas. No menos importante fue la contribución para la postulación de un marco regulatorio de la IA desde la ética a nivel nacional, participando en la iniciativa de una estrategia mexicana de la Inteligencia Artificial.

La detección de las pautas éticas más concurrentes en el diálogo internacional ha permitido distinguir las primordiales consideraciones éticas y una taxonomía, lo que ha generado que se participe en la divulgación científica en proyectos académicos y, además, de la sensibilización de desarrolladores de IA y tomadores de decisión para la implementación de políticas públicas.

1.3 Expectativa de la IA

Tras reconocer que la IA nos interpela, es necesario considerar cuáles son las expectativas que tenemos de dicha tecnología para generar un horizonte. De este modo, cuando hablamos de mímesis tecnocefálica se encuentra una polarización entre quienes conciben que es una amenaza, no sólo para la humanidad sino para todo el planeta. Por otro lado, existen argumentos a favor de que la IA es un bien que nos puede ayudar a resolver algunos de los problemas que se nos presentan. Por ello, las posibilidades de que acontezcan escenarios favorables o no para la humanidad derivados del uso de la IA deben de examinarse desde un ámbito ambivalente.

La clasificación de la Inteligencia Artificial en la actualidad se encuentra limitada a: Inteligencia Artificial Estrecha, Inteligencia General Artificial y Superinteligencia Artificial (Al-Imam, A. et al., 2020). En la primera categoría (ANI por sus siglas en inglés) se encuentran aquellos sistemas que realizan tareas muy limitadas, por eso también se le conoce como IA débil, pero no significa que no esté dentro de las metodologías racionales, sino por el contrario, que justo pueden concertar reglas lógicas pero enfocadas en la especialización de una tarea. Por su parte, la IA General (AGI por sus siglas en inglés) es toda aquella mímesis tecnocefálica que puede operar de igual o mejor manera que una inteligencia humana promedio y que, por lo tanto, puede resolver problemas en un contexto determinado. Por último, la IA superinteligente (ASI por sus siglas en inglés) es de orden superior a cualquier mente humana, incluso, a cualquiera de las más brillantes que han existido a lo largo de la historia, ya que en teoría este tipo de software podría ser 100% autónomo y consciente por replicar las capacidades humanas, pero de forma exponencial. El estado actual en el que se encuentra actualmente cualquier IA confirma que sólo hemos desarrollado Inteligencia Artificial estrecha y que las dos categorías que le siguen son mera especulación ya que no se ha logrado desarrollar ninguna de ellas (Bradley, P., 2020).

Lo que es una realidad por ahora, es que este tipo de tecnología ha causado revuelo por los saltos cuantitativos de mejora que se han realizado en los últimos años, lo cual invita a generar perspectivas o escenarios no gratos para la vida humana. Desde luego, la ciencia

ficción ha contribuido a que el imaginario colectivo ponga en puerta situaciones en las que la Inteligencia Artificial supera al humano o bien lo pone en conflicto. Los libros o películas muestran escenarios catastróficos donde las máquinas dominan a la humanidad o están en combate para extinguirla gracias a la singularidad tecnológica. Este tipo de ideas son posibles porque se parte del presente para ir al futuro, pero con un valor predominante: el miedo generado por la desconfianza. Toda sensación de desconfianza que conlleva a creer que llegará alguna negatividad, es considerado como un miedo, mismo que está impulsado por una angustia real de un peligro latente. En el caso de IA, se crean ocasiones en las que se ha generado un impulso del temor hacia los softwares, que está fundamentado en la Superinteligencia Artificial. Cuando Hans Jonas (2019) hablaba de la heurística del temor, se refería al ejercicio práctico de reflexionar sobre lo que en este mismo momento es el humano y lo que le espera de él en el futuro, al crear una vinculación entre el presente y el mañana para generar una responsabilidad tras la interiorización de la incertidumbre de la estabilidad (Jonas, H., 2019). Desde luego, este escritor judío hablaba desde un horizonte ecológico, pero sin duda este concepto puede ser interceptado para el tema que nos ocupa. La heurística del temor puede llegar a ser una expectativa de irrupción de la tranquilidad humana en un futuro, que se desprende del desarrollo de la Inteligencia Artificial (Tibaldeo, R. F., 2015). El miedo que se señala es la consecuencia de la capacidad de la tecnología de modificar el mundo de manera severa. Por analogía, se cree que si hubo un desastre una vez, puede volver a suceder de nuevo y por tanto es viable la amenaza de una IA potencialmente dañina (que ha sido en el pasado) y que puede repetirse, cuando ese “daño” que ha creado a personas ha sido a causa de los humanos que la han programado de dicha forma. ¿La IA ha causado daño a las personas? Sí, por ello es válido hablar de una heurística del temor.

¿Cuáles son los riesgos reales de que la IA nos pueda aniquilar? Cuando se habla de una Superinteligencia Artificial se debe de entender que es un acto metafísico, es decir, es una conclusión basada en argumento de que los humanos han creado una tecnología que, en su momento, serán incapaces de controlar cuando se pueda integrar en una *singularidad* tecnológica, es decir, la creación de una tecnología totalmente autónoma del humano con la capacidad de manipular distintos recursos para lograr objetivos de su interés, según las definiciones de Alan Turing. Por estas razones se puede afirmar que es posible que se puedan

concentrar varios programas informáticos de manera asincrónica para dañar a los seres humanos sin que estos puedan generar un control (Alfonseca, M. et al., 2021). En ese sentido, la única forma de minimizar estos desencuentros es usando la ética dentro de la programación computacional para que los sistemas aprendan y siempre se mantengan en una línea de beneficio hacia el humano.

Si bien es cierto que la heurística del temor ha sido alimentada sobre todo por Hollywood, se debe de observar el lado opuesto a esta situación. En efecto, la actualidad muestra que la Inteligencia Artificial Estrecha (o débil) que prevalece hoy en día sería incapaz de hacer daño al humano, apelando a la primera ley de Asimov, puesto que se pondera un optimismo que hace concebir a esta tecnología como una herramienta que contribuye para el desarrollo del ser humano. Contrario al temor, se genera una confianza basada en una "domesticación". Al respecto, se pueden considerar la noción de Antoine de Saint-Exupéry sobre *domesticar*, con la cual se refería a "*crear lazos*" ("*apprivoiser"*) y cuya finalidad es conocer o descubrir el mundo.

En el caso de la IA, la humanidad se encuentra en un proceso de desarrollo y creación, el cual es muy probable que las generaciones venideras no lo tomen de esta manera, sino que para ellos sea algo "*dado en el mundo"* o bien "*natural*". Ante ello, la vinculación debería de irse formando de manera histórica como una tecnología que procura el bien hacia el ser humano, porque en efecto, no hay pruebas científicas de que la mímesis tecnocefálica quisiera hacernos daño. ¿Por qué una Inteligencia Artificial Suprema desearía hacer mal al humano si es programada desde la responsabilidad? Es por ello que la evolución que tenga la IA deberá de ser acorde a las necesidades del humano y no al revés, acercándose a una relación armónica como la que se ha dado con animales de compañía potencialmente peligrosos para los humanos y esa situación no ha llevado a la catástrofe mundial, porque al día de hoy "*nuestros gatos y perros no están planeado matar a todos los humanos*". El pesimismo de una IA queda suprimido por la realidad de que dichos sistemas no se modificarán a sí mismos para llegar a ser superinteligentes a pesar del crecimiento exponencial del software y hardware que se desarrolle en los próximos años (Bentley, P. J. et al., 2018).

Una consideración que pude ser errónea es que la IA quita empleos a las personas (Oppenheimer, A., 2018), debido a que no puede suceder en un primer momento porque en el trayecto la citada tecnología sólo es capaz de realizar pequeñas tareas, es decir, automatizar procesos repetitivos como clasificar, reconocer, tareas mecánicas o rutinarias. En todo caso, se puede asumir a la IA como una ayuda colaborativa al humano, ya que una vez que entra en funciones, el operador que se hacía esas tareas, se podría dedicar a asuntos más complejos dentro del mismo proceso que antes no le daba tiempo para ser asumidos, lo cual podría implicar también que ese humano desarrolle nuevas capacidades-habilidades más elevadas. Por lo tanto, la IA mejora al humano al ser su complemento porque lo ayuda -invita- a ser mejor. Ahora bien, hay quienes incluso hablan de que, cuando la IA haga la mayor parte de las actividades humanas, el *homo sapiens* podrá tener una vida de recreación (Bentley, P. J. et al., 2018).

Ni el pesimismo y optimismo en torno a la IA deben ser posturas contundentes, debido a que crean especulaciones. Por ello se propone la adopción de ambas perspectivas en la construcción de una mímesis tecnocefálica con perspectiva ética. No puede haber alguna consideración ética si no se piensa en los posibles riesgos a futuro (heurística del temor: miedo, como principio operativo) a partir de la examinación de las capacidades científicas existentes que podrían incrementarse para generar una tecnología responsable (optimismo) partiendo desde la realidad científica. La tecnología no debería de levantar la sospecha de que vienen tiempos peores, ya que es un instrumento que ahora mismo se usa para generar atajos a procesos mentales-computacionales de manera razonable orientados a resolver problemas o tomar decisiones anticipadas al futuro. La IA no es una desgracia para la humanidad, no se debería de tener fobia alguna siempre y cuando existan marcos éticos que garanticen certeza de procedimientos, máxime si creamos acciones desde ahora que nos "protejan" contra ella (acciones preventivas). El miedo cuando es tomado como una emoción negativa, pero que puede enseñarnos a preguntarnos qué *somos* y qué podemos *ser*, mientras que la responsabilidad es el método para autoreconocernos y crear (avanzar). Al observar esta realidad, se puede afirmar que hay una dialéctica entre realidad y miedo, que potenciaría las buenas prácticas entorno al desarrollo y uso de la IA. La heurística enseña a tomar conciencia de ello, pero enseguida debe de venir la calma por el estatuto de que el humano

es quien está desarrollándola. El miedo del que hablaba Jonas (2019) no es aquel que se detiene al momento actuar, sino el que invita a transformar o a hacer algo, por lo que el temor crea conciencia frente a la IA ya sea simbólica o subsimbólica.

Recientemente se ha puesto de moda el concepto de algoritmo como una situación que es capaz de generar grandes peligros y beneficios para la humanidad. Incluso se puede constatar que hay una confusión entre algoritmos e Inteligencia Artificial, ya que en ocasiones se toman como sinónimos, sin transparentar que los primeros son la metodología para que se genere lo segundo (Hill, R. K., 2016). En efecto los algoritmos y la inteligencia han existido antes que cualquier sistema electrónico. Por un lado, un algoritmo es una serie de procesos para concluir un problema, mientras que la inteligencia está asociada a la construcción lógica para abordar una situación. Entonces, el cuerpo humano cuenta por naturaleza con algoritmos para subsistir, como lo es el proceso de respiración, digestión o percepción del mundo. De allí la distinción entre la inteligencia y propiamente un algoritmo.

Cuando se habla de algoritmos en el terreno de la IA, es adecuado definirlos como una serie de reglas que ayudan a resolver problemas concretos por medio del procesamiento de datos en un ordenador imitando la función humana de la inteligencia (Turner, R., 2013). Entonces, más allá de la ciencia ficción presentada por Hollywood y de la heurística del temor anunciada por medios de comunicación y autores, la Inteligencia Artificial por ahora está limitada a la focalización de actividades específicas y no al dominio o destrucción de la humanidad. Lo que también es una realidad, es que dicha tecnología se hace cada día más presente en la vida cotidiana del ser humano, lo que no es novedoso porque ya desde hace varias épocas la ciencia ha posibilitado un entorno de instrumentos no naturales, que ayudan o colaboran con el ser humano. Para el año 2026, el mercado de la IA alcanzará los 23,426.3 millones de dólares a nivel global (Fortune Business Insights™, 2020), lo cual no es asunto de menoscabo porque se calcula que existen 347 millones de teléfonos celulares en el mundo (Canalys, 2021) que usarán algún tipo de Inteligencia Artificial.

Un ejemplo de lo anterior, se constata en el sector salud donde existen programas computacionales que ayudan a detectar enfermedades de manera muy precisa, incluso

superando al humano mismo (McKinney S. M. et al., 2020). De igual manera, ya están puestos en marcha diversos softwares que ayudan en el control del tráfico humano en las ciudades (Bustos, C. et al., 2021). En el terreno militar existe la preocupación creciente de obtener una ventaja tecnológica basada en IA para establecer batallas (Human Rights Watch, 2020). También hay plataformas o aplicaciones que ayudan a los humanos para desplazarse de un lugar a otro de manera más eficiente, como también se ha desarrollado Inteligencia Artificial que organizan la vida de los usuarios desde la alimentación, el pernoctar, la inversión económica, orientación en compra de vivienda y hasta la planeación de un embarazo.

Así pues, adicional a los beneficios pragmáticos, subyacen también algunos desafíos como el de la privacidad, la toma de decisiones o la autonomía del humano frente a los sistemas computacionales. Todas estas cuestiones son atravesadas transversalmente por el tema de la ética, por la cual se ha fomentado un diálogo para generar bases regulatorias y tener una Inteligencia Artificial responsable.

La interpelación del ser humano respecto a la Inteligencia Artificial en el mundo cotidiano ha provocado que, más que una novedad, exista una dependencia de esta mímesis tecnocefálica en la vida diaria derivada del creciente uso en diferentes entornos. Esta situación ha transparentado una serie de conflictos que antes ya existían pero que ahora se migran al ámbito computacional. En efecto, la aparición de los sesgos al momento de utilizar la IA es ya un asunto que preocupa a diversos sectores de la sociedad, especialmente, a los comprometidos con los Derechos humanos.

El sesgo es considerado como uno de los problemas relacionados con la recopilación o el procesamiento de datos que podrían resultar en decisiones prejuiciosas (Ntoutsi, E. et al., 2020). Así pues el sesgo es la poca capacidad de generar un sistema de IA para contemplar un panorama amplio, es decir, no considerar todas las posibles variantes al momento de generar un proceso dentro de la tecnología referida y que, en el caso de la ética, vulnera principalmente la dignidad humana. De forma novedosa, surgen dilemas que son plausibles

de abordar para generar una mayor comprensión de las dificultades al momento de deliberar en la interpelación de la IA cuando ella se apropia de la humanidad.

Sobre lo anterior, Google Vision® es un software que utiliza la IA para analizar imágenes a través de Aprendizaje Automatizado y clasificar los archivos en categorías determinadas. En abril de 2020, Nicolas Kayser-Bril mostró que dicha Inteligencia Artificial al procesar dos diferentes imágenes de manos humanas sosteniendo un termómetro, donde el único contraste era el color de piel, el programa reconocía como una pistola al termómetro en la imagen de un afroamericano, mientras que determinaba que era un aparato para detectar la temperatura en la imagen de la mano de la persona de color blanco (Kayser-Bril, N., 2020). Más allá de las conclusiones inmediatas de esta revelación, hay que decir que el software ya está siendo utilizado no sólo por particulares sino por empresas y gobiernos para fines de control de la información, lo que supone la existencia de otros sesgos más en el terreno real pero que aún no se revelan. A decir verdad, el algoritmo no es por sí mismo discriminatorio, sino que migró los prejuicios raciales que existen en la sociedad, especialmente, el de los programadores que la desarrollaron. Entonces, el abordaje ético se considera necesario porque de hecho la tecnología va develando lo que los seres humanos venimos cargando desde la *normalidad* y que se reproduce en una codificación, a veces binaria.

A raíz del confinamiento generado por el COVID-19, en agosto del 2021, se dio a conocer que la empresa Xsolla instaurada en el negocio de los videojuegos, despidió a 150 de sus empleados, por medio de un análisis realizado por una Inteligencia Artificial en su desempeño laboral (Yuzbekova, I. & Tairov, R., 2021). Según se puede distinguir en el comunicado que se les hizo llegar a las personas desafortunadas que perdieron su empleo, la decisión fue tomada por medio de algoritmos en el monitoreo de correos electrónicos, conversaciones, documentos, así como otro tipo de actividad y la tecnología etiquetó como "*empleado poco comprometido e improductivo*", lo cual representó su separación de la compañía. A primera vista surge el dilema de los límites de la supervisión del empleador hacia sus empleados en contraste con la privacidad. Si bien es cierto que se monitorean datos generados en horas del trabajo y sus propias tareas, ¿Cuál es la frontera entre lo que es propio de la intimidad del empleado, aunque trastoque sus funciones dentro de la empresa? ¿Hasta

qué punto la compañía puede meterse en la actividad personal de sus colaboradores independientemente si están o no están en su hora de trabajo? Por su lado, ¿Qué cosas se deben alinear a lo no productivo? ¿Qué datos son significativos para valorar un bajo desempeño laboral? ¿Bajo qué parámetros se etiqueta lo productivo de lo no productivo? ¿Quién determina qué es productivo?

En el mes de agosto del año 2021, la empresa Apple anunció la creación de una Inteligencia Artificial que ayuda a limitar la propagación del material de abuso sexual infantil (pornografía) (Apple, 2021), por medio de la examinación de imágenes y videos que se pueden llegar a reproducir en los dispositivos. De forma consensuada, se puede asentir al Artículo 34 de la *Convención sobre los Derechos del Niño* (UNICEF, 2006), y del Artículo 1 del *Protocolo Facultativo de la Convención sobre los Derechos del Niño* (UNICEF, 2000), donde se pondera como bien mayor la dignidad del menor, superando la privacidad que podría llegar a tener el usuario de un sistema operativo, lo cual apela a la ética de mínimos (Cortina, A., & Conill, J., 2014), pero queda la pregunta ¿Si hay interrupción de la privacidad por medio de la IA para un bien ético universal, es posible el mismo asecho digital para infringir la intimidad de cualquier persona con el afán de salvaguardar la soberanía nacional, detección de fraudes, combate a enfermedades, lucha contra el terrorismo o cualquier otra cuestión consensuada? ¿Quién puede frenar a un gobierno, por ejemplo, que busca oprimir minorías o dar dirección a las opiniones de sus ciudadanos usando IA en el marco de una dictadura digital? (Harari, Y. N., 2018).

Sumado a lo anterior, es posible examinar nuevos escenarios éticos derivado de la IA que se postulan ahora mismo y de los nuevos que vendrán, como es el caso de la patente US010853717 de Microsoft© (Microsoft Technology Licensing, 2020) que tiene por objetivo crear un *chatbot* que pueda comportarse como una persona por medio de la estela digital que va dejando una persona en concreto: video, mensajes, correos electrónicos, fotos y toda clase de información de dominio público encontrada en internet. ¿Hasta qué punto se puede usar la información dejada en la red (consentida o no) para imitar a una persona de forma digital? ¿Qué repercusiones éticas tendrá esta tecnología si se usa para recrear digitalmente a personas fallecidas y así generar una cercanía con sus seres queridos aún

vivos? ¿Cuáles serán las consecuencias legales cuando se use esta tecnología (y otras más) para construir una imagen “verdadera” ante los demás como es el caso del *deepfake*?

No obstante, de todos estos ejemplos de desarrollo tecnológico, se pueden encontrar otras realidades que demuestran el uso responsable de la mímesis tecnocefálica y que van más allá de la apertura de dilemas éticos.

Para aprovechar el creciente uso del celular, algunos investigadores de la Universidad Autónoma de Madrid desarrollaron un sistema inteligente que detecta de forma anticipada la aparición de alzhéimer o párkinson, ya que al sostener e interactuar con el dispositivo móvil se emiten ciertas frecuencias que son capaces de hacer comparativas históricas de datos del usuario, lo cual se traduce en alertas tempranas de que existen riesgos de padecer alguna de las patologías ya mencionadas y, por consecuencia, tomar decisiones a corto plazo en beneficio de la calidad del paciente. La creación de patrones a partir de parámetros e información biométrica, y del Aprendizaje Automático y Redes Neuronales Profundas, permite una opción tecnológica que beneficia la salud de las personas (Tolosana, R. et al., 2019). Desde luego, se podría potenciar el uso de sistemas similares si se combinan otros desarrollos emergentes como cámaras, sensores o cualquier otra tecnología emergente que recoja datos del usuario y pueda ayudar en pro de la salud.

El cáncer de próstata es una patología en la cual células anómalas dividen y destruyen los tejidos de quienes son afectados. Para hacer frente a esta realidad, se desarrolló un algoritmo que detecta el cáncer de próstata con una sensibilidad clínica para esta enfermedad de un 98,46% y especificidad de 97,33%. Esto se pudo lograr gracias al reconocimiento y contraste de imágenes extraídas de muestras de tejidos de pacientes ya diagnosticados, mismas que después sirvieron para entrenar una IA. La alta precisión del algoritmo implementado genera la certeza de que este tipo de ejercicios pueden ser usados para un diagnóstico confiable y por lo tanto para el bien del paciente (Pantanowitz, L. et al., 2020).

Con el uso de Redes Neuronales y cámaras se pudieron analizar y procesar imágenes de arañas que tejían su telaraña para determinar ciertos patrones de comportamiento de estos

artrópodos. La información numérica y de patrones del sistema implementado orientó a la comprensión del cerebro de estos animales, debido a que se descodificaron las reglas neuronales que operan al construir las estructuras de seda bajo la influencia de ciertas sustancias inducidas. Esto podría permitir desarrollar medicamentos y técnicas en beneficio de los humanos más adelante, sobre todo porque el acto de "tejer una telaraña" es ya un algoritmo complejo que se compone por fases y concluye en una arquitectura robusta. La Visión Artificial de este proyecto siguió el comportamiento de los arácnidos de forma muy precisa gracias al enfoque de agrupación no supervisado (Corver, A. et al., 2021). Por lo tanto, se puede llegar a la propuesta de que generar Inteligencia Artificial con una perspectiva de responsabilidad advierte beneficios para la humanidad a mediano y largo plazo.

Para valorar el rendimiento estudiantil dentro de algunos institutos de educación, se utilizó la IA para medir los riesgos de deserción de los estudiantes. Gracias al uso del Big Data, se ha podido predecir el rendimiento académico en escuelas de Portugal para tomar decisiones oportunas y generar políticas públicas. El modelo estructurado bajo la IA se enfocó en entrenar Redes Neuronales para predecir valores, contrastando los datos de entrada y de salida que proporcionaron los alumnos registrados en centros educativos. Con los resultados no sólo se pudo identificar los factores de riesgo que forjan una mayor incidencia para el abandono escolar, sino también aquellos alumnos potencialmente propensos a estar en esta situación (Cruz-Jesus, F. et al., 2020).

En el año 2017, el Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires lanzó un sistema de IA para contribuir en las tareas jurídicas de los funcionarios públicos. El software fue alimentado con datos del departamento judicial para que se pudiera generar una conversación con los funcionarios y así tratar de agilizar los trámites burocráticos en la deliberación de procesos judiciales. En este sentido, el *chatbot* ha permitido agilizar las tareas administrativas, al reducir los tiempos de sentencias hasta en más de 170 días, ya que el tiempo promedio de respuesta del sistema es de 2 minutos (Corvalán, J. G., 2018). Este tipo de iniciativas se han logrado gracias a que se han ponderado los Derechos Humanos en las propuestas tecnológicas dentro de la esfera de la vida pública, por lo que es posible de hablar de una IA predictiva al servicio de la justicia.

Si bien es cierto que podríamos seguir enlistando ejemplos de uso responsable y emergencias de dilemas éticos entorno a la Inteligencia Artificial como la gestión de capital humanos que pondera estereotipos de género, discriminación racial en sistemas policíacos o la polarización de los ciudadanos y genera tendencias en la esfera pública, es necesario concluir, por ahora, que se ocupa una perspectiva ética en el desarrollo y uso de la IA.

Por todo lo anterior, se considera que la mímesis tecnocefálica crea una expectativa en dos sentidos: los posibles riesgos y los beneficios. Para que se tenga confianza en el futuro sobre esta tecnología, se recomienda no solamente proceder de manera técnica desde la ingeniería computacional sino abordar el sentido ético para que se pueda suscitar confianza del software frente al humano.

1.4 De la coyuntura científica

En cuanto a que la Inteligencia Artificial forma parte de las ciencias, específicamente, de la ingeniería computacional y toca en varios sentidos la vida de los seres humanos, es necesario indagar sobre la correspondencia entre la ciencia y la ética. Es así que el propósito de la presente investigación es reflexionar desde la ética y sus posibles bases normativas para un desarrollo y uso de la tecnología. En ese sentido, se ha hablado de la posibilidad de construir una ética de la IA que sea aplicable a nivel global. Los esfuerzos de las organizaciones como la Unión Europea, IEEE, UNESCO u OCDE han iniciado el debate de la universalidad de pautas éticas de esta mímesis tecnocefálica, sin embargo, ¿es posible establecer una universalidad de principios? ¿Es suficiente basarse en convenios generales como los Derechos Humanos para perfilar una ética de la IA?

En ese orden de ideas, la consideración general es materia de un campo de conocimiento científico, y por lo cual es necesario indagar sobre el horizonte que conlleva ciertas afirmaciones, por ejemplo, la racionalidad, el empirismo, lo útil y lo cuantificable. Así pues, el neopositivismo que se ha permeado en esta época advierte que todo discurso que pretende

ser creíble o aceptado, necesariamente, debe de contar con al menos una de dichas características, es decir, que tenga en su base a la ciencia.

Descartes, Comte y más pertinentemente Kant, entre otros, han puesto de manifiesto que la condición de la ciencia es ser universal (Hertogh, C. P., 2016), es decir, que procure explicar los fenómenos que circundan al humano desde ese tipo de conocimiento racional. Por ejemplo, Kant advierte que se puede alcanzar la universalidad por medio de la Razón, es decir, por un imperativo categórico que esté basado en la autorreflexión como propios legisladores considerado ir hacia los demás (Hoffman, M. et al., 2015). Con ello, los imperativos hipotéticos evitan toda caracterización contingente del mundo y se deberían de considerar como lo opuesto a lo universal.

Con ello, toda postura empírica o material percibida desde la ética, sólo funcionaría con base en la experiencia y estaría limitada a lo que cada uno pueda recoger de ella, por lo que la universalidad se vería mermada al *espacio* y el *tiempo* (*a posteriori*). En cambio, aquello que sólo ocupa para *ser* es lo universal, es decir, en sí mismo el enunciado contiene la verdad sin necesidad de constatarlo en la realidad con los sentidos. Por ejemplo, “el triángulo tiene tres lados” y “actuar como yo quisiera que actuaran los demás” son dos proposiciones que contienen tautologías en sí mismas y, por lo tanto, siempre serán verdad. Por lo que la universalidad en la ética tendría que ser así mismo el deber por el deber; un criterio que todos compartan en cualquier parte independientemente de las categorías de *espacio* y *tiempo*, lo cual se podría postular para una ética que abarque a la IA.

Un problema de los universales desde esta visión es la caracterización, es decir, realizar denominaciones a través de la descripción de las cosas que percibimos, por ejemplo, una vez que ya hemos consensuado que nadie puede ser discriminado por su raza, sexo, color de piel, se puede diseñar una IA que no acentúe realizando preferencias por sexo, pero mientras sólo prevalezca una definición de qué es el sexo, se programará dicha tecnología para que funcione así. pero cuando deje de operar esa caracterización, es decir, cuando el paradigma de la sexualidad cambie y adquiera otra connotación o cuando se aplique dichos algoritmos programados desde la visión de un lugar del mundo distinto, ¿Podrán replicarse los mismos

códigos de distinción de sexo? (Little, A. C. et al., 2002) Por tanto, la descripción de lo universalizado (la forma) parece tornar una paradoja: se ocupa decir cómo son las cosas, pero al momento de describirlas, las particularizamos, es decir, le quitamos su universalidad.

Al parecer, subyace una nueva dificultad a la universalidad: la forma. En efecto, todo lo que aspira a ser universal debe de crear un molde donde "*lo parecido*" entre allí, por lo que las normas morales o toda pauta ética debería de estar caracterizada desde una idea que compartan todas las demás acciones-cosas similares. De esta manera, se puede crear un postulado ético como el siguiente: "toda IA que sea con fines bélicos, debe de ser prohibida", a lo cual podríamos preguntarnos ¿Qué es lo bélico? ¿Se limita a armas para la ejecución de la guerra? ¿A caso las *fake news* no son en esta época un asunto que promueven las polarizaciones, es decir, la guerra entre humanos? (Roozenbeek, J., & Van Der Linden, S., 2019). Y si un arma es todo aquel artefacto que daña a las personas, entonces los algoritmos que impulsan a las personas a ser compradores compulsivos y a generar deudas, las cuales por estrategia están cateterizadas para que así las instituciones crediticias puedan ganar y los clientes tengan que incluso llegar a quitarse la vida (Odgers, C., 2018), ¿No sería también esa mímesis tecnocefálica caracterizada como una arma más letal y pasiva que una pistola?

Por lo tanto, si lo universal se basa en las formas y todo lo parecido que pueda caber en ello, lo particular que hace referencia al contenido debe de ser antes caracterizado, desprendiéndose de lo universal para ser reconocido y saber si encaja o no en dicha forma. El dilema está que siempre lo universal comparte propiedades de lo particular, de lo pequeño y contingente al mismo tiempo. Para el caso de la ética, la situación se torna más compleja por un factor radical: la vida, ¿cómo una vida particular puede acoplarse a lo general?

En ese sentido, cabe preguntarse ahora si una ética de la IA que aspire a ser aplicable para todos, necesariamente debe basarse en principios universales como los Derechos Humanos. En otras palabras ¿Es suficiente postular a los Derechos Humanos como principio de una ética de la IA que sea aplicable en todo el orbe? Más puntualmente en el sentido que se ha presentado la argumentación en este texto, ¿Los Derechos Humanos son universales?

Si partimos de que lo universal es aplicable en todo momento y en todas las personas, entonces aquellos Derechos del humano no son necesariamente parte de la vida diaria de todos los humanos ahora mismo, ni mucho menos en otros tiempos, porque, de hecho, los Derechos Humanos son "*de un tiempo para acá*", porque pertenecen a la historia, es decir, tienen un principio, un desarrollo y un fin. En efecto, es posible que dichos postulados de la *Declaración de los Derechos Humanos,* al no ser infalibles o cercanos propiamente a la naturaleza, podrían ser suplantados por un nuevo consenso al ser abordados desde un paradigma y realizar otros Nuevos Derechos Humanos, por ejemplo, los digitales. A razón de que se equivoque el talente francés encaminado para los Derechos fundamentales, se podría preguntar ¿La libertada, igualdad y fraternidad serán siempre como las conocemos? Más concretamente, ¿La libertad, igualdad y fraternidad siempre han sido concebidas de la misma forma como ahora se nos presentan? Muy seguramente de los siglos XVI al XIX, los esclavos africanos sometidos por la corona española no coincidirían con la universalidad de los Derechos Humanos, ni mucho menos los uigures que forman parte de una minoría étnica en china y que son vigilados por medio de algoritmos de reconocimiento facial por su gobierno (Wright, N., 2018).

Entonces, en un mundo caracterizado por la diversidad de las particularidades, los Derechos Humanos difícilmente podrían ser universales aunque tengan como base el imperativo categórico kantiano. Tampoco serían una base para una ética global en el terreno de la IA al ser insuficientes porque no tienen la concurrencia de "*para todos en todo momento*".

Si esta realidad es así, ¿Por qué la insistencia de una ética de la IA global? ¿Por qué esmerarse en encontrar principios que direccionen cómo desarrollar y usar la IA de una forma afortunada para todos? La respuesta a estas preguntas está anidada en la correspondencia de IA y la ciencia. En efecto, la ciencia busca lo universal y como la IA se encuentra ubicada en el campo semántico de las ciencias computacionales, entonces se busca con insistencia alcanzar el idealismo, es decir, un mecanismo, una reducción matemática que pueda resolver los problemas que acosan al uso no responsable de la IA. Este anhelo es una ilusión de los profesores que enseñan IA a los alumnos para mitigar los impactos negativos de la IA y, también, es un deseo de los Comités de Investigación de las instituciones formadoras de

talentos en dichas tecnologías, ya que se espera formar de un solo golpe el esquema con el cual se logre anular toda acción ética dañina. La realidad es que ni la ciencia es universal, ni la ética es un formulario o un *check list*, a pesar de los esfuerzos importantes de crear herramientas para análisis exploratorios en la ejecución de proyectos de sistemas de IA (Denis, G., et al., 2021) o bien listas de verificación para el equipo técnico (Sánchez-Ávalos, R. et al., 2021).

En este momento, algunos intelectuales podrían aducir que al menos existen intentos de querer establecer directrices, por ejemplo, los Derechos Humanos, pero de allí a que se puedan tomar como universales es un tramo más amplio, debido a que los Derechos fundamentales siempre son condicionales, entre tantas cosas, a la soberanía. Sí, se puede leer en dicha Carta que toda persona es libre, excepto cuando el Estado decida caracterizar la pérdida de libertad en ciertas circunstancias, por ejemplo, el homicidio, la traición a la patria o la corrupción. ¿Qué es el derecho a la propiedad enmarcada en la Declaración Universal de los Derechos Humanos sino un deseo amplio a gozar de bienes materiales bien habidos? Esta realidad es aplicable al menos si la legislación local no despoja de sus bienes a las personas a causa de una expropiación o un bien superior de la nación. El reconocimiento biométrico desarrollado con IA es un claro ejemplo de lo que aquí se menciona, ya que diversos gobiernos han prohibido el uso de esta tecnología, excepto cuando se trata de un asunto de seguridad nacional, allí sí el derecho positivo encuentra una razón de existir al dotar a la soberanía de toda la facultad para suspender o anular cualquier universalidad de derechos (Holzgrefe, J. L., & Keohane, R. O., 2003).

Entonces, ¿Qué se puede hacer con estas directrices "universales" en un entorno de regulación ética de la IA? En el entendido de que los Derechos Humanos no son naturales, sino una invención avalada por el consenso de los representantes de los Estados, entonces, esas herramientas normativas deben de ser un punto de partida para el diálogo de cómo actuar, pero no la razón de ser para dominar en una ética global de la IA, porque siempre será ello un consenso, un convenio que atestigua las voluntades de algunas personas (Epstein, G. et al., 2014), pero no de todas. Así pues, "*No hay dioses en el universo, no hay naciones, no hay dinero, ni derechos humanos, ni leyes, ni justicia fuera de la imaginación común de los*

seres humanos" (Harari, Y. N., 2014), todo parece ser un discurso para tratar de dar orden al mundo, para deshacerse del caos que tanto mal genera a la paz en los humanos. En efecto, lo universal (metafísica) trata de dar tranquilidad al caos de la diversidad y, en terrenos éticos, procura encasillar toda forma posible de *ser*-actuar de manera *adecuada*. He allí el dilema de la ciencia, que aspira a ser general, pero se enfrenta con lo particular; trata de plantear reglas de programación para determinar de manera universal el actuar humano.

Así, la misma ciencia apuntala que es difícil reducir las decisiones éticas a un puñado de procesos computacionales posibles. El ya clásico ejemplo de los coches autónomos, más recientemente transformado en el experimento de Moral machine (Cunneen, M. et al., 2020), nos permite distinguir en los resultados que no siempre en las mismas circunstancias todo ser humano actúa de la misma forma. Aún hoy en día los sistemas de IA "fallan" en ese tipo de pruebas éticas, porque se comete el error de leer el caso como si fuera universal, cuando de hecho una misma persona puede tomar una decisión ética diferente al volante, tan sólo por sus experiencias, su aprendizaje acumulado, su desarrollo neuronal, su estado de ánimo, en fin. Evidentemente, la neurociencia nos explica que todas esas características cefálicas pueden ser cuantificables a través -quizá- de actuadores, pero de hecho la capacidad de los neurotransmisores como la dopamina determina el actuar ético, y ello depende de condiciones genéticas, emocionales, alimenticias y hasta de prácticas religiosas o culturales (Moratalla, N. L., & Sueiro, E., 2010). Es así que surge la proposición en la ciencia: "*es más probable que…*", la cual advierte que los programadores siempre están en búsqueda de la Verdad, a pesar de que están imposibilitados, más allá de las cuestiones técnicas computacionales, por el choque de realidades del sistema y por la carga de tantas variables éticas. De hecho, un acto ético puede ser bien apreciado para algunos como salvar la vida a una persona, pero para otros dependería del mérito de ser salvados en una situación de muerte.

Al ser atravesada la ingeniería computacional por el tema ético, se puede llegar a preguntar ¿Vale la pena caminar con este tipo de prejuicios? A lo que se puede responder con que no hay otro camino, ya que es el único: un intento de sistema de IA que todo el tiempo está luchado contra los defectos de la realidad amorfa, divergente y cambiante en la que se

encuentra el humano ubicado como el principal problema de la ecuación. En efecto, es viable que las personas que se encargan de desarrollar habilidades en la IA, ahora mismo, se vean decepcionadas por no alcanzar la Verdad o, dicho de otra manera, aceptar la imposibilidad de tener un sistema acabado y perfecto porque justamente el estado científico pretende encontrar el absoluto dentro del mar de divergencias, la multiplicidad de conciencias y las formas de actuar que siempre están atravesadas por una moral circundante que se inclina a prejuicios y realidades. Por ello se considera que se debe de hacer un ejercicio de reconocimiento de esta realidad para generar un estado abierto ante cualquier desarrollo, implementación o uso de mímesis tecnocefálica, con el afán de cambiar, modificar o anular los programas que sean incompatibles en el terreno ético. Por lo tanto, no debería de existir una decepción de los ingenieros en IA al comprender que cualquier sistema tiene la potencialidad de ser poco ético, muy por encima de su correcta programación. En este sentido los Derechos Humanos son un claro ejemplo de que hay motivación para reconocer, al menos, de que no son universales, y así ser punto de partida en el diálogo en torno a una ética de la IA independientemente de la latitud donde se discutan estas realidades.

Ahora bien, de los temas sobre filosofía, ciencia, IA y ética que nos conciernen, salta una problemática más aguda en torno a los Derechos Humanos en su carácter metafísico de presentarse como universales. Como ya se ha dicho, la forma y el contenido son características que distinguen lo universal de lo no universal y, en efecto, todas las cosas que quepan en la idea general, según el planteamiento platónico, son necesariamente aplicables para todos, lo cual supone que un esbozo de comportamiento universal debería de presentarse como un molde para hacer, pensar y actuar. Esto último crea por sí mismo la sospecha, al menos en ingeniería computacional como en otros campos del saber, sobre que los comportamientos humanos morales deberían de ser ubicados o clasificados por los sistemas de acuerdo con la aceptación de una mayoría, pero no necesariamente la mayoría o el indicador cuantitativo implica lo correcto en términos éticos (Peces-Barba, G., 1994). Así pues, podría darse el caso de que estos moldes, que no sólo los sistemas computacionales tienen sino en general las sociedades, traten de obligar, sujetar o adaptar los comportamientos a la forma que se llama universal. Entonces, la sentencia de "*todos los humanos son libres*" es necesariamente universal en la medida en que la libertad no se suspenda por el orden

jurídico o moral de una sociedad. Por lo tanto, no todos entran en el molde de lo que se puede estipular por universal, porque incluso estas formas que se presentan como algo ideal son desconocidas, o incluso contrarias a algunas prácticas locales o de subjetividades de los individuos. Éste es el mismo problema que aqueja al Principalísimo o cualquier sistema de pautas éticas, ya que minimiza la diversidad, lo que se convierte en un modelo idéntico de operar al de los Derechos Humanos: partir de conceptos indiscutibles, que encasillan-reducen a los seres humanos en directrices únicas a seguir. ¿Existe la autonomía o la justicia como puntos de partida para todo el mundo? En Noruega, por ejemplo, hay una alta valoración por la gente anciana lo que implica cuidados especiales (Daatland, S. O., 2015), sin embargo, sacrificar a un humano por los dioses era aceptable (Harner, M., 1977); en otras latitudes y en otros tiempos ha sido altísima la condena por asesinar a una persona adulta (Radin, M., 1920). La diferencia de estas realidades estriba en diferenciar los tipos de relativismos, sumergidos en el *espacio* y *tiempo*, que se han propuesto por diversos autores.

Si pensamos al relativismo como una postura que consciente la diversidad de puntos de vista, entonces se rechaza cualquier presentación de Verdad, basándose en la idea de que los individuos poseen características o circunstancias que los hacen percibir al mundo y actuar de diferentes maneras (Chaichian, M. et al., 2005). Para el relativismo, la Verdad no existe. Bajo esta concepción se puede entender que cualquier modo de abordar la epistemología siempre será incompleta o al menos imposibilitada para ser acabada. Por el contrario, todo conocimiento humano debe ser complementado desde posturas no consideradas en su inicio. Esta realidad no es una implicación menor en el desarrollo de una ética de la IA, ya que, si se considera este abordaje de realidades, se puede permitir estar alerta a los posibles sesgos no descubiertos en el momento del desarrollo o implementación de los sistemas en lo que refiere al terreno ético. Nuevamente aparece una paradoja en la construcción de la presente argumentación, debido a que la IA, al ser parte de la ciencia y en su pretensión de ser universal, debería de aceptar la contingencia, o al menos percibir la probabilidad en la dimensión no en términos cuantitativos, sino de la ruptura de paradigmas de acuerdo con las circunstancias donde se desarrolla o usa dicha tecnología.

Para clarificar aún más esta situación, es menester distinguir entre los diferentes tipos de relativismo. En un primer momento nos encontramos con el llamado relativismo radical, que es llamado así porque postula que los fenómenos presentados al humano no son iguales ante sí, sino que cada uno posee, independientemente de su forma o naturaleza, una interpretación diferente y válida sin que exista vinculación por los actos derivados de la toma de decisiones. Ni lo bueno ni lo malo existen como algo en sí mismos. Esta postura suele ser criticada de forma contundente porque da pie a que cada sujeto haga lo que mejor le convenga, sin pensar en las consecuencias de vivir con los demás, lo cual implica que nunca será posible ponernos de acuerdo, debido a que todo el tiempo se está respetando las diferencias de abordar un hecho. Por el contrario, el relativismo moderado indica que, por encima de las diferentes percepciones del fenómeno, existe la posibilidad de encontrar puntos de partida para direccionar la deliberación, es decir, es más flexible en el abordaje de respetar las diversas opiniones o formas de comprender el mundo. En tanto que, el relativismo epistemológico indica que hay una reducción a la conciencia en torno a la verdad o la falsedad, independientemente si proviene de una comunidad o persona. Muy particularmente, Chris Gowans (2004) presenta dos maneras de clasificar el relativismo. La primera de ellas es considerada como un relativismo descriptivo que está acotado por la discrepancia permeada en la cultura o creencias que se van formando colectivamente a lo largo de la historia, es decir, tienen como condición al *tiempo* en tanto un factor que fragua una idea que es válida para una sociedad, la cual va migrando estas posturas a las nuevas generaciones. En tanto que, el relativismo metaético asume que existen ideas que comparten grupos de personas a partir de la conciencia y construcción de sus propios argumentos, sin necesidad de recurrir completamente a consideraciones históricas, por lo que podrían elaborar sus propios conceptos para deliberar de una forma u otra (Gowans, C., 2004).

Aquí es importante aclarar que la ciencia no cree en la Verdad, pero pretende la Verdad, lo cual quiere decir que está buscando a través de sus metodologías alcanzar lo universal, lo ideal o absoluto, mientras existe un desplazamiento de paradigmas. Por ello es difícil que, dentro la comunidad científica que discute temas de ética de la IA, se pueda partir de relativismos, ya que esa realidad complica la metodología de análisis y clasificación de datos por la multiplicidad de variables y formas de concluir los sistemas.

Para ejemplificar lo anterior, se puede recurrir al discurso de la alteridad y otredad, porque si se considera al humano no como un extraño sino como "*un idéntico a uno mismo fuera de nosotros*", se tendría que deberíamos de considerar las formas de actuar para no perjudicar a los demás. De esta forma, se apela a que la singularidad de diversos abordajes ideológicos desemboca en "*el cuidado del otro*" como si fuera un "*yo fuera de mí*". Para la visión oriental de Confucio, el postulado se plantea como "*lo que no deseas para ti, no lo hagas a los demás*", en tanto que para el cristiano se formula "*amarás a tu prójimo como a ti mismo*", mientras que Mahoma indica "*desea a los demás, lo que deseas para ti mismo*" y Buda estipula "*no le haré a otro lo que no deban de hacerme a mí*" (Amnistía Internacional, 2021). Aunque se podrían citar otras visiones de cómo actuar frente a los demás, por ahora con esta información se puede mostrar que en el relativismo moderado se encuentran mínimos que pretenden ser universales para definir una ética global.

En efecto, las posturas de tratar a los demás como una consideración propia implican que "*el otro*" que "*soy yo*" es un fin y no un medio, al posibilitar al menos las máximas de: no matar, no mentir, no robar y no abusar sexualmente de los demás (Weltethos, S., 1993). Entonces, si se trataran de aplicar estos principios para guiar una ética en la IA se podrían, en la pretensión de universalidad científica, tener puntos de partida para guiar el diálogo regulatorio, siempre a reserva de que la aparición de la contingencia que nos trae el tratamiento de la vida, puede también cambiar el horizonte de las discusiones o postulados. Los empresarios o inversionistas que cuentan con el poder de desarrollar algoritmos computacionales que se manipulen a través de la perfilación de personalidades para ofrecer publicidad focalizada, ¿Desean que otro humano en su condición haga lo mismo que él hace? Es aquí cuando el imperativo categórico kantiano se muestra en aprietos o cualquier pretensión de universalidad ética, ya que las acciones dependen de una microfísica que incluso es inconsciente para los tomadores de decisiones: el cúmulo de voluntades para desarrollar la IA no necesariamente tiene una aceptación ética (Sylvia IV, J. J., 2020).

La cuestión científica de la ética en la IA es la universalidad, la cual no es imposible en enunciados analíticos, pero, por ejemplo, los Derechos Humanos o cualquier clase de

principios al ponerlos en práctica siempre resultarán insuficientes ante la diversidad de formas de *ser* y *actuar*. Cuando John Locke (Rollin, B. E., 2007) hablaba de una injerencia por naturaleza de valores en los humanos, pensaba en la vida, libertad y propiedad como *a priori*, pero en todo caso proponía consensuar principios mínimos razonables para enfocar el fenómeno de la interacción con los demás. Al retomar nuevamente a Kant, lo universal es *a priori* porque pertenece al ámbito de la Razón y siempre es principio y fin en sí mismo, aunque justamente esta idea es una propuesta razonable de que existen los universales, y esto se convierte en una exhortación para que los demás se convenzan de ello, es decir, se busca un consenso unificado de cómo tratar el tema. Salta ahora, lo consensual, los acuerdos o asentimientos que se tienen al sostener un diálogo con los demás. En este caso, se apela a la mayoría porque le parece lo suficientemente razonable para que el fenómeno sea abordado o aceptado de esa forma. La *Declaración Universal de los Derechos Humanos*, postulada por las Organización de las Naciones Unidas (1948), no es otra cosa que el resultado del choque de diferencias e igualdades, que se basan en la creencia de que tal cosa conviene más a la mayoría. El verbo *creer* queda aquí muy *ad hoc* porque, de hecho, el consenso se basa en apostar al futuro sin que aún sea posible. ¿Por qué lo universal tiene que ser consensuado? ¿Si es universal, no deberíamos de tener impresa la idea dentro de la naturaleza de esos principios que establecen algunas organizaciones o autores? Si tenemos que discutir las cosas para descubrir lo universal, entonces estamos dejando a la suma de voluntades racionales lo que es mejor para todos. En la IA, el asunto es aún más claro, ya que de acuerdo con la base de datos que se tenga, del análisis y la suma de reglas que se establezcan, el resultado siempre será lo que es más probable que se acerque a la Verdad, es decir, a lo universal. De este modo muchos ingenieros han creído que entre más robusta se encuentre la base de datos y mientras más adecuadas sean las reglas con las que se discriminan esos datos en programación, el resultado necesariamente debería ser más ético, porque se acerca a un consenso, lo cual no necesariamente es así. Por lo tanto, aquellas personas que dicen que sólo hay que partir de mínimos morales como los Derechos Humanos para guiar a la IA en el terreno ético, deberían de examinar la dificultad que esto representa para acercarse a la Verdad. En este sentido, los Derechos Humanos no deberían de ser descartados como base o principio de una ética en la IA, sino que, por su vocación de universalidad y por la paradoja de dicha universalidad, se

conforman como una ocasión para conversar sobre lo que conviene de momento a la sociedad en el desarrollo y uso de la mímesis tecnocefálica.

Existe un elemento más a considerar en esta cuestión científica de la IA: la ética está relacionada con la moral. Esta realidad muestra que la moral está inclinada a los actos que determinan lo bueno y lo malo, que devienen de una conciencia colectiva y, por ello se puede decir que es relativa ya que es incompatible con otras formas de ser en diferentes espíritus. Entonces, si la ética es un acto reflexivo atravesado por la moral, y si la moral es divergente a las categorías de *espacio* y *tiempo*, entonces no podría existir una ética universal. Si la moral es subjetiva, todo acto de reflexión queda predestinado a la opinión en el devenir de la historia, por eso Yuval Noah Harari tiene razón al afirmar que "*Dentro de cien años, nuestra creencia en la democracia y en los derechos humanos quizá les parezca igualmente incomprensible a nuestros descendientes*" (Harari, Y. N., 2016) porque la cuestión científica de la imposibilidad de universalidad en el terreno ético muestra que los principios pueden ser parte del método para deliberar en la IA cuando se habla de una ética global.

Por su parte, es necesario considerar que la ciencia es un concepto complejo de definir, pero para fines prácticos se puede decir que es el estudio de la estructura de las cosas naturales y la forma en que se comportan (Dictionary Cambridge, 2008). No obstante, se debe aclarar que la ciencia es un conocimiento, uno entre tantos más existentes el mundo y que se ha puesto por encima de otros porque ha explicado de buena forma el mundo y ha dado beneficios a la humanidad. La exhortación de Kant sobre la liberación de la tutela de la inteligencia de otros con su *sapere aude* es ya una ponderación en la Razón y en el alejamiento de la metafísica (Kant, I., 1784). Para Philippe Lacoue-Labarthe y Jean-Luc Nancy existe una imposición de la diosa Razón que representa un despojo de lo místico en un sentido vengativo (Nancy, J. L., & Labarthe, P. L., 2002), por ello la credibilidad del conocimiento científico sobresale sobre cualquier otro en la actualidad. No obstante, la promesa de la realización o felicidad que proyectó la ciencia no ha alcanzado a la mayoría de los humanos, ya que si bien es cierto que se ha prolongado la vida por medio de vacunas y nos aproximamos al futuro con predicciones más acertadas, por otro lado se han creado instrumentos de destrucción como las armas nucleares, entre otros que han traído desdichas.

Ahora bien, dentro del campo semántico de la ciencia, existe una disciplina que preocupa a la comunidad internacional: la Inteligencia Artificial. A partir del análisis clasificatorio de la IA de Russell y Norvig (2013), esta tecnología está basada en la imitación de la facultad humana de pensar (Russel, S., & Norvig, P., 2013). Al relacionar el pensamiento con la actividad cerebral, se puede decir por ello que es una imitación de las funciones cefálicas humanas como el pensar o razonar. Por tanto, el presente trabajo propone identificar a la IA con el concepto de mímesis tecnocefálica (*μίμησις-mímēsis*, *τέχνη-téchnē*, *κεφαλή- kephalḗ*), y así como una noción unificadora entre lo científico y filosófico, debido a que el discurso relacionado con "pensar" se refiere a la epistemología y, a su vez, dicha capacidad se entrelaza con lo lógico (argumentación, razonamiento y validez), que deviene de una tradición enraizada en Aristóteles y que el positivismo ha tomado como estandarte. Es conveniente preguntarse si la IA es ya una tecnología que ha superado al humano como principio de imitación de la sinapsis y permutación, sobre todo a la altura de los tiempos de desarrollo de computadoras cuánticas, procesadores, capacidad de almacenamiento, metodologías de Aprendizaje Automático, inmensas bases de datos y repositorios disponibles en la red.

Derivado de lo anterior, se puede afirmar que el impulso que ha tomado la ciencia se reforzó con el positivismo, creando no sólo una institucionalización de ese tipo de conocimiento, sino también un aprecio por lo físico (*φυσικός*): lo pragmático, cuantitativo y útil. No obstante, en la era actual, tras el desencanto de tal saber científico, está brotando una tolerancia entre la ciencia y las humanidades. Un ejemplo de ello es la manifestación que hizo Potter en *Bioethics: bridge to the future*, donde estableció el nacimiento de la bioética como la preocupación por la reflexión en torno a la vida como mecanismo para mejorar el ser de lo humano (Potter, V. R., 1971). La ética emerge como una necesidad vital y, en diversidad de casos, como un asunto normativo-obligatorio que ha llegado hasta lo punitivo.

En las razones expuestas, se puede encontrar el sentido al por qué en la actualidad diversas organizaciones a nivel global han estipulado marcos éticos en la IA, pero sin aún haber llegado a acuerdos generales. Según lo muestra la Universidad de Harvard, son más de 30

propuestas robustas de ética en torno a esta tecnología que destacan por su nivel de representatividad institucional en el orbe (Fjeld, J. et al., 2020). Nuevamente la divergencia irrumpe en el afán de la unificación de directrices éticas para la mímesis tecnocefálica.

Si se vive en un mundo hiperconectado, si hay una predominación de la Razón, si es ya es tolerada la unión de las ciencias y las humanidades, si se reconoce que la tecnología puede ser perjudicial como una heurística del temor, ¿Por qué no hemos encontrado acuerdos globales éticos para la IA?

Al realizar un análisis en la historia de occidente, sobresalen algunas características que configuran los acuerdos como humanitarios. En primer lugar, la Razón es el elemento por excelencia para mostrar y demostrar que justamente basándose en la herencia aristotélica se confirma la herencia de la lógica como un lenguaje que tiene premisas, sostiene una conclusión y muestra una coherencia si se siguen las reglas de la validez; la ciencia y las humanidades tienen sustentado su conocimiento en tal mecanismo, ya que nadie en dichos terrenos admitiría una propuesta ilógica o irracional como algo verdadero. En segundo lugar, gracias al idealismo platónico, se ha dado una ponderación a los conceptos como fuente para una realidad, ya que se nos ha dicho que para conocer, explicar y pensar algo, se ocupa primero realizar definiciones, al ser solo procesos mentales entrelazados y, por lo tanto, abstracciones. Luego, se devela un punto crucial, la aparición de la metafísica, ya que en el intento de objetivar la realidad se subjetiva todo, es decir, cada humano que quiere explicar "*la realidad*" primero tiene que pensarla, construirla desde su mente en categorías, por lo que la metafísica se convierte en el sustento de "*lo real*", así la lucha histórica por desprenderse de la metafísica lleva a ocuparla, por lo que la síntesis kantiana crea problemas en sí misma. En cuarto lugar y relacionado con lo anterior, la Verdad ha sido quitada del centro del conocimiento y se niega que exista porque, para una mirada científica, cualquier afirmación absoluta limita el progreso, al configurar la proposición de que "*no existe la Verdad*", es una Verdad y ello implica la negación de los universales, es decir, enunciados que son aplicables para todo y para todos. Es allí cuando surge la quinta consideración: la de la contingencia, que no es otra cosa que la contradicción de lo verdadero-lógico-coherente-racional en el mundo, lo cual crea confusión y extrañamiento ya que es justo donde la divergencia hace su

aparición, bajo el postulado de Nietzsche de "*no hay hechos, sólo interpretaciones*"; se abre la puerta para admitir la multiplicidad de realidades (Heit, H., 2018).

Aparecen ahora preguntas cruciales: ¿Cómo lo razonable puede echar mano de lo contingente para explicar la realidad? ¿Por qué la ciencia rechaza la metafísica haciendo metafísica? ¿Cómo es que se quiere hacer una ética de la ciencia apelando a lo universal, pero rechazando la Verdad? ¿Por qué lo abstraído es fuente de Verdad si el mundo físico es cambiante? Y más aún ¿Cómo se espera lograr un consenso ético global de la IA en un mundo divergente?

Si tratamos de ir al fondo del asunto, nos encontraremos con esa tolerancia a "*lo diferente*", porque como nadie tiene la Verdad, en todo momento todo puede ser Verdad. Esto implica que sobresalga la tolerancia como un baluarte en la actualidad, pero más que un logro a ratos pare ser un obstáculo porque extiende, irrumpe o inhibe el diálogo. El respeto a la libertad y autonomía de todos parece ser lo que subyace a este postulado, pero es justo poner en un mismo escenario a dos libertades o dos autonomías y eso crea conflictos. Esta confrontación de libertades es una disputa de razonamientos y entendimiento, de procesamiento de datos y de un acto de fe. A este encuentro de subjetividades se puede denominar una colisión de conciencias, toda vez que el choque entre autonomías sobrepasa el umbral de lo visiblemente consensuado o hasta tolerable.

En una sociedad, el consenso es la unificación de voluntades, es la democratización de las decisiones en una sola representación (Olsthoorn, J., 2020). Cuando se aspira a acuerdos, se añora lo universal. Una ética de la IA, al provenir del conocimiento científico, debe de estar basada en la pretensión de la universalidad y en el esfuerzo por alcanzar, como ya se mencionó, el estandarte de la Verdad. Para Habermas, la ética del discurso podría ser la respuesta, pero el problema en la ética de los universales consiste en que la democratización disminuye aún más a las minorías, por lo que esas particularidades que son aplastadas por las mayorías, quedan exceptuadas de dicha universalidad, por ejemplo, de los principios o pautas máxime si los sujetos diferentes son sometidos a la obediencia legislativa. La base de este razonamiento la podemos encontrar en la tautología: A + B = AB, donde en sí misma la proposición es adecuada (validez), porque "*toda figura geométrica que tiene tres lados es un*

triángulo". Para el caso que nos ocupa, todo algoritmo es una programación sistematizada que procura la Verdad en sí misma.

En efecto, un sistema computacional es "*bueno en sí mismo*" ya que produce conclusiones derivadas de una secuencia lógica en un acto moral (eso es la ética: una expresión elocuente, por lo que la IA parece ser un discurso-lenguaje también). Por lo que la discusión no debería de estar en las máquinas, sino en eso que está afuera del sistema computacional: el mundo-real y el humano que se nos presenta como una interpretación lingüística.

Así pues, las máquinas crean una moral, pero no tienen moral en sí mismas, sino que los procesos que se les asignan en una programación son el conjunto de reglas que le indica el humano, basado en la realidad misma que está determinada por la historicidad y plexos de referencia. Si el mundo es contingente y el humano es cambiante y determinado, ¿Qué se puede esperar de una IA basada en todas estas fragilidades? Más aun ¿Cómo se puede esperar que una IA sea buena en términos morales si la base para su creación es la corrupción y concupiscencia? Esta situación amorfa del mundo hace concluir a O'neil, C. (2016) que todo algoritmo es una opinión del programador, es decir, las instrucciones digitales expresan las opiniones del mundo de quienes las crean (O'neil, C., 2016). Por lo tanto, si la fuente de alimentación de la IA son los datos que recoge un programador de un mundo dinámico, no se podrá obtener una universalidad por más precisos-robustos que sean los *inputs* y los procesos consecuentes.

Cabe destacar que la explicación que realiza Derridá de la mano de Heidegger, en la que se menciona que la realidad está compuesta por interpretaciones del ser humano y que tiende a clasificar conceptos derivados de una contemplación de la naturaleza para explicar el mundo donde vive por medio del lenguaje (Bowden, B., 2020). Las matemáticas han sido culmen de esta clasificación por medio de un lenguaje específico que nos ayuda a ordenar la realidad. No por nada Leibniz consideraba que la lógica-matemática podría explicar y resolver la realidad (Chaitin, G. J., 2003). En el caso de la IA, el procedimiento es el mismo, ya que los datos que se recogen para hacer un algoritmo provienen de la realidad (fìsica-*φύσις*) pero se transforman en un lenguaje numérico (de programación) en el cual el sistema puede entender

(de forma lógica) una serie de instrucciones para concluir una predicción (futuro y decisiones). Entonces, la realidad (mundo físico) se transforma en metafísica (mundo simbólico) que se adentra al problema cartesiano de la *res cogitans* y *res extensa* donde subyace esta mímesis tecnocefálica que está ordenada por un binarismo que actúa en dicha realidad y el humano lo interpreta como un acto ético ("bueno" o "malo"), lo asocia a la moral dada en su subjetividad, y con ello devela la caída heideggeriana porque nos damos cuenta que en el fondo, no hay fondo, sino que hay una estructuración de divergencias que complican alcanzar lo universal. Todo esto nos lleva al mismo punto: una intersección de conceptos que dificultan la propuesta de principios globales para una ética de la IA.

¿Qué son los principios sino estandarizaciones para establecer lo universal? En bioética, la aparición del principialísmo no vino a resolver los dilemas éticos revisados porque de hecho la "excepción" se presenta en las discusiones. En efecto, toda pauta o principio no es universal porque en tal pretensión se encuentra con que "*existe por lo menos uno que…*" o "*no en todos los casos…*", así mientras que para unos el aborto debería de ser un asunto legislado en caso de violación, para otros el mismo acto constituye un atentado contra la dignidad humana gestante. ¿La no maleficencia, beneficencia, autonomía y justicia son aplicables en todos los dilemas éticos presentados? Si la respuesta es afirmativa, ¿Por qué desde el surgimiento del principialísmo no todas las personas lo aplican como método estandarizado?, ¿Por qué han surgido otras formas de análisis como la casuística?

La razón es sencilla porque el principialísmo no es universal. En esencia, lo que busca la universalidad-ciencia en el plano ético (más allá de una ética utilitaria) es que todos los humanos seamos iguales (estandarizados) a través de medias, algoritmos, ecuaciones y toda clase de distinción de datos, y también de una descripción lingüística de conceptos. Este último caso es exactamente el mecanismo de funcionamiento de la IA: discriminación de información para generar *outputs*.

Las predicciones que se generan por medio de la IA deberían de ser tomadas con cautela, puesto que la dificultad de normar éticamente el desarrollo e implementación de dicha tecnología, sumado a las intersecciones conceptuales que están detrás de ella, supone un

acercamiento a la propuesta cartesiana de claridad y distinción para no dar nada por verdadero antes que no se haya examinado su evidencia (Hatfield, G., 2008), evitando todo acto de fe (automatismo humano) al asumir la contingencia del mundo a pesar de que detrás de las decisiones de la mímesis tecnocefálica haya un proceso lógico-razonable que pretenda la Verdad o una condición necesaria, sobre todo porque la realidad es cambiante y en ella permea en un concepto aún más cuestionable: la vida.

1.5 Consideraciones ético-normativas de la IA

En vista de que vivimos en una sociedad que ocupa mecanismos para regular el comportamiento de sus integrantes, es necesario también indagar sobre las formas jurídicas en las que se está ensamblando la ética de la Inteligencia Artificial alrededor del mundo.

En la actualidad, no existe unificación internacional de directrices éticas que puedan regular el diseño y uso de la IA, pero sí se han logrado diferentes entidades, pautas o recomendaciones para interactuar con dicha tecnología.

En este tenor, el caso más sobresaliente es el de la Unión Europea, quien ha tomado el tema de la IA de forma institucional, al realizar discusiones desde el ámbito parlamentario. Tal fue el caso que el 16 de febrero del 2017, el Parlamento Europeo emitió una "*Resolución con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica [2015/2103(INL)]*" (Unión Europea, 2017), en la cual se exploran temas como: Principios generales relativos al desarrollo de la robótica y la IA para uso civil; Principios éticos; Una agencia europea -de AI-; Derechos de propiedad intelectual y flujo de datos; y, Medios de transporte autónomos; entre otros. El 10 abril del 2018, se firma la Estrategia europea sobre la Inteligencia Artificial denominada "*European Commission, "EU Member States Sign Up to Cooperate on Artificial Intelligence (AI)*" (Unión Europea, 2018), donde se realiza una trazabilidad de las acciones prácticas a seguir en materia de Inteligencia Artificial. Derivado de lo anterior, en junio de ese mismo año, se crea el Grupo de expertos de alto nivel sobre Inteligencia Artificial (AI HLEG), que está compuesto por conocedores de la materia que

encabezan las reflexiones de dicho tema. Tan sólo 6 meses después de este hecho, publicaron su primer "*Proyecto de Directrices éticas sobre una IA confiable*", que sirvió como elemento de análisis hasta abril del 2019, cuando se publicaron las "*Directrices éticas para una IA fiable*" en su versión final y de las cuales se estipulan como se sigue: Una IA fiable, Destinatarios y alcance, IA lícita, IA ética, IA robusta y El marco. En ese mismo mes, derivado de tal documento, la Comisión Europea inicia una prueba piloto para recolectar opiniones de las Directrices y así encontrar un consenso mundial respecto de la ética y la Inteligencia Artificial. Para que la Prueba piloto se pueda llevar a cabo, se propusieron 3 enfoques:

1. Siete requisitos esenciales para lograr una Inteligencia Artificial fiable;
2. Fase piloto a gran escala con los socios;
3. Creación de consenso internacional para la Inteligencia Artificial centrada en el ser humano.

Además, la misma Comisión, en ese mes, generó una ruta de reflexión por medio de "*Building Trust in Human-Centric Artificial Intelligence*".

Actualmente, como resultado de este intento de consenso mundial por parte de la Unión Europea, se están llevando a cabo foros en diversas partes del mundo para analizar las directrices éticas de la IA. Para abril del 2021, se proponen las primeras normas entorno a la ética de la IA en su Artificial Intelligence Act. (21 April 2021), para dirigir la confianza humana a la tecnología ya citada, considerando los riesgos que ella conlleva y estipulando estándares. De esta forma, toda posibilidad de la implementación de la Inteligencia Artificial se valora respecto de la magnitud de los daños o peligrosidad que se produzcan. Son cuatro las clasificaciones de estas situaciones:

1. Riesgo inadmisible: sistemas diseñados para manipular el comportamiento de los humanos con base en el uso de los datos personales.

2. Riesgo Alto: tecnología que vulnere infraestructuras críticas, educación, salud, trabajo, servicios públicos y privados, migración y fronteras, justicia y democracia, pero especialmente los sistemas de identificación biométrica a distancia.
3. Riesgo limitado: aquellos sistemas que requieren notificar a los usuarios que se está interactuando con IA.
4. Riesgo mínimo o nulo: aquellos que son inofensivos para los humanos y que forman parte de la vida cotidiana pero que no vulneran los derechos de los ciudadanos.

Estas clasificaciones se ocupan de que los sistemas de Inteligencia Artificial sirvan a las personas en general, pero no se ocupan de temas delicados como el uso de esta tecnología en el terreno militar. Es necesario destacar que estas clasificaciones no son un reglamento sino directrices (Floridi, L., 2021).

Más recientemente, en marzo del 2020, la UNESCO empieza a abordar el tema y designó a 24 expertos (por sus siglas en inglés: AHEG) para estipular recomendaciones globales sobre la ética de la Inteligencia Artificial, basándose en el estudio preliminar sobre los aspectos técnicos y jurídicos relativos a la conveniencia de disponer de un instrumento normativo sobre la ética de la IA. En el proceso de elaboración de las directrices éticas, se pudo integrar voces internacionales de la sociedad en general. A partir de lo anterior, el 25 de noviembre del 2021, fue aprobado el "*Proyecto de recomendación sobre la ética de la inteligencia artificial*" en el marco de la 41ª Conferencia General. Este *Informe de la Comisión de Ciencias Sociales y Humanas (SHS)*, en su Debate 3 y 4, Punto 8.2, establece lo siguiente (UNESCO, 2021):

Valores:

- Respeto, protección y promoción de los Derechos Humanos, las libertades fundamentales y la dignidad humana (dignidad como centro de acción).
- Prosperidad del medio ambiente y los ecosistemas (preocupación por el entorno).
- Garantizar la diversidad y la inclusión (proteger la diversidad e inclusión).
- Vivir en sociedades pacíficas, justas e interconectadas (promover la paz humana, justicia e interconexión).

Principios:

- Proporcionalidad e inocuidad: Adecuar todo desarrollo e implementación al contexto para lograr su legitimación, garantizar los Derechos Humanos; decisiones fiables en humanos; y, que la IA no debe de usarse para calificación social o vigilancia masiva.
- Seguridad y protección: Prevenir y evitar daños en implementación de IA.
- Equidad y no discriminación: Promover en desarrolladores y usuarios la justicia social, salvaguardar la equidad y luchar contra todo tipo de discriminación.
- Sostenibilidad: Considerar los Objetivos de Desarrollo Sostenible en implementación de IA.
- Derecho a la intimidad y protección de datos: Generar el respeto a privacidad en recopilación, usos y almacenamiento de datos, con ayuda de gobernanza.
- Supervisión y decisión humanas: Ponderar la trazabilidad para generar responsabilidades éticas y jurídicas para rendir cuentas.
- Transparencia y explicabilidad: Mostrar de manera sencilla cómo funcionan y toman decisiones los sistemas de IA.
- Responsabilidad y rendición de cuentas: Asumir consecuencias y atribuir responsabilidades derivadas de la operación de los sistemas de IA.
- Sensibilización y educación: Socializar conocimiento de IA y sus impactos en esfera pública.
- Gobernanza y colaboración adaptativas y de múltiples partes interesadas: Adoptar proyectos multidisciplinarios de IA.

Así pues, la UNESCO pondera cuatro dominios fundamentales: Derechos Humanos, medio ambiente, sociedad justa, inclusión y diversidad. Desde luego, la aplicación de estas disposiciones es voluntaria mediante la adopción de las medidas adecuadas en particular aquellas del orden legislativos o de otra índole que puedan ser necesarias. Para ello, se debe de considerar que los Estados Miembros establezcan medidas eficaces como marcos o mecanismos normativos de evaluación del impacto; la previsión y dispositivos eficaces de protección así como todo mecanismo que propicie marcos éticos; el fomentar buenas prácticas en sector público y privado; la elaboración de estrategias de gobernanza en

desarrolladores de IA; la promoción de datos abiertos; la priorización de ética de la IA por parte de Estados Miembros y las empresas transnacionales; la contribución para lograr la igualdad de género (no estereotipos de género y los sesgos discriminatorios); el aliento de la cultura y conocimiento; el acceso a información; el mejorar la economía y trabajo; y, la promoción de salud y bienestar social (UNESCO, 2021).

En el año 2017, el *Institute of Electrical and Electronics Engineers* (IEEE), a través de *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, publicó su primera edición de "*Ethically Aligned Design*" (IEEE 2017). Este documento proporciona recomendaciones e ideas directrices centradas en el ser humano para los generadores de tecnología, docentes y gobernantes que tengan como eje central la Inteligencia Artificial. En 2019, se presenta la Segunda edición de dicho documento denominado "*Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (A/IS)*" (IEEE, 2019), en el cual el IEEE no sólo rescató la esencia de su primer escrito, sino que abrió la discusión para generar reflexiones sobre las pautas éticas globales de la IA y, al mismo tiempo, propuso un estándar (IEEE P7000™) y la certificación de programas (Spiekermann, S., 2017). En esencia, este tipo de acreditaciones busca el fortalecimiento de una sociedad universal más allá de las diferencias de pensamiento o creencias de cada individuo y es recomendado a creadores de tecnología, instituciones de educación y organismos de planeación para proteger los Derechos Humanos, generar bienestar en las personas, realizar una buena gestión de datos, promover la efectividad de los sistemas, crear sistemas transparentes, ponderar la responsabilidad, no usar este tipo de sistemas de forma indebida, coadyuvar con la formación integral de los programadores, fomentar la sustentabilidad, proteger los datos personales, crear estrategias para informar a la sociedad y estipular valores (Vakkuri, V. et al., 2019)

Al ser un tema contemporáneo, se están generando otros esfuerzos para proponer pautas éticas respecto a la IA, por ejemplo, el *Future of Life Institute* en 6 de enero del 2017 en California, a través de la *Conferencia de Asilomar 2017* (Asilomar, A. I., 2019), dio a conocer sus propias pautas, también conocidas como los *Principios de Asilomar*, que comprenden 23 directrices para ser consideradas en los próximos años por la comunidad internacional en

torno a la IA. Estas pautas fueron redactadas por expertos a nivel mundial de la iniciativa privada, así como académicos.

Otro caso similar es el de la "*Declaración de Montreal para un desarrollo responsable de la inteligencia artificial 2018*" (Montreal Declaration Responsible AI, 2018), que establece un marco ético en el desarrollo e implementación de IA para incentivar los beneficios derivados de la IA, así como recopilar opiniones a nivel mundial sobre el desarrollo equitativo, inclusivo y ecológicamente sostenible de la IA.

Ya para el año 2017, la UNI Global Union publica "*Top 10 principles for ethical artificial intelligence*" (Union, U. G., 2017), documento que estipula una visión de directrices éticas que deben de tener los creadores de AI.

El 23 de mayo del 2019, se aprobaron las "*Recommendation of the Council on Artificial Intelligence*", también llamados "*Principios de la OCDE sobre la Inteligencia Artificial*", por parte de la Organización para la Cooperación y el Desarrollo Económico, en las cuales se establecen principios de políticas internacionales para alienar normas que beneficien a la humanidad con un planteamiento de los sistemas de IA que se diseñen de manera robusta, seguros, imparciales y fiables. Cabe destacar que es uno de los dos documentos internacionales que de momento interpela directamente y de forma institucional a México por ser miembro de la OCDE. Los cinco principios que estipula este organismo son: crecimiento inclusivo, desarrollo sostenible y bienestar; valores y equidad centrados en el ser humano; transparencia y explicabilidad; robustez, seguridad y protección; y, responsabilidad (Instruments, O. L., 2019).

Una mirada interesante al hecho sincrónico fue lo que sucedió en junio del 2019, cuando el Ministerio de Ciencia y Tecnología de China y el Gobierno municipal de Pekín, en conjunto con universidades como la Academia de Inteligencia Artificial de Pekín y la iniciativa privada (Baidu, Alibaba y Tencent), publican sus *Principios sobre la IA de Pekín*. En dicho documento, se muestran las directrices que el gobierno comunista perfila desde la privacidad humana, la dignidad, la libertad, la autonomía y los derechos, y que deben ser suficientemente

respetados. Por sí mismo, dicho manifiesto luce muy occidental y en contradicción con idearios comunistas, pero de momento es un buen referente para acercarse a la reflexión del tema. Para septiembre del 2021, el gobierno chino presentó el documento "*Especificaciones Éticas para Inteligencias Artificiales de Nueva Generación*", en el cual se estipulan seis principios:

1. Mejorar el bienestar de la humanidad.
2. Promocionar la equidad y la justicia.
3. Proteger la privacidad y la seguridad.
4. Asegurar la controlabilidad y la confiabilidad.
5. Fortalecer la rendición de cuentas.
6. Mejorar la educación en ética.

Con ello, se pretende evitar prejuicios que conlleven a la discriminación, se permite gestionar de forma adecuada la IA por parte de los desarrolladores, así como asegurar la privacidad y la no filtración de información (National Governance Committee for the New Generation Artificial, 2021).

De manera particular en México, se realizó un esfuerzo multisectorial encabezado por C-Minds que dio como fruto la primera *Agenda nacional mexicana de inteligencia artificial*, en la cual se destacan algunas consideraciones éticas en la IA:

- Resguardar los Derechos Humanos por un organismo autónomo.
- Establecer estándares mínimos y mecanismo de auditoría.
- Suscitar monitoreo de IA por parte de la sociedad.
- Promover ecosistema de ética en IA (comités de ética y asignaturas de ética en Programas Educativos).
- Crear herramientas para seguimiento de acuerdos internacionales en la materia.
- Promover inclusión de minorías y grupos vulnerables.

De este modo, la coalición ciudadana denominada IA2030Mx ha establecido directrices generales para el quehacer público de una ética de la IA (Del-Pozo, C. M. et al., 2020).

Del mismo modo, es necesario destacar que en México la Dirección General de Datos Abiertos presentó la "*Estrategia de Inteligencia Artificial de México 2018*" para cumplir los acuerdos internacionales, que incluye un marco de gobernanza, un reconocimiento de necesidades en la materia, una generación de liderazgo en OCDE y D7, una publicación de propuestas derivadas de consultas abiertas y la inclusión de consideraciones de expertos. Como producto de esta iniciativa, se presentan los resultados de la "*Consulta pública referente a los principios y guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la administración pública federal*", con el objetivo de "*fortalecer el uso responsable y ético de la AI bajo Principios Generales y la Guía de Análisis de Impacto para el desarrollo y uso de sistemas con elementos de Inteligencia Artificial en la Administración Pública Federal en México*" (Dirección General de Datos Abiertos, 2018).

En abril del 2018, se publica "*Hacia una Estrategia de IA en México: Aprovechando la Revolución de la IA*" por parte de diversas instituciones encabezadas por la Embajada Británica en México, donde se enmarcan las áreas de oportunidad para que México camine en la ruta global de la IA y se incluyen las recomendaciones en cinco áreas: gobierno y servicios públicos; datos e infraestructura digital; investigación y desarrollo; capacidad, habilidades y educación; y ética (CIty, B. E. M., 2018).

El gobierno federal de México implementó en noviembre del 2013 una *Estrategia Digital Nacional* que proyectaba una serie de acciones para consolidar un país digital con ayuda de la tecnología y la innovación para alcanzar metas como nación. Fueron cinco los objetivos de dicha planeación: conectividad, inclusión de habilidades digitales, interoperabilidad, marco jurídico y datos abiertos. Fue este marco institucional el que permitió abordar a diferentes instituciones en México el tema de la tecnología como un punto para potenciar el desarrollo de la sociedad. Para el año 2015, México se suscribió a la *Carta Internacional de Datos Abiertos* con la finalidad de adentrarse en la transformación global del aprovechamiento y el potencial de las tecnologías y así impactando de forma positiva a la economía y en general a la sociedad. No fue hasta el año 2019, cuando la Secretaría de economía consolidó *Data México*, como un instrumento de concentración de datos públicos

y base para una política pública. El 15 de diciembre del 2021, el Congreso de la Unión a través de la Cámara Federal de Diputados aprobó las reformas a la *Ley de Ciencia y Tecnología* en sus Artículos 2, 5 y 6 para que el desarrollo científico en el país sea desde un terreno ético y respete a los Derechos Humanos. Esta acción permitirá comenzar a crear marcos regulatorios de la IA desde el terreno ético (Cámara de Diputados, 2021).

De forma general, ha habido trabajos que influyen en el horizonte de la IA y la ética, pero que no deberían de ser considerados como estudios especializados porque hablan sobre el diseño y uso de software, entre ellos: *Principios de innovación* de la UNICEF (2009); *Principios de Greentree de mHealth* (2010); *Principios de diseño de servicios digitales del gobierno del Reino Unido* (2012); el Foro de *Principios para el Desarrollo Digital y los Principios Digitale*s por parte de organizaciones como Fundaciones, SIDA, UNICEF, PNUD, Banco Mundial, USAID y OMS (del 2015 al 2017) y el *Programa de Generación IA* de la UNICEF (2018).

Por otro lado, la Organización Mundial de la Salud (OMS), en junio del 2021 hizo del dominio público su guía denominada "*Ética y Gobernanza de la Inteligencia Artificial para la Salud*" en la cual permite mostrar criterios sobre cómo debe de implementarse la IA en torno a la salud para que pueda tener un sentido ético, debido a que dicha tecnología se ha presentado en el área del bienestar y equilibrio humano. En total, son seis principios que se presentan a continuación:

1. Proteger la autonomía humana.
2. Promover el bienestar y la seguridad humanos y el interés público.
3. Garantizar la transparencia, la explicabilidad y la inteligibilidad.
4. Fomentar la responsabilidad y la rendición de cuentas.
5. Garantizar la inclusión y la equidad.
6. Promover una IA que sea receptiva y sostenible.

Con estas pautas, se procura dar asistencia sanitaria con el uso de la tecnología en el trayecto de buscar la salud universal en el diagnóstico, atención, tratamiento y seguimiento de los pacientes (World Health Organization, 2021).

Para el año 2019, el Departamento de Defensa de los Estados Unidos consolidó sus principios denominados "*Principios de IA: Recomendaciones sobre el uso ético de la Inteligencia Artificial por parte del Departamento de Defensa*", en los cuales se propone un diseño, desarrollo y despliegue desde la responsabilidad sin dejar de enfatizar en la seguridad del país. Los cinco principios son: responsabilidad, equidad, trazabilidad, fiabilidad y gobernanza. Con ello, las fuerzas armadas se proponen reflexionar en los albores de una sociedad tecnologizada, pero actuando con responsabilidad (Board, D. I., 2019).

En otro caso, el Vaticano también ha postulado sus propios principios en alianza con la iniciativa privada. En febrero del 2020, se publicó el "*Llamamiento de Roma para una Ética de la IA*", donde se pueden apreciar seis criterios para tener una IA responsable: transparencia, inclusión, responsabilidad, imparcialidad, fiabilidad, seguridad y privacidad. Con estas bases se procura potencial una mejor convivencia social y bienestar personal a partir del aumento de las capacidades humanas desde una *algorética* (enfoque de ética desde el diseño de IA) promoviendo una IA responsable (Pontificia Accademia per la Vita, IBM e Microsoft, 2020).

No obstante, de las iniciativas sobre ética en la IA ya mencionadas, existen otras tantas organizaciones e iniciativas privadas, así como centros educativos y de investigación como el Instituto Tecnológico de Massachusetts y la Universidad de Navarra, lo cual advierte un diálogo para reflexionar en torno a la mímesis tecnocefálica, sobre todo por la irrupción en la autonomía del ser humano desde el terreno ético. En el horizonte de la multiplicidad de propuestas, sobre salen la privacidad, gobernanza de datos, transparencia, uso responsable y dignidad humana, ya que son los conceptos que más se entrelazan en las propuestas institucionales arriba mencionadas. En el marco del *Reglamento General de Protección de Datos* (Unión Europea, 2021b) y de la *Directiva sobre la privacidad y las comunicaciones electrónicas de la Unión Europea* (Unión Europea, 2021a), la privacidad es la gestión

adecuada de datos de los usuarios para que no le causen ningún perjuicio en su seguridad y confidencialidad, por lo que se establece que toda Inteligencia Artificial debe de contar con un soporte que no vulnere la intimidad del individuo y que no lo ponga en riesgo. Por su parte, si consideramos que los datos son la principal fuente de accionar de los algoritmos, entonces la gobernanza de datos es todo un mecanismo que procura una calidad de estos para generar un control adecuado de dicha información y para evitar posibles sesgos. En tanto que, la IA es un conocimiento técnico que puede llegar a ser muy elaborado al crear las llamadas cajas negras, es importante generar transparencia en las rutas para determinar cómo actúa la programación realizada y que en última instancia modifica el mundo real. Debido a que el bien común es una meta en la cual existe concurrencia global, se ha propuesto que toda IA esté focalizada para contribuir en la resolución de problemas que aquejan a la humanidad y no para provocar maleficencia. Y, por último, se apela que toda mímesis tecnocefálica tenga como centro al ser humano y que sea eje transversal en el desarrollo y uso de dicha tecnología, tomando como referencia al menos los Derechos Humanos.

Para complementar lo anterior, se propone crear y observar mecanismos de certificación o acreditación sobre ética de la IA, sobre todo para los desarrolladores e instituciones de educación; además, se considera necesario una mayor capacitación en ética dentro de los Programas de Estudio donde se forman los programadores de IA; no menos importante, es garantizar la claridad y distinción de los axiomas de autonomía y libertad para separarse de la manipulación; de igual manera, se cree conveniente que las entidades desarrolladoras de IA cuenten con códigos de ética debidamente discutidos al interior de las organizaciones observando las normativas que se aplican en el tema; no menos importante es el empuje que tiene la socialización de la ética de la IA para que en la dispersión de información se pueda educar al ciudadano y así garantizar un blindaje epistémico; y, de forma primordial, actuar con responsabilidad social en el desarrollo y uso de IA, lo cual debe de llevar a las personas, sobre todo a los programadores, a preguntarse: ¿Cómo afecto a los demás con el diseño e implementación de la IA?, ¿Para qué me sirve la IA? ¿En qué se va a ocupar la IA?

Puesto que la única vía para alcanzar los acuerdos de cualquier marco regulatorio es el lenguaje, es necesario ponderar los mecanismos para entablar un diálogo en la multiplicidad

de formas de abordar la ética en la IA para tratar de dar respuesta a la pregunta ¿Si ya la humanidad ha consensuado valores mínimos, por no llamarlos “universales”, por qué en la realidad no nos ponemos de acuerdo con la gestión tecnológica que se nos presenta ahora? Porque la única forma de acceder a la realidad es el lenguaje que goza de dos características fundamentales: es *impersonal* y es *único*. Lo anterior genera pensar que cuando se discute sobre IA y ética, los discursos están elaborados desde el horizonte y desde los intereses de las personas y, una vez que se tiene planteada la propuesta a enunciar, se da uno cuenta que esas manifestaciones “universales” realmente no lo son, sino que pertenecen a sólo un grupo de personas que piensan como uno mismo (comunidad), porque siempre el lenguaje-discurso tiene más que un sólo significado y no le pertenece a nadie. Por eso el razonamiento de Hobbes es muy plausible al considerar que las leyes son un medio para resguardar los intereses diferentes entre los ciudadanos; cada uno puede actuar conforme a sus instintos, pero debe de haber algo que sea depositario de ese control social (Pinilla, K. F., & Cordero, J. M. C., 2017)

En lo práctico, aunque existen diferentes formas de abordar esta situación, se propone al menos considerar dos cuestiones. En primer lugar, la llamada soberanía que el derecho positivo procura, ya que, más allá de los acuerdos globales, existe la venia de las propias regulaciones del Estado, por lo que suele ser un impedimento para aplicar los marcos internacionales. Por su parte, aquello que llamamos interés, es una premisa para la actuación en cualquier contexto ético, más cuando el tema que nos ocupa atraviesa lo económico, privado y público, la cultura, historia, legislación y moralidad, por lo que la IA en su desarrollo e implementación está relacionada con la utilidad que la subjetividad determina o, dicho de otra forma, con el grado de beneficio de las particularidades (personal y organizacional). Si se consideran estas consideraciones -además de los mínimos morales-, es posible generar una comprensión que postule la apertura de las subjetividades.

Adicional a lo anterior, en los grupos donde se habla sobre ética de la IA, se deben de considerar algunas cuestiones filosóficas como puntos de partida, a saber:

a) Razón: Sistema de interpretación que filtra la validez de un lenguaje propositivo en el mundo digital.

b) Abstracción: Realización de conceptos provenientes del mundo material para la generación de un mundo digital basando en el idealismo.
c) Metafísica: Discurso que da sustento al mundo físico-real, desde el terreno computacional, para generar tranquilidad del ser humano en el mundo digital.
d) Verdad: Proposición basada en el aforismo de sí mismo y que genera una aporía en la pretensión de universalidad o tautología.
e) Contingencia: Extrañamiento del ser humano por el mundo que lo circunda, basada en la contradicción, confusión, divergencia y dinamismo de la vida y que hace crear divergencia o sesgos.

En consecuencia, de manera muy concreta se propone actuar desde dos hilos conductores para alcanzar una ética de la IA. El primero de ellos se vincula con la pregunta ¿Quién crea a la IA? La respuesta es: el programador, por lo que cualquier irrupción ética que devenga de los desarrollos que realice, considerando la base del interés arriba mencionado, siempre sus resultados informáticos tienen una estrecha relación con su forma de ver el mundo, lo cual postula que no hay moral algorítmica, sino una transparencia de la realidad al mundo digital ya que si la IA da como resultado alguna discriminación, es porque en la sociedad existen datos que alimentan las reglas de programación de esa forma. Por lo que se propone apuntalar la formación ética de los desarrolladores de IA para disminuir controversias causadas por dicha tecnología. Por su parte, para abordar la pregunta ¿Quién usa la IA? está la otra entidad que actúa en este binomio: los usuarios, que son todas las personas que están siendo interpeladas por la IA y que pueden llegar a desconocer la forma de actuación que crea problemas éticos, por lo que se ocupa una socialización del conocimiento del tema, para que se puedan generar una mayor autonomía en la toma de decisiones en la colisión de las conciencias provocada por el encuentro disruptivo de dicha tecnología (Rodríguez-Reséndiz, H. 2020b).

Por su parte, si se piensa en la Inteligencia Artificial dentro de las políticas públicas, es necesario considerar dos vertientes: el uso de sistemas de IA para ayudar a la sociedad; y, la IA en un marco de política pública regulada. En el primer abordaje, se debe de considerar a la IA como algo útil ya que sirve como herramienta para solucionar problemas sociales en el

ejercicio de las políticas públicas, como puede ser el caso de un sistema para ayudar en la educación de estudiantes de niveles básicos. En ese caso, se debe de considerar las etapas del ciclo de vida de la política pública y su intersección con la IA:

1. Detección y delimitación del problema.
2. Planteamiento de la política y sus acciones. Aquí es donde expertos y tomadores de decisión deben concertar o no en la implementación de un software con Inteligencia Artificial para contribuir con el bien social.
3. Implementación de la política pública.
4. Evaluación de la política pública.

En tanto que, es necesario considerar los ciclos de vida de los sistemas de IA:

1. Conceptualización y diseño: Establecer de objetivos y planeación de acceso a los datos suficientes y necesarios con calidad para operar el sistema.
2. Recolección y procesamiento de los datos: Verificar la disponibilidad y calidad de los datos.
3. Desarrollo del modelo y validación: Proponer un sistema robusto y variado que satisfaga los objetivos alcanzados por el modelo algorítmico.
4. Uso y monitoreo: Ejercer la propuesta en marcha de la propuesta, realizando evaluaciones para alcanzar mejoras en el desarrollo.
5. Rendición de cuentas: Generar informes del sistema en la implementación del mismo.

La suma de estas dos variantes, más el contexto global de la ética de la IA, pueden permitir acercarse a la respuesta sobre la Inteligencia Artificial y su posible ayuda para resolver problemas comunes en la sociedad, al considerar siempre a la ética como la guía de actuación, especialmente, y al poner atención en aquello que sucede en China, Estados Unidos y Europa, y sus realidades económicas, importancia del sector privado, nivel redesarrollo tecnológico, acumulación de datos de las personas y sus robustas recomendaciones éticas, máxime que al parecer el tema legislativo-normativo se queda atrás respecto del avance tecnológico.

CAPÍTULO SEGUNDO

2. Una epistemología de la IA

Ahora mismo, y en un futuro previsible, no somos lo suficientemente inteligentes para crear inteligencia. No entendemos cómo funciona el cerebro biológico. No sabemos por qué funcionan algunas de nuestros mejores métodos de IA. No sabemos cómo mejorarlos.

Peter J. Bentley

2.1 Colisión de conciencias

Cuando se habla de Inteligencia Artificial, es común encontrar palabras clave que nos adentren al concepto de esta mímesis tecnocefálica (Rodríguez-Reséndiz, 2020a): Sistema, algoritmo, tecnología, computación o autonomía. Particularmente la autonomía genera un doble sentido que ayuda a orientar las reflexiones en torno a la ética, ya que se puede referir al mismo sistema o también a aquella capacidad inherente de las personas para actuar por sí mismas. La autonomía del sistema y la autonomía humana se diferencian particularmente por la capacidad de actuar por sí misma (tecnología) y la voluntad (persona). Se presentan estas dos realidades y se genera una discordia al momento de actuar o transformar la realidad ya que ambas procuran influir en otra. Mientras que la IA con sus procesos algorítmicos produce resultados o conclusiones que afectan las decisiones de las personas, estas conglomeran datos que sirven para el sistema como orientación para determinar las acciones o variaciones en el mundo real. Este ejercicio de sometimiento no es exclusivo de ambas entidades, sistema y personas, sino que entre los pares iguales surge también fuerzas que desembocan en violencias para estar encima de otras entidades: sistemas contra sistemas y humanos contra humanos. Estas disputas que aspiran a ser superiores, generan a través de un buen uso de la razón el convencimiento de que lo que se produce como efecto en el otro “es lo mejor”. De esta manera, existen los sistemas que someten la autonomía de las personas porque a estas les parece sumamente confiables las decisiones que se le presentan en el uso de esta

tecnología, mientras que el sistema lleva a cabo su proceso de discriminación de posibles respuestas gracias a la singular clasificación de las características dadas por el usuario (datos personales). La oposición derivada del choque de las autonomías se le puede denominar colisión de conciencias, puesto que el nivel de conocimiento que se genera en estos escenarios de ambas entidades resulta de una historicidad y reconocimiento de una circunstancia. En un mismo lugar de actividad, se pueden provocar dilemas o problemas éticos, máxime cuando se encuentran en lugares antagónicos de argumentación. La colisión de conciencias es la fuerza ejercida en un *espacio* y *tiempo* por las autonomías que discrepan en su forma de percibir y ser en el mundo. ¿Cómo se puede definir y abordar la autonomía humana y no humana desde un terreno ético de la IA para acercarse a una tecnología responsable?

En efecto, por sí misma, la autonomía ha sido uno de los problemas tradicionales de la ética desde la filosofía que incluso toca el umbral con la psicología, el derecho y la sociología. Además, en estos terrenos del conocimiento no se pueden separar la autonomía de la libertad, la conciencia, alteridad, lo externo y lo social. Para Aristóteles y Platón, la autonomía es un estado-idea al que se puede acceder por medio del ejercicio del intelecto donde la soberanía, ciencia y sabiduría están estrechamente ligadas, pues:

> "[…] *de las ciencias, aquella que se escoge por sí misma y por amor al conocimiento es sabiduría en mayor grado que la que se escoge por sus efectos. Y queja más dominante es la sabiduría en mayor grado que la subordinada: que, desde luego, no corresponde al sabio recibir órdenes, sino darlas, ni obedecer a otro, sino a él quien es menos sabio*." (Aristóteles, 1994)

Así, se es más autónomo mientras más se participe de la virtud, la contemplación y el bien en lo general-universal, más allá si las esencias-ideas están en sí mismas o en otro mundo, es decir, del grado de liberación que goce el individuo de las cosas físicas y más cerca de uno mismo (voluntades y autogobierno).

Una consideración importante en el recorrido del concepto de la autonomía es la que se manifiesta en la Edad Media, donde se distingue la voluntad interna (uno mismo) y la externa (la divinidad), que se unen para determinar la libertad con la que se actúa (Trego, K., 2005). Así pues, para Agustín de Hipona, la distinción entre libertas y *liberum arbitrium* está determinada por la voluntad y la capacidad de decidir, es decir, la autonomía define las acciones de uno mismo de las que son de Dios (Bello, H., 2019), lo que ocasiona que se pase de un fin a uno mismo (Grecia); a la condición de posibilidad de libertad como asunto teleológico (Edad Media).

Desde luego, en Kant existe el rechazo de lo teológico ya que los fines son contingentes para las personas. De tal forma, la autonomía se transforma en una capacidad de la voluntad para obrar con independencia de los estímulos que genera la sensibilidad, por lo que uno mismo debe de ser su propio legislador moral, actuando según las creencias para los demás como fines en sí mismos: "*Obra sólo según aquella máxima por la cual puedas querer que al mismo tiempo se convierta en una ley universal*" (Kant, I., 2012), que no es otra cosa que el imperativo categórico que establece un axioma que no está sujeto al mundo sensible y contingente, es decir, al actuar con autonomía (Schubbach, A., 2019). Por eso no es de extrañar que el pensamiento kantiano para determinar la voluntad se fundamente en dos principios: la razón (actuar libremente: autonomía) y la inclinación (actuar con la sensibilidad -apetitos humanos-: heteronomía).

Para Hegel, todo sucede por una razón y cualquier cosa racional es real, siempre manifestada por el espíritu (aquello que está en la posibilidad de serlo), con un autoreconocimiento de uno mismo y de la realidad circundante, por lo que gozando de esa disposición y usando la dialéctica (amo y el esclavo) se alcanza la autoconciencia del espíritu (Padial, J. J., 2017), es decir, la conciencia social o de "*el otro*", por tanto, la autonomía deviene de la lucha de cocientes en la historia: "*la autoconciencia sólo alcanza su satisfacción en una autoconciencia distinta, en otra autoconciencia*" (Hegel, G. W. F., 2009).

Otro enfoque de autonomía del sujeto es la que se deriva del orden jurídico, que puede reducirse a la capacidad política que tiene cada persona para decidir con base en la limitación

regulatoria establecida en una sociedad de forma libre (expresamente), pero que siempre atrae obligaciones (Iosa, J., 2017).

A su vez, la autonomía en el marco de la tecnología es la capacidad que tiene cada persona para actuar de forma consecutiva al uso de software o hardware, es decir, la libertad está determinada por la heteronomía que deviene de aquello que denominamos técnica, por lo que siempre existirá el supuesto de "*el uso*" para "*poder ser*" afectado o "*determinado*" (Lemmens, P., 2017).

Por todas estas razones conceptuales, se puede llegar a entender que la autonomía es una postura dialógica entre el sujeto, la norma ético-jurídica y, para los fines de este trabajo, la tecnología en tanto la toma de decisiones. Estos tres elementos condicionan la autonomía en un sentido ético ya que, en la interpelación de la tecnología (creador y usuario) que considera la moral circundante, modifican la realidad en un sentido ontológico: las posibilidades de *ser*. Por tanto, el problema ético que subyace con esta argumentación es el de la creación e implementación de la tecnología en particular de la IA.

Los horizontes pertinentes de examinación entonces se convierten en: ¿Quién crea? (perfil), ¿Para qué la crea? (propósito, interés, finalidad) y ¿Cómo crea? (metodología). La primera pregunta no sólo se refiere a la educación técnica del creador, sino al conjunto de determinaciones morales y éticos que enmarcan al sujeto; la segunda cuestión está orientada al ejercicio de interiorización de proyección y autoconciencia; y, por último, cabe referirse a la forma en la que la programación funciona (trazabilidad).

Por su parte, es posible preguntarse desde la ética frente a la IA ¿Quién la usa? (operatividad), ¿A quién afecta el uso? (alcance operativo-formas de interpelación) y ¿Para qué la usa? (objetivo o propósito operacional). En esta disputa entre entidades (sistemas y personas) siempre sobresale una de las dos propuestas: la IA está cargada con los datos y prejuicios del programador o bien que el usuario pueda inyectar su información para que el sistema influya en la decisión que se le presentará en el uso de dicha tecnología. En definitiva, el móvil de convencimiento de esta colisión de conciencia es el uso de reglas que ambas entidades

establecen al generar un diálogo en su comunicación y que procuran conmover al “otro”. En efecto, el uso del lenguaje, que el sistema o el humano utilizan para interactuar entre ellos, procura mover de una posición a otra a la entidad interpelada. De este modo el sistema orienta al humano hacia cómo actuar en sus decisiones o bien decide por él mismo, mientras que los usuarios permiten dirigir al sistema hacia las salidas con las cuales fue programado el sistema.

El caso de Cambridge Analytica resulta ser emblemático para ejemplificar la colección de las conciencias, puesto que los algoritmos diseñados por los programadores de esta empresa sirvieron para generar un dominio de los votantes tras la recopilación de datos obtenidos de redes sociales como Facebook. El interés era satisfacer la necesidad de un cliente (orientar las votaciones), por eso se diseñó la focalización de perfiles que fueran aptos para que votarán por un candidato en específico en los comicios estadounidenses del año 2018. La autonomía del sistema de Inteligencia Artificial operada se convierte en una actuación por sí misma con las reglas diseñadas para la discriminación de datos de acuerdo con el enfoque de posibles votantes. Mientras que la autonomía de los cibernautas estaba rodeada por su voluntad histórica creada por su afinidad. Esta colisión de conciencias ha llevado a preguntar de forma legal, hasta qué punto los algoritmos vulneran por sí mismos la autonomía del sujeto creando un control en el comportamiento bajo el uso de técnicas de programación computacional que han llevado a la segmentación de la población objetivo para así disminuir la capacidad crítica de los usuarios del sistema. Si la colisión de conciencias acarrea un uso retórico para conmover ¿cuáles son las posibles salidas a estas realidades en el ámbito ético?

Cuando un programador de IA genera una codificación para que actúe en el sistema, lo hace bajo diferentes móviles que se pueden resumir en: voluntad, interés y conocimiento del impacto. En este sentido la Inteligencia Artificial comienza con una libertad del humano para decidir qué o no hacer en el diseño del software, siempre al ser direccionado por la utilidad o el valor que de estas acciones devengan y al buscar un beneficio deseado, basándose en una acumulación de la información previamente entendida. Por estas razones, aparecen escenarios de colisión de conciencia en el uso de la mímesis tecnocefálica, puesto que existen diferentes voluntades intereses y formas de percibir el mundo, las cuales son los alcances éticos de los sistemas diseñados. Para disminuir la colisión de conciencias se propone generar

espacios o momentos de reflexión del programador para reconocer las voluntades, intereses o impactos que se producirán en la implementación de IA. La única puerta para lograr este encomiable deseo es la formación crítica de los aprendices programadores durante el proceso de generación de capacidades técnicas e intelectuales que estén ajustadas a mínimos morales como el bien común. Por lo tanto, esta propuesta impulsa el refuerzo de planes de aprendizaje o programas de estudio que contengan una mayor cantidad de elementos de ética como un asunto transversal para generar al menos dos preguntas elementales: ¿Por qué? y ¿Para qué?

En el reconocimiento del extenuante proceso de creación de un marco ético de IA, se ha observado que la colisión de conciencias impide lograr acuerdos que dejen satisfechos del todo a todos. Las realidades del sector económico, privado, gubernamental y de la sociedad civil, a ratos parece no encajar en el andamiaje de un diálogo unificado y se presentan desarticulaciones que promueven desencuentros. Por ejemplo, ¿Se puede diseñar una Inteligencia Artificial que genere beneficios para una empresa que se ocupa de recopilar datos biométricos con o sin consentimiento informado para después hacer uso de ese historial de un perfil, y quebrantar la autonomía del usuario tras direccionar o limitar las decisiones? Por ello, el reconocimiento de las voluntades, intereses y conocimientos del impacto en torno a la IA es fundamental en la articulación del diálogo para el entendimiento y producción de marcos referenciales de ética de la IA. La situación a la que se refiere esta idea es la actitud de los dialogantes en el efecto comunicativo al procurar relacionarse con los demás, al generar posturas de apertura respetuosa y al reconocer los ruidos de los interlocutores y los propios. Este referente dialógico de Habermas, asume que toda acción está direccionada hacia un entendimiento basado en la racionalidad del lenguaje, pero es necesario poner atención en las voluntades, intereses y consideraciones éticas de los interlocutores en sus exposiciones sobre las ideas del andamiaje de una IA responsable.

En este sentido, todo discurso está fundamentado en las ideas; una idea representada no es igual para todos. En cuanto al acuerdo o discordancia, es posible que trastoque la colisión de conciencias, porque la multiplicidad de asumir el mundo tiene su fundamento en las particularidades de la diferencia, contingencia o incluso el relativismo. Es por ello, que este escenario de colisión de conciencias es un encuentro que se da en el aspecto metafísico con

la acentuación ya mencionada de conmover. La fuerza aplicada en dicho encuentro es la referenciada por Huntington como "*clash*", es decir, el ejercicio argumentativo de imponer la universalidad propia en los demás (Huntington, S. P., 1998). Sin duda esta fuerza es derivada del convencimiento propio de asumir el bien y tratar de propagarlo en los demás, considerando que este dinamismo siempre transforme la realidad. Por todas estas razones, el diálogo para una ética de la Inteligencia Artificial debe de atravesar el aspecto metafísico de la colisión de conciencias, entendiendo que no todo conflicto es negativo, sino que es un paso necesario para la unificación o emancipación de ideas, acuerdos o pautas, máxime sí se está hablando de marcos referenciales globales. Esta violencia metafísica es equiparable a la propuesta antropológica de Girard, ya que la paz es un proceso catártico donde se confrontan formas de *ser-pensar* y que por momentos parece ser sólo la herencia irracional de no suscribir la autocrítica con los valores dados a uno mismo, cayendo en la imitación establecida en la historicidad. La promoción del diálogo de ideas diferentes no sólo debe de motivar una escucha reglamentaria o políticamente correcta, sino presentar la oportunidad de autoreflexionar sobre el reconocimiento del "*otro*" y su deseo de *conmover*, recordando que dicha palabra tiene su significado etimológico en "*poner en movimiento*". Cuando se transita por esta idea, se debe tener un especial cuidado en la distinción del relativismo ya que puede impedir los acuerdos.

2.2 Dispersión de umbrales

La irrupción de la IA en la vida del ser humano ha traído la aparición de dilemas éticos por acuñar conceptos o modificaciones de los ya existentes. La naturaleza y lo artificial son nociones que han sustentado discursos filosóficos y científicos para generar lo que se denomina progreso. A lo largo de la historia se ha hecho una diferenciación de las cosas naturales y de las artificiales. Este argumento en la actualidad pudiera no ser sostenido debido a que las condiciones del ser humano se van ajustando a nuevas realidades. Los problemas subyacentes por la distinción entre lo natural y lo artificial se solucionan con una vindicación de un antropocentrismo.

La percepción del ser humano hacia el mundo suele tener diversos enfoques, así también sucede con el autodiscernimiento. La historia nos ha demostrado que la humanidad va cambiando su forma de conceptualizarse, validando así acciones morales y éticas. Esta es la explicación del por qué en su momento histórico la esclavitud era del todo aceptada, puesto que había distinciones entre el concepto de humano y no humano, diferencia a raíz del alma, la razón, la libertad o la salud mental. En la época actual, el concepto de naturaleza se disuelve entre lo llamado artificial, porque de hecho "*lo natural*" es un concepto que humano que va cambiando y se normaliza una homogeneización entre la *natura* y aquello que tiene intervención de la técnica, así que no existe más una distinción entre lo natural y lo artificial.

Justo ahora, tras el reconocer la diversidad del autoconocimiento como humanos, aflora la forma de entender la humanidad y, por ende, su naturaleza misma. La promoción de la técnica ha impulsado aún más esta multiplicidad de naturalezas, ya que ahora la ciencia puede "*crear*" naturaleza, y se va advirtiendo que los individuos cambian el mundo y la forma de entendernos.

Por ejemplo, se ha incidido de manera clara en el medioambiente, es decir, en eso que llamamos "*mundo natural*" o planeta Tierra, al sufrir grandes estragos por la actividad humana. Algunos estudios como el de Eissa y Zeky (2011) muestran los cambios que ha tenido el entorno derivado de una apropiación del mundo y que han sido perjudiciales para todas las especies vivas (Eissa, A. E., & Zaki, M. M., 2011). Hasta ahora la responsabilidad mayor de estos cambios ha sido atribuida al ser humano porque se considera que ha actuado sólo en un sentido: su propio beneficio. A esta unidireccionalidad se le conoce como antropocentrismo porque se pone en el centro al humano de todo posible beneficio, dejando sin consideración a otros entes que acaso son "menos humanos" como plantas y animales. Así pues, cabe preguntarse si el antropocentrismo es un factor que afecta la naturaleza y lo artificial.

Los investigadores no se han puesto de acuerdo para identificar quién fue el autor que primero delimitó al antropocentrismo; algunos hablan de que fue Protágoras con su sentencia de "*el hombre es la medida de todas las cosas*", donde da pauta para ir generando un camino hacia

esta forma de ordenar el mundo (Chen, C. H., 2018), pero el concepto toma mayor fuerza en la Edad Media cuando se empezaron a producir atisbos de que el hombre (*humos*: que viene de la tierra) debería de ser ponderado sobre otras *creaciones* o sobre la naturaleza. Giovanni Pico Della Mirandola (2018), en su *Discurso de la dignidad del hombre*, reflexionó sobre este hecho y pudo encaminar sus argumentaciones a favor de la supremacía del humano sobre el resto de las cosas (Della-Mirandola, G. P., 2018). En efecto, tomando las Sagradas Escrituras, desdeña el pecado cometido por los guardianes del Edén, pero sin dejar de pensar que siguen siendo ellos, varones y mujeres, parte de la divinidad: imagen y semejanza de Dios; así, ninguna otra cosa creada puede igualarse al ser humano y, por ello, él mismo se distingue de entre las demás cosas naturales, pero con la supremacía de la herencia divina.

En su momento, el talante de Francis Bacón de acentuar esta distinción *natural* de lo *humano* y "*lo demás*" toma principal fuerza con el argumento de la apropiación de la *naura* gracias a la técnica (Bacon, F., 2006). Así pues, la historia va fraguando poco a poco el estereotipo del humano como ser supremo entre las demás especies, aumentando un desequilibrio en su entorno por el hecho de pensarse como "*una diferente naturaleza*" pero dentro de la misma naturaleza.

Es allí donde aparece la idea del antropocentrismo, planteado por Hayward (Hayward, T., 1997) que pondera, exclusiva o arbitrariamente, los intereses del ser humano por encima de los de otras representaciones de naturaleza. Esta es la visión conceptual que ha prevalecido hasta el día de hoy, inclinando la percepción de que el hombre es fuente de maldad al ponderar su voluntad sobre otras especies vivas, sin límites ni conciencia. Al perecer este tipo de visiones, se piensa al antropocentrismo (*άνθρωπος*, *κέντρον*, el humano al centro y *ισμός*, forma de ser o doctrina) como algo totalmente negativo, clasificándolo como una idea que se debe de quitar del accionar para lograr una armonía entre "*lo demás*". En efecto, al poner al humano al centro de la creación, se hila que todo está a la voluntad de él mismo. Pero ¿Qué pasaría si no pensamos el antropocentrismo como un problema sino como una virtud para alcanzar la conciencia? ¿Qué pasaría si entendemos la frase "*dominar la tierra*" por "*cuidar la tierra*" en el relato de Génesis como algo diferente? Por ejemplo, quizá, entonces podríamos ver que el origen de los siniestros ambientales no es la causa del egoísmo

del humano, sino de su falta de conciencia y percepción de "*lo otro*"; quizá también podríamos concluir que hemos cuidado mal esa "*natura fuera de nosotros*", y que entonces se ocupa generar acciones para asistir las necesidades del medioambiente a través de la autorreflexión basada en la razón, usando por ejemplo IA. ¿Qué es lo natural? ¿Qué es lo opuesto a lo natural?

Asumimos que lo contrario a la naturaleza es lo artificial, luego, se identifica entre los dos conceptos, donde la distinción suele diluirse porque lo natural está mezclado con lo artificial, llevando a creer que un elemento artificial es parte de la naturaleza o al revés. Por ello, el concepto que se traslada hasta el día de hoy de naturaleza es: *todo aquello que no tiene intervención humana*. ¿Qué cosa en este mundo no tiene intervención humana? Por ejemplo, alguien podría afirmar que el cuerpo humano es natural, y lo mismo con bosques, mares, especies vivas en latitudes alejadas de las civilizaciones, pero la imposibilidad conceptual ocasiona distinguir que no necesariamente las cosas son así. Para que un humano pueda ser procreado, ya no se ocupa necesariamente a la naturaleza antigua, sino que con técnicas de reproducción asistida se puede crear un ser, incluso con realidad aumentada; también se usan instrumentos clínicos para extraer a los humanos del vientre de las madres y muchos agentes externos como son los medicamentos; luego, se ingieren alimentos procesados y manipulados por el ingenio humano como los químicos; después, el aire que sostiene el organismo humano está lleno de alteraciones por contaminantes creados por el mismo humano como lo es la combustión de hidrocarburos; se usan todo tipo de cremas, vacunas, *brackets*, marcapasos, intervenciones quirúrgicas, decolorantes para cabello, rectificación de retinas con rayos láser, etc. Entonces, ¿Quién puede decir que el ser humano en la actualidad es natural y que no está intervenido por lo artificial? Lo mismo sucede con esas otras formas en las que representamos a la naturaleza: bosques, mares, animales, cielo, subsuelo, todo tiene algún tipo de intervención humana directa, indirecta o incluso sólo lingüística (nombrar las cosas implica una posesión de ello, aunque no se tenga físicamente frente a uno).

Es entonces cuando la situación se complica, ya que tanto la existencia de lo natural como de lo artificial son significados que parecen ser antagónicos, y ahora se especula que no hay un umbral visible para distinguir uno de otro. Así pues, toda naturaleza tiene algún grado de

artificialidad y toda artificialidad contiene naturaleza. El mismo hombre es naturaleza y artificialidad puesto en un medioambiente, ya que somos producto de una naturaleza y de una técnica médica, sobrevinimos por los impulsos del cuerpo, pero también por la ciencia de la salud. Por lo tanto, la *natura* deja de ser unidireccional porque de hecho ya no hay nada completamente natural, puesto que las acciones humanas ya han afectado todo el entorno. Por ejemplo, la huella ecológica que se deja en una parte del planeta afecta un ecosistema en otra latitud, sin duda. De allí la razón del por qué el humano debe de ocuparse de otras formas de naturaleza, porque al no haber frontera de lo natural y lo artificial todo se vuele uno. Para esto, el concepto manejado desde la filosofía para hacer explicativo este fenómeno es la interpelación.

Aquello natural que no es humano está hablando al humano, lo está interpelando para que voltee a verlo de forma simbólica y es por eso que, en este acto de diálogo y gracias a la intervención de la ciencia, el humano se siente interpelado, es decir, llamado a responder a lo que esos otros seres naturales le dicen, lo que hace posible una ética (Dussel, E., 1994): el sujeto es interpelado por el entorno, por la naturaleza.

A propósito de esta afirmación, cabe distinguir tres formas de naturaleza desde el mundo de la filosofía, específicamente, desde Aristóteles. La primera de ellas se encuentra basada en la concepción de lo físico (*φυσικός*), que se estipula como la *esencia* de las cosas para que *sean*. Otra forma de percibir la naturaleza es lo tangible, es decir, la materia física que da *sustento* a lo *real*. Un tercer significado es "*aquello que se le dota*" a las cosas por *efectos* de algún creador. Justo estas tres representaciones son las que hacen distinguir lo natural de lo artificial. Pero ¿No es acaso que en la actual época las definiciones colapsan porque hay artificialidad y naturaleza mezcladas?

Por su puesto, en la historia se ha manifestado aún más este fenómeno, ya que antes el *homo sapiens* veía y distinguían por completo la naturaleza por fuera de sí mismo. Este discernimiento hacía diferenciar dos entes: lo humano y lo natural. En esta separación, existía un respeto por el desconocimiento en su forma de dominar esa *natura*. No obstante, la ciencia y la técnica se sumaron para que dicha distinción de entes se nulificara, ya que la apropiación

de lo natural lleva hasta reproducirlo, y ha creado lo *compuesto*. Esta realidad provoca un acercamiento mayor: el humano es naturaleza y técnica, por lo tanto, somos híbridos responsables y afectados de lo que está pasando en la universalidad de lo natural. En el terreno ético, ¿Cuál es la responsabilidad más interpelante que el ser humano tiene con "*la demás creación*"? la repuesta a esto son las decisiones.

¿Cómo la presente propuesta de vindicar el antropocentrismo como parte de una nueva realidad de naturaleza puede ayudar a contribuir a una mejora del trato con "*lo que no es humano*" y tener repercusiones positivas en el establecimiento de una IA responsable? La respuesta de dirigir las acciones-elecciones de lo que pasa en la naturaleza se vuelca hacia los humanos que han creado cambios en ella empezando por su representación, ya que sólo el *homo sapiens* posibilita la realización desde la *razón-reflexión-análisis*: la conciencia. Ningún otro ser natural se ocupará de algún cambio porque no tiene la conciencia para crear técnica razonada, conocimiento o acciones. Así, la respuesta a este hecho es una misma: poner al centro de las acciones al humano, es decir, vindicar un antropocentrismo basado en la conciencia de poder ser desde la interpelación del mundo al hombre. Por ello, la propuesta de tomar decisiones a favor de toda la naturaleza es no signar con maldad al humano en el centro de las especies vivas, sino crear una ética que aspire a implementar una moral de responsabilidad en las elecciones que tomamos, ya que, en la época actual, gran parte de la naturaleza depende de las decisiones humanas.

La responsabilidad social parece encajar muy bien aquí, toda vez que en la indistinción de lo natural y lo artificial, subyace un diálogo de dependencia: lo humano ocupa de lo natural y viceversa, es decir, la *naturaleza* requiere de otras naturalezas e incluso echa mano de lo artificial para *ser*. Aquí y ahora, ¿Lo natural puede subsistir por sí mismo? Difícilmente. Por tanto, sobresale nuevamente la responsabilidad e iniciativa de aquella *otra naturaleza*, la humana como principal promotora de cambio. De hecho conviene generar este nuevo planteamiento de lo humano al centro con conciencia, echando mano de la técnica, por ejemplo, la desprendida de la ingeniería computacional.

En efecto, existe una conveniencia al abrazar al antropocentrismo para generar un equilibrio con las demás cosas naturales. Para ello, puede haber diversos caminos, sin embargo, el que aquí se propone es el de la ética. No obstante, al haber multiplicidad de definiciones de ética, es necesario volver a citar a esta diciplina como la que se desprende de la filosofía y que procura un discurso reflexivo entorno de la moral, usando a la Razón para evaluar los paradigmas que se presentan al sujeto. Por tanto, todo ejercicio ético es lenguaje crítico situado en torno al humano que lo postula.

La sustitución del paradigma que se propone es meramente conceptual con implicaciones prácticas: asumir el antropocentrismo y actuar con responsabilidad. Desde luego, en el fondo lo que se formula es, por medio de la ética, se plantee una moral para que a su vez normalice la obligatoriedad de una conducta a favor de lo llamado natural, hasta que la actuación sea intrínseca en la sociedad. Así pues, esta propuesta de conceptualización no es dejar al humano al centro de todo como algo aislado, sino como una experiencia ética de vivir con los demás seres vivos desde lo humano.

Para efectos de ello, es necesario indicar que la conciencia es "*la época fenomenológica*" (*ἐποχή*), según Husserl, del humano donde se pregunta qué es algo de ese mundo que se le aparece, haciendo uso de la libertad para indagar la relación entre lo que se *le presenta a uno* con aquello que *se cree que es* (Gutland, C., 2018), y que así la filosofía aproxima con su criticidad a un estado de reflexión al abstraer un acto ético para después generar razón práctica (Bakhtin, M. M., & Ponzio, A., 1997). Además, como ya se ha dicho, el aporte de Hans Jonas sobre el reconocimiento del mal, especialmente, sobre otras especies vivas, también genera una autoconciencia de los actos, al ser ocasionada por una heurística del temor, es decir, del miedo a una catástrofe venidera a causa de los actos de los humanos (Dinneen, N., 2014).

En suma, es conveniente considerar la vindicación del antropocentrismo desde una consideración filosófica no negativa, porque otorga conciencia al sujeto que realiza un ejercicio ético y, por ende, debería de estar por encima de otras bases ético-humanas, por ejemplo, de la dignidad. Antes de reconocer que uno es digno, primero hay que realizar el

acto de conciencia de serlo, es decir, ponerse en el centro para hacer las valoraciones correspondientes, porque de hecho, es poco fiable valorar que otras especies vivas tengan conciencia y aún hoy en día los humanos siguen siendo el único *ser* que al tener autoreconocimiento pueden generar un cambio en favor de otras especies con vida.

Por su parte, se puede afirmar que al no haber una distinción de la dicotomía naturaleza y artificialidad, la naturaleza del ser humano también se disuelve, por lo que es conveniente preguntarse si el *homo sapiens* debe de alejarse de esos tiempos donde tal binarismo metafísico daba tranquilidad (clasificar las cosas entre A y B genera paz al ser humano) (Bowden, B., 2020). Por tanto, ya no existe más una naturaleza humana, sólo una mezcla de lo natural y la técnica del humano que se debe de sostenerse desde un horizonte antropocéntrico, sobre todo porque la artificialidad cada día más orilla a reinventar de forma constante lo que somos como humanos y cuando la naturaleza ya ocupa de manera necesaria a la artificialidad para *poder ser*.

Estas realidades no son del todo negativas porque la novedad de que no hay fronteras entre lo natural y lo antinatural, más que confusión, es una oportunidad para replantearse qué somos ahora para resolver problemas éticos.

Por todo ello, no se puede definir en la presente época el concepto de naturaleza humana, porque ya de suyo mismo es algo cambiante y adaptable a cada momento histórico; la complejidad que ahora se presenta radica en la aprobación de la diversidad, ya que toda definición del humano deja fuera a algunos humanos y no podrá ser universal, a no ser de que pongamos a lo natural en un lado y al otro extremo lo artificial, y en medio al humano reflexivo, es decir, un antropocentrismo basado en la conciencia.

Toda esta reflexión es para argumentar que, con la aparición de la IA en todas las esferas de la vida humana, se está perdiendo el umbral entre lo natural y lo sintético, lo cual habilita nuevas formas de entender y abordar problemas éticos como la autonomía, libertad, dignidad, verdad o realidad. ¿No será que la autonomía ahora debe de entenderse como una cosa independiente a los sistemas computacionales y como una herramienta complementaria para

la toma de decisiones? ¿La libertad no estará ahora enmarcada por las opciones que nos presenta la IA derivada de procesos en la vida cotidiana? ¿La dignidad no estará sujeta a la posibilidad de interactuar con mímesis tecnocefálica para generar paz, salud o bienestar? ¿La Verdad no estará suscrita por la segmentación de datos o noticias que nos presentan los algoritmos ahora? Y ¿La realidad no será una conjunción entre lo digital y el mundo físico? Estas singularidades promueven más que un transhumanismo, una ruptura de conceptos en la adaptación de circunstancias que nos trae la irrupción de la IA.

Finalmente, es necesario indicar que el humano no se da cuenta de los riesgos de la IA, porque vive tecnologizado, es decir, ha naturalizado la tecnología en la vida diaria. Si por alguna razón una persona de la antigua Grecia viniera a la presente época, seguramente vería magia a su alrededor, pero para el *homo sapiens* es cotidiano prender un sol con su voz al ordenar al asistente personal que encienda los focos del hogar o capturar una lectura en un papel en segundos.

2.3 Determinación de la proyección (*concurrencia del ahora*)

Si no se tuviera dudas de lo que va a pasar en un futuro, no se ocuparían decisiones en la vida diaria. En el caso de la IA, una de sus principales funciones es el tomar partida frente al *futuro* que necesariamente modifica el *ahora*. En ingeniería computacional, la decisión es la resolución de un proceso de análisis de información para exponer una probabilidad bajo circunstancias determinadas. Desde luego, las decisiones han existido antes de que apareciera la Inteligencia Artificial, ya que la fascinación por el futuro es una característica social, antropología y hasta neurológica (Ekman, M., Kok, P., & de Lange, F. P., 2017). El pensar en el futuro advierte acercarse a la tranquilidad y alejarse del desastre de lo que puede suceder, pero hacerlo con ayuda de la tecnología ahora se convierte en algo no sólo cotidiano, sino esencial por el grado de objetividad que de la ciencia se desprende.

Anterior a la tecnología, el humano basaba sus decisiones en deseos, es decir, en querer que pasaran las cosas de una manera determinada. Ahora con los sistemas de IA, el deseo se va

construyendo con base en una experiencia de acumulación de datos en el vestigio de la estela digital que va dejando el humano. En efecto, el análisis de datos acerca de una certidumbre matemática, donde los sucesos previstos son destacados de forma cuantitativa, genera una sensación de placer y control (Alba, J. W., & Williams, E. F., 2013).

Todos los días el humano toma decisiones en intersección con la IA, ya que ella puede sugerir o bien a actuar directamente en las opciones presentadas para acercarse al futuro. Desde la medicina hasta creación de políticas públicas, la mímesis tecnocefálica ha servido como apoyo para acercarse de manera satisfactoria a una reducción de posibilidades. En ese sentido se puede afirmar que para que un sistema de IA pueda generar decisiones locales/particulares, necesariamente debe de ser entrenado (Molnar, C., 2020), sin olvidar que detrás del adiestramiento destacan los algoritmos y los datos con los que se llega a los objetivos deseados. Entonces, para conocer cómo los sistemas de IA resuelven, es necesario considerar la explicabilidad o interpretabilidad para descubrir su entramado de procedimientos.

De manera reducida, la explicabilidad es conocer por qué y cómo pasan las cosas dentro del sistema, no sólo para un proceso de mejora continua sino para detectar fallas que pueden vulnerar a los humanos. No podemos estar todo el tiempo preguntándonos cómo y por qué es que pasan las cosas, pero la tecnología al interpelar fuertemente al humano, lo mueve o perjudica. Por ejemplo, un candidato que desea obtener la libertad que se le ha anulado por algún delito, puede preguntarse "*¿Por qué a mí no me selecciona el juzgado para salir libre cuando a otras personas en condiciones similares lo han hecho?*", y es que es posible que un algoritmo esté detrás de dichas decisiones por medio de la medición del riesgo de reincidencia de los presos (Karimi-Haghighi, M., & Castillo, C., 2021). Entonces, al igual que a una persona que el banco le negó un crédito, se ocupa que los sistemas puedan tener clara la ruta para saber cómo pudieron llegar a tomar decisiones, porque los humanos necesitan explicaciones cuando el mundo le interpela y porque a través de estos mecanismos se genera la tranquilidad por medio del entendimiento (argumentación-razonamiento-coherencia-lógica). Una vez que esto llega, se consigue la confianza hacia los sistemas de manera particular para después saltar (recomendaciones y puntajes) a una aceptación social.

Así pues, toda legitimación de la IA debe de estar sustentada bajo una explicabilidad que dirija hacia la normalidad atravesada por la ciencia.

Como ya se ha mencionado en párrafos anteriores, la ciencia se caracteriza por ser razonable, de modo que si una IA pretende ser explicable al llegar a sus decisiones, primero debe de mostrar en las pruebas de entrenamiento que es equitativa, es decir, que se acerca a decisiones imparciales, a la privacidad (datos protegidos), fiabilidad o robustez (cambios pequeños en los sistemas no afectan decisiones grandes en predicción), causalidad (existencia de relaciones causales) y confianza (que el humano pueda tener certeza del uso del sistema) (Doshi-Velez, F., & Kim, B., 2017). También es necesario considerar a la interpretabilidad *post hoc* (método complejo de interpretar donde se valida después de haber entregado el modelo, por ejemplo, la permutación), la intrínseca (método simple de interpretar como arboles de decisión cortos o modelos lineales dispersos con reglas binarias: sí o no), la regresión lineal, no lineal, análisis discriminante, árboles de decisión, cuantificación vectorial, Redes Neuronales Profundas, por mencionar algunos; siempre es bueno saber qué método se usará en el diseño de IA con base en el análisis del contexto donde se encuentre, ya que la interpretabilidad debe ser un punto central al momento de generar la factibilidad en el rastreo de cómo fue que el sistema llegó a generar las conclusiones-predicciones dadas (Samek, W., Wiegand, T., & Müller, K. R., 2017).

No se puede dejar de afirmar que las decisiones con o sin IA están atravesadas por la voluntad, porque es justo al direccionamiento (diseño) hacia donde uno se redirige. No obstante, de esta realidad, hay que sumar la posibilidad latente de la determinación en el proceso de decidir con o sin tecnología, ya que la delimitación de opciones a seguir por un sistema de IA o de un humano, siempre está sujeta a las condiciones dadas en el contexto que pasa desde lo histórico hasta lo lingüístico. Tanto los datos están cargados de determinación como también lo está la conciencia del humano, lo que influye en los resultados. En el caso del tema ético es más clara la cuestión porque está alineado a lo contextual de la aplicación de la IA. De modo que en ciertas latitudes es más probable que la programación de un coche autónomo sea ética si se salvan vidas de estratos, por género o apariencia, ya que todo esto es parte de un fenómeno social y no particular (Jaques, A. E., 2019). El fenómeno social

tampoco es aislado, ya que depende de una condición más para acercarse a las formas de tomar partida: tanto las decisiones de los humanos como las de las máquinas están sujetas a tiempo, pues se sostienen en el pasado, referenciando el ahora y proyectándose a futuro, todo en un mismo momento, diferenciándose únicamente por los métodos para llegar a las conclusiones. Si bien es cierto que la IA genera automatización, control de procesos, detección y predicción, con las argumentaciones hechas hasta ahora, parece que lo fundamental está en acercarse al futuro con esta tecnología en un sentido matemático.

Para dar sentido a lo mencionado, se destacan tres tipos de análisis de datos que usa la IA: descriptivo, predictivo y prescriptivo. La analítica descriptiva es la usada por la IA para informar a los usuarios sobre algo dado, es decir, su límite es el pasado y no se asoma al porvenir. Por su parte, la analítica predictiva se clasifica en Modelos Probabilísticos, Aprendizaje Automático/Minería de Datos, Programación Matemática, Computación Evolutiva, Simulación y Modelos Basados en Lógica y Análisis Estadístico, los cuales se proyectan en acciones humanas en la toma de decisiones y disminuyen el juicio humano lleno de subjetividades (Lepenioti, K. et al., 2020). Por ello, la IA busca soluciones en el conjunto de variables del fenómeno analizado para desembocar en la probabilidad de que un hecho o situación suceda después del "*ahora*". En efecto, el juicio humano está "*contaminado*" por diversas variables que incluso no se ven la toma de decisiones, sino que solamente se asumen por el hecho de lo habitual o repetición, en cambio, los modelos matemáticos toman datos objetivos históricos para mostrar decisiones mucho más adecuadas a lo racional, es decir, a lo que *naturalmente* se debería hacer en una situación determinada. En tanto que la analítica prescriptiva ocupa datos históricos para cambiar el futuro, por lo que no sólo se contempla el tiempo como una proyección sino como una realidad moldeable al deseo de la voluntad de las conclusiones asignadas por el sistema.

Justamente, el problema ético dentro de la toma de las decisiones es la falta de objetividad de los humanos y su constante cambio en su forma de abordar los dilemas que se le presentan. Por su parte, un coche autónomo con IA, programado con analítica descriptiva, personalizada con el conjunto de datos del dueño del vehículo o bien por las estadísticas generales de la sociedad donde se desenvuelve, puede acercarse a lo que "*en casos similares históricos se*

habría hecho", pero el ser humano está sumergido en la contingencia de la conciencia, es decir, cambia fácilmente de decisiones según los contextos, experiencias vividas o situaciones derivadas del comportamiento neurológico (Savulescu, J., & Persson, I., 2012). La situación se complica más por una serie consideraciones que emergen como el aumento de la capacidad de las máquinas para procesar información en tiempo real, la calidad de los datos con que son alimentados los sistemas, el despliegue de modelos metaheurísticos, la dicotomía de lo probabilístico y el determinismo, lo cual no sólo puede propiciar decisiones indeseadas al sujeto que usa IA, sino daños a terceros.

En relación con las decisiones, se puede preguntar ¿Qué es la historia, el pasado, presente y futuro sino sólo una forma de abordar el fenómeno del tiempo a través del lenguaje? El lenguaje y tiempo son dos factores que se ven atravesados por la forma de deliberar, es decir, tomar una postura. Esta "*forma*" es una constitución lógica: la argumentación. Es difícil pensar que las decisiones éticas o de los sistemas de IA no sean con una estructura lingüística heredada. A esa estructura se le denomina *proposición* (Corcoran, J., 2009), la cual requiere un análisis hermenéutico para vislumbrar si es pertinente argumentar al momento de abordar las decisiones éticas.

La lingüística ocupa de una interpretación como un acto de atribución de características a los fenómenos que se pueden describir, con la finalidad de que el hablante pueda dar claridad a su expresión recibida por un "*otro*", llamado receptor o interprete. Esto nos lleva al problema de alteridad, donde se puede separar el significado del signo (lo que se entiende y cómo se dice) donde el emisor y el receptor se apropian de un campo semántico que los hace actuar (hablar y oír; expresar y entender), lo cual lleva a significados particulares e interpretaciones diversas (Bowerman, M., & Choi, S., 2001). Para superar esta problemática, se requiere que las partes realicen una contextualización para dar sentido o coherencia y se pueda generar un diálogo. Por tanto, el camino del entendimiento es posible si las partes se apropian del significado desde los plexos de referencia que lo circundan: sus propias circunstancias que los determinan. Entonces, es siempre el lenguaje -argumento- el que construye el sentido y, por tanto, la deliberación moral siempre está direccionada por el sujeto.

Pero, ese lenguaje usado no es unidireccional, sino que es un diálogo donde aflora el escuchar y el ser escuchado y donde de forma simultánea existen dos discursos que respetan su momento: exponer y ser receptivo. Es aquí justamente donde cobra importancia la metodología de "*hablar bien*" para "*ser entendido*" que se destaca por generar premisas y argumentar adecuadamente para concluir algo. En efecto, para tomar decisiones, hablamos (con otros o con nosotros mismos), lo cual lleva a pensar que lo verdaderamente importante es la conclusión (resolución final) (Henle, M., & Michael, M., 1956) porque ello da validez con base a lo razonable y después se convierte en "*lo mejor*".

Esta herencia de la Ilustración de ponderar la racionalidad como fundamento de la ciencia, también ha afectado al terreno ético y tecnológico, llevando a dicha disciplina filosófica a una institucionalización donde sólo lo razonable es aceptado. Los programas de estudio educativo, los Comités de ética, Códigos, pautas, lineamientos y todo mecanismo de ética pasa por esa criba del neopositivismo: un lenguaje "*bien*" hecho, que convenza y sea suscrito por un análisis, una elección de significado (postura) y una construcción de un sentido (actuación). Esto es la coherencia, un mecanismo para afirmar un "*aquí y ahora*" (Avens, R. S., 1982) frente a los dilemas de la IA.

La distinción más aceptada para este ejercicio lingüístico es convencer y saber expresar las ideas, lo cual se determina con el nombre de retórica, que permanece intacta desde la época medieval, por mucho tiempo siguiendo el modelo de elocuencia y convencimiento. La ruptura de este esquema devino con Nietzsche quien postuló el fin de la Verdad y la apertura al sentido que proviene del sujeto quien crea estrategias discursivas para interpelar al "*otro que escucha*" desde la validez lógica-argumentativa, pero tomando como referencia un contexto determinado para ser entendido. Esta situación remite a un problema platónico: nombrar o ser nombrado (la palabra nombra al mundo o la palabra crea el mundo). De allí que se considere que "*la palabra*" juega un papel importante en la toma de decisiones ya que el emisor pronuncia una palabra que le significa a él, pero el receptor le da el significado entre tantas otras palabras para comprender (dialogar), todo ello en el marco de unas reglas lógicas o una institucionalización del lenguaje. Esas reglas no son otra cosa que "*la historia*", un "*otro*" que antecede a la pronunciación de manera que determina el hablar y escuchar con

el rigor de un contexto gramatical: sujeto, verbo y predicado. Se deja ver que en el diálogo no existe sólo un sujeto que pronuncia, sino una diversidad de entidades que participan en la postura pronunciada y, por tanto, las ideas de uno mismo nunca serán originales, de modo que se debe de asumir en las decisiones esta situación para no crear empachos y para resolver en la originalidad del convencimiento.

Así, se debe de tener presente esta triada de "*el que habla*" (pronuncia discurso), "*el que pone las reglas*" (institucionalización del lenguaje) y "*el que escucha*" (el que da sentido al discurso), por ahora con especial énfasis en el tema de IA, que justo es por el lenguaje y sus problemas que subyacen y tienen intersección con esta realidad. Sobresale ya un elemento transversal: la pretensión de la Verdad, toda vez que cuando uno habla con cierta estructura para ser entendido o aplicar un sistema de IA, se busca demostrar la Verdad (validez) pero sin afirmar la Verdad (lo universal a causa de la diversidad de entendimientos y sentidos). Esta extraña adhesión a la Verdad es un aforismo, puesto que en fin de la Verdad se enuncia una Verdad: no existe la Verdad, o lo que es lo mismo: *no hay hechos, sólo interpretaciones* (Neves, J., 2019), por lo que las decisiones propuestas no sólo por humanos sino por sistemas de IA tomarían fuerza en la *probabilidad de ser*, pero no en la pretensión de la ciencia: ser consideraciones universales, contundentes y con contenido tautológico.

Así pues, dado que la racionalidad es un mecanismo para validar afirmaciones y dichas conclusiones son por medio del lenguaje (léxico validado), entonces, existen una racionalidad imperante que toca lo que pretende ser un discurso reflexivo, por ejemplo, la ética. Es justo aquí donde aparee una digresión: ¿Si toda ética es racional, no se puede hacer una ética sin apegarse a un lenguaje institucional que desdeña la Verdad, pero se apodera de la validez? ¿Alguien que no es racional, puede hacer ética? ¿Toda persona no racional no puede generar reflexiones éticas? ¿Sólo los discursos en torno a la ética de la IA deben de ser adecuados a la lingüística establecida?

La racionalidad no es universal. Sobran ejemplos en la historia donde se puede comprobar que los discursos racionales, con una estructura lingüística apropiada, no generan consensos que tengan el visto bueno de toda la humanidad. Incluso, ha pasado que el lenguaje racional

ha persuadido para generar una ética y tomar de decisiones que no son benéficas para otros discursos, por ejemplo, el nacismo. ¿El nacismo fue construido bajo un lenguaje racional? Sí, pero al no ser aceptado por otros discursos, se anula toda posibilidad de universalidad.

La IA no es inteligencia humana, sólo se basa en la naturaleza para funcionar (mímesis), por lo que, si el humano pensara de otra forma, probablemente el sistema binario (0101) no estructuraría el lenguaje computacional, porque, de hecho, el lenguaje informático es un reflejo (construcción social-histórica), que después migra sus reglas a la programación.

Para acercarse a una mayor persuasión se requiere de un diálogo (lenguaje) donde no se asuma que el hablante está determinado por circunstancias para construir el entendimiento y decisiones acercadas al consenso en la apertura del reconocimiento de "*el otro*" (Kögler, H. H., 2005). Es allí donde encajan las deliberaciones morales: la armonía entre lo lingüístico, el entendimiento y lo histórico. Esto es la Verdad, la pretensión de la comunión entre las partes de un diálogo y por esta razón, se ocupa regresar a ella.

Cuando se habla de ética de una IA global, se debe de asumir la pretensión de la Verdad, de lo contrario se camina a ciegas y se genera incertidumbre, por lo que se debe de tener en cuenta, en este supuesto, que cada paso para deliberar, puede *no-ser* (falsedad), lo que desvirtúa y genera "*muchos caminos*" (ambigüedad), incluso puede que en el trayecto de destino se llegue a *la nada* (vacío o sin decisión). Por su parte, se ocupa tener presente que la incertidumbre es un estado de ánimo-sentimiento, pero también es un carácter: seguir sobre un mismo tema, especializarse y adentrarse al asunto con determinación. De allí que Descartes se pronuncie por "*una moral provisional*" que se asuma como Verdadera lo que de momento con la razón se ha concluido con *claridad* y *distinción* hasta llegar el momento donde, por argumentos (método) se pueda probar lo contrario y rectificar o abrazar otra postura moral, siempre *temporal* (Cumming, R., 1955).

En cualquiera de los casos, siempre se decide en presente, porque, de hecho, toda decisión está sujeta al tiempo. Así las cosas, se reconoce que hay dos formas de abordar el tiempo: de manera objetiva y subjetiva. En ambas concepciones, el ser humano es elemento decisivo,

puesto que el tiempo está al interior de él o es autónomo de toda humanidad. Independientemente del tipo de enfoque (Aristóteles, Newton, Kant, Einstein o Hawking), lo que es un hecho, es que la postura objetiva es la que prevalece en los presentes días, donde se asume al tiempo como algo lineal, como un fenómeno y, por lo tanto, se puede disponer de él por medio de la cuantificación (Scott, D., 2006). De este absolutismo se deriva una situación ontológica: el tiempo *es* (el *Ser* está en el tiempo). Con esta afirmación, es posible que el humano asuma que el tiempo es de él ya que lo puede medir y, con ello, se toma lo verdadero de lo que *fue*, está *siendo* y lo que *será* (este es un reflejo más del neopositivismo científico). Si esto es así, ¿Por qué existe la discrepancia de lo que fue la historia, lo que está pasando ahora y lo que sucederá para la humanidad? Allí radica el problema puesto que, desde Heidegger, "*el tiempo no es*", sino que "*el tiempo se da*" (Edwards, P., 1975), lo cual habilita que las decisiones que se deriven del humano o la tecnología sean un hecho en la proyección que *está siendo* en cada momento.

De forma general, la concepción agustiniana del tiempo nos remite al presente, pasado y futuro como tres conceptos diferentes (Knuuttila, S., 2001). Esta situación cimentó el carácter ontológico ya descrito y que llevó a aprehender el tiempo para el hombre. No obstante, cada instante que se acaba se convierte en pasado y al mismo tiempo se acerca al futuro, pero en realidad tanto pasado como futuro *nunca son*, es decir, no existen porque están anclados al presente por son referencia del "*ahora*". Nadie puede decir "*aquello fue*" sin estar en un "*aquí*" o bien, ninguna persona puede hablar del futuro sin estar en el presente. Por tanto, todo pasado y futuro son presente, son un "*aquí y ahora*", lo cual imposibilita deliberar en pasado y en futuro ya que esos dos estadios del tiempo que están condicionados al *no-ser* (error) y sólo se puede hablar en presente (conjunción del pasado y del futuro). Por ello, cada decisión es una afección al futuro con la presencia del pasado (únicamente podemos decidir de lo que precede para proyectarnos -experiencias y deseos-).

La anterior condición posibilita generar decisiones considerando "*lo que uno es y lo que uno será*", para indicar "*qué soy*" (decidir), es decir, pensar, asomarse a la ventana del pasado y arrojarse al futuro para analizar asumir el ahora. Si la ética es un discurso reflexivo, entonces

tomar en cuenta que "*todo tiempo es ahora*" ayudará en el propósito de deliberar, incluyendo los momentos de interpelación de la IA.

Con todo lo dicho se puede concluir que el "*ahora*" es el sujeto, es decir un individuo que está determinado por su pasado y que no puede acceder a un futuro cierto si no es bajo la ciencia que, en este caso, se vale de la IA para generar una probabilidad. De hecho, el sujeto es la manifestación más clara del "*aquí y ahora*", ya que "*es*" *por* el pasado y *para* el futuro, porque en cada instante *está siendo*. De allí que al momento de generar decisiones en una ética de la IA, se puedan concebir por un contexto determinado proveniente del pasado y enfocado al futuro.

Cabe destacar que la IA puede alterar la percepción de espacio y tiempo del individuo, ya ha llevado a creer que se está en lugares que no están en el entorno (Realidad Aumentada) o bien "*disminuir*" el tiempo de trámites al simplificar procesos arbolados (arbolado: reconocimiento de rutas, variables y posibles soluciones).

La consideración de una ética de la IA debe de tomar en cuenta el análisis de la observación, pero al ser algo visible se ocupa abordarlo desde algo diferente al humano: el tiempo *a priori* (dentro de nosotros como categoría). La problemática de analizar una decisión ética en la IA es que el tiempo se percibe y, por ende, se constata por lo que ocupa ser verificado. Una ética pragmática o consecuencialista parece ajustarse a la IA, porque en la estructuración de la programación del sistema ("*antes*" como tiempo), es necesario hacer pruebas ("*después*" como tiempo), pero si el tiempo avanza ¿Cómo se puede hacer un análisis de algo que está en constante cambio-movimiento? ¿Cómo podemos captar las decisiones éticas de la IA si nosotros mismos pertenecemos al movimiento? La respuesta es realizar revisiones periódicas al sistema para evitar los sesgos que se vayan entretejiendo en el tiempo.

2.4 Conocimiento de los datos

Cuando Kant mencionaba que todo conocimiento iniciaba con la experiencia pero que no se limitaba a ella, trataba de exponer que las fuentes de donde nos alimentamos para construir realidades o ideas, pueden ser desde diferentes orígenes. En la actualidad y más con la aceleración tecnológica que trajo el COVID-19, la realidad que percibimos está condicionada por la ciencia computacional, entre ella la Inteligencia Artificial. Antes del confinamiento social ya usábamos dispositivos que nos ayudaban a entender lo que pasaba a alrededor (afuera de nosotros), sobre todo a través de las pantallas de los dispositivos móviles, ordenadores o televisores inteligentes. Cada uno de estos artefactos funcionan por el diseño del sistema con el cual está configurado, pero más puntualmente, detrás de los *displays*, la mímesis tecnocefálica contribuye a crear una realidad gracias a los algoritmos que procesan la información que el *homo sapiens* percibe. Como ya se ha mencionado, los datos son la fuente con la cual se alimenta la IA, por lo que es necesario generar una reflexión de ellos y su forma de ser interpretados para generar realidades, porque de ello se pueden realizar consideraciones entorno de la ética.

El *bit* es una señal electrónica usada dentro de la ingeniería computacional que ayuda a generar un entramado proceso organizado para que el sistema pueda existir y entenderse. El conjunto de *bits* crea un lenguaje entendible que se llama información, misma que orienta el flujo de las acciones del sistema. Para el caso que nos ocupa, el desarrollador de la IA es quien selecciona los *bits* que serán no sólo información sino ruta para ser procesados, toda vez que el diseño y elección de método de Inteligencia Artificial juega un papel fundamental para que en su implementación se construyen realidades. El paso de un *bit* a una incidencia en el mundo es el proceso explicable que la ética ocupa para poder generar una vinculación en el caso que esa “transformación de realidad” cause daños a los humanos. ¿Cómo elige un desarrollador de IA o una empresa la información con la cual se va a generar una modificación del mundo real? La respuesta debería de ser la conciencia, pero es posible que la fuente y motor de las creaciones tecnológicas sean el beneficio o la utilidad que de ellas devengan en su momento (Ordine, N., 2017). Este esquema pragmático es originado por la

interpretación creadora del sistema, porque es justo en la atribución de un significado donde se origina la transformación del mundo o construcción de realidad.

Para el quehacer que nos ocupa en este texto, la realidad es el conjunto de información dada en una estructura sistemática que describe, conduce y/o afirma la Verdad (lo adecuado, lo mejor, lo bueno, etc.). Cuando en el año 2018 Amazon implementó un sistema de Inteligencia Artificial para contratar a nuevos empleados, algunos expertos se dieron cuenta que los candidatos seleccionados por el sistema eran mayoritariamente hombres a pesar de que había mujeres aspirantes. Estas últimas eran rechazadas por el algoritmo por el diseño de esa IA, por lo tanto, se estableció la información (órdenes) para que se valorara la comparativa histórica de otros elementos de trabajo contratados anteriormente por la empresa, dando como resultado una mayor puntuación a los candidatos hombres que a las mujeres (Davenport, T. et al., 2020). Aunque en este ejemplo es evidente un sesgo que refuerza la distinción de género, es necesario puntualizar que el origen de esta situación es la interpretación de los arquitectos de IA (personas e instituciones), quienes propician el filtrado de datos y las condiciones en las cuales se tratan esa información, por lo que difícilmente se podría afirmar si fue buena o mala la acción que realizaron en su momento al proponer dicho sistema, toda vez que "*es el mundo*" que conocen y desde donde parten para estructurar el plan y la implementación. Si existiera una realidad objetiva conocida por todos, los programadores, empresas, gobiernos, personas e instituciones involucradas en la IA, caminaríamos todos al unísono hacia allá, pero la realidad en sí misma no existe, sólo hay una trama de interpretaciones de información dada.

De lo anterior se desprende que los niveles de realidad están asociados a la forma en la que se interpreta la información que se recibe y se orienta a la Verdad. Para Platón, las ideas estaban separadas del mundo sensible el cual refleja una multiplicidad de interpretaciones, por lo que la existencia de las "*cosas en sí mismas*" era no solamente necesaria sino alcanzables a través del razonamiento, lo que resulta en una ponderación de la subjetividad. En el caso de Aristóteles, lo verdadero era una correspondencia entre el razonamiento y la materialidad, por lo que la interpretación podría ser objetiva. La postura de Einstein está encaminada a que los humanos no podemos captar la realidad en sí misma, sino que a través

del mundo físico se pueden crear representaciones, por lo que la apertura de diversas realidades simultáneas estriba en la elección de la interpretación del sujeto. En cualquiera de los casos metafísicos, pragmáticos o relativistas, siempre el mundo se construye gracias a la información que recogemos pero que es procesada por la interpretación (categorías) que nosotros asumimos, es decir, damos dirección voluntaria, involuntaria, consciente o inconscientemente a la construcción de la Verdad.

Toda realidad se construye con información misma que está organizada por argumentos, por lo que el lenguaje es el hilo que entreteje las interpretaciones del mundo, porque hasta donde sabemos, eso que está "*fuera de nosotros*" permanece idéntico así mismo (sin necesidad del humano) como un fenómeno (*φαινομενον*) hasta que lo afectamos (modificamos). El campo semántico de dicha afección es lo material como son las cosas, objetos o personas, pero también conciencias como son los modos de ser, pensar o ética. Si la información es lenguaje y crea realidades, y si la Inteligencia Artificial es un lenguaje, entonces esta tecnología construye realidades a partir de la interpretación de los desarrolladores. Existen algoritmos creados para generar *fake News* (Giansiracusa, N., 2021), que tienen un propósito establecido en la interpretación del desarrollador cuyo trayecto está atravesado por el tema ético: ¿La construcción de realidades derivadas de la orientación de las conciencias humanas por efecto de esas noticias falsas es adecuado o no adecuado? En este mismo momento el problema se complica más, porque la información antes dada era mayoritariamente por textos lo cual era posible que despertara ciertas dudas, pero ahora con *deepfake* (vídeos falsos), aumenta la credibilidad y el establecimiento de realidades puede ser más contundente, violento, silencioso y dogmático, lo cual puede llegar a nombrarse como manipulación digital (Maksutov, A. A. et al., 2020).

¿Los datos/información revelan o construyen toda la realidad? No, pero al menos en la interpretación y en el uso hacen creer o percibir un todo y que ese todo haga que el humano valore cómo deben de ser las cosas. De allí la importancia que una Inteligencia Artificial sea planeada, diseñada e implementada con una correspondencia de construcción de la realidad que beneficie no sólo en su momento al usuario sino a terceros involucrados, porque una interpretación de esta manera genera seguridad y fiabilidad en las personas.

En este sentido el perspectivismo nietzscheano no genera un relativismo al negar "*las cosas en sí mismas*", sino que es posible que ayude a crear un juicio crítico al aceptar la multiplicidad de interpretaciones sobre los hechos dados alrededor de la IA, porque ni el realismo ingenuo o idealismo platónico sustentan completamente la correspondencia entre la mente humana y la realidad, sino que todo es una creación del *homo sapiens* edificada por su capacidad lingüística de nombrar y describir todo lo que está afuera de él por medio de las categorías, por ejemplo, de *espacio* y de *tiempo*. ¿Por qué para la Unión Europea el marco de una ética de la IA considera sanciones económicas y las pautas de la UNESCO en el mismo tema sólo son recomendaciones? Es por la forma de interpretar lo vinculante de lo no vinculante en la ética de la IA que incluso tiene diferentes significados, porque la construcción de las realidades deriva de la forma en la que es interpretado el mundo.

¿Qué tan adecuadas son las realidades que los datos nos dan por medio de la IA en la actualidad? ¿Es sólo la IA la que construye estas realidades? Byung-Chul Han (2017) habla de la técnica que genera el *poder* para *poder* procesar los datos que se desprenden de la conciencia humana para ser usados como información, controlar a los humanos y convertirlos en sujetos (Han, B. C., 2017). La afirmación de esta conjetura tiene su origen en la prescripción foucaultiana de la microfísica (Foucault, M., 2000), ya que no existe "*el gran poder*" o "*el gran algoritmo*", sino un entrelazado de pequeños poderes que van construyendo la realidad por medio de la interpretación que le van dando el mundo. Tanto una estrategia nacional de IA suma a una ética, como la inversión en fibra óptica en determinadas rutas por parte del sector privado. La voluntad de poder de la cual se hace referencia es la de la conquista de los datos, de la información que genera el humano en la estela digital por la interacción con los sistemas directa o indirectamente. Por lo tanto, la construcción de la Verdad, realidad, mentalidad del y en el humano, pasa por el cedazo que el poder direcciona: sólo conocemos lo que se *permite* filtrar hasta la conciencia.

En la condición humana de interpretar por naturaleza todo el tiempo, se decide cómo es la realidad, lo cual debería de generar una actitud de apertura o disposición a realidades ajenas o discrepantes con las que uno tiene y se ha sumado. Esta situación no es menor, porque en

la preocupación de una ética de la IA global, se posibilita en diálogo con los demás y, por ende, acuerdos sin caer en relativismos bajo el método que se quiera elegir: mínimos, dialogismo, científico, negociación, etc. Uno de los beneficios de esta actitud de apertura es la disminución de sesgos, porque a nivel ligústico se consideran diversos horizontes, toda vez que no existe una sola realidad, sino múltiples construidas por el lenguaje, que es la única vía para expresar "*la realidad*" que tenemos construida a nivel metafísico.

No obstante, es necesario mencionar que en el entramado de los datos de las interpretaciones se realizan desde quien las hace, por lo cual migra sus propios sesgos, en este caso, desde la IA y que vulnera a las personas, especialmente a los usuarios. En efecto, si el programador asume la Verdad objetiva, afirma que hay "*hechos cerrados*", realidades concluyentes y acabadas, crea problemas al momento de postular un sistema. Por el contrario, aquí se recomienda percibir el mundo como una interpretación más de los datos que nos hacen edificar la realidad, para no caer en el dogmatismo. Entonces, el desarrollador de IA tiene su propia limitación de acceso a la Verdad, por lo que sólo traza rutas de acuerdo con visión, datos, reglas y métodos lo que después se convierten en soluciones, pero que traen cargando problemas o dilemas éticos como ya se ha venido destacando.

De manera general, los sesgos están presentes en la vida humana, por lo que la IA no está exceptuada de esta situación. En este caso, el sesgo es la falta de consideraciones de realidades humana en el horizonte computacional, lo cual crea errores sistemáticos generados por prejuicios (Mehrabi, N. et al., 2021). Por ejemplo, el sesgo implícito genera un valor a las personas o grupos determinados desde estereotipos, para lo cual se ocupa generar conciencia por medio de la educación y ser sensibles ante la diversidad y los marcos normativos. En tanto que el sesgo de muestra se da cuando los datos aleatorios de la muestra no representan de manera adecuada a la totalidad, favoreciendo las decisiones hacia subconjuntos, desfavoreciendo a minorías, por lo que se debe de evitar estas realidades revisando que los cálculos matemáticos sean proporcionalmente adecuados. Por su parte, el sesgo del tiempo indica que el sistema es obsoleto en la base de datos con la cual opera porque cuando se creó estaba enmarcado en otras realidades que se vuelven no vigentes y carecen de adaptación a nuevos contextos; en este sentido, se ocupa realizar actualizaciones

periódicas de los sistemas y de los datos. Por estas razones, conviene cimentar una ética en la revisión de los resultados de los sistemas de IA en las pruebas, sobre todo en aquellos resultados que se encuentran en los umbrales o valores atípicos. El resultado de estas observaciones es la generación de confianza, máxime si se cuenta con personas que gestionan los riesgos (valorar posibles sesgos) y con grupos multidisciplinarios (ampliar a diversas visiones). De igual manera se destaca que el sesgo no es sinónimo de prejuicio, ya que el primero es un procedimiento estadístico que favorece respuestas/decisiones determinadas desde las bases de datos, mientras que el segundo es una actitud de las personas. Así pues, la IA puede contener sesgos con prejuicios, pero no todo prejuicio es un sesgo, porque de hecho se puede tener el prejuicio de que los humanos viven, pero no de que todos los humanos están vivos o son hombres.

Desde luego, existían sesgos antes de la aparición de la IA, el problema ahora es el daño causado por la masificación en el acto de usar y replicar de forma automática-sistemática. ¿Por qué la IA tiene sesgos? Porque los humanos tienen sesgos y los datos que se ocupan para generar dicha tecnología no son objetivos en sí mismos, sino que son tratados con la interpretación que orientan su accionar. Por su puesto, dicha interpretación está siendo operada por métodos de IA que todo el tiempo discriminan (clasifican) datos, lo cual por sí mismo ya crea un problema desde el orden metafísico que se detallará más adelante. Por ahora, la gobernanza de los datos no significa sólo una responsabilidad de quienes desarrollan mímesis tecnocefálica, sino también de los gobiernos que deben de auditar los sistemas que interpelan en gran medida a la ciudadanía, así como también otras organizaciones que pueden postular buenas prácticas de gestión de datos como los estándares para tener y procesar información de calidad y al menos preguntarse ¿Cómo, para quién y para qué se recolectan los datos? En ese sentido, es indispensable indicar que una ética de la IA no se reduce al tratamiento de los datos, también al "*para qué*" se recolectan, toda vez que el "*cómo se procesan*" son las técnicas que el programador usa en su forma de ver-entender el mundo.

CAPÍTULO TERCERO

3. Fenomenología digital

¿Qué vamos a hacer con estos detectores [sistemas inteligentes] *ahora que los tenemos entre nosotros*?

Joseph Redmon

3.1 Moral sintética

Varios diccionarios coinciden que la moral es un conjunto de normas dadas por el entorno que son clasificadas como buenas o malas. Del latín *mos, moris* que significa "*costumbre*", la moral se relaciona con un contexto en el cual se *es*. A diferencia de la ética, según lo ya enunciado, la moral se ocupa de una distinción práctica de los actos, resultando un cúmulo de efectos. De igual manera, dado que la IA transforma el mundo y afecta al humano, es pertinente cuestionarse si esta tecnología es se acerca a la moral o si de hecho tiene moral en sí misma.

En recientes debates se ha planteado la cuestión de que, si la tecnología es o no neutral, porque la sospecha se acrecienta debido a que, desde su planeación y desarrollo, la mímesis tecnocefálica es creada por alguien y para algo. El autor y el propósito son ya estimulados para llegar a creer que la IA no puede ser neutral en sus inicios puesto que contiene toda la carga afectiva que los programadores o instituciones concentran desde que nace. Sin embargo, existen posturas que indican que la tecnología no es buena ni mala, ya que es inerte y no actúa por sí misma en el sentido de que es una herramienta del humano, quien la usa con el propósito (bueno o malo) que él decide. La mímesis tecnocefálica siempre tiene intencionalidad porque todo el tiempo carga consigo los prejuicios (no en un sentido negativo) de los humanos, es decir, las bases históricas que se ha heredado, por ejemplo, que todo sistema de IA tiene que mejorarse de forma constante porque el progreso es bueno.

Ahora, la IA siempre cae en el mundo, por lo que está afectada por quien la diseña, crea, solicita, está dirigida y usa. ¿Cómo negar que los algoritmos no contienen intencionalidad de programador, la empresa, el cliente, la sociedad, la ciencia o la economía? Por todo esto se puede afirmar que, en algún lugar del trayecto, la IA asume la moral de la circunstancia en la que fue creada y, por tanto, no es neutral.

Para Kant la moral tiene una condición metafísica, la afirmación de que el humano es libre, ya que si no es de esta manera seríamos como otras especies de animales que actúan por instinto, pero toda vez que gozamos de autonomía por la razón podemos obrar según principios que adoptamos siempre y cuando se considere a la "*otredad*" en el proceder, así como la universalidad del acto a realizar esperando que entre todos alcancemos ese bien. De esta forma se consigue la ponderación de la dignidad, fundamento de los Derechos Humanos, de actuación del deber por el deber. Entonces, un sistema de IA que discrimina a las personas por su color de piel o condición económico-social puede ser agente moral, ya que carga dentro de sí la promoción de valores con los cuales se le han entrenado. En ese caso, las máquinas tendrían moral porque resguardan, operan costumbres y normas por medio de los datos y reglas, además de que realizan juicios y decisiones en torno al comportamiento, también distinguen entre lo bueno y lo malo según se le haya programado (sin referirse a la conciencia) y, no menos importante, fomenta un equilibrio en la sociedad gracias a los algoritmos destinados a hacer cosas que son percibidas como buenas o malas como colectivo. Por otro lado, hay quienes argumentan a favor de la intuición como una característica que impide a la IA ser moral y, por su parte, porque también la moral es una consideración de estar en contacto con los demás que son los que ayudan a distinguir lo bueno y lo malo (solo se es moral gracias a la comunidad); y, quienes indican que los sistemas de IA no tienen moral es porque justo la moral está destinada exclusivamente a las personas y mientras los humanos actúan por sí mismos, las máquinas están sujetas a la programación (determinada).

¿Los humanos somos libres y no actuamos con programación? No, estamos determinados no por algoritmos que nos inyectaron con ingeniería computacional al momento de ser concebidos, sino por el hecho de *estar-caídos-en-el-mundo*, es decir, por estar vivos y pertenecer a una especie en un momento preciso de la historia. El símil es contundente, el

sujeto está *sujeto*, el humano está condicionado por el *aquí* y el *ahora*, no por el gran programador denominador, *tiempo* repartido (pasado, presente y futuro) y *espacio* (contexto). Cuando escribo estas letras lo hago de una forma que se me ha enseñado, usando reglas gramaticales, siguiendo una estructura para lograr una comprensión y, en su caso, modificar (afectar: crear efecto) en un *otro* que lo lee. Eso es un algoritmo, el cual fue una solicitud del cliente llamado *modernidad* unida a la requisición-metodología de la Razón-Ciencia para crear un producto denominado *entendimiento* que creará realidades-Verdades. ¿A quién no se le ha cargado el algoritmo de vestirse adecuadamente, respetar las leyes o ayudar a los demás? al menos en occidente, se observa un sistema operativo que articula una moral predominante, la del humanismo. No somos maquinas, las maquinas son como nosotros. Todo se parece a su creador -*dueño*- y hereda sus cosas porque es objeto de su pertenencia. Es así que el deseo de todo programador de IA es crear un sistema exitoso, no estúpido.

Hasta aquí, queda muy fuerte la sospecha de que la IA tiene moral, pero mientras la claridad llega, podemos afirmar un asunto más que circunda: debemos cambiar el concepto de moral. La tecnología ha sido creada para ayudar al humano potenciando, en un primer momento, sus capacidades físicas. El automóvil ayudó a los pies y su capacidad de promover la movilidad; los anteojos perfeccionaron los ojos, llegando hasta la finura de tener una vista de los virus que no pensábamos que existieran por su diminuto tamaño; las computadoras han servido para desarrollar la capacidad de pensar y resolver problemas, superando ahora con la IA algunos procesos simples repetitivos, los cuales ocupan gran cantidad de información que la memoria humana no puede almacenar. Creamos tecnología para perfeccionar el cuerpo, ahora la tecnología es parte del cuerpo, es decir, se *incorpora* a nosotros como una extensión (celular), da soporte a al cuerpo (exoesqueletos) o incluso da vitalidad al cuerpo (marcapasos), y todo ello tiene mímesis tecnocefálica. Por estas razones, si no se quiere asumir que las máquinas tienen moral, al menos hay que cambiar el paradigma de la enunciación de la moral, porque ya ahora es parte de nosotros y ella influye en las decisiones, especialmente toda IA que ha creado vínculos con humanos que ahora el *homo sapiens* emiten añoranza en el caso del vacío o ausencia de esta tecnología.

Una moral sintética es posible pensado en que la tecnología ha estrechado lazos en la concurrencia del ahora, en la unión de los elementos: humano e IA. Algunos estudios demuestran que las redes sociales no sólo sirven para comunicarse, sino que crean comportamientos éticos y morales, y afectan las opiniones y elecciones de las personas (Yang, K. C. et al., 2019). Y si a esa circunstancia le sumamos los crecientes actuadores y sensores en la vida diaria que recopilan datos, las mejoras las técnicas en los sistemas computacionales, la obtención datos biométricos, el aumento del conocimiento de la neurociencia, la mayor velocidad de procesamiento y almacenamiento de información, entonces, el libre albedrío estaría en tela de juicio, al menos ya no sería tan útil porque la objetividad de la IA podrá decidir por el humano en términos éticos y morales (Melhado, F., & Rabot, J. M., 2021). Si la ética y la moral toman a la experiencia (datos) para valorar qué conviene más frente a los dilemas que se presentan, entonces es posible que la mímesis tecnocefálica pueda resolver de manera práctica y objetiva entorno de la conciencia o intuición, ya que el humano está condicionado por la fatalidad de la contingencia que apremia el dinamismo de la vida. De esta forma, los dilemas como el de Heinz o del tranvía utilizados en la enseñanza de toma de decisiones con principios morales, pasarán de ser dilemas a ser problemas, toda vez que las bases de datos recopiladas durante toda una vida servirán como punto de partida para alinear las decisiones éticas y morales, incluso mejor que la de los humanos. La IA sabrá de forma cuantitativa cuál es la mejor decisión ética en un espacio y tiempo, al hacer medias o configuraciones particulares (traje a la medida moral) que convenga según una determinada situación, o al menos los sistemas de IA podrán hacer cálculos para generar equilibrios probables y de gravedad en la gestión de riesgos éticos (Krügel, S., & Uhl, M., 2022).

No debemos de temer a estas situaciones, ya que todo pasará a la normalidad de una moral híbrida o totalmente sintética. En su momento, los mismos escándalos sucedieron con la aparición de la radio y la televisión, avecinando gente manipulada y hueca. Nada más contrario a lo que se vive hoy en día ya que, tanto como las radiofrecuencias como los *displays* también ayudan -y mucho- a construir conciencia en la gente, tal como sucedió en la primavera árabe en Egipto (Kile, F., 2013). Desde luego que existen peligros entorno de ese control sistemático, por eso es buena la oportuna coyuntura de que a nivel global exista

un debate sobre la ética de la IA, porque esto al menos socializará la ayuda colaborativa o la heurística del temor que de ella se puedan desprender, generando oportunidades para compartir los beneficios o perjuicios actuales, generando procesos de apropiación de la tecnología para que en su momento la moral sintética nos indique suscribirnos a una ética, al menos provisional, y configurar el entorno digital para que el sistema decida por los humanos, o bien que el sistema detecte que sea de otro modo por una sugerencia.

Pensar en una moral sintética no debería de ser un problema si se asume que se puede ayudar a simplificar la toma de decisiones a través de los dispositivos con IA que generen conocimiento *tecnoafectivo* (Melhado, F., & Rabot, J. M., 2021), acercándose al umbral de una vigilancia que podría ser escandalosa para algunos pero que no necesariamente sería algo negativo para otros (Hagerty, A., & Rubinov, I., 2019), toda vez que dicho control disciplinario (Deleuze, G., 2017) podría crear un bien social que desemboque en beneficios particulares (Russell, S. et al., 2015). La transformación tecnológica no es más que una adaptación a las realidades que van surgiendo, en este caso desde el ámbito *tecnomoral* (Keulartz, J. et al., 2002).

Considerando que un algoritmo es un conjunto de reglas secuenciadas para llegar a un determinado lugar (objetivo), en la actualidad se le está dando relevancia a este concepto dada la trascendencia computacional de estos días donde los sistemas de IA no sólo son útiles, sino hasta necesarios. El mecanismo de hacer algoritmos sintéticos, ya existía antes de la aparición de la IA, porque hacíamos algoritmos en la agricultura (serie de pasos para producir) o incluso ya existían los algoritmos en el organismo humano (modos de realizar el metabolismo), sólo que no los habíamos nombrado de esa forma; la cuestión es que ahora se prenden las alarmas, ya que los algoritmos computacionales toman mayor parte de las decisiones o determinaciones.

Del mismo modo se tiene que considerar que, por ahora, la IA con la que contamos sólo interpreta símbolos, números, pero no es capaz de procesar "*todo lo demás*" ya que los sistemas son primitivos (Inteligencia Artificial Estrecha), sólo hacen una o pocas funciones limitadas.

3.2 Cuestión metafísica

Hay diferentes formas de abordar la metafísica, por ahora conviene acercarse a la definición aristotélica, la cual considera que es una disciplina encargada de estudiar los primeros principios, no referirse a los históricos, sino al fundamento de las cosas (Gasser-Wingate, M., 2020). La relación que existe entre la metafísica y la IA es la síntesis y construcción de la realidad, es decir, la afección-transformación del mundo.

La metafísica como disciplina que busca el conocimiento de lo que las cosas son en sí mismas, lleva a la ontología, el estudio del *ser*. Opuesto a ello, tenemos las cosas que están en el mundo y que se caracterizan por poseer ese *ser*, y que siempre se materializan. De tal modo, subsisten por distintos lugares lo metafísico y lo real. En lo metafísico encontramos todos los conceptos de las cosas y sus características que los hacen diferenciarse de otras representaciones que tenemos; en ese sentido, autores como Platón y Descartes configuran su discurso en las ideas y en la razón, que son por sí mismas. Por su parte, el mundo real es la representación de esas ideas, es la materialidad de lo que se concibe y donde se realizan todos los fenómenos, ponderaciones que promovieron tanto Aristóteles como Hume. Durante mucho tiempo empirismo y racionalismo se mantuvieron distantes hasta que Kant pudo unir las dos formas de abordar el problema, ya que el conocimiento es el encuentro entre categorías *a priori* del pensamiento y la experiencia de las *intuiciones sensibles*, surgiendo una síntesis de lo que trae cargando el sujeto (*categorías*) y todo lo que está fuera de él (realidad material), al concluir que la *forma* ya no está en la realidad, sino que está en el sujeto. La metafísica, que es lo *a priori*, es necesaria para la ciencia basada en la experimentación (*a posteriori*). Ni realidad sin metafísica, ni metafísica sin realidad.

La IA está sustentada en los *bits*, los cuales están estructurados y representados por los números que proveen las matemáticas y que son ordenados por ciertas reglas. Cuando hablamos de mímesis tecnocefálica, se está concibiendo dicha síntesis entre la imitación de un proceso neurológico (*a posteriori*) por medio de la lógica (*a priori*). ¿Dónde están los números que componen un sistema binario si no en los conceptos del humano? ¿Por qué entonces los modelos matemáticos afectan al mundo real? Porque todo algoritmo es

metafísica, sólo un conjunto de representaciones/interpretaciones que se han ido fraguando en la razón humana durante la historia. La IA es conocimiento porque tiene de su lado a la razón, pero ocupa del mundo sensible para *ser*. Y es que, en el proceso de crear un modelo matemático se busca la *universalidad* y la *necesidad*, las dos características esenciales de la metafísica, porque cuando se planea el sistema se espera cubrir todos los casos posibles. Justo esas son características de la ciencia y su despliegue, el de querer siempre "*aplicar para todo*" y "*ser comprobable*", lo cual confirma que la IA está dentro de la ciencia computacional, porque toda programación nunca busca resolver un problema aislado, sino que por el contrario se dispone a ser replicada en condiciones similares a otros casos. Entonces, no podemos conocer qué son los números, pero sabemos que son útiles para entender y transformar el mundo, en tanto que ello significa generar una realidad por medio de la interpretación que se le da.

Los números son expresiones cuantitativas que revelan una unidad dentro de los sistemas y sin que ellos se puedan conocer por estar en el lado metafísico, son la causa de todas las causas de la IA. En efecto, el 0 y el 1 son abstracciones que configuran un lenguaje que es entendido por las computadoras. Puesto que el humano es el creador de esos significantes, toma de ellos la forma lingüística para poder operar: una estructura determinada que se conoce como reglas o gramática. Este ordenamiento es un legado de la lógica aristotélica que tiene como principio axiomas de argumentación para concluir y, en el caso de la IA, tomar decisiones.

Mientras Wittgenstein y el círculo de Viena rechazan toda metafísica (Stern, D., 2007), aquí se afirma que es necesaria para la construcción de la realidad en la que vivimos, porque al estar rodeados de tecnología, se vindica que el fundamento de ser de la IA es la metafísica al funcionar desde la abstracción de los *bits*. ¿Cuál es la relación pragmática entre la IA y la metafísica? Inputs, outputs y las decisiones; datos, reglas y resoluciones.

Sea cual sea el camino para decidir, se tiene que asumir la estructura prevaleciente para deliberar: una composición binaria del mundo conceptual-lenguaje, más en el mundo digital. En efecto, Derrida (Hepburn, A., 1999) nos muestra que el mundo está compuesto por

opuestos binarios sustentados por conceptos que determinan el proceder en una decisión. La aseveración de tal situación lleva a comprender que por cada dilema se abre la dicotomía: sí o no; bueno y malo; correcto e incorrecto; falso y verdadero; etc. En este sentido, incluso el no asumir una postura frente a un dilema moral, es tomar postura (se actúa o no se actúa), por lo cual la necesidad de decidir es un impulso que posibilita la realización de una deliberación. Mientras más claro se tenga este argumento lingüístico, es mayor facilidad de acercarse a la actuación moral justo en el momento adecuado.

Por su cuenta, uno de los principios del racionalismo metafísico es el *Análisis* propuesto por Descartes el cual se cuenta como un segundo paso precedido solo por la duda metódica. En ese contexto, el *Análisis* es la reducción de lo complejo en partes más sencillas y pequeñas, por eso es por lo que la IA procura acotar y simplificar la complejidad de la realidad ordenándola en conjuntos simbólicos, para generar procesos direccionados hacia una solución.

De la misma manera, la metafísica se ocupa de ordenar la realidad por medio de las clasificaciones del "todo" (cosmos). Esto es muy representativo para la IA, toda vez que esa tecnología realiza agrupaciones de datos y discrimina los que no le son convenientes para llegar a la solución planeada. Desde luego, vivimos en un mundo desordenado y caótico, pero para poderlo comprender y manipular, hace falta generar esquemas con los cuales se pongan las cosas en su lugar, no por el hecho de ser ordenados simplemente, sino porque está en el humano una orientación a lo clasificado y que genera tranquilidad. La metafísica ayuda a liberarnos de la angustia, es el fármaco (*φρμακον*) informático de los tiempos actuales. Las etiquetas de los datos en la IA, procuran acercar al programador a la paz de sentir que todo está en su lugar. Llegar puntual a una cita gracias a las aplicaciones con IA, obtener un diagnóstico médico preciso con software con IA, tener una ganancia en la bolsa de valores por a las decisiones que toma la IA, hacen sentir al humano que todo está bien, lo cual puede denominarse satisfacción. En tanto que el descubrimiento de falsos negativos y positivos son la perdida de la tranquilidad por la aparición del error, en ontología-metafísica se trata de no caer en el *no-ser*, en la nada o el vació. Y como ya se ha dicho que el humano da sentido al mundo, entonces toda programación computacional está orientada a vindicar el *ser*, es decir,

evitar el error por conceptos puros, necesarios y universales, tomando en cuenta que la realidad *no es* (hecho), sino que es siempre *posibilidad de ser*. El resultado que tiene esta digresión para el tema ético es trascendente porque de hecho cuando se habla de una IA con responsabilidad, se está buscado no hacer el mal como principio y fin en sí mismo, transformando para el programador esto en una pregunta ¿Lo que se desarrolla es bueno para todas las personas en todos los casos similares? Por lo tanto, la metafísica sustenta a la IA y además es necesaria para orientar el actuar ético porque en el camino se descubre que el "*yo*" tiene una relación de responsabilidad con "*el mundo se le presenta*".

Por otro lado, la metafísica crea el mundo, da sustento a las creencias o conjeturas del *homo sapiens*. En su momento, dicha diciplina filosófica sustentaba el buen comportamiento de los humanos en la tierra para alcanzar el cielo. A raíz de estos horizontes se conquistaron territorios, se firmaron acuerdos, se aprendió a usar vestimenta, se crearon los Derechos Humanos y se sostiene que lo mejor es el bien común. Pero ni las instituciones, ni el dinero existen, todos son constructo del humano, de esa capacidad de abstraer cosas para crear el mundo (Harari, Y. N. 2014). Lo mismo pasa con la IA, porque con los *bits* y el binarismo se crea un lenguaje de programación que se transforma en filtros para fotografías, tendencias de noticias, desahogo de procesos gubernamentales, software para hacer videollamadas, es decir, el humano crea el mundo, una realidad con la tecnología. Entonces, se tiene una IA que afecta al mundo y, el mundo, afecta al *homo sapiens* por un principio metafísico, porque las decisiones que de esta tecnología devienen, sus preguntas, angustias, sentimientos, obligaciones, movilidad, cambio, conformidad. Todo esto tiene un sentido ético, porque de hecho ninguna tecnología es independiente al humano por más autónoma que se predique, al menos hasta ahora no sabemos que una máquina haya creado *totalmente* a otra máquina.

Surge una cuestión ética de la IA con un referente metafísico: la discriminación. Ya que la separación de datos por las etiquetas asignadas forma parte del ordenamiento conceptual, al final la analítica resume a símbolos breves: sí o no. Desde la ontología, este sistema binario hace resaltar la "*diferencia*" (Käufer, S., 2005) por lo que se refiere a la distinción de los entes que están unidos-relacionados entre sí y la tecnología que hace que se separen las mezclas diluidas por medio del 0 y 1. Este ordenamiento de la realidad supone una

disminución de lo diverso, lo que deja de fuera a lo que no es etiquetado de tal o cual manera y esto puede llegar a ser una discriminación en un sentido de los Derechos Humanos. Lo particular queda fuera de las clasificaciones, el sesgo de minorías subyace dando sentido a la realidad y a la vida. A veces ni la vida ni la realidad tienen sentido, sino que se basan en el absurdo de no ser igual a la etiqueta que la moral establece. Entonces, la realidad de la realidad es que no es ordenada, es caótica por naturaleza. Allí está una de las angustias de los programadores, el de crear algo que esté fuera de control, que no funcione, que promueva colisiones de conciencias y que se pueda llegar hasta una situación legalmente vinculante. El dilema ético-metafísico está en que entre más se procura el orden-clasificación, más se acerca a la simplificación del mundo y a su entendimiento, pero esta separación es difícil de establecerse porque ninguna cosa es idéntica en sí misma, todo está ligado y más en la ética, porque somos con los demás, no estamos aislados, pertenecemos a una convivencia social y solo puede ser comprendida y realizada en relación con otros semejantes. Gracias a la metafísica, la mímesis tecnocefálica busca cosas en común en entre los entes del mundo para generar tranquilidad, pero en esa reducción de agrupaciones aparecen las minorías, contingencias y anomalías que quedan fuera pero que ocupan también formar parte de ese "*ser*" computacional.

Ésta es la respuesta a la pregunta del por qué cuesta tanto a la humanidad ponerse de acuerdo cuando se habla de ética en la IA (y quizá de tantos temas diversos), porque la metafísica contiene al *Ser* y el *homo sapiens* quiere imponer al *Ser* en las cosas, en las acciones y éstas diversas tal que construyen realidades múltiples. Un ejemplo contundente de ello son las diversas cantidades de lenguaje computacional usados para programar, porque ese "*código de lenguaje*" es la representación metafísica de un programador, pero como hay diversas manifestaciones del lenguaje, entonces hay diversas instrucciones-secuencias en la IA (Python, R, C, C++, C#, Ruby, SQL, Visual Basic, Java, PHP, etc.). El humano hace IA y él es quien migra todos los problemas que le subyacen, en este caso el de no asumir la *diferencia ontológica* y querer ser universales por medio de la metafísica. Se ha creado una informática a partir de categorías que pretenden acercarse a la *naturaleza* del humano, en este caso mímesis tecnocefálica como una imitación de funciones cerebrales, y con la ayuda de esa

metafísica se procesan los datos por medio de palabras (conceptos), algoritmos y reglas que crean un sin número de realidades en el mundo.

Una persona no es un *bit* o "*un solo número*", no se reduce a un dato. El humano es "*muchos números*" y siempre está cambiando su codificación (devenir), quitando y sumando nuevas variables, porque de hecho las personas son vida. Por estas razones, la metafísica podría servir para caminar a los universales desde el terreno computacional y ético.

3.3 Conciencia 2.0

Hasta ahora, el único *ente* que se ha demostrado que tiene conciencia es el humano. El uso de Redes Neuronales Profundas en la IA invita a pensar que los sistemas informáticos podrían llegar a tener una conciencia, toda vez que se estima que las neuronas pertenecen exclusivamente al cerebro, donde regularmente se aloja la conciencia. ¿Podrá llegar a tener conciencia un sistema de IA? ¿Será posible alcanzar una conciencia diferente a la biológica conocida?

Es complicado definir la conciencia de forma simple, pero por ahora se puede considerar como la autopercepción de los actos que se realizan o de los cuales se es testigo y que permite delimitar lo que es bueno o malo por medio de sensaciones subjetivas (Searle, J., 2009). La conciencia no es ética y ni moral, toda vez que dicho reconocimiento es quien propicia a la ética como reflexión desde un entorno moral. En ese sentido, la conciencia es el único acceso a la ética, pero no necesariamente a la moral, recordado que la primera es un ejercicio reflexivo y las segunda es la condición de estar con los demás (*ser-en-el-mundo*).

Bajo estas consideraciones, es posible indagar sobre la Conciencia Artificial, sobre todo por la antesala que nos ha puesto la tecnología, la ciencia y el transhumanismo, pero especialmente la neurociencia y la biología. Por un lado la neurociencia ayuda al descubrimiento a la conciencia porque usa las técnicas de resonancias magnéticas funcional (IRMf), la electroencefalografía (EEG), imágenes de resonancia magnética funcional o

resonancia magnética funcional (IRMf), cámaras, sensores, electrodos y todo un software cargado con IA, que revelan una actividad cerebral en el acto consciente porque se puede ver en qué parte del cerebro y bajo qué circunstancias el humano se autopercibe (Hobson, J. A., & Pace-Schott, E. F., 2002). Por su parte, la biología ha estudiado la composición de la naturaleza, especialmente lo orgánico y sus procesos de creación de vida, de modo que ha llegado un punto que dicho conocimiento ha servido para crear vida por medio de la biología sintética, combinando en laboratorios componentes orgánicos dando la existencia a realidades nuevas vivas (Benner, S. A., & Sismour, A. M., 2005). Por eso es por lo que hasta aquí se ha hablado de mímesis tecnocefálica, por la asociación que hemos impulsado desde la ciencia y tecnología para procurar la imitación de las funciones cerebrales, porque al parecer el origen y precursor de la conciencia está en la región cefálica humana, que no es lo único para la autopercepción, pero posiblemente lo más importante porque es el conducto de interpretación de lo que el *homo sapiens* es, piensa, hace y cómo se proyecta.

Ahora bien, si consideramos que técnicamente todo lo que es metafísicamente cuantificable, puede ser representado en un *bit* y luego en un algoritmo para generar IA, y si la conciencia es completa o parcialmente cuantificable, entonces es posible tener una Conciencia 2.0, la que viene después de sistemas biológicos. Por ello se considera que existe la inclinación de crear una Conciencia Artificial, que sea *un otro* (la tecnología), aunque lo más cercano por ahora es que la conciencia humana pueda ser apoyada por la IA, la neurociencia y la biología. A cualquiera de estas dos formas, se le podría nombrar Conciencia 2.0 porque sería el desprendimiento de los sistemas biológicos sobre la autocomprensión de los actos para hibridar la antigua naturaleza con la actual naturaleza (2.0), favoreciendo una moral sintética.

Lo anterior es posible porque se considera que, para tener conciencia, primero se ocupa pensar, crear razonamientos específicos de manera articulada y coherente. Este paso ya lo tienen los sistemas de IA, ahora la siguiente meta es generar múltiples acciones de pensamiento de los sistemas, que vayan más allá de resolver problemas de patrones, utilizando la intuición para abordar el mundo, así como el aprendizaje por sí mismo de nuevos conocimientos (Lake, B. M. et al., 2017). Luego, la neurociencia revela que los actos de conciencia podrían ser predecibles-cuantificables en combinación con las sensaciones-

impulsos, lo que puede crear mediciones-*bits* que puedan ser interpretados en lenguaje de programación computacional gracias a los descubrimientos de la incidencia de la oxitocina, melatonina, dopamina, vasopresina que conllevan a la conciencia, moral y hasta ética. Este umbral de determinismo ha motivado a que se piense en neuroderechos (Yuste, R. et al., 2021) e incluso ha gestionado la primera discusión parlamentaria y constitucional en Chile.

Para sumar a estos argumentos, hay que puntualizar que la conciencia promueve la actuación por uno mismo, por ejemplo, aprender (voluntad de conocimiento), y esto es lo que ya hace la mímesis tecnocefálica con su Machine Learning o Aprendizaje Automático, que habilita la posibilidad de que el sistema computacional aprenda por sí mismo por medio del auto entrenamiento (Zoph, B. et al., 2020). Si este autoaprendizaje lleva a un autoreconocimiento, se pueden crear sistemas como el robot diseñado en la Universidad Tecnológica de Nanyang, Singapur (NTU Singapur), que es capaz de percibir el dolor y reparte a sí mismo, por medio de sensores locales (mini cerebros) que mandan datos a la unidad central donde se procesan para deliberar (John, R. A. et al., 2020). Esta situación habilita las alarmas que la Unión Europea considera importante detener, ya que el reconocimiento del dolor artificial (fenomenología sintética) es una experiencia subjetiva que no puede ser en el mundo (Bentley, P. J. et al., 2018), aunque no necesariamente eso es algo negativo porque el dolor biológico ayuda el cuidado y la salud. Esta capacidad de aprendizaje de autocuidado implica un reconocimiento del cuidado humano como la predisposición de un compromiso del bienestar humano, por lo que dicha Conciencia 2.0 puede ayudar al *homo sapiens* desde el horizonte de la conciencia biológica del reconocimiento de la satisfacción o realización, que de hecho ya hace la mímesis tecnocefálica por medio de los sensores con su percepción y actuación de los cambios en su entorno. Ahora bien, ¿Se debe de prohibir la fenomenología sintética? ¿Por qué el humano debe de ser el único *ser* con conciencia? Si el humano posee el monopolio de la conciencia ¿Por qué se les enseña a las mascotas a "*portarse bien*" procurando la imitación del *homo sapiens*?

Hasta ahora, se había pensado que una de las características de la conciencia humana era la exclusividad del libre albedrío, pero ahora el umbral de dicha situación se puede desvanecer ya que investigaciones de la Universidad de Columbia y el Instituto Politécnico Rensselaer,

demuestran que es posible programar sistemas de IA para que sean capaces de generar sus propias determinaciones al momento de elegir, por medio de la enseñanza robusta de una lógica (Bringsjord, S., 2008), llegando a la autoconciencia artificial (Bringsjord, S. et al., 2015) en las experimentaciones en sistemas computacionales (Licato, J. et al., 2016). Entonces, si se considera la concentración de conocimiento entorno de la conciencia (especialmente el que se entrelaza con la neurociencia y neuroética), además de que si el grado instrumental para transparentar dicha conciencia es cada vez más posible por el *hardware* y *software* (dispositivos para representar el cuerpo humano en imágenes; y la IA como medio de interpretación), y si la metodología implementada de transformar la realidad en *bits* para ser "decodificada", presentada y procesada por sistemas cuantitativos digitales, entonces estaremos hablando de que es posible alcanzar una conciencia sintética o Conciencia 2.0.

La conciencia es uno de los puntos centrales de la ética, que a su vez está ligada a la subjetividad. ¿Qué es lo que está debajo del sujeto (*sub-iectum, sub-iacere*)? Su historia (datos), lo que representan para él (interpretación) y lo que desea ser (proyección y decisiones). Entonces, en algún grado la mímesis tecnocefálica es consciente porque recoge algunas de estas características, diferenciando que no es lo mismo IA que Conciencia Artificial.

Aunque de forma reciente se empezaron a crear sistemas de IA que generan reflexiones éticas como el sistema muy criticado de Delphi (Jiang, L. et al., 2021) o el robusto sistema de Megatron (The Conversation, 2021), se espera crear después de estos avances, nuevas funciones que no se encuentran en la antigua naturaleza a partir de la metafísica de la ingeniería computacional, haciendo que la Conciencia 2.0 sea posible en el sentido de reprogramar los sistemas biológicos para complementar el autoconocimiento o crear entes que son ajenos al *homo sapiens* con capacidades de percibirse a sí mismos.

En la frontera del transhumanismo y los efectos éticos-morales que devienen de la mímesis tecnocefálica, Jean-Luc Nancy nos invita a pensar bajo la pregunta ¿Quién soy yo? (Nancy, J. L., & Piazza, V., 2006). Si ese "*yo*" está conformado por algo, y puede ser la IA desde un

sentido sintético. Si hay algo más que "*no soy yo dentro de mí*", entonces eso "*otro*" es un extraño, un extranjero que no sólo es un intruso en mi territorio que llega de manera violenta y me toma por sorpresa, con astucia, sino que me ayuda a sobrevivir como lo hace un marcapasos en el corazón o un trasplante de corazón. Ese "*otro*" es la tecnología, la IA. ¿Qué consideraciones va a tener la humanidad con dicho extranjero?

CAPÍTULO CUARTO

4. Conclusiones y trabajo futuro

Los futuros sujetos artificiales de experiencia no tienen representación en el proceso político actual, no tienen estatuto jurídico y sus intereses no están representados en ningún comité de ética.

Thomas Metzinger

4.1 Conclusiones

La Inteligencia Artificial es una tecnología que ha llegado por las manos del hombre al mundo y que, de hecho, le interpela de una manera muy fuerte, al grado de tener una gran necesidad para poder *ser*, pero que en el trayecto se ha creado problemas y dilemas morales que postulan una reflexión ética. Esta realidad ha llevado a que diversas organizaciones globales postulen marcos éticos de la IA que ayuden a tomar reservas en el horizonte tecnológico de la responsabilidad para llegar a una mímesis tecnocefálica fiable. Para que esto se pueda dar, se ha propuesto una taxonomía que incentiva la reflexión ética en la Inteligencia Artificial, realizando un análisis jurídico, científico y de ingeniería computacional que posibilitan una apertura de los sujetos en el diálogo para crear principios y resaltando lo que hasta ahora en diversas latitudes resulta ser más homologado, tomando como base la delimitación de una ética de la IA. Dicha taxonomía comprende principios conceptuales (Mímesis tecnocefálica, Estela digital, Heurística del temor, Optimismo Digital, Colisión de conciencias, Naturaleza 2.0, Moral Sintética y Conciencia 2.0) y principios dialógicos (Razón, Abstracción, Metafísica, Verdad y Contingencia). Por su parte, la dispersión de conceptos que nos trae esta tecnología como la autonomía, crea problemas en el choque de las conciencias, rozando un relativismo que puede ser peligroso para generar acuerdos éticos. No obstante, se postula abordar nuevos conceptos, como el de la naturaleza y conciencia, para entender y alinearse a los desafíos éticos de la Inteligencia Artificial. Tomando en cuenta que los datos son la principal fuente de movimiento de dicha tecnología,

se puede crear una epistemología y fenomenología subyacente, la cual apertura una moral híbrida basada en la comprensión de la metafísica como base de la Inteligencia Artificial, que puede llevar a una nueva conciencia, que se interseca con cuestiones éticas hasta ahora universales como: privacidad, gobernanza de datos, transparencia, uso responsable y dignidad humana.

4.2 Trabajo a futuro

Se espera profundizar en la taxonomía propuesta en este documento, especialmente, en las reflexiones en torno a la ética y a la tecnología desde la filosofía en correlación con la ciencia, para alcanzar comprensiones más amplias alrededor de la forma en que se deben de abordar los desarrollos e implementaciones de la IA, y para crear instrumentos de apoyo para los proyectos de investigación y desarrollo, que le sean útiles a las instancias gubernamentales comités de investigación e instituciones de educación, que pretendan deliberar acerca de la Inteligencia Artificial y la ética. De igual manera se podrá trabajar sobre estándares para el sector privado, así como para Instituciones de Educación Superior. En el caso de México, se puede fomentar la construcción del marco legal que llevará el Congreso de la Unión para la ciencia y tecnología.

Bibliografía

Sciutti, A., & Sandini, G. (2017). Interacting with robots to investigate the bases of social interaction. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(12), 2295-2304. https://doi.org/10.1109/TNSRE.2017.2753879

Ahuja, A. S. (2019). The impact of artificial intelligence in medicine on the future role of the physician. *Peer J*, *7*, e7702. https://doi.org/10.7717/peerj.7702

Alba, J. W., & Williams, E. F. (2013). Pleasure principles: A review of research on hedonic consumption. *Journal of consumer psychology*, *23*(1), 2-18. http://dx.doi.org/10.1016/j.jcps.2012.07.003

Alfonseca, M., Cebrian, M., Anta, A. F., Coviello, L., Abeliuk, A., & Rahwan, I. (2021). Superintelligence cannot be contained: Lessons from Computability Theory. *Journal of Artificial Intelligence Research*, *70*, 65-76. https://doi.org/10.1613/jair.1.12202

Al-Imam, A., Motyka, M. A., & Jędrzejko, M. Z. (2020). Conflicting opinions in connection with digital superintelligence. *IAES International Journal of Artificial Intelligence*, *9*(2), 336. http://doi.org/10.11591/ijai.v9.i2.pp336-348

Bradley, P. (2020). Risk management standards and the active management of malicious intent in artificial superintelligence. *AI & SOCIETY*, *35*(2), 319-328. https://doi.org/10.1007/s00146-019-00890-2

Amnistía Internacional (2021). Historia de los Derechos Humanos. No hagas a otro lo que no quieras que te hagan a ti. Recuperado de*: http://www.amnistiacatalunya.org/edu/es/historia/inf-intro.html*

Apel, K. O. (2017). Globalisation and the need for universal ethics. In *Public Reason and Applied Ethics* (pp. 135-151). Routledge. https://doi.org/10.1177/13684310022224732

Apple (2021). Expanded Protections for Children. Recuperado de: *https://www.apple.com/child-safety/*

Aristóteles (1994). *Metafísica*. Traducción Tomas Calvo Martinez. Gredos. España: 1994. 982a, 15.

Asilomar, A. I. (2019). Principles, Future of Life Institute, 2017.

Avens, R. S. (1982). Heidegger and archetypal psychology. *International Philosophical Quarterly*, *22*(2), 183-202. https://doi.org/10.5840/ipq198222218

Bacon, F. (2006). *Nueva atlántida* (Vol. 129). Ediciones AKAL.

Bakhtin, M. M., & Ponzio, A. (1997). *Hacia una filosofía del acto ético: y otros escritos* (Vol. 100). Anthropos Editorial.

Bello, H. (2019). Desafíos para la elaboración de una idea de libertad El aporte de la oboeditio fidei en Dei verbum 5. *Teología y vida*, *60*(4), 525-550. http://dx.doi.org/10.4067/S0049-34492019000400525

Benner, S. A., & Sismour, A. M. (2005). Synthetic biology. *Nature Reviews Genetics*, *6*(7), 533-543. https://doi.org/10.1038/nrg1637

Benthall, S., & Haynes, B. D. (2019, January). Racial categories in machine learning. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 289-298). https://doi.org/10.1145/3287560.3287575

Bentley, P. J., Brundage, M., Häggström, O., & Metzinger, T. (2018). *Should we fear artificial intelligence?: in-depth analysis*. European Parliament.

Bergoglio, M. (2018). Audiencia a los participantes en la Conferencia Internacional promovida por el Pontificio Consejo para la Cultura, 28.04.2018. Recuperado de: https://press.vatican.va/content/salastampa/es/bollettino/pubblico/2018/04/28/conf.html

Board, D. I. (2019). AI Principles: Recommendations on the ethical use of artificial intelligence by the Department of Defense. *Supporting document, Defense Innovation Board*.

Bowden B. (2020) Jacques Derrida: Cosmopolitan Critic. In: Bowden B., Muldoon J., Gould A.M., McMurray A.J. (eds) *The Palgrave Handbook of Management History*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-62114-2_79

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. *Language acquisition and conceptual development*, *3*, 475-511. https://doi.org/10.1017/CBO9780511620669.018

Bringsjord, S. (2008). The logicist manifesto: At long last let logic-based artificial intelligence become a field unto itself. *Journal of Applied Logic*, *6*(4), 502-525. https://doi.org/10.1016/j.jal.2008.09.001

Bringsjord, S., Licato, J., Govindarajulu, N. S., Ghosh, R., & Sen, A. (2015, August). Real robots that pass human tests of self-consciousness. In *2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN)* (pp. 498-504). IEEE. https://doi.org/10.1109/ROMAN.2015.7333698

Brown, N., & Sandholm, T. (2019). Superhuman AI for multiplayer poker. *Science*, *365*(6456), 885-890. https://doi.org/10.1126/science.aay2400

Bustos, C., Rhoads, D., Solé-Ribalta, A., Masip, D., Arenas, A., Lapedriza, A., & Borge-Holthoefer, J. (2021). Explainable, automated urban interventions to improve pedestrian and vehicle safety. *Transportation research part C: emerging technologies, 125*, 103018. https://doi.org/10.1016/j.trc.2021.103018

Cámara de Diputados (2021). La Cámara de Diputados aprobó reformas a la Ley de Ciencia y Tecnología. Recuperado de: *https://comunicacionsocial.diputados.gob.mx/index.php/boletines/la-camara-de-diputados-aprobo-reformas-a-la-ley-de-ciencia-y-tecnologia#gsc.tab=0*

Canalys (2021). Global smartphone market Q1 2021. Recuperado de: *https://www.canalys.com/newsroom/canalys-worldwide-smartphone-market-Q1-2021*

Castelfranchi, C. (2013). Alan Turing's "Computing machinery and intelligence". *Topoi*, *32*(2), 293-299. https://doi.org/10.1007/s11245-013-9182-y

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': the US, EU, and UK approach. *Science and engineering ethics*, *24*(2), 505-528. https://doi.org/10.1007/s11948-017-9901-7

Chaichian, M., Prešnajder, P., & Tureanu, A. (2005). New concept of relativistic invariance in noncommutative space-time: twisted Poincaré symmetry and its implications. *Physical review letters*, *94*(15), 151602. https://doi.org/10.1103/PhysRevLett.94.151602

Chaitin, G. J. (2003). Leibniz, information, math and physics. *arXiv preprint math/0306303*.

Chen, C. H. (2018). Subjectal Scale and Micro-biopolitics at the End of the Anthropocene. *Mosaic: an interdisciplinary critical journal*, *51*(3), 179-198.

CIty, B. E. M. (2018). Hacia una Estrategia de IA en México: Aprovechando la Revolución de la IA.

Corcoran, J. (2009). Aristotle's demonstrative logic. *History and Philosophy of Logic*, *30*(1), 1-20. https://doi.org/10.1080/01445340802228362

Cortina, A., & Conill, J. (2014). La responsabilidad ética de la sociedad civil. *Mediterráneo económico*, *26*, 13-29.

Corvalán, J. G. (2018). Inteligencia artificial: retos, desafíos y oportunidades-Prometea: la primera inteligencia artificial de Latinoamérica al servicio de la Justicia. *Revista de Investigações Constitucionais*, *5*, 295-316. https://doi.org/10.5380/rinc.v5i1.55334

Corver, A., Wilkerson, N., Miller, J., & Gordus, A. G. (2021). Distinct movement patterns generate stages of spider web-building. *bioRxiv*. http://dx.doi.org/10.1016/j.cub.2021.09.030

Cruz-Jesus, F., Castelli, M., Oliveira, T., Mendes, R., Nunes, C., Sa-Velho, M., & Rosa-Louro, A. (2020). Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country. *Heliyon*, *6*(6), e04081. https://doi.org/10.1016/j.heliyon.2020.e04081

Cumming, R. (1955). Descartes' Provisional Morality. *The review of Metaphysics*, 207-235.

Cunneen, M., Mullins, M., Murphy, F., Shannon, D., Furxhi, I., & Ryan, C. (2020). Autonomous vehicles and avoiding the trolley (dilemma): vehicle perception, classification, and the challenges of framing decision ethics. *Cybernetics and Systems*, *51*(1), 59-80. https://doi.org/10.1080/01969722.2019.1660541

Daatland, S. O. (2015). Cuidados de larga duración en Noruega: legados, tendencias y controversias. *FJ Moreno Fuentes, y E. Del Pino Matute (Coords.), Desafíos del Estado de Bienestar en Noruega y España. Madrid: TECNOS*, 31-53.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, *48*(1), 24-42. https://doi.org/10.1007/s11747-019-00696-0

Davis, M. (2018). *The universal computer: The road from Leibniz to Turing*. AK Peters/CRC Press. https://doi.org/10.1201/9781315145839

De Asúa, M. (2018). The "Conflict thesis" and positivist history of science: a view from the periphery: with James C. Ungureanu, "Relocating the Conflict between Science and Religion at the Foundations of the History of Science"; and Miguel de Asúa,"The 'Conflict Thesis' and Positivist History of Science: A View from the Periphery.". *Zygon®*, *53*(4), 1131-1148. https://doi.org/10.1111/zygo.12467

Montreal Declaration Responsible AI (2018). The Montreal Declaration for the Responsible Development of Artificial Intelligence Launched. Recuperado de: *https://www.montrealdeclaration-responsibleai.com/reports-of-montreal-declaration*

Del-Pozo, C. M., Gómez-Mont & Martínez-Pinto, C. (2020). *Agenda Nacional de IA de México. México: IA2030Mx. Agenda nacional mexicana de inteligencia artificial*.

Deleuze, G. (2017). *Postscript on the Societies of Control* (pp. 35-39). Routledge.

Della-Mirandola, G. P. (2018). Discurso sobre la dignidad del hombre. UNAM, Dirección General de Publicaciones y Fomento Editorial.

Denis, G, Hermosilla, M., Aracena, C., Sánchez-Ávalos, R., González-Alarcón, N. & Pombo, C. (2021). *Uso responsable de IA para política pública: manual de formulación de proyectos*. Banco Interamericano de Desarrollo.

Dhariwal, P., Jun, H., Payne, C., Kim, J. W., Radford, A., & Sutskever, I. (2020). Jukebox: A generative model for music. *arXiv preprint arXiv:2005.00341*.

Dictionary Cambridge (2008). Cambridge Academic Content Dictionary. Cambridge University Press.

Digiampietri, L. A., Roman, N. T., Meira, L. A., Filho, J. J., Ferreira, C. D., Kondo, A. A., ... & Goldenstein, S. (2008). Uses of artificial intelligence in the Brazilian customs fraud detection system. In *Proceedings of the 2008 international conference on digital government research* (pp. 181-187).

Dinneen, N. (2014). Hans Jonas's Noble 'Heuristics of Fear': Neither the Good Lie Nor the Terrible Truth. *Cosmos and History: The Journal of Natural and Social Philosophy*, *10*(2), 1-21.

Dirección General de Datos Abiertos (2018). *Consulta pública referente a los principios y guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la administración pública federal*, México, Unidad de Gobierno Digital de la Secretaría de la Función Pública, Secretaría de la Función Pública.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Dussel, E. (1994). La razón del otro. La interpelación como acto-de-habla. *E. Dussel [Compilador]: Debate en torno de la ética del discurso de Apel. Diálogo filosófico Norte-Sur desde América latina. Iztapalapa: Siglo Veintiuno*, 55-89.

Edelman (2019). Edelman Research Highlights Perception About Artificial Intelligence Future. Artificial Intelligence Survey.

Edwards, P. (1975). Heidegger and Death asPossibility'. *Mind*, *84*(336), 548-566.

Eissa, A. E., & Zaki, M. M. (2011). The impact of global climatic changes on the aquatic environment. *Procedia Environmental Sciences*, *4*, 251-259. https://doi.org/10.1016/j.proenv.2011.03.030

Ekman, M., Kok, P., & de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*, *8*(1), 1-9. https://doi.org/10.1038/ncomms15276

Epstein, G., Bennett, A., Gruby, R., Acton, L., & Nenadovic, M. (2014). Studying power with the social-ecological system framework. In *Understanding society and natural resources* (pp. 111-135). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-8959-2_6

Unión Europea (2017). Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103 (INL)). Recuperado de: *https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_ES.pdf*

Unión Europea (2018). EU Declaration on Cooperation on Artificial Intelligence. Recuperado de: *https://ec.europa.eu/jrc/communities/en/node/1286/document/eu-declaration-cooperation-artificial-intelligence*

Unión Europea (2021a). Directiva 2006/24/CE del Parlamento Europeo y del Consejo, de 15 de marzo de 2006, sobre la conservación de datos generados o tratados en relación con la prestación de servicios de comunicaciones electrónicas de acceso público o de

redes públicas de comunicaciones y por la que se modifica la Directiva 2002/58/CE. Recuperado de: *https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32006L0024&from=ES*

Unión Europea (2021b). Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales ya la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) (Texto pertinente a efectos del EEE). Recuperado de*: https://www.boe.es/doue/2016/119/L00001-00088.pdf*

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, (2020-1).

Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, *7*(1), 1-9. https://doi.org/10.1057/s41599-020-0494-4

Floridi, L. (2021). The European Legislation on AI: a Brief Analysis of its Philosophical Approach. *Philosophy & Technology*, 1-8. https://doi.org/10.1007/s13347-021-00460-9

Fortune Business Insights™ (2020). *"Artificial Intelligence (AI) in Retail Market Analysis-2026"* (ID: FBI101968). Recuperado de: *https://www.fortunebusinessinsights.com/artificial-intelligence-ai-in-retail-market-101968*

Foucault, M. (2000). *Vigilar y castigar: nacimiento de la prisión*. Siglo XXI.

Garvie, C., Bedoya, A. M., & Frankle, J. (2019). *The perpetual line-up. Unregulated police face recognition in America*. Georgetown Law Center on Privacy & Technology. Recuperado de: *https://www.perpetuallineup.org/*

Gasser-Wingate, M. (2020). Conviction, Priority, and Rationalism in Aristotle's Epistemology. *Journal of the History of Philosophy*, *58*(1), 1-27. https://doi.org/10.1353/hph.2020.0001

Giansiracusa, N. (2021). *How Algorithms Create and Prevent Fake News*. https://doi.org/10.1007/978-1-4842-7155-1

Gowans, C. (2004). Moral relativism. Stanford Encyclopedia of Philosophy. Recuperado de: *https://seop.illc.uva.nl/entries/moral-relativism/*

Gutland, C. (2018). Husserlian phenomenology as a kind of introspection. *Frontiers in psychology*, 9, 896. https://doi.org/10.3389/fpsyg.2018.00896

Hagerty, A., & Rubinov, I. (2019). Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence. *arXiv preprint arXiv:1907.07892*.

Hajian, S., Domingo-Ferrer, J., & Farràs, O. (2014). Generalization-based privacy preservation and discrimination prevention in data publishing and mining. *Data Mining and Knowledge Discovery*, *28*(5), 1158-1188. https://doi.org/10.1007/s10618-014-0346-1

Han, B. C. (2017). *Psychopolitics: Neoliberalism and new technologies of power*. Verso Books.

Harari, Y. N. (2014). *Sapiens. De animales a dioses: Una breve historia de la humanidad*. Debate.

Harari, Y. N. (2016). *Homo Deus: breve historia del mañana*. Debate.

Harari, Y. N. (2018). *21 lecciones para el siglo XXI*. Debate.

Harner, M. (1977). The ecological basis for Aztec sacrifice 1. *American ethnologist*, *4*(1), 117-135. https://doi.org/10.1525/ae.1977.4.1.02a00070

Hatfield, G. (2008). René Descartes. Stanford Encyclopedia of Philosophy. Recuperado de: *https://plato.stanford.edu/entries/descartes/?utm_source=orilliamatters.com&utm_campaign=orilliamatters.com&utm_medium=referral*

Haydari, A., & Yilmaz, Y. (2020). Deep reinforcement learning for intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2020.3008612

Hayward, T. (1997). Anthropocentrism: a misunderstood problem. Environmental Values, 6(1), 49-63. https://doi.org/10.3197/096327197776679185

Hegel, G. W. F. (2009). *Fenomenología del espíritu*, Edición y traducción: Manuel Jiménez Redondo, España, Pre-textos, 2009, pp. 281.

Heidegger, M. (1993). *El ser y el tiempo*, trad. de José Gaos, Segunda edición, México, Fondo de Cultura Económica, 1993, pp. 82.

Heit, H. (2018). "*there are no facts...*": Nietzsche as Predecessor of Post-Truth?. Studia Philosophica Estonica, 44-63.

Henle, M., & Michael, M. (1956). The influence of attitudes on syllogistic reasoning. *The Journal of Social Psychology*, *44*(1), 115-127. https://doi.org/10.1080/00224545.1956.9921907

Hepburn, A. (1999). Derrida and psychology: Deconstruction and its ab/uses in critical and discursive psychologies. *Theory & Psychology*, *9*(5), 639-665. https://doi.org/10.1177%2F0959354399095004

Hertogh, C. P. (2016). Thought Experiment Analyses of René Descartes' Cogito1. *Trans/Form/Ação, 39*, 9-22. https://doi.org/10.1590/S0101-31732016000300002

Hill, R. K. (2016). What an algorithm is. *Philosophy & Technology*, *29*(1), 35-59.

Hobson, J. A., & Pace-Schott, E. F. (2002). The cognitive neuroscience of sleep: neuronal systems, consciousness and learning. *Nature Reviews Neuroscience*, *3*(9), 679-693. https://doi.org/10.1038/nrn915

Hoffman, M., Yoeli, E., & Nowak, M. A. (2015). Cooperate without looking: Why we care what people think and not just what they do. *Proceedings of the National Academy of Sciences*, *112*(6), 1727-1732. https://doi.org/10.1073/pnas.1417904112

Holzgrefe, J. L., & Keohane, R. O. (2003). *Ethical, legal, and political dilemmas*. Cambridge: Cambridge University Press.

Human Rights Watch (2020). As Killer Robots Loom, Demands Grow to Keep Humans in Control of Use of Force. Recuperado de: *https://www.hrw.org/world-report/2020/country-chapters/global-0#*

Huntington, S. P. (1998). El choque de las civilizaciones y la reconfiguración del orden mundial. *Cuadernos de estrategia*, (99), 239-248.

IEEE (2017). Ethically Aligned Design, Estados Unidos, *Institute of Electrical and Electronics Engineers*.

IEEE (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (A/IS), Segunda Edición, Estados Unidos, *Institute of Electrical and Electronics Engineers*.

Instruments, O. L. (2019). Recommendation of the Council on Artificial Intelligence. *Organization for Economic Cooperation and Development*.

Iosa, J. (2017), "Libertad negativa, autonomía personal y constitución", *Revista Chilena de Derecho*, Vol. 44, Número 2, 495 - 518, Pontificia Universidad Católica de Chile, Chile, agosto de 2017, pp. 499-506.

Jaques, A. E. (2019). Why the moral machine is a monster. *University of Miami School of Law*, *10*.

Jiang, L., Hwang, J. D., Bhagavatula, C., Bras, R. L., Forbes, M., Borchardt, J., Liang, J., Etzioni, O., Sap, M., & Choi Y. (2021). Delphi: Towards machine ethics and norms. *arXiv preprint arXiv:2110.07574*.

John, R. A., Tiwari, N., Patdillah, M. I. B., Kulkarni, M. R., Tiwari, N., Basu, J. Bose, S. K., Ankit, Yu, C. J., Nirmal, A., Vishwanath, S. K., Bartolozzi, C., Basu, A., & Mathews, N. (2020). Self healable neuromorphic memtransistor elements for decentralized sensory signal processing in robotics. *Nature communications*, *11*(1), 1-12. https://doi.org/10.1038/s41467-020-17870-6

Jonas, H. (2019). The heuristics of fear. In *Ethics in an age of pervasive technology* (pp. 213-221). Routledge.

Kaiser, B. (2019). *La dictadura de los datos: La verdadera historia desde dentro de Cambridge Analytica y cómo el Big Data, Trump y Facebook corrompieron la democracia, y cómo puede volver a pasar*. HarperCollins Mexico.

Kant, I. (1784). *¿Qué es la Ilustración?* Filosofía de la historia, 25-38.

Kant, I. (2012). *Fundamentación para una metafísica de las costumbres*. Versión castellana y estudio preliminar por Roberto R. Aramayo. Alianza Editorial. El libro de bolsillo, Madrid. Segunda Edición, 2012, A 52, pp. 126.

Karimi-Haghighi, M., & Castillo, C. (2021). Enhancing a recidivism prediction tool with machine learning: effectiveness and algorithmic fairness. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law* (pp. 210-214). https://doi.org/10.1145/3462757.3466150

Käufer, S. (2005). The Nothing and the Ontological Difference in Heidegger's What is Metaphysics?. *Inquiry*, *48*(6), 482-506. https://doi.org/10.1080/00201740500320047

Kelly, M. G. (2013). Foucault, subjectivity, and technologies of the self. *A companion to Foucault*, 510-525. https://doi.org/10.1002/9781118324905.CH26

Keulartz, J., Korthals, M., Schermer, M., & Swierstra, T. (2002). Oragmatist ethics fora technological culture. *SpringerScience+ BusinessMedia Dordrecht. OriginallypublishedbyKluwerAcademicPublishers*. https://doi.org/10.1007/978-94-010-0301-8

Kile, F. (2013). Artificial intelligence and society: a furtive transformation. *AI & society*, *28*(1), 107-115. https://doi.org/10.1007/s00146-012-0396-0

Knuuttila, S. (2001). Time and creation in Augustine. *The Cambridge Companion to Augustine*, 103-115.

Kögler, H. H. (2005). Recognition and difference: The power of perspectives in interpretive dialogue. *Social Identities*, *11*(3), 247-269. https://doi.org/10.1080/13504630500257082

Krügel, S., & Uhl, M. (2022). Autonomous vehicles and moral judgments under risk. *Transportation Research Part A: Policy and Practice*, *155*, 1-10. https://doi.org/10.1016/j.tra.2021.10.016

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, *40*. https://doi.org/10.1017/S0140525X16001837

Lemmens, P. (2017). "Social Autonomy and Heteronomy in the Age of ICT: The Digital Pharmakon and the (Dis)Empowerment of the General Intellect", *Foundations of Science*, Vol. 22, Issue 2, 287-296. https://doi.org/10.1007/s10699-015-9468-1

Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, *50*, 57-70. https://doi.org/10.1016/j.ijinfomgt.2019.04.003

Licato, J., Bringsjord, S., & Rensselaer, A. I. (2016). PAGI World: A Physically Realistic, General-Purpose Simulation Environment for Developmental AI Systems.

Little, A. C., Jones, B. C., Penton-Voak, I. S., Burt, D. M., & Perrett, D. I. (2002). Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1496), 1095-1100. https://doi.org/10.1098/rspb.2002.1984

Lynch, J. (2020). Face off: Law enforcement use of face recognition technology. *Available at SSRN 3909038*.

Mahesh, B. (2020). Machine Learning Algorithms-A Review. *International Journal of Science and Research (IJSR). [Internet]*, *9*, 381-386. https://doi.org/10.21275/ART20203995

Maksutov, A. A., Morozov, V. O., Lavrenov, A. A., & Smirnov, A. S. (2020). Methods of deepfake detection based on machine learning. In *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)* (pp. 408-411). IEEE. https://doi.org/10.1109/EIConRus49466.2020.9039057

Manheim, K. M., & Kaplan, L. (2018). Artificial intelligence: risks to privacy and democracy. *Yale Journal of Law and Technology*, 21, 106.

McCarthy, J. (1987). Generality in artificial intelligence. *Communications of the ACM*, *30*(12), 1030-1035. https://doi.org/10.1145/33447.33448

McConwell, A. K., & Currie, A. (2017). Gouldian arguments and the sources of contingency. *Biology & Philosophy*, *32*(2), 243-261. https://doi.org/10.1007/s10539-016-9556-9

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., Ledsam, J. R., Melnick, D., Mostofi, H., Peng, L., Reicher, J. J., Romera-Paredes, B., Sidebottom, R., Suleyman, M., Tse, D., Young, K. C., De Fauw, J., & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89-94. https://doi.org/10.1038/s41586-019-1799-6

McLuhan, M., Gordon, W. T., Lamberti, E., & Scheffel-Dunand, D. (2011). *The Gutenberg galaxy: The making of typographic man*. University of Toronto Press.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1-35. https://doi.org/10.1145/3457607

Melhado, F., & Rabot, J. M. (2021). Sentiment Analysis: From Psychometrics to Psychopolitics. *Comunicação e Sociedade*, *39*, 101-118. https://doi.org/10.17231/comsoc.39(2021).2797

Microsoft Technology Licensing (2020). *Creating a conversational chat bot of a specific person. US010853717*: Estados Unidos.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com. 9780244768522

Moratalla, N. L., & Sueiro, E. (2010). Cerebro ético, atajo emocional ante dilemas. *Diario de Noticias Navarra*, *18*.

Nancy, J. L., & Labarthe, P. L. (2002). El mito nazi, Barcelona. *Anthropos*, 28.

Nancy, J. L., & Piazza, V. (2006). *El intruso*. Amorrortu.

National Governance Committee for the New Generation Artificial (2021). AI Ethics and Governance at Institute of Automation, Chinese Academy of Sciences. Recuperado de: *http://www.most.gov.cn/kjbgz/202109/t20210926_177063.htm*

Neves, J. (2019). Nietzsche for physicists. *Philosophia Scientiæ. Travaux d'histoire et de philosophie des sciences*, (23-1), 185-201. https://doi.org/10.4000/philosophiascientiae.1855

Kayser-Bril, N. (2020). Google apologizes after its Vision AI produced racist results. *AlgorithmWatch. Retrieved August*, *17*, 2020. Recuperado de: *https://algorithmwatch.org/en/story/google-vision-racism/*

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., Broelemann, K., Kasneci, G., Tiropanis, T., & Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1356. https://doi.org/10.1002/widm.1356

Odgers, C. (2018). Smartphones are bad for some teens, not all (vol 554, pg 432, 2018). *Nature*, *555*(7698), 580-580. https://doi.org/10.1038/d41586-018-02109-8

Olsthoorn, J. (2020). Leviathan Inc.: Hobbes on the nature and person of the state. *History of European Ideas*, *47*(1), 17-32. https://doi.org/10.1080/01916599.2020.1779466

O'neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Oppenheimer, A. (2018). *¡Sálvese quien pueda!: El futuro del trabajo en la era de la automatización*. Debate.

Ordine, N. (2017). *La utilidad de lo inútil: manifiesto* (Vol. 36). Acantilado.

Padial, J. J. (2017). La herencia aristotélica en la teoría hegeliana de la sensación como «encontrar-se» vital del espíritu. *Contrastes. Revista Internacional de Filosofía*, *22*(3). https://doi.org/10.24310/Contrastescontrastes.v22i3.3757

Pantanowitz, L., Quiroga-Garza, G. M., Bien, L., Heled, R., Laifenfeld, D., Linhart, C., Sandbank, J., Shach, A. A., Shalev, V., Vecsler, M., Michelow, P., Hazelhurst, S., & Dhir R. (2020). An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and

deployment study. *The Lancet Digital Health*, *2*(8), e407-e416. https://doi.org/10.1016/S2589-7500(20)30159-X

Park, S. H., & Han, K. (2018). Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*, *286*(3), 800-809. https://doi.org/10.1148/radiol.2017171920

Peces-Barba, G. (1994). El principio de las mayorías desde la Filosofía del Derecho. *Anuario de la Facultad de Derecho de Alcalá de Henares*, 1993-1994, vol. 3, p. 41-[56]. ISSN 1134-9492.

Pinilla, K. F., & Cordero, J. M. C. (2017). Thomas Hobbes: The modern economist. *Praxis Filosófica*, (44), 221-250.

Pinnaparaju, N., Indurthi, V., & Varma, V. (2020, September). Identifying Fake News Spreaders in Social Media. In *CLEF (Working Notes)*.

Pontificia Accademia per la Vita, IBM e Microsoft (2020). Rome Call for AI Ethics. Recuperado de: *https://www.romecall.org/*

Potter, V. R. (1971). Bioethics: bridge to the future.

Radin, M. (1920). The lex Pompeia and the poena cullei. *The Journal of Roman Studies*, *10*, 119-130. https://doi.org/10.2307/295798

Randhawa, G. S., Soltysiak, M. P., El Roz, H., de Souza, C. P., Hill, K. A., & Kari, L. (2020). Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *Plos one*, *15*(4), e0232391. https://doi.org/10.1371/journal.pone.0232391

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Palavecino, C. R. (2017). Autonomía y democracia en Cornelius Castoriadis. *Mutatis Mutandis: Revista Internacional de Filosofía*, *1*(9), 65-88.

Rodríguez-Reséndiz, H. (2020a). *Consideraciones Éticas en la Inteligencia Artificial. Mes de la bioética*. Investigaciones a quince años de la Declaración Universal Bioética y de los Derechos Humanos. ISBN 978-607-513-527-4; 978-607-513-536-6.

Rodríguez-Reséndiz, H. (2020b). *Ética de la Inteligencia Artificial en el contexto global*. Mes de la bioética. Investigaciones a quince años de la Declaración Universal Bioética y de los Derechos Humanos. ISBN 978-607-513-527-4; 978-607-513-536-6.

Rollin, B. E. (2007). Animal mind: science, philosophy, and ethics. *The Journal of Ethics*, *11*(3), 253-274. https://doi.org/10.1007/s10892-007-9018-3

Roozenbeek, J., & Van Der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1), 1-10. https://doi.org/10.1057/s41599-019-0279-9

Russel, S., & Norvig, P. (2013). Artificial intelligence: a modern approach. Pearson Education Limited.

Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. *Ai Magazine*, *36*(4), 105-114. https://doi.org/10.1609/aimag.v36i4.2577

Samek, W., Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*.

Sánchez-Ávalos, R., González, F. & Ortiz, T. (2021). Uso responsable de la IA para las políticas públicas: manual de ciencia de datos. Banco Interamericano de Desarrollo.

Sarsanedas, A. (2015). *La filosofía de la tecnología*, España, Editorial UOC, 2015, pp. 37.

Savulescu, J., & Persson, I. (2012). Moral enhancement, freedom and the god machine. *The Monist*, *95*(3), 399.

Schubbach, A. (2019). Judging machines: philosophical aspects of deep learning. *Synthese*, 1-21. https://doi.org/10.1007/s11229-019-02167-z

Scott, D. (2006). The "concept of time" and the "being of the clock": Bergson, Einstein, Heidegger, and the interrogation of the temporality of modernism. *Continental Philosophy Review*, *39*(2), 183-213. https://doi.org/10.1007/s11007-006-9023-4

Searle, J. (2009). La conciencia. *J. González. Filosofía y ciencias de la vida*, 60

Shen, S., Jiang, H., & Zhang, T. (2012). Stock market forecasting using machine learning algorithms. *Department of Electrical Engineering, Stanford University, Stanford, CA*, 1-5.

Spiekermann, S. (2017). IEEE P7000—The first global standard process for addressing ethical concerns in system design. *Multidisciplinary Digital Publishing Institute Proceedings*, *1*(3), 159.

Stern, D. (2007). Wittgenstein, the Vienna Circle, and Physicalism: A Reassessment. The Cambridge Companion to Logical Empiricism, 305–331. https://doi.org/10.1017/CCOL0521791782.013

Sylvia IV, J. J. (2020). The Biopolitics of Social Distancing. *Social Media+ Society*, *6*(3). https://doi.org/10.1177/2056305120947661

The Conversation (2021). We invited an AI to debate its own ethics in the Oxford Union – what it said was startling. Recuperado de: *https://theconversation.com/we-invited-an-ai-to-debate-its-own-ethics-in-the-oxford-union-what-it-said-was-startling-173607?fbclid=IwAR3Tp5wwBaRV0mPWEj7EfA7JosOqbv2bFfp1mPquRFvOQK9aI01LLWow3o0*

Tibaldeo, R. F. (2015). The Heuristics of Fear: Can the Ambivalence of Fear Teach Us Anything in the Technological Age?. *Ethics In Progress*, 6(1), 225-238. https://doi.org/10.14746/eip.2015.1.9

Tolosana, R., Gomez-Barrero, M., Busch, C., & Ortega-Garcia, J. (2019). Biometric presentation attack detection: Beyond the visible spectrum. *IEEE Transactions on Information Forensics and Security*, *15*, 1261-1275. https://doi.org/10.1109/TIFS.2019.2934867

Trego, K. (2005). From the Ethics of Wisdom to the Ethics of Freedom. *Revue des sciences philosophiques et theologiques*, *89*(4), 641-653.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *J. of Math*, *58*(345-363), 5. https://doi.org/10.3917/rspt.894.0641

Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the turing test* (pp. 23-65). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-6710-5_3

Turner, R. (2013). The philosophy of computer science. In Zalta, E.N. (Ed.) *The Stanford encyclopedia of philosophy*. Fall 2013 edition.

UNESCO (2021). Conferencia General, 41st, 2021 [795]. Informe de la Comisión de Ciencias Sociales y Humanas (SHS). Proyecto de recomendación sobre la ética de la inteligencia artificial. Recuperado de: *https://unesdoc.unesco.org/ark:/48223/pf0000379920_spa*

UNICEF (2000). Protocolo facultativo de la Convención sobre los Derechos del Niño relativo a la venta de niños, la prostitución infantil y la utilización de niños en la pornografía. Recuperado de: *https://www.ohchr.org/sp/professionalinterest/pages/opsccrc.aspx*

UNICEF (2006). Convención sobre los Derechos del Niño. Recuperado de: *https://www.unicef.org/es/convencion-derechos-nino/texto-convencion*

Union, U. G. (2017). Top 10 principles for ethical artificial intelligence. *The future world of work*.

Vakkuri, V., Kemell, K. K., Kultanen, J., Siponen, M., & Abrahamsson, P. (2019). Ethically aligned design of autonomous systems: Industry viewpoint and an empirical study. *arXiv preprint arXiv:1906.07946*.

Webb, A. (2019). *The big nine: How the tech titans and their thinking machines could warp humanity*. Hachette UK.

Weltethos, S. (1993). Hacia una ética mundial: Una declaración inicial. *Declaración del II Parlamento de las Religiones del Mundo*, Chicago 1993.

Wiriyathammabhum, P., Summers-Stay, D., Fermüller, C., & Aloimonos, Y. (2016). Computer vision and natural language processing: recent approaches in multimedia

and robotics. *ACM Computing Surveys (CSUR)*, *49*(4), 1-44. https://doi.org/10.1145/3009906

World Health Organization (2021). Ethics and governance e of artificial intelligence for health: WHO guidance. Recuperado de: *https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf*

Wright, N. (2018). How artificial intelligence will reshape the global order. *Foreign Affairs*, *10*.

Yang, K. C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., & Menczer, F. (2019). Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, *1*(1), 48-61. https://doi.org/10.1002/hbe2.115

Yuste, R., Genser, J., & Herrmann, S. (2021). It's Time for Neuro-Rights. *Horizons: Journal of International Relations and Sustainable Development*, (18), 154-165.

Yuzbekova, I. & Tairov, R. (2021). “Nada cambiará si no existen”: el responsable de la startup Perm explicó los despidos masivos de empleados. Recuperado de: *https://www.forbes.ru/newsroom/biznes/436639-nichego-ne-izmenitsya-esli-ih-ne-budet-glava-permskogo-startapa-obyasnil*

Zoph, B., Ghiasi, G., Lin, T. Y., Cui, Y., Liu, H., Cubuk, E. D., & Le, Q. V. (2020). Rethinking pre-training and self-training. *arXiv preprint arXiv:2006.06882*.

Acciones derivadas de la investigación

i. Participación para la construcción de un Marco ético de la Inteligencia Artificial

2020. Participación en Foro: *Un acercamiento a la Inteligencia Artificial*. Secretaría de Relaciones Exteriores y Comisión de Ciencia y Tecnología del Senado de la República, Ciudad de México.

2020. Contribución en la *Agenda Nacional de IA de México. México: IA2030Mx. Agenda nacional mexicana de inteligencia artificial*.

2020. Participación: *Consultation: Ethics of Artificial Intelligence (the first global standard-setting instrument on the ethics of artificial intelligence in the form of a Recommendation*) de la UNESCO.

ii. Ponencias

2020. *Consideraciones Éticas en la Inteligencia Artificial.* Mes de la bioética. Investigaciones a quince años de la Declaración Universal Bioética y de los Derechos Humanos. Congreso Internacional de Bioética. México.

2020. *Ética de la Inteligencia Artificial en el contexto global*. Mes de la bioética. Investigaciones a quince años de la Declaración Universal Bioética y de los Derechos Humanos. Congreso Internacional de Bioética. México.

2020. *Ética en la Inteligencia Artificial*. Seminario interdisciplinario. Centro de Investigación en Ciencias Aplicadas y Tecnología Avanzada del Instituto Politécnico Nacional.

2020. Presentación de Poster: *La intersección filosófica en la ética de la Inteligencia Artificial*. Foro de Investigación y Posgrado de la Facultad de Derecho 2020, Universidad Autónoma de Querétaro.

2021. *Ética en el emprendimiento, una mirada desde la Inteligencia Artificial*. III Congreso Internacional Emprendimiento UAQ. México.

2021. *Perspectiva de la bio / ética en la Inteligencia Artificial*. 7o Congreso internacional de Bioetica. Mesa Bioética y tecnología. Colombia.

2021. *Irrupción de la Inteligencia Artificial y enajenación de la autonomía humana en la toma de decesiones*. 1° Coloquio internacional de Investigación y Posgrado en ciencias sociales y humanidades de la Facultad de Derecho 2021, Universidad Autónoma de Querétaro.

2021. *Ética en la Inteligencia Artificial. ¿Los algoritmos nos controlan?* International program Learn to Lead by Ericsson. Ericsson México (Telefonaktiebolaget LM Ericsson).

iii. Publicaciones

Rodríguez-Reséndiz, H. (2020). *Consideraciones Éticas en la Inteligencia Artificial. Mes de la bioética*. Investigaciones a quince años de la Declaración Universal Bioética y de los Derechos Humanos. ISBN 978-607-513-527-4; 978-607-513-536-6.

Rodríguez-Reséndiz, H. (2020). *Ética de la Inteligencia Artificial en el contexto global*. Mes de la bioética. Investigaciones a quince años de la Declaración Universal Bioética y de los Derechos Humanos. ISBN 978-607-513-527-4; 978-607-513-536-6.

Zamora-Antuñano, M. A., Rodríguez-Reséndiz, J., Rodriguez Segura, L., Cruz Pérez, M. Á., Altamirano Corro, J. A., Paredes-Garcia, W. J., & Rodríguez-Reséndiz, H. (2021). Analysis of Emergency Remote Education in COVID-19 Crisis Focused on the Perception of the Teachers. *Sustainability*, *13*(7), 3820.

Villegas-Mier, C. G., Rodriguez-Resendiz, J., Álvarez-Alvarado, J. M., Rodriguez-Resendiz, H., Herrera-Navarro, A. M., & Rodríguez-Abreo, O. (2021). Artificial neural networks in MPPT algorithms for optimization of photovoltaic power systems: A review. *Micromachines*, *12*(10), 1260.

García Guerrero, J., Rodríguez Reséndiz, J., Rodríguez Reséndiz, H., Álvarez-Alvarado, J. M., & Rodríguez Abreo, O. (2021). Sustainable Glass Recycling Culture-Based on Semi-Automatic Glass Bottle Cutter Prototype. *Sustainability*, *13*(11), 6405.

Zamora-Antuñano, M. A., Rodríguez-Reséndiz, J., Cruz-Pérez, M. A., Rodríguez Reséndíz, H., Paredes-García, W. J., & Díaz, J. A. G. (2022). Teachers' Perception in Selecting Virtual Learning Platforms: A Case of Mexican Higher Education during the COVID-19 Crisis. *Sustainability*, *14*(1), 195.

Rodríguez-Reséndiz, H. & Zepeda, H. R. (2021). *Perspectiva de la bio / ética en la Inteligencia Artificial*. Memorias del 7o Congreso internacional de Bioetica. Mesa Bioética y tecnología. (Pendiente)

Rodríguez-Reséndiz, H. (2021). *Estructura Institucional de la diplomacia científica en México*. UNESCO. (Pendiente)

Pérez-Carvajal, M., Carretero-Farfán, M. A., Zamora-Antuñano, M. A., Domínguez-Olano, D. C. & Rodríguez-Reséndiz, H. (2022). Análisis del aprendizaje con software matemático en una asignatura de ingeniería mediante el uso de TIC´s. *Revista CPU-e*. (Pendiente)

iv. Becas obtenidas

2020. *Curso de publicaciones científicas*. Universidad Autónoma de Querétaro.

2021. *Seminario "Derecho penal y comportamiento humano: Avances desde la neurociencia y la inteligencia artificial"*. Universidad de Castilla-La Mancha, Facultad de C.C.J.S de Toledo, España.

2021. *Curso de Postgrado Regional "Diplomacia Científica aplicada a las Neurociencias"*. UNESCO, Ministerio de Educación y Cultura de Uruguay, Instituto de Investigaciones Biológicas Clemente Estable, Programa de Desarrollo de las Ciencias Básicas y SciDip GLOBAL.

2021. *Programa de Formación multidisciplinario de Inteligencia Artificial. Impulsando el ecosistema de Inteligencia Artificial en América Latina*. Universidad de Buenos Aires, Facultad de Derecho, Argentina.

2021. *Algoritmos Éticos, Responsables y Transparentes*. Iniciativa fAIr LAC del Banco Interamericano de Desarrollo y Universidad Adolfo Ibáñez, Chile.

2021. *Seminario intensivo de Ética de la Investigación: "Ética y Covid-19: investigación, distribución de recursos, asistencia y más"*. Facultad Latinoamericana de Ciencias Sociales, Argentina.

2021. *Biopoder y transhumanismo. Cátedra Michel Foucault: Lenguajes de poder*. Centro de Estudios Críticos en Cultura Contemporánea de la Universidad Autónoma de Querétaro.

v. Certificaciones obtenidas

2021. *Diplomacia científica aplicada a las neurociencias*. UNESCO.

2021. *IDB44x: ¿Cómo hacer uso responsable de la inteligencia artificial en el sector público?* Banco Interamericano de Desarrollo.

2021. *Inteligencia Artificial en América Latina y el Caribe: para que nadie quede atrás.* UNESCO.

2021. *AI Business Sessions, Programa de formación sobre ética de datos del Proyecto Algoritmos Éticos, Responsables y Transparentes*. Universidad Adolfo Ibáñez, Chile.

2021. *Salud e Inteligencia Artificial*. Programa de Formación Multidisciplinaria en Datos, Programación e Inteligencia Artificial de la Universidad de Buenos Aires.

2021. *Gobernanza de datos*. Programa de Formación Multidisciplinaria en Datos, Programación e Inteligencia Artificial de la Universidad de Buenos Aires.

2021. *Inteligencia Artificial y Educación*. Programa de Formación Multidisciplinaria en Datos, Programación e Inteligencia Artificial de la Universidad de Buenos Aires.

2021. *Tecnoética y mundo digital*. Universidad Carlos III de Madrid.

vi. Certamen:

2021. Finalista en *9no Encuentro de Jóvenes Investigadores del Estado de Querétaro, "Ideas jóvenes para impulsar el crecimiento del país"*. Universidad Autónoma de Querétaro, Consejo de Ciencia y Tecnología del Estado de Querétaro. (Pendiente evaluación final)