



# Derechos humanos, ética y transparencia algorítmica

HUMAN RIGHTS, ETHICS AND ALGORITHMIC TRANSPARENCY

**Evelyn Téllez Carvajal**

Investigadora del Centro de Investigación e Innovación en Tecnologías de la Información y Comunicación (INFOTEC), México.

Profesora de la Facultad de Ciencias Políticas y Sociales de la UNAM.

[evelyn.tellez@infotec.mx](mailto:evelyn.tellez@infotec.mx)  0000-0001-6136-6821

Recibido: 02 de mayo de 2021 | Aceptado: 16 de junio de 2021.

## RESUMEN

Hoy en día el uso de algoritmos que resuelven o predicen problemáticas en el entorno digital hacen uso de información de los individuos quienes con sus datos nutren las bases que a su vez son el insumo de los análisis de los grandes cúmulos de datos. Sin embargo, es necesario reconsiderar si la información que se recolecta es adecuadamente recogida y protegida.

¿Qué tanto conoce un diseñador de algoritmos sobre la protección de datos personales y del adecuado manejo de los mismos?, ¿cuáles son los límites en el entrenamiento de los algoritmos?, ¿cuáles deberían ser los principios que rigen el diseño de los algoritmos base de la Inteligencia Artificial? La presente contribución busca entender lo que es la ética y la transparencia algorítmica así como la importancia de comprender el manejo de la información que atañe a los individuos cuando se trata del tema de Inteligencia Artificial.

## ABSTRACT

Nowadays, the use of algorithms that solve or predict problems in the digital environment make use of information from individuals who, with their personal data, nourish the data bases that are the input for the data mining analysis. However, it is necessary to reconsider whether the personal information is properly collected and protected.

What does an algorithm designer know about the protection of personal data and its proper handling? Are there any limits in algorithms trading? What are the principles that must rule the Artificial Intelligence? This contribution seeks to understand what is ethics and algorithmic transparency as well as the importance of understanding the handling of information that concerns individuals when it is related to Artificial Intelligence.

## PALABRAS CLAVE

Algoritmo  
Ética algorítmica  
Inteligencia Artificial Fiable  
Derechos Humanos  
Principios de Inteligencia Artificial Fiable

## KEYWORDS

Algorithmic Ethics  
Reliable Artificial Intelligence  
Human Rights  
Principles of Reliable Artificial Intelligence

## I. INTRODUCCIÓN

Con los avances científicos y tecnológicos hemos sido testigos de la transformación de los medios de comunicación, el cambio de los dispositivos analógicos a los digitales, hasta llegar a los aparatos autónomos como los aviones y autos, los drones y los robots, entre otras cosas.

Internet ha sido una herramienta que sin lugar a dudas también ha tenido un impacto en esta transformación. Desde su creación en 1969 tuvo una gran relevancia al permitir que la información no fuera almacenada en un solo lugar, nos referimos a la época de ARPANET en la que el acceso a la red solo estaba a la disposición de ciertas universidades y comunidades de investigación. Con el paso del tiempo en la década de los años noventa comenzó el acceso público a la red con evidentemente muy pocos usuarios alrededor del mundo, pero de esa fecha al año 2018 más del 50% de la población a nivel global ya posee el acceso a internet. (Banco Mundial, 2019).

Gracias al uso público de internet, los usuarios han alimentado con su información bases que hoy constituyen un gran volumen de datos que se encuentran almacenados en distintos lugares, desde bases de datos en posesión de agentes gubernamentales hasta aquellas bases que están en manos de empresas como Facebook, Google, Amazon, Uber, entre muchas otras; pero que también podrían estar siendo recolectados por individuos de manera independiente sin que podamos conocer quién lo está realizando y con qué finalidad lo hace.

Es una realidad que en los próximos años los análisis de grandes cúmulos de datos serán fundamentales tanto en la industria como en el sector gubernamental; determinando un incremento en las actividades que se realizarán de manera automatizada, por lo que se requerían de más especialistas en este campo del conocimiento. Esta condición llama la atención a ocuparse de los principios éticos que se están reforzando para contar con una ética algorítmica que no violente los derechos humanos de los individuos en estas sociedades hiperconectadas.

Para poder hacer esto realidad es pertinente entender con claridad cómo funciona la minería de datos, los análisis de Big Data, la Inteligencia Artificial, el aprendizaje de las máquinas, pero sobre todo entender la relevancia de la ética en el manejo transparente de los algoritmos que involucran a los seres humanos. A este respecto:

“La minería de datos es una rama relativamente joven dentro de las ciencias de la computación, en la cual a través del uso de campos como inteligencia artificial, aprendizaje automático, estadística y bases de datos es posible descubrir patrones e inferir información nueva a partir de grandes colecciones de datos. Hoy día se aplica en áreas de investigación muy diversas, por ejemplo, para el estudio del clima o la predicción de incendios forestales, en biomedicina para la detección temprana de enfermedades, en ingeniería genética para el estudio del genoma. Se ha usado y se usa en el campo de los negocios para detectar, por ejemplo, patrones de fuga de clientes o fraudes con tarjetas de crédito. También se utiliza para examinar redes eléctricas para diagnosticar problemas y prevenir fallas totales o parciales de la red”. (Bogado-García, 2012, p. 2).

De esta manera el propósito de este ensayo es dar claridad de los conceptos de transparencia y ética algorítmica en el uso de los análisis de grandes cúmulos de datos en el que se involucren datos de los individuos.

Para poder llegar a dicha claridad conceptual se analizó material bibliográfico que refiriera estudios sobre discusiones previas y el estado del arte de la ética en el diseño de los algoritmos que se utilizan en el análisis Big Data, pues estos temas son relativamente recientes y aún hay mucho que estudiar al respecto. Es importante destacar que los Estados tienen la responsabilidad de garantizar que con el uso de estas nuevas tecnologías los seres humanos no verán violentados sus derechos. Por esta razón resulta importante garantizar, entender y exigir un diseño transparente y ético de los algoritmos.

## II. CONCEPTOS BÁSICOS SOBRE ÉTICA, ALGORITMOS Y TRANSPARENCIA ALGORÍTMICA

Para comprender la ética algorítmica es necesario entender qué es la ética y cómo debería estar presente en particular entre las personas que trabajan cotidianamente con el entrenamiento de algoritmos que tienen que ver con los datos de los seres humanos.

Gustavo Ortiz (2016) señala que la ética proviene

del griego, y puede tener dos distintas etimologías que son complementarias. Una primera etimología nos dice que proviene de ἔθος (εοσ, τό, ἔθω) que significa “hábito”, “costumbre”, “estar acostumbrado” [...]. Según esta etimología, el término “ética” compartiría sus raíces con el término “etología”, del griego ἔθος, costumbre, y λόγος, razonamiento, estudio o ciencia. La etología estudia el comportamiento de los animales, sus hábitos y sus costumbres, en el [...] ambiente o en el lugar que se encuentren. Curiosamente, la palabra griega ἦθος, que es la otra posible raíz del término, también significaba “lugar acostumbrado”, “morada”, “refugio”, “guarida”, [...]

Una segunda etimología del término “ética” lo haría provenir de ἦθος (êthos) que significa “carácter”, y [...], vincula con ἔθος, hábito o costumbre. La vinculación de estos dos términos es clara dentro de la ética aristotélica: el carácter se forma a través del hábito o la costumbre. Por ejemplo, es a través de la repetición de acciones virtuosas que éstas se vuelven un hábito y se forma el carácter virtuoso (ἀρετή). (pp. 117 y 118).

Así, ética es tanto un hábito o costumbre en el comportamiento de los individuos como también el carácter que se forma y moldea por esta misma costumbre que pretende tener como resultado acciones virtuosas. No podemos apartarnos de que la ética hace referencia al control del propio temperamento y que se encuentra presente en el momento que se deben de tomar decisiones para actuar de una manera o de otra.

La ética si bien se practica en lo individual se crea en lo colectivo pues es dentro de la sociedad en donde se gestan los valores sobre las conductas correctas o incorrectas que se toman con base de las experiencias que se han vivido en cada una de las sociedades en particular.

Por su parte la ética en la Inteligencia Artificial (IA) es entendida como:

un subcampo de la ética aplicada que estudia los problemas éticos que plantea el desarrollo, despliegue y utilización de la IA. Su preocupación fundamental es identificar de qué modo puede la IA mejorar o despertar inquietudes para la vida de las personas, ya sea en términos de calidad de vida o de la autonomía y la libertad humana necesarias para una sociedad democrática. (Grupo independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, 2018, p. 13).

En este tenor es importante entender cuál es la función de los algoritmos que son utilizados para el desarrollo de la IA. Manzano et. al. (2016) explica que “[l]os algoritmos son una secuencia lógica y detallada de pasos para solucionar un problema. Su campo es amplio y dinámico e intervienen directamente en la vida de las organizaciones resolviendo problemas mediante programas de computadora en las distintas áreas [...]”. (p. 5)

Por su parte Robin Hill señala que “es un constructo matemático con una estructura de control finita, abstracta y efectiva de acción imperativa para cumplir un propósito dada una serie de criterios” (en Monasterio, 2017 p. 185). Monasterio señala que “una definición más comprehensiva y operativa que permita entender y englobar las distintas instalaciones de un algoritmo que puede estar detrás de la conducción de un vehículo autónomo, plataforma on line, agentes o sistemas de Inteligencia Artificial [...] etc., es la siguiente: *un código software que procesa un conjunto limitado de instrucciones*”. (pp. 185-186) (Cursivas del original)

Para profundizar sobre este tema, se consultó a diversos expertos y especialistas. De ellos quienes participaron fueron el Dr. Feliú Davino Sagols Troncoso quien es investigador del Departamento de Matemáticas del Centro de Investigación y de Estudios Avanzados (Cinvestav) del Instituto Politécnico Nacional, y que al momento de entrevistarse se encontraba a cargo de la Dirección Adjunta de Innovación y Capital Humano del INFOTEC, Centro de Investigación e Innovación en Tecnologías de la Información y la Comunicación y el Dr. Guillermo Benito Morales Luna, quien también es Investigador del Cinvestav y se destaca por haber estudiado en la Academia Polaca de Ciencias y trabajar actualmente en los fundamentos matemáticos de la computación como la complejidad y los algoritmos.

De las opiniones recibidas el Dr. Feliú Sagols (2020) opina que:

Formalmente un algoritmo queda especificado en el autómata finito que controla las funciones de una máquina de Turing. Está es una definición muy técnica y hay otras que se ven diferentes, pero son igual de técnicas. En realidad tomó mucho tiempo entender qué es lo que se puede hacer con un dispositivo físico programable. Donald Knuth en “The art of computer programming” dice que un algoritmo es algo que se parece a una receta de cocina sólo que sin las ambigüedades propias de las recetas. Así decir “una pizca de sal” en un algoritmo no es válido y debería transformarse en algo como “3.52344 gramos de sal” para evitar imprecisiones.

Por su parte el doctor Guillermo Morales Luna (2020) señaló que:

Desde el punto de vista formal, un algoritmo es un procedimiento, es una receta. Describe la manera en la que se ha de obtener un resultado final partiendo de un dato inicial. Por ejemplo, el dato inicial puede ser una consulta, y el resultado es la respuesta

a esa consulta. Un algoritmo consta de una serie de pasos, en cada uno de los cuales se efectúa una acción básica, o primitiva. Es pues un artificio matemático susceptible de ser implementado en un dispositivo mecánico, electrónico o físico.

El mismo doctor Morales Luna explica que hay diferentes tipos de algoritmos que él identifica:

Hay algoritmos deterministas (el paso siguiente a cada paso actual de cómputo es uno único), indeterministas (puede haber varias posibilidades a ser asumidas como paso siguiente en cada paso actual), probabilistas (el paso siguiente se determina como una función aleatoria del paso actual) y cuánticos (cada paso actual es una superposición de varios posibles).

Hay diversas variantes de esos tipos: los algoritmos heurísticos, los genéticos, los miméticos, los voraces, etc.

- Desde el punto de vista de implementación, hay algoritmos seriales (sólo se ejecutan en un procesador), distribuidos (varios procesadores realizan diversas partes del algoritmo) y paralelos (se ejecutan de manera sincronizada por varios procesadores).
- Desde el punto de vista de su composición hay algoritmos recursivos, iterativos, de fuerza bruta, aleatorizados, de estrategias basadas en “divide y vencerás”, de programación dinámica, etc.
- Desde el punto de vista de las aplicaciones, hay algoritmos sintácticos, numéricos, de Optimización, de Cálculo Simbólico, de Teoría de Gráficas, de Geometría Computacional, de Bases de Datos, de Redes, de Inteligencia Artificial, de Grandes Cúmulos de Información, etc.

Las preguntas en esta encuesta buscan explorar la relación entre algoritmos y sociedades humanas. Así pues, nos concentraremos en el aspecto de aplicaciones de los algoritmos, en particular en aquellos relacionados con temas de Inteligencia Artificial y de Grandes Cúmulos de Información. El aspecto formal de los algoritmos queda ajeno a esta encuesta y lo dejamos en la quimera platónica de idealidad matemática. (2020)

Como se puede observar estamos frente a un lenguaje técnico que debe ser comprendido ya no solo por ingenieros, físicos o matemáticos, sino que ahora se los científicos sociales también deben entender qué es un algoritmo, así como su repercusión en las sociedades actuales.

Monasterio (2017) considera que “la progresiva introducción de la tecnología de las máquinas en todas las facetas de la vida [...] hace que esperemos y confiemos más en ellas que en las propias personas” (p. 186) y complementa con que esta revolución algorítmica tiene uno de los grandes desafíos y amenazas ya las máquinas y las tecnologías en general se han hecho más complejas tanto que no se tiene la capacidad de entender cómo funcionan ni se pueden anticipar comportamientos inesperados o las brechas de seguridad y pone como ejemplo los 2 billones de líneas de código que en 2015 se usaron para programar Google frente a los 145,000 líneas de código usadas en 1969 para hacer el viaje a la Luna. Continúa con que “los sistemas algorítmicos son laberintos incomprensibles, en muchos casos hasta para los ingenieros, matemáticos y físicos que las escribieron”. (p. 186)

Respecto a la relación que existe entre el tema de los derechos humanos y el diseño de algoritmos el Dr. Sagols comentó que no existe una relación entre el diseño de algoritmos y los derechos humanos ya que:

A un diseñador de algoritmos se le tiene que pedir como parte de los requerimientos del problema que va a resolver que considere los derechos humanos como restricciones funcionales. También se le podría dar un manual sobre “derechos de los gatos” para que haga algoritmos amables con estos felinos.

Podemos observar que para el Dr. Sagols los algoritmos en sí no tienen relación directa con un tema en particular como los derechos humanos, sino que al momento de diseñar los algoritmos es cuando se les dota de restricciones para que puedan desempeñar determinadas respuestas, o dar resultados a determinados planteamientos que pueden o no estar relacionados con los individuos y por consiguiente con sus derechos humanos.

Por su parte el Dr. Morales Luna es de la opinión de que

el diseño de algoritmos conlleva una buena educación matemática y computacional. En estos tiempos en que se llega a ver al acceso a la ciencia como un derecho humano, el diseño de algoritmos puede ser visto también como inherente a los derechos humanos.

Por otro lado, desde las aplicaciones hay muchos puntos a considerar. La Algorítmica es esencial a la Computación y ésta lo es a las comunicaciones, a la Administración Pública, a las relaciones comerciales y prácticamente a cualquier aspecto de la vida moderna. Así, todos estos aspectos conllevan a la Algorítmica. Conllevan, asimismo, y a su vez, aspectos de privacidad, de libertad individual, de gobierno y de relaciones sociales.

La mejoría de la administración pública es proporcional a la cantidad de información con que se cuente, pero ésta es inversamente proporcional a la privacidad y a nociones, digamos canónicas, de libertad individual. Diversos países han procurado conciliar estos dos intereses (libertad individual y eficiencia administrativa) pero esta polémica está lejos de ser resuelta. La organización de los servicios médicos actuales en los EUA y en China son ejemplos extremos de otorgamientos de prioridades en esa dicotomía.

También se tiene que las relaciones sociales no son completamente deterministas, por lo que al intentar mecanizarlas se las restringe. Por ejemplo, si se quisiese reconocer un rasgo social, digamos potencialidad para obtener un grado académico, se puede precisar ese rasgo al nivel de tener un procedimiento mecánico para decidir si un individuo lo cumple o no, digamos un examen de admisión y algunos criterios socio-económicos. Con ello, una decisión negativa de cumplimiento podría considerarse como un acto de discriminación, aunque tal vez no lo fuera. Es en este sentido que puede haber afectación a derechos humanos mediante la Algorítmica.

Es precisamente esta dicotomía que plantea el Dr. Morales Luna lo que motivó este análisis pues si bien como se anotaba al inicio, las TIC pueden ser benéficas para la vida de los individuos, representan a su vez riesgos que se pueden plasmar en actos de segregación, discriminación, vulnerabilidad de la libertad de expresión, de acceso a la información por mencionar algunas.

Entendiendo el concepto de algoritmo, los tipos de algoritmos y en particular la algorítmica que se utiliza para realizar actividades en la sociedad como puede ser el gobierno electrónico, el comercio en línea, la relaciones vía redes sociales, etcétera es preciso prever que con este tipo de usos de la información no se violenten los derechos de los seres humanos y haya un respeto por su dignidad la cual:

contiene en sí la idea de que todo ser humano posee un «valor intrínseco» que jamás se debe menoscabar, poner en peligro ni ser objeto de represión por parte de otros (ni de las nuevas tecnologías, como los sistemas de IA). En el contexto de la inteligencia artificial, el respeto de la dignidad humana implica que todas las personas han de ser tratadas con el debido respeto que merecen como sujetos morales, y no como simples objetos que se pueden filtrar, ordenar, puntuar, dirigir, condicionar o manipular. En consecuencia, los sistemas de IA deben desarrollarse de un modo que respete, proteja y esté al servicio de la dignidad física y mental de los seres humanos, el sentimiento de identidad personal y cultural y la satisfacción de sus necesidades esenciales. (Grupo independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, 2018, p. 15).

En este tenor existe la complejidad de ajustar el diseño de algoritmos a la protección de los derechos humanos. Sin embargo, resulta complejo para los Estados y las autoridades poder conocer la ética con la que se diseña un algoritmo y máxime si esta acción se realiza por parte de entes privados. Entonces hay que abonar por una transparencia algorítmica. Al cuestionar a los expertos sobre qué debemos entender por esta el Dr. Sagols considera que

“Esta siempre existe. Un algoritmo puede ser auditado de manera estricta, de modo que bajo las mismas entradas y condiciones de operación es posible reproducir todo lo que el algoritmo hace”.

Sin embargo, esta auditoría que refiere no necesariamente se realiza en todos los algoritmos que se diseñan en un tiempo, espacio y época determinada. Más bien surge la interrogante sobre la responsabilidad que atañe a los propios diseñadores de algoritmos o quienes solicitan su diseño respecto a las violaciones de derechos humanos que pudieran derivarse de un determinado algoritmo.

El Dr. Morales Luna por su parte considera que

“La noción de transparencia algorítmica se ha entendido como el derecho de cualquier individuo a conocer en detalle los criterios mecánicos que le son aplicados para decidir si posee o no un rasgo determinado, para decidir si cumple o no con un criterio predicado. Por supuesto que la transparencia algorítmica es una condición deseable, pero no siempre puede ser otorgable y esto por causas que van desde la falta de interés o de los conocimientos técnicos de los individuos, hasta las reservas de confidencialidad de las partes decisorias.

Por ejemplo, en Criptografía, o sea en las comunicaciones seguras, el principio de Kerckhoffs, postulado en el siglo XIX, estipula que un protocolo seguro debe serlo aun cuando sean de conocimiento público los procedimientos que utiliza, y su robustez o seguridad, debe depender únicamente de las claves que se usen. La Criptografía actual lo ha asumido, y los procedimientos de comunicaciones seguras de Internet no sólo son de conocimiento público, sino que han sido estandarizados de manera que

esas comunicaciones sean susceptibles de ser realizadas por cualesquiera dos participantes en la red. Sin embargo, diversas empresas financieras (bancarias, hacendarias, de tarjetas de crédito o de seguros) desarrollan algunos procedimientos propios y los mantienen en secreto. Estas últimas empresas plantean reservas también para hacer plenamente públicos sus criterios para clasificar a sus clientes o para decidir si se les otorga o no créditos o servicios de cierto tipo.

Así, la aplicación de criterios mecánicos de decisión puede afectar a los individuos, y puede reclamarse como un derecho de éstos el conocerlos. La dicotomía aparece entre ese derecho y la confidencialidad y reserva de las partes decisorias. Por un lado, el derecho a la información se ha consagrado casi de manera universal, así como también la posibilidad de que exista información reservada. Diversos enfoques han sido asumidos, en diversos contextos, gubernamentales o no, para caracterizar lo que ha de considerarse como información reservada, y estos criterios han de aplicarse a sí mismos”.

La Colisión “The Public Voice” del Centro de Privacidad de Información Electrónica integrada por la asociación para Comunicaciones Progresivas, el programa de Derechos Digitales Europeos, el Observatorio de la Sociedad de la Información de la UNESCO, entre otros agentes, en sus Lineamientos Universales para la Inteligencia Artificial de 2018 señala como una de sus directrices el derecho a la transparencia que se refiere a que “todas las personas tienen derecho a conocer la base de una decisión de AI que les concierne. Esto incluye el acceso a los factores, la lógica y las técnicas que produjeron el resultado”. (Directrices Universales sobre la Inteligencia Artificial”. (Cuevas, 2018)

Ahora bien, por un lado encontramos el derecho a la transparencia y por otro es menester preguntar a los expertos sobre el concepto de ética algorítmica. El Dr. Sagols y el Dr. Morales Luna coinciden en que no existe tal. Sin embargo, el Dr. Morales Luna profundiza en la temática señalando que si bien los algoritmos no pueden ser buenos ni malos, pues son neutrales, no lo es la aplicación de los mismos y señala que esto si representa un tema de discusión toda vez que por ejemplo:

“Isaac Asimov, en los años 40, postuló unas “leyes de la robótica”, propias de la ficción científica, que tenían que ver con el comportamiento ético de los robots. En la práctica, las leyes de Asimov han sido ilusas y ahora existen drones, sin tripulantes humanos, que realizan labores bélicas, vale decir, de exterminio de enemigos humanos.

Las Convenciones de Ginebra han excluido el uso de ciertos tipos de armamento, aún en condiciones de guerra, pero a la fecha no se tiene un acuerdo que prohíba armas mecánicas autónomas en combate contra humanos. Pero ésta es una situación extrema.

En los esquemas de comunicaciones seguras, hay atacantes que buscan quebrantar protocolos y hay defensores que buscan protegerlos. Basta con que un atacante encuentre una debilidad, para que él tenga éxito en su cometido. El defensor, en cambio, ha de prever toda posibilidad de ataque y protegerse adecuadamente contra cualquiera de esos posibles ataques. La defensa, en este sentido, es más difícil que el ataque. A pesar de las connotaciones (maligna en el caso del participio activo “atacante” y benigna en el de “defensor”), la calidad de bueno o malo, de ético o no ético depende de quiénes sean los usuarios finales. Si el defensor es el sistema bancario y el atacante es alguien que busca cometer fraudes, puede pensarse que el defensor está de parte de la ética. Si el defensor trabaja para el crimen organizado y el atacante es el sistema policíaco, es el atacante quien podría estar del lado ético.



Los algoritmos son elementos neutros desde el punto de vista de la ética. Son los usuarios quienes deben someterse a la ética en el uso de ellos. Un dilema similar se dio al fin de la Segunda Guerra Mundial con el desarrollo del Proyecto Manhattan de la bomba atómica por renombrados científicos físicos y matemáticos”.

Los algoritmos permiten elaborar respuestas a problemáticas e incluso hacen posible que las máquinas puedan ser un agente que pueda interactuar con los seres humanos y con otras máquinas. La IA para poder funcionar necesita analizar su entorno y accionar para un determinado fin como puede ser un asistente como Siri o Cortana, “... programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones de internet de las cosas)”. (Belloso Martín, 2018, p. 91)

Eliezer Yudkowsky (2008), citado por Monasterio (2017), señala que propuso “diseñar una ‘IA amigable’. Él entiende esto como la idea de que ‘valores de tolerancia y respeto por el bienestar de los seres humanos sean incorporados como elementos nucleares en la programación de agentes de IA’ Es decir, agentes artificiales de todo tipo tengan implementados valores humanos que conviertan a la IA en un factor positivo en lugar de uno negativo a la hora de contemplar riesgos globales”. (p. 188)

Por su parte Belloso resalta que la “Declaración de Barcelona para el Uso Apropiado de la Inteligencia Artificial”, celebrado el 18 de marzo de 2017, es un manifiesto que plasma las preocupaciones respecto al uso malicioso, o inapropiado e incluso prematuro de las nuevas tecnologías que además de alertar sobre esta situación plantea un código de conducta basado en seis puntos a saber: prudencia, fiabilidad, transparencia, responsabilidad, autonomía restringida y el papel humano. (Belloso Martín, 2018, p. 94-95)

El reporte de IA del organismo AI Now Institute, del año 2019 concluye que a pesar de que se han incrementado los marcos éticos sobre los sistemas de IA se continúan implementando dominios sobre temas tales como la medicina, la educación, el empleo, la justicia penal sin el adecuado cuidado en las estructuras en las que se basa y aún es poco claro el tema de las responsabilidades que se derivan de un mal manejo de la IA que van desde el daño ambiental como por ejemplo en la extracción desmedida de minerales de la tierra, o bien la dependencia de los sistemas de salud en la IA que podría traer como consecuencia la muerte de los pacientes.

La participación de reguladores, investigadores, trabajadores y defensores señala el mismo reporte, forjan nuevos lazos de solidaridad que son necesarios para establecer las condiciones apropiadas sobre el uso y progreso de la IA que garantice buenos augurios en el futuro. (AI Now Repot, 2019, pp. 58 y 59)

Es así como se llega a estudios como los presentados en el año 2019 del Grupo de Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial de la Unión Europea que han trabajado sobre una Guía al respecto de cómo lograr una IA Fiable y que algunos de los elementos presentados son atractivos a este estudio como se explica a continuación.

### III. ALGUNOS PRINCIPIOS DE LA IA FIABLE PARA EVITAR LA VULNERACIÓN DE DDHH

Los derechos humanos que pudieran verse vulnerados con el uso de la IA, es decir que no sea creada teniendo en consideración los aspectos éticos básicos para no dañar la dignidad humana son varios.

Básicamente el primero de los derechos que se ven relacionados en el tema de los análisis de big data se encuentra la privacidad que aunque si bien no existe un concepto de privacidad único o cerrado, se observa que al menos en ordenamientos jurídicos como el norteamericano se suele utilizar el concepto del “derecho a no ser molestado”, o el “derecho a la intimidad”, “derecho a ser dejado en paz”; esto derivado del contexto e incluso el idioma al que se hace referencia sobre la privacidad encontramos en el francés el uso del concepto de vida privada (*vie privée*) en lugar de privacidad y en el idioma inglés se traduce en derecho a la privacidad o derecho estar solo (*right to privacy* o *right to be alone*), o en italiano reserva de la información (*riservatezza della informazione*). Así todas estas concepciones alrededor de privacidad vuelven aún más complejo el dar una definición. Sirva por ahora con identificar a la privacidad como aquel derecho que poseen las personas para que no haya una interferencia en su actuar cotidiano, la ausencia de intromisión de terceros en la vida de las personas.

Sin embargo, se debe tomar en cuenta que no solo es la privacidad de las personas el tema que se ve comprometido cuando se hace uso de la IA. Ya en las aportaciones del doctor Morales Luna se trataba el tema de la discriminación, que es otro de los aspectos en los que se puede incidir si se carece de una ética de la inteligencia artificial, también se pueden generar discursos de odio, e inclusive la elaboración de listas negras en franca vulneración de los derechos de los individuos. (Téllez-Carvajal, 2020)

Afortunadamente “una nueva generación de filósofos está militando en el activismo global poniendo el énfasis y ayudando, en la introducción de la sensibilidad hacia los valores humanos a la hora de diseñar sistemas de IA, (ro)bots o cualquier software que tenga un potencial impacto en las personas”. (Monasterio, 2017, p. 189) Estos impactos los señala Monasterio como la discriminación social, económica de acceso libre a la información y privación de la libertad, así como la discriminación y abuso de control.

Algunos de estos ejemplos se han observado en algoritmos que por ejemplo predecían la reincidencia de delincuentes en el sistema norteamericano encontrando un sesgo en el algoritmo al arrojar datos de una mayor incidencia de reincidir en actos ilícitos por el hecho de pertenecer a la raza negra. Esto evidentemente tenía un sesgo. “Cuando procedimientos o protocolos automatizados (algoritmos) deciden por los seres humanos y encima lo hacen de manera sesgada y en contra de derechos y libertades civiles que las personas poseen se produce un fenómeno ético particular: el daño causado tiene difícil identificación para rendición de cuentas y/o responsabilidad, la complejidad de la programación de los algoritmos impide corregir o enmendar...” (Monasterio, 2017, p. 197)

Para tratar de mitigar entonces cualquier vulneración a los derechos humanos es que el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial de la Unión Europea

plantean la necesidad de contar con una ética fiable de la Inteligencia Artificial que será garantizada siguiendo determinados principios como es el principio de explicabilidad, el principio de equidad, el principio de diversidad y el de no discriminación que es definido por los mismos expertos como:

### **Principio de explicabilidad**

“significa que los procesos han de ser transparentes, que es preciso comunicar abiertamente las capacidades y la finalidad de los sistemas de IA y que las decisiones deben poder explicarse —en la medida de lo posible— a las partes que se vean afectadas por ellas de manera directa o indirecta. Sin esta información, no es posible impugnar adecuadamente una decisión. No siempre resulta posible explicar por qué un modelo ha generado un resultado o una decisión particular (ni qué combinación de factores contribuyeron a ello). Esos casos, que se denominan algoritmos de «caja negra», requieren especial atención. En tales circunstancias, puede ser necesario adoptar otras medidas relacionadas con la explicabilidad (por ejemplo, la trazabilidad, la auditabilidad y la comunicación transparente sobre las prestaciones del sistema), siempre y cuando el sistema en su conjunto respete los derechos fundamentales. El grado de necesidad de explicabilidad depende en gran medida del contexto y la gravedad de las consecuencias derivadas de un resultado erróneo o inadecuado”. (Grupo independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, 2018, p. 16).

### **Principio de equidad**

“El desarrollo, despliegue y utilización de sistemas de IA debe ser equitativo. Pese a que reconocemos que existen muchas interpretaciones diferentes de la equidad, creemos que esta tiene tanto una dimensión sustantiva como procedimental. La dimensión sustantiva implica un compromiso de: garantizar una distribución justa e igualitaria de los beneficios y costes, y asegurar que las personas y grupos no sufran sesgos injustos, discriminación ni estigmatización. Si se pueden evitar los sesgos injustos, los sistemas de IA podrían incluso aumentar la equidad social. También se debería fomentar la igualdad de oportunidades en términos de acceso a la educación, los bienes los servicios y la tecnología. Además, el uso de sistemas de IA no debería conducir jamás a que se engañe a los usuarios (finales) ni se limite su libertad de elección. Asimismo, la equidad implica que los profesionales de la IA deberían respetar el principio de proporcionalidad entre medios y fines, y estudiar cuidadosamente cómo alcanzar un equilibrio entre los diferentes intereses y objetivos contrapuestos. La dimensión procedimental de la equidad conlleva la capacidad de oponerse a las decisiones adoptadas por los sistemas de IA y por las personas que los manejan, así como de tratar de obtener compensaciones adecuadas frente a ellas. Con este fin, se debe poder identificar a la entidad responsable de la decisión y explicar los procesos de adopción de decisiones”. (Grupo independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, 2018, p. 15).

## **Principio de diversidad, no discriminación**

“Para hacer realidad la IA fiable, es preciso garantizar la inclusión y la diversidad a lo largo de todo el ciclo de vida de los sistemas de inteligencia artificial. Además de tener en cuenta a todos los afectados y garantizar su participación en todo el proceso, también es necesario garantizar la igualdad de acceso mediante procesos de diseño inclusivos, sin olvidar la igualdad de trato. Este requisito está estrechamente relacionado con el principio de equidad”. (Grupo independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, 2018, p. 23).

“La diversidad y la inclusión desempeñan un papel esencial al desarrollar los sistemas de IA que se utilizarán en el mundo real. Es crucial que, a medida que los sistemas de IA vayan desempeñando una mayor cantidad de tareas por sí mismos, los equipos encargados de la adquisición o del diseño, desarrollo, ensayo, mantenimiento y despliegue de estos sistemas reflejen la diversidad de los usuarios y de la sociedad en general. Esto contribuye a garantizar la objetividad y la toma en consideración de las diferentes perspectivas, necesidades y objetivos. Lo ideal es que la diversidad no solo se materialice en los equipos en términos de género, cultura y edad, sino también de antecedentes profesionales y conjuntos de competencias”. (Grupo independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, 2018, p. 29).

Así, el talento humano que interviene en la creación de la IA debe considerar que aquellos encargados de la recolección de datos tienen y siguen códigos de ética, que se apegan a la privacidad de los individuos y que siguen los principios de la protección de datos como es que se recolecten con fines lícitos, que los individuos conocen y autorizan el uso de su información.

A este respecto se hace necesario conocer cómo se está formando la ética en la IA en el talento humano, en qué carreras universitarias aparte de ciencias de la computación debería impartirse obligatoriamente formación sobre los derechos humanos, la protección de datos y la privacidad tendiente a reflexionar sobre los comportamientos éticos y la vulneración de derechos por medio del manejo inadecuado de la información de grandes cúmulos de datos.

Los comportamientos éticos en torno a la IA y el manejo de grandes cúmulos de datos es un tema vigente y necesario en las agendas de los Estados pues por su propia naturaleza, no puede haber un control sobre la recolección de la información que se obtiene, suprime, modifique o borre en la red de redes por ello un acercamiento a este respecto es precisamente incidir en la formación del talento humano que habrá de trabajar con estas valiosas informaciones.

## **IV. ÉTICA Y TRANSPARENCIA EN LA IA COMO PARTE DE LA CULTURA DE RESPETO A LOS DERECHOS HUMANOS**

Es entendible que si no hay un control sobre quienes diseñan algoritmos puede existir un algoritmo que en estos momentos se esté utilizando en franca violación a los de-

rechos humanos y que sin embargo no tenemos conciencia de pudiera inclusive estar haciendo uso de nuestros propios datos.

“Como bien dan cuenta Julia Powles y Clarissa Veliz (2016) en un artículo para *The Guardian* [..existen] las maniobras de colusión y/o monopolio de las empresas tecnológicas de Silicon Valley para ordeñarnos como “vacas digitales” y así extraernos todos los datos posibles que solo ellos controlaran como señores feudales; [por ello] no solo es necesario una legislación fuerte como el Reglamento Europeo de Protección de Datos (2016), sino la ética y la cultura”. (Monasterio, 2017, p. 192)

En este aspecto es precisamente fundamental la ética que se ha gestado ya en torno al tema de la IA y por supuesto la consecuente cultura que se comienza a adquirir por parte de todos los involucrados en el tema de las nuevas tecnologías, el manejo de datos y todo el tema relacionado a los derechos humanos y sus vulneraciones por el uso de tecnologías de la información y comunicación,

“Hay que entender y saber quién decide, quién gobierna y quién distribuye la información; esencial para las libertades individuales y nuestras democracias. Las implicaciones éticas de una sociedad cada vez más gobernada por los algoritmos obligan a la industria, academia e instituciones públicas a buscar alianzas para crear una gobernanza transparente, ética y justa de la caja de Pandora que puede ser la IA” (Monasterio, 2017, p. 191)

Como sociedades alrededor del mundo se hacen intentos para que estos temas permeen al interior de los Estados pero que al mismo tiempo sean temas que lleguen hasta los ciudadanos de a pie de quienes también se recaban, tratan, almacenan y en general tratan datos o que son usuarios de tecnologías de IA que incluso podrían estar influenciando sobre qué comprar, a dónde ir, o incluso qué pensar.

Belloso señala que la

“progresiva opacidad de los algoritmos y concentración de la tecnología en manos privadas [son un problema]. Mientras que la Inteligencia Artificial en forma de aprendizaje automático (*machine learning*) y en especial de ‘aprendizaje profundo’ (*Deep learning*), alimentado por Big Data, se hace cada vez más poderosa y se aplica en un número creciente de nuevos productos y servicios digitales en los sectores público y privados, su funcionamiento interno se está haciendo extremadamente difícil, si no imposible, de seguir, explicar y evaluar críticamente. Además, estas capacidades avanzadas se están acumulando en gran parte en manos privadas”. (Belloso Martín, 2018, p. 106)

No menos relevante es reflexionar sobre las personas, instituciones, organizaciones que se ven involucradas en la IA y sus obligaciones, pero también sus responsabilidades cuando se vulneran derechos humanos por usar dicha IA. Es justamente este un tema que ha cobrado relevancia entre los licenciados en derecho que discurren sobre las responsabilidades por ejemplo de la empresa que vende un auto que se maneja solo, o bien la compañía que diseñó el software o bien el propio ingeniero que estuvo a cargo de la elaboración de este.

“La investigación, el diseño y el desarrollo de IA, la robótica y los sistemas “autónomos” deben guiarse por una auténtica preocupación por la ética de la investigación, la responsabilidad social de los desarrolladores y la cooperación académica global para proteger los derechos y valores fundamentales, diseñando tecnologías que los respalden en vez de restarles valor”. (Belloso Martín, 2018, p. 108)

Belloso resalta que:

“La falta de transparencia es clamorosa –como se pone de manifiesto en el diseño de los algoritmos y las dificultades para seguir su trazabilidad- de manera que se acaba al albur de las “mentes” que diseñan la inteligencia artificial –tanto por intereses económicos o por otras finalidades espúreas-”.(Belloso Martín, 2018, p. 112)

Finalmente hay dos temas que deben ser considerados para lograr una cultura que se relacione con la ética y transparencia de la IA y que por consiguiente se encuentre armónicamente con el respeto a los derechos humanos que es la democracia en la IA y el Estado de Derecho y la responsabilidad de todos los actores que intervienen en el proceso de creación de IA.

Belloso (2018) identifica que se entiende por Democracia en este contexto:

“Democracia: Las decisiones clave sobre la regulación del desarrollo y la aplicación de la inteligencia artificial deben ser el resultado del debate democrático y la participación pública. Un espíritu de cooperación global y de diálogo público sobre el tema garantizará que las decisiones al respecto se tomen de manera inclusiva, informada y con visión de futuro. El derecho a recibir educación o de acceder a información sobre nuevas tecnologías y sus implicaciones éticas facilitará que todos entiendan los riesgos y las oportunidades, y estén facultados para participar en los procesos de toma de decisiones que determinen de manera crucial nuestro futuro”. (Belloso Martín, 2018, p. 109)

Y señala también al

“Estado de derecho y responsabilidad: El Estado de Derecho, el acceso a la justicia y el derecho a una reparación y un juicio justo proporcionan el marco necesario para garantizar el respeto a los derechos humanos en la posible regulación específica de la IA. Esto incluye la protección contra los riesgos derivados de los sistemas ‘autónomos’ que podrían infringir los derechos humanos, como la seguridad y la privacidad. En este sentido, los gobiernos y las organizaciones internacionales deberían aumentar sus esfuerzos para aclarar con quién recae la responsabilidad por los daños causados por el comportamiento indeseado de los sistemas “autónomos”. Además, deberían existir sistemas efectivos de mitigación de daños” (Belloso Martín, 2018, p. 109)

## DISCUSIÓN

El avance de las tecnologías en particular la minería de datos, los análisis de grandes cúmulos de datos así como la utilización de Inteligencia Artificial tienen como uno reto reconocer y respetar estándares éticos mínimos ya que desconocer el tema nos lleva a riesgos y consecuencias adversas para los individuos en particular y las sociedades en

general al ignorar las consecuencias que se pueden derivar de un mal manejo de estas tecnologías como puede ser la violación a derechos humanos como incitando la segregación, incidiendo en las libertades como de opinión y comunicación e incluso llegar a manipular a los individuos sin darnos cuenta.

Algunos expertos en el manejo de las matemáticas, ingenierías y programación en particular señalan que el entrenamiento de algoritmos es una secuencia lógica por medio de la cual se diseña alguna función determinada con la finalidad de resolver una problemática, hasta este punto se puede observar las bondades de esta actividad humana, sin embargo al permitir que estas secuencias lógicas puedan derivar en la Inteligencia Artificial existen riesgos de que las soluciones planteadas lejos de favorecer a los individuos puedan perjudicar sus vidas al no contar con una reflexión sobre las posibles vulneraciones que se derivan de resultados de algoritmos que no pueden reconocer necesidades particulares o estándares distintos pues no se tomaron en cuenta durante su programación, situación que nos hace vulnerables ante estos diseños algorítmicos.

Se observa que al entender a los algoritmos como una serie de pasos para que las máquinas puedan realizar determinadas actividades puede darnos una falsa percepción de que alrededor de estas acciones no puede haber fallos y que la Inteligencia Artificial solo trae aparejados beneficios para los individuos, estas falacias pueden provocar verdadero daño a los individuos ya que los algoritmos no solo crean soluciones sino problemas si no se tiene restricciones funcionales como las que señala el Dr. Sagols en la entrevista.

Por su parte el Dr. Morales resalta con claridad que la Algorítmica ha permeado casi todos los aspectos de la vida moderna por lo que también trastoca temas como las libertades, la privacidad y en general las relaciones sociales.

Es así que en la discusión del tema el Grupo independiente de Expertos de Alto Nivel sobre Inteligencia Artificial desde hace ya casi tres años ha señalado que la Inteligencia Artificial debe desarrollarse siempre teniendo en cuenta la dignidad de los seres humanos, esto incluye su identidad, su cultura, sus derechos y por supuesto la satisfacción de sus necesidades esenciales. En sus aportaciones se menciona la Inteligencia Artificial fiable que se puede alcanzar tomando en consideración el principio de explicabilidad, el principio de equidad, el principio de diversidad y el de no discriminación.

Infortunadamente la realidad como se ha dejado en evidencia aquellos que diseñan los algoritmos no necesariamente tienen claridad en los límites que estos diseños deben de tener y de allí deriva la riqueza de la discusión de este tema pues se deja al descubierto la falta de una adecuada incorporación del tema de la ética algorítmica y la Inteligencia Artificial fiable que es dejada entre aquellas personas que diseñan algoritmos.

## CONCLUSIONES

El uso de las tecnologías representa grandes aciertos para la humanidad pero también grandes desafíos. A lo largo de este trabajo se ha dejado de manifiesto que si bien los

individuos utilizamos las tecnologías no necesariamente estamos conscientes de lo que la algoritmia poco ética puede traer aparejado para nosotros.

Acorde a la información brindada se puede indicar que el tema de la ética algorítmica y la Inteligencia Artificial fiable dependen en gran medida de la concientización que se tiene del tema entre los diversos participantes en el tema que van desde los diseñadores de algoritmos pero también de aquellos responsables de tratamiento de datos, aquellos que encargan el diseño de una secuencia lógica para atacar un problema específico y por ende las universidades que brindan la educación y capacitación en el tema, los Estados a través de sus gobiernos que prevean las posibles vulnerabilidades y riesgos así como estar preparados para las posibles consecuencias y eventualidades, las empresas que pueden estar recolectando información personal, y tratándola sin dejar de lado las buenas prácticas que se tengan para proteger los derechos humanos por encima de las secuencias lógicas que pretenden brindar soluciones a problemáticas específicas.

Así es como hoy el Grupo independiente de Expertos de Alto Nivel sobre Inteligencia Artificial de la Unión Europea busca que la Inteligencia Artificial se vuelva fiable siguiendo los principios de explicabilidad, el principio de equidad, el principio de diversidad y el de no discriminación.

Un tema que no debe ser dejado de lado es lo que respecta a la transparencia con que se obtienen, tratan, conservan y en general utilizan los datos de los individuos ya que la minería que se hace de los datos de los individuos no es necesariamente conocido por los individuos quienes somos al final del día quienes aportamos la materia prima para el diseño de estos mismo algoritmos y que esta minería no necesariamente resulta manejarse del todo éticamente.

Finalmente es pertinente señalar que si bien el tema de la regulación para lograr la fiabilidad de la Inteligencia Artificial se ha abordado por los expertos de alto nivel de la Unión Europea nuestra región latinoamericana también deben de involucrarse y pronunciarse al respecto.

## BIBLIOGRAFÍA

- Banco Mundial, "Personas que usan Internet (% de la población). Disponible en <https://datos.bancomundial.org/indicador/IT.NET.USER.ZS?end=2018&start=1998>.
- Bellosos Martín, Nuria. (2018) La necesaria presencia de la ética en la robótica: La roboética y su incidencia en los derechos humanos. Cuadernos de Programa de Pós-Graduaçã. Vol. 13, núm. 2.
- Bogado, Joaquín y García, María Betriz (2012) Reflexiones iniciales acerca de la validez ética de la utilización de técnicas de minería de datos sobre datos personales en la búsqueda de terroristas, XVIII Congreso Argentino de Ciencias de la Computación, Red de Universidades con Carreras en Informática (RedUMCI). Disponible en <http://sedici.unlp.edu.ar/handle/10915/23867>
- Cuevas, Verónica. (2018) Firmamos las Directrices Universales sobre la Inteligencia Artificial" de la Coalición "The Public Voice". Noviembre 4. Disponible en [https://sontusdatos.org/2018/11/04/directrices\\_inteligencia\\_artificial/](https://sontusdatos.org/2018/11/04/directrices_inteligencia_artificial/)



- Crawford, Kate, Roel Dobbe, Theodora Dryer *et al.* (2019) AI Now 2019 Report. New York. AI Now Institute. Disponible en [https://ainowinstitute.org/AI\\_Now\\_2019\\_Report.html](https://ainowinstitute.org/AI_Now_2019_Report.html).
- Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, creado por la Comisión Europea (2018), "Directrices éticas para una IA fiable".
- Irazabal, E. (2020) Contra la planificación jurídica europea: el caso de la IA. Disponible en A definitivas, <https://adefinitivas.com/arbol-del-derecho/nuevas-tecnologias/inteligencia-artificial-fiable/>
- Manzano Peñaloza, G., Montesano Brand, R. y Zúliga López, L. F. (2016). Análisis, diseño e implantación de algoritmos. México, UNAM.
- Monasterio Astobiza. (2017) "Ética algorítmica: Implicaciones éticas de una sociedad cada vez más gobernada por algoritmos", Dilemata, año 9 número 24.
- Ortiz Millán, Gustavo, "Sobre la distinción entre ética y moral", Isonomía, núm. 45, octubre 2016. Disponible en <http://www.scielo.org.mx/pdf/is/n45/1405-0218-is-45-00113.pdf>.
- Téllez Carvajal, E. (2020). Análisis documental sobre el tema del big data y su impacto en los derechos humanos. Derecho PUCP, (84)
- The Public Voice (2018) Universal Guidelines for Artificial Intelligence, 23 October, Brussels, Belgium. Disponible en <https://thepublicvoice.org/ai-universal-guidelines/>