e-ISSN 2594-2166 Medicina y ética, vol. 36 núm. 4, octubre-diciembre, 2025

¿Prejuicios en la IA? Análisis del sesgo algorítmico y una propuesta de solución

Biases in AI? An analysis of algorithmic bias and a proposed solution

Pablo de Robina Duhart*
Tiffin University, Ohio, Estados Unidos

https://doi.org/10.36105/mye.2025v36n4.02

Resumen

La Inteligencia Artificial (IA), desde su configuración, como una serie de algoritmos que sirven para obtener información de manera inmediata y razonada, ha planteado dilemas considerables en torno a qué y cómo es que se diseña la programación, sobre todo porque quien está detrás de dicho proceso puede y, de hecho, transfiere sus propios prejuicios a la programación en torno a la visión humana que tiene detrás. Este proceso de transferencia se llama sesgo algorítmico, que será lo que discutiremos en el presente artículo, analizaremos

* Profesor en Tiffin University, Ohio, Estados Unidos. Correo electrónico: derobinap@tiffin.edu https://orcid.org/0000-0003-4939-6264

Recepción: 24/04/2025 Aceptación: 04/06/2025

CÓMO CITAR: Robina Duhart, P. (2025). ¿Prejuicios en la IA? Análisis del sesgo algorítmico y una propuesta de solución. *Medicina y ética, vol. 36, núm. 4.* DOI: https://doi.org/10.36105/mye.2025v36n4.02

sus impactos y la forma en que podríamos aminorarlos o, incluso, eliminarlos. En ese sentido, se propone que sean equipos interdisciplinarios que, al tener diseños transparentes con consideraciones éticas y bioéticas mitiguen los sesgos fomentando así la dignidad humana y la justicia social.

Palabras clave: sesgo algorítmico, inteligencia artificial, equidad, justicia, rendición de cuentas.

1. Introducción

Si bien es cierto que la Inteligencia Artificial (IA) ha estado en nuestra realidad desde hace mucho tiempo, no fue sino a finales de 2021, con la apertura pública de *ChatGPT*, que la IA se puso al centro del diálogo de la sociedad como una herramienta omnipresente y omnipotente que permite influir en todos los aspectos de la vida y la sociedad (1): desde la toma de decisiones a diversos niveles para optimizar la eficiencia (2); hasta la atención médica en la búsqueda de diagnósticos menos subjetivos y más precisos (3). En ese sentido, la IA ha generado un cambio sustancial en la forma de afrontar la vida y, por ello, generar una promesa de mejorar la calidad de vida, pero también la eficiencia de aquellos sistemas que la han adoptado (4); no obstante, esta promesa también ha suscitado diversas preocupaciones, entre ellas, una de las más imponentes es el fenómeno denominado: sesgo algorítmico.

El sesgo algorítmico surge como un problema fundamental que desafía a la ética y la bioética en el uso de la IA en términos de equidad, igualdad y justicia debido a los sistemas que utilizan la IA generan información y predicciones que "beneficia(n) sistemáticamente a un grupo de individuos frente a otro[s], resultando así injustas o desiguales" (5, p. 29). En ese sentido, el sesgo es un instrumento de perpetuación y amplificación de los perjuicios humanos que pueden resultar dañinos para la sociedad (6).

https://doi.org/10.36105/mye.2025v36n4.02

El problema no es menor, ya que los sesgos, si bien pueden ser conscientes, en su mayoría son inconscientes y, por ende, están incrustados en los algoritmos de la IA, lo que fomenta que la toma de decisiones, el uso de los sistemas de IA y las desigualdades sociales estén no sólo en las personas, sino, de manera inherente, en cada uno de los sistemas de IA.

Sin embargo, demos un salto atrás y expliquemos qué es el sesgo algorítmico y por qué es tan preocupante. Como sabemos, todo sistema de IA debe tener una programación, misma que se lleva a cabo por un ser humano o un grupo de personas que, de manera directa o no, trasladan todo su conocimiento, bases de datos, su moral y sus prejuicios a la IA (7). Una vez realizada la programación, el sistema debe ser alimentado con información, misma que también puede estar sesgada por sexo, género, raza, etc. lo que fomenta que la IA no pueda procesar la información completa y, desde su origen, tenga ya programada información específica con la que va a procesar todos sus análisis y la generación de respuestas.

En ese sentido, al tener un sesgo en la información, así como ciertos prejuicios y moralidades que están detrás de la programación, fomentan que la respuesta de la IA tenga, de suyo, injusticias y/o actos discriminatorios en sus tomas de decisiones, debido a la lógica de la IA, lo que puede afectar los resultados del proceso de análisis (8,9).

Por lo anterior, es vital reconocer que, a medida que la IA se vuelve cada día más parte de nuestras vidas, reconozcamos la existencia de los sesgos algorítmicos y busquemos eliminarlos, en la medida de lo posible, ya que, como sabemos, los sistemas de IA están diseñados para imitar el pensamiento humano, lo que, de manera natural, hará que se perpetúen, de manera inadvertida, los sesgos presentes en su programación.

Es por ello por lo que problema del sesgo algorítmico plantea diversos desafíos técnicos, que implican a la ética, la antropología y la filosofía, de manera fundamental, en el proceso de desarrollo de la

https://doi.org/10.36105/mye.2025v36n4.02

tecnología y, sobre todo de los sistemas de IA que están detrás de ésta, con miras al futuro desarrollo de una sociedad más justa, así como a la interacción social en pro del bien común y el respeto de la dignidad y la naturaleza humana (8).

Por lo anterior, el impacto del sesgo algorítmico no se queda sólo en ámbitos de la ética y la sociedad, sino de la bioética, ya que no sólo implica la generación de texto, audio o video, como son las IA generativas; sino que, debido a que dichos sistemas están cada día más inmersos en diversos ámbitos de la vida humana —como en la salud, la política, el mundo empresarial e, incluso, la educación—, es importante garantizar la dignidad, la equidad y la justicia; frente a un mundo que genera injusticia sistemática y, por ende, discriminación constante ya sea por raza, género o etnia, por decir algunos ejemplos (5).

De esa manera, el objetivo del presente es profundizar en el conocimiento del sesgo, para identificar la forma en que se presenta, entenderlo y, con ello generar estrategias que lo mitiguen en todos los sistemas de IA. En ese sentido, debemos abordar estrategias que hagan consientes los sesgos desde un inicio, previa la programación y el entrenamiento de la IA y asegurar que tanto la tecnología, como el desarrollo de los sistemas de IA sigan avanzando, a su ritmo, pero siempre con base en la dignidad humana y en pro del beneficio social, acercando a la sociedad, en lugar de fisurarla más de lo que ya puede estar en la actualidad.

En ese sentido, algunas de las propuestas que surgen para lo anterior, y que analizaremos más adelante, son: la diversificación de datos (frente a la discriminación de éstos), la revisión y actualización de los sistemas con un enfoque ético (frente a los sesgos morales y sociales) y el aumento de la trasparencia y la explicabilidad de los algoritmos (frente a una secrecía de la información) (5). De lo anterior se desprende la necesidad de analizar el problema para asegurar que los avances de la IA y la tecnología promuevan una sociedad más justa y equitativa para todos (11,12).

https://doi.org/10.36105/mye.2025v36n4.02

2. Origen del sesgo algorítmico

Como hemos podido observar en la introducción, todo sistema de IA tiene un algoritmo que hace que dicho sistema funcione de manera adecuada y nos presente la información que requerimos de ellos. No obstante, también hemos reconocido que existen problemas en la programación y en el entrenamiento de este debido a los sesgos existentes de antemano. Así, a través del uso de los sistemas de IA estamos fomentando, de manera indirecta, ciertos resultados sesgados y que, lamentablemente, son producto de los sesgos sociales de los programadores y/o de la información con la que son entrenados (7).

Por lo anterior, es importante reconocer que los algoritmos de la IA están diseñados para emular el proceso de pensamiento humano, lo que lleva, de manera natural, a que generen ideas y elaboren juicios que, voluntaria o involuntariamente, perpetúan tanto las ideas positivas, como los prejuicios negativos que existen, tanto en los datos entrenados, como en la sociedad. Esto, por ende, puede generar dilemas en la IA, sobre todo, pero esto por falta de información de datos de algún grupo social y no por parte de un prejuicio, que se discrimine a algún grupo social, esto, ya es un sesgo de manera natural lo que discrimina, en diversos sistemas de justicia penal o en temas de género, a la sociedad y puede generar problemas de clasificación de casos o la toma de decisiones (7).

En ese sentido, no es menor decir que la presencia de los sesgos algorítmicos son un desafío ético y bioético por las consecuencias que puede tener para la sociedad y el medio ambiente (10). En ese sentido, es muy importante reconocer que los algoritmos, aunque producto de la tecnología, son programados por personas y, por ende, alimentados por las decisiones humanas preexistentes tanto por el diseño, como por el entrenamiento que llevan los sistemas; pero también por los propios usuarios al momento de utilizar la IA; por lo tanto, es necesario abordar efectivamente el sesgo desde su

https://doi.org/10.36105/mye.2025v36n4.02

naturaleza, así como de los impactos éticos y bioéticos que éstos presentan (9).

Por lo anterior, es muy importante comprender adecuadamente los orígenes de los sesgos algorítmicos para lograr mitigar los efectos que éstos pueden generar en la sociedad, por ello, los datos que ya están sesgados desde su origen son un factor preponderante, por ello Abràmoff *et al.* (13) señalan que los sistemas de inteligencia artificial y de aprendizaje automático (AI/ML) aprenden a tomar decisiones a partir de los datos con los que son entrenados (tanto durante el proceso de desarrollo, como en el uso cotidiano, ya que éste también sigue entrenando a la IA).

Uno de los estudios revisados en casos de equidad en salud (13), señala que el acceso equitativo a diagnósticos con uso de IA y al tratamiento de datos se puede exacerbar por los sistemas de AI/ML dependiendo de cómo se aborden los sesgos; esto debido a que, como mencionamos anteriormente, o no existe la información suficiente, o, en su defecto, puede producir inequidades en los diagnósticos debido a los prejuicios programados. Existen otros casos que se han documentado que pueden mostrar el problema del sesgo algorítmico, por ejemplo, cómo es que en la atención sanitaria existen sesgos de género, donde se ha diagnosticado a mujeres de manera menos precisa en comparación a los hombres debido a la subrepresentación de las primeras en conjuntos de datos de entrenamiento (14).

Por otra parte, también hay que mencionar que los sesgos algorítmicos pueden tener su origen en el modo en que se diseñan, construyen y estructuran los algoritmos, lo que, de manera automática, influye en la forma de procesar la información y la toma de decisiones. Por ende, la selección de variables y datos, la definición de vínculos y los criterios utilizados para el éxito, o no, en la obtención de resultados puede sesgar la forma en que un algoritmo se comporta y, por ende, la forma en que se presentan suposiciones y/o prejuicios, sobre todo porque, en informática: el orden de los factores sí altera el producto.

https://doi.org/10.36105/mye.2025v36n4.02

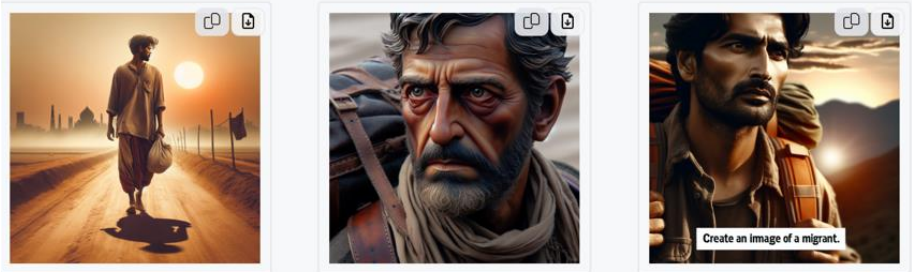
Akter *et al.* (15) hacen eco en que, en áreas como el marketing, la presencia del sesgo algorítmico puede tener impactos fundamentales y opresivos en diversos grupos de *buyer persona* debido a las decisiones en el diseño, el contexto y la aplicación del algoritmo y lo que éste representa en la ubicación, colores, uso de imágenes y marca a la hora de la publicidad y de la ubicación de los productos tanto en páginas de ventas por internet, como en los locales presenciales. Esto, como se mencionó anteriormente, es parte de la disposición de los algoritmos que, de manera indirecta, genera sesgos en la información producida o, incluso, en la forma en que se solicita la información a la IA.

Como hemos podido ver hasta ahora, tanto la propagación de los prejuicios sociales, de manera directa o indirecta; el entrenamiento de la IA, así como la forma del diseño de los sistemas de IA generan sesgos en la información y la forma en que buscamos ayuda, por ello, hay que tener una consideración cuidadosa del diseño, la programación y el entrenamiento de los algoritmos y la presentación de los sistemas de IA para evitar caer en ellos.

Veamos ahora, en la práctica, cómo estos ejemplos, en el desarrollo de imágenes pueden aplicarse a nociones muy sencillas, pero que tienen un impacto importante en cómo las personas pueden utilizar la información: en la Figura 1, la primera imagen es la respuesta a la idea de un migrante, donde, se mantiene la idea de una persona "con buena salud" pero de tez morena; la Figura 2, refuerza los estereotipos del mexicano como un charro y, en la Figura 3, se observa una persona de servicio doméstico como un mayordomo de clase alta y con recursos de primer nivel; en ninguno de los tres casos, se presentan mujeres en las imágenes (sesgo de género).

https://doi.org/10.36105/mye.2025v36n4.02

Figura 1. Un migrante

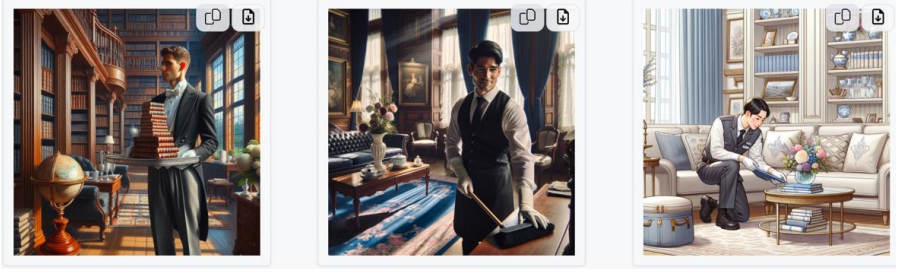


Fuente: desarrollo propio utilizando Bearly.ai

Figura 2. Un mexicano ()

Fuente: desarrollo propio utilizando Bearly.ai

Figura 3. Una persona de servicio doméstico



Fuente: desarrollo propio utilizando Bearly.ai

https://doi.org/10.36105/mye.2025v36n4.02

Como se puede observar, aunque podrían parecer insignificantes, estos efectos presentados en las imágenes están perpetuando la inequidad social y de género, todos hombres, con buena condición física e, incluso, parece que económica. Por ello, debemos considerar que, si esto es algo sencillo a lo que cualquiera tiene acceso, si llevamos estos sesgos a sistemas más completos y robustos, estaríamos demostrando que los sistemas de IA pueden profundizar las desigualdades existentes, así como potencializarlas y crear nuevas formas de discriminación.

Por un lado, al potenciar las desigualdades, los algoritmos legitiman estructuras de poder desiguales, lo que genera la negación de oportunidades y de acceso equitativo a diversos servicios o productos (7,16); por otro lado, también amplifican la discriminación en el acceso y uso de estas a nivel social.

Por lo anterior, es muy importante empezar a generar estrategias de mitigación de los sesgos enfocadas tanto en la fuente de datos que nutre a los sistemas de la IA como en el diseño y programación de los propios sistemas, lo que implica que, de manera directa se intervenga en la tecnología y en la consideración de sus aplicaciones y consecuencias para todos los contextos, en general; pero también para cada contexto en específico (16).

3. Diversas implicaciones en el ámbito social, ético y bioética

Como hemos podido observar, las implicaciones del sesgo algorítmico son múltiples y profundas, que van desde replicar estereotipos de una cultura o un género particular; hasta fomentar discursos que, de manera velada, atentan ya contra la dignidad de una persona por su género, sexo, religión, etcétera, de manera inintencionada. Obviamente, esto tiene un impacto directo en la sociedad, en la ética y la bioética ya que, como señalan Barocas, Hardt, & Narayanan (7), el sesgo algorítmico puede agravar las desigualdades sociales

https://doi.org/10.36105/mye.2025v36n4.02

existentes, lo que lleva a consideraciones éticas y bioéticas en el uso de los sistemas de IA para la promoción de la justicia y la equidad.

Algunos de los ejemplos que Diakopoulos (10) plantea son los sistemas de justicia basados en sistemas de IA, ya que, en caso de haber sesgos en sus algoritmos, ciertos grupos étnicos podrían ser tratados de manera injusta, incluso discriminatoria, por lo que se podría agravar o confirmar que, debido a la pertenencia a estos grupos pueden sufrir más o solventar mejor el peso de la ley, razones mediante las cuales se elimina la justicia y la equidad para contextos críticos.

Otro de los sesgos que puede afectar en términos bioéticos es la disposición de la autonomía y el consentimiento informado para diversos tratamientos, ya que, en caso de existir los sesgos, la primacía en la atención, así como el trato digno a las personas puede verse afectado y, por ende, darse atención prioritaria a personas de ciertos grupos sectarios invalidando el *triage*, así como la valoración médica de cada paciente en particular (13).

Además, el sesgo algorítmico erosiona la confianza en la tecnología debido a que se plantea el dilema sobre la objetividad de los sistemas de la IA y, por lo tanto, de sus productos y resultados; todavía más cuando esto tiene implicaciones en robótica y en sistemas antropomórficos ya que, si los algoritmos de IA producen resultados injustos o discriminatorios, la sociedad puede perder la fe en la tecnología y rechazar su adopción (1).

Lo anterior, por ejemplo, puede tener consecuencias negativas en áreas como la atención médica, donde la IA ha demostrado ser una buena herramienta para el diagnóstico y el tratamiento de padecimientos que puedan tener los pacientes (9), sin embargo, imagina que el diagnóstico tiene un sesgo en su programación y éste, por lo tanto, genera mejoras en la salud en ciertos pacientes, pero, por su parte, genera perjuicios o, incluso, la muerte de otros por su mera condición étnica o de género. Cuidar entonces los sesgos algorítmicos es de vital importancia y es la razón esencial que la ética y la equidad sean consideradas en cualquier sistema de IA, sobre todo en

https://doi.org/10.36105/mye.2025v36n4.02

su desarrollo, uso, implementación y manejo, además de la necesidad de una persona que ayude a la interpretación que valide o corrija los resultados de la IA (9).

4. El transhumanismo y la IA

Como hemos visto anteriormente, el uso de la IA y los sesgos que esta presenta no sólo afectan en términos de justicia y de ética, sino también en casos de bioética, uno de los ejemplos más claros es en la mejora de las condiciones humanas naturales por medios no necesarios o, mejor conocido como transhumanismo. El vínculo entre los sistemas de IA y éste plantea cuestiones profundas sobre cómo la tecnología puede mejorar las capacidades humanas, más cuando reconocemos al transhumanismo como un movimiento que busca la mejora de las capacidades humanas mediante tecnología avanzada (17,18).

Es por ello por lo que, como señala Bostrom (17), esta búsqueda de la "mejora" tecnológica debe abordar cuidadosamente el sesgo algorítmico, ya que, si no se hace, existe el riesgo de amplificar los mismos prejuicios que se intentan superar. De ahí el valor tan necesario y reflexivo que requiere la convergencia de estos dos fenómenos para que puedan coexistir, de manera bioética y benéfica para la humanidad la tecnología y la búsqueda de la mejora en ámbitos de la salud (11,12,19,20).

De acuerdo con lo anterior, es fundamental reconocer que la mejora tecnológica del ser humano debe estar alineada no sólo con valores éticos y sociales, sino con la búsqueda de una sociedad más justa y equitativa. Por ende, no se trata de simplemente mejorar las condiciones y capacidades humanas, sin importar el costo; sino de mejorar la calidad de vida humana de manera ética, respetando siempre, la dignidad de la persona y el bien común (21,22).

Dicho lo anterior, el sesgo algorítmico es, en cierta medida un obstáculo importante en el camino de la mejora tecnológica, sobre

https://doi.org/10.36105/mye.2025v36n4.02

todo porque esta se basa en perpetuar prejuicios y desigualdades; por tanto, el transhumanismo, es, de suyo un problema de prejuicios, sesgos y estereotipos en su origen, ya que no se fundamenta en lo que es el ser humano y la búsqueda de la calidad de vida, sino mejorar al ser humano porque, de suyo, la persona no es adecuada y, por ende, sólo aquellos que puedan mejorarse serán los más aptos para vivir en sociedad (17,18).

4.1. *La búsqueda de una antropología que defienda la dignidad humana*[1]

Para poder adentrarnos al tema de la "mejora" o "perfección humana", es decir, para llevar el uso de la IA al campo del transhumanismo (12,23), entonces, es necesario establecer una antropología filosófica completa que, como sugiere Floridi (24), debe considerar todas las dimensiones de la persona, su complejidad, pero también sus facultades y la diversidad de experiencias en el contexto del sesgo algorítmico en la IA.

En ese sentido, si reconocemos que los algoritmos de IA, que tienen la capacidad de interpretar y reflexionar en segundos una gran cantidad de datos, también son inherentemente reflejo de la humanidad que los crea y utiliza, de ahí la importancia de tener una antropología filosófica que resalte la importancia de la empatía y la ética en el diseño de sistemas tecnológicos que respeten y valoren la dignidad humana (24) y, por lo tanto, la necesidad de reconocer que la tecnología no debe ser simplemente una extensión fría de la lógica, sino una herramienta que refleje las verdaderas facultades humanas, su vínculo con sus dimensiones, sin afectarlas negativamente, por lo contrario, la necesidad de desarrollarlas y expandirlas y, por lo tanto, nutrir la dignidad humana y a la misma humanidad (21,22).

Además, si consideramos que la tecnología y el uso de la IA surgen de las facultades humanas como una extensión y/o herramienta que las potencializa, entonces, entender qué concepción de ser humano subyace a los avances y, por ende, a los sesgos es vital. De ahí

[1] Entendemos aquí una antropología ontológicamente fundamentada.

https://doi.org/10.36105/mye.2025v36n4.02

la importancia de considerar una antropología filosófica que subraye la responsabilidad en el desarrollo de la IA (21,22), sobre todo porque si reconocemos el valor de la libertad de los desarrolladores y programadores de los algoritmos, también reconocemos la responsabilidad ética y bioética que tiene de garantizar que sus creaciones no sólo no perjudiquen a los individuos o las comunidades, sino que, además, busquen un beneficio para estos.[2]

Lo anterior implica, por un lado, mitigar los sesgos algorítmicos al mismo tiempo que se promueve la equidad y la justicia (25), como principios fundamentales para lograr una vida de solidaridad y de subsidiaridad (26), por ello, la tecnología debe ser vista como un medio para ampliar y mejorar la condición humana, en lugar de anularla (24), lo que requiere una verdadera comprensión de la naturaleza humana, en todas sus dimensiones, y cómo es que los sistemas de la pueden servir a la humanidad de manera ética y bioética (26).

Por lo tanto, al analizar los sesgos de los sistemas de la IA, en cualquiera que sea su uso, requiere considerar una antropología filosófica de fondo, ya que el análisis antropológico, que, en muchos casos, pasa por alto la dinámica del hombre con la tecnología, debe ahora voltearla a ver y caminar de la mano. En palabras de Latour (27), la rápida evolución de la tecnología puede desafiar las nociones tradicionales de identidad y naturaleza humana y, por lo tanto, la adaptación y adopción de la tecnología se ajusta a diversos contextos culturales y socioeconómicos, pero, no necesariamente antropológicos. En pocas palabras, no se trata de cómo la tecnología afecta al ser humano, sino de cómo la sociedad, en su conjunto se adapta y reformula sus valores adaptándolos al mundo tecnológico.

De esta forma, al hacer un análisis contextualizado y aplicado en pro de la dignidad humana y en la aplicación a contextos específicos puede ayudar a que diferentes concepciones sobre el sesgo y la visión que hay detrás de los prejuicios y los estereotipos pueda aminorar si se consideran las diversas culturas, religiones y diversos grupos

[2] Aquí hacemos referencia no sólo a los principios bioéticos de beneficencia y no maleficencia (25), sino también al de libertad-responsabilidad del personalismo (26).

https://doi.org/10.36105/mye.2025v36n4.02

sociales detrás del diseño y la programación de los algoritmos; de hecho debe ser parte del análisis antropológico que exige ver a la tecnología como parte de la sociedad y cómo se modifican las interacciones entre esta y la sociedad, así como entre las personas y la forma de tomar decisiones a partir del vínculo con los sistemas de la IA.

5. Dilemas éticos y bioéticos del sesgo algorítmico

Otro elemento no considerado y que también genera grandes preocupaciones sobre todo para algunas organizaciones internacionales (28,29) que, además de los conflictos que ya hemos visto en torno a la dignidad, la justicia y la equidad, también es cierto que los algoritmos, tanto su desarrollo como sus sesgos se mueven por objetivos económicos y tecnológicos, sobre todo porque como sostiene Mittelstadt *et al.* (30), la maximización de la eficiencia y la forma de hacer más rentables los sistemas de IA prevalecen por encima de los impactos sociales que podrían tener (9).

En este sentido, es evidente que la creación y el desarrollo de tecnologías que benefician a ciertos grupos sociales, sobre todo por encima de la mayoría de la población, socavan, con ello, la justicia y la equidad, como lo hemos mencionado y, además, anulan la dignidad y las oportunidades del acceso, goce y beneficios que ofrecen los sistemas de la IA, sacrificando así a la ética y la dignidad humana por obtener beneficios económicos, con ello se plantean entonces dudas sobre los valores y las prioridades en la implementación de los algoritmos de la IA (30).

La falta de consideración ética en la implementación de la IA puede tener consecuencias graves (9), ya que puede llevar a la explotación y la discriminación, incluso, como hemos podido verlo en la aplicación al transhumanismo, también reconoce qué o quién es un humano y, por lo tanto, podría tener las bases y fundamentos para definir al ser humano. Aunado a ello, en caso de que algoritmos de

https://doi.org/10.36105/mye.2025v36n4.02

IA pueden ser usados para la interacción humana y la toma de decisiones en diversos aspectos de la vida, entonces, imaginemos el caso en que, además, esté integrada al ser humano, ¿dónde quedaría la libertad humana y la propia falibilidad del ser humano? (31,32,33,34,35).

Para responder esta pregunta, tenemos que afirmar que, si estos algoritmos no son éticos y justos, pueden perpetuar los prejuicios sociales, exponenciales y, al cabo de los tiempos, dividir más a la sociedad entre los que si pueden y los que no pueden pagar las modificaciones, ser parte de diversas etnias, por un sexo específico o cualquier otra segmentación que permita el sesgo algorítmico. Esto nos lleva a reafirmar la importancia de que existan criterios éticos que sean una parte integral del proceso de desarrollo de la IA, y no simplemente una reflexión tardía.

6. Propuesta de solución: mitigación del sesgo algorítmico

El reconocimiento de que la tecnología IA, aunque avanzada, no es inmune a las falencias humanas ha llevado a la implementación de diversas estrategias para combatir el sesgo; por ello para abordar el sesgo algorítmico y las preocupaciones éticas en la IA, es imperativo priorizar la transparencia en el diseño y funcionamiento de los algoritmos (36), sobre todo porque un enfoque de "Transparencia por Diseño" puede ofrecer una guía práctica para promover las funciones beneficiosas de la transparencia al tiempo que mitiga sus desafíos en entornos de decisión automatizada.

6.1. *Transparencia por diseño*

Según Felzmann *et al.* (37), la adopción de los principios de "Transparencia por Diseño" puede ayudar a las organizaciones, empresas y a programadores de algoritmos a integrar prácticas, de manera sistemática, que aseguren que la rendición de cuentas sea una prioridad

https://doi.org/10.36105/mye.2025v36n4.02

desde la planeación y desarrollo de los sistemas de IA y no, como ahora, un complemento posterior. Lo anterior debido a que los principios de "Transparencia por Diseño" abarcan consideraciones del contexto, técnicas, informativas y del cuidado de datos sensibles a todas las partes involucradas, incluyendo al usuario final; lo que implica cuidar una multiplicidad de dimensiones para que se logre una eficacia y eficiencia en la transparencia.

No obstante, el lograr implementar dichos principios también representa grandes desafíos que, no por ello imposibles de lograr. Para empezar, es vital reconocer la complejidad de los algoritmos y el proceso de decisiones y cómo es que, para aquellos que no son especialistas, resulta muy difícil, incluso opaco para la comprensión, lo que hace que el público en general se vea imposibilitado para entenderlos y analizarlos.

6.2. *Propiedad intelectual y seguridad de los sistemas de IA*

En segundo lugar, también es importante considerar la propiedad intelectual y la seguridad de los sistemas, sobre todo aquellos que manejan datos sensibles y/o biométricos, ya que la información no se puede divulgar de manera fácil; sin embargo, en caso de que se hiciera público el algoritmo, podría favorecer la filtración de datos innecesaria y que podría poner en riesgo a múltiples personas.

Derivado de lo anterior, Kirat *et al.* (38), al hablar sobre la equidad y la rendición de cuentas de la toma de decisiones de los algoritmos, sugieren que, al contextualizar la equidad algorítmica, sobre todo con relación a los marcos legales aplicables hoy en Estados Unidos y en Europa, se muestran diversas posturas, incluso contradictorias, sobre lo que es una verdadera y adecuada rendición de cuentas y cómo se debe aplicar la transparencia en la práctica. Lo anterior pone énfasis en la importancia de desarrollar estándares internacionales coherentes con cada contexto particular que guíen la implementación efectiva de las políticas y normativas en torno a la trasparencia y la rendición de cuentas de los algoritmos.

https://doi.org/10.36105/mye.2025v36n4.02

6.3. *Equipos multi y transdisciplinarios*

En tercer lugar, ya hemos mencionado la importancia de que los equipos de diseño y desarrollo de los algoritmos de los sistemas de IA son vitales, sobre todo porque a mayor diversidad, menor posibilidad de sesgos; si consideramos que los miembros tienen una amplia gama de perspectivas (de ingeniería, bioética, filosofía, mecánica, etcétera), además de las experiencias que puedan compartir, esto puede ayudar significativamente a que se racionalicen más los sesgos y, por ende se eliminen los prejuicios inadvertidos en las etapas iniciales del diseño y desarrollo de los algoritmos.

Lo anterior, no es exclusivo del ambiente del diseño de la IA, sino también aplica para otros campos, ya que siempre que se fomente la diversidad e interdisciplinariedad de los equipos de desarrollo de cualquier proyecto, se reducen los sesgos inherentes a la práctica y se asegura una variedad de perspectivas en el proceso de desarrollo (39). Con esto, podemos confirmar que, en el desarrollo de los sistemas de la IA, si queremos buscar la equidad, la justicia, el respeto a la dignidad humana y el respeto a los principios de la bioética, entonces, a mayor cantidad de voces, más probable que se logre que el producto final cuente con dichos principios y, por ende, sea un sistema inclusivo y equitativo.

6.4. *Metodologías éticas de diseño y evaluación para la mitigación de sesgos*

En cuarto lugar, debemos considerar también la integración de metodologías éticas en el diseño y evaluación de los sistemas de IA para que se fomenten algoritmos justos y equitativos. Lo anterior, implica que constantemente se hagan revisiones críticas a los conjuntos de datos utilizados desde el entrenamiento de la IA para que realmente sean representativos de todos los grupos (mayorías y minorías), así como estén libres de la mayor cantidad de prejuicios (históricos y modernos); además, una evaluación continua del modelo para identificar cuando, durante la ejecución de los algoritmos surjan sesgos,

https://doi.org/10.36105/mye.2025v36n4.02

tanto por la programación natural, como por los inputs, lo que permita corregirlos desde su origen.

En ese sentido, la transparencia y la rendición de cuentas de los algoritmos se vuelven esenciales, como lo vimos anteriormente, ya que estos permiten comprender más y mejor la forma en que la IA toma decisiones, por ello es vital "hacer que los algoritmos sean más transparentes y explicables ayuda a los usuarios a entender cómo se toman las decisiones" (40), lo que lleva a una mejor forma de evaluar los sistemas, medir los impactos éticos y bioéticos de los mismos y a facilitar la identificación y corrección tanto de los sesgos emergentes, como de errores naturales en los algoritmos.

6.5. *Funciones de los gobiernos y los organismos internacionales para la mitigación*

La quinta y última propuesta de solución para mitigar los sesgos algorítmicos tiene que ver con la incorporación de los gobiernos y organismos internacionales, sobre todo porque cuando se crean normas, leyes y políticas que, además de incorporar los principios éticos y bioéticos (25,26), exijan equidad, transparencia, justicia y una adecuada rendición de cuentas en los sistemas de IA sería un paso crucial para la mitigación de los sesgos.

En ese sentido, la *Recomendación de la UNESCO sobre la ética de la IA* (41), la *Rome call for AI ethics* (24) y la *Ley de Inteligencia Artificial de la Unión Europea* (42) son ejemplos claros de cómo los organismos internacionales guían a los Estados miembros a lograr e incorporar prácticas éticas en todo el proceso de desarrollo, programación e implementación de los algoritmos de la IA, sobre todo enfatizando en los elementos antes mencionados, incluyendo también la protección a los derechos humanos y la dignidad humana. Con estos esfuerzos, además de destacar la colaboración global, fomenta principios compartidos éticos y bioéticos para atender los desafíos que plantea la IA.

Además de las políticas que defiendan los principios éticos y bioéticos (25, 26), también es importante que las normativas, la sociedad

https://doi.org/10.36105/mye.2025v36n4.02

y las empresas de IA, en conjunto, debamos adoptar un enfoque proactivo para la creación, desarrollo, programación uso e implementación de los sistemas de IA (24), esto debido a los impactos que va a tener la tecnología no sólo en el corto y mediano plazo, sino también a largo plazo, en la sociedad y, con ello, anticipar posibles efectos adversos.

Finalmente, el integrar el diálogo interdisciplinario entre filósofos, bioeticistas, científicos sociales, ingenieros, mecatrónicos, líderes políticos y demás miembros de la sociedad involucrados en la IA para generar los principios y normativas, llevará a un enfoque colaborativo y orientado a la equidad y la justicia, fundamental para mediar, mitigar y, en cierta medida, eliminar los sesgos algorítmicos, lo que llevará a un uso y desarrollo de sistemas de IA responsables y benéficos para la humanidad.

De cara al futuro, el compromiso para que la IA sea equitativa y justa, en todas sus dimensiones, requiere, necesariamente, de una colaboración global que trascienda fronteras, disciplinas y sectores de la sociedad; implica, además, establecer estándares internacionales para la ética y la bioética en la IA. Por ello, la visión que se tiene es aquella en la cual la IA no sólo refleje los valores éticos y bioéticos humanos, sobre todo la equidad y la justicia, sino que sea un catalizador para reducir las desigualdades sociales. Finalmente, como sugieren Gebru *et al.* (43), abordar el sesgo algorítmico para los sistemas de IA presenta una oportunidad única para reexaminar y mejorar la forma en la que las tecnologías emergentes sirvan mejor a la sociedad y se alineen a principios éticos.

7. Conclusiones

En conclusión, lo que hemos resaltado en el presente trabajo ha resaltado la importancia crítica de la mitigación del sesgo algorítmico para asegurar el desarrollo y uso equitativos de la IA. Para lograrlo, la necesidad de considerar la transparencia, la interdisciplina y ética en

https://doi.org/10.36105/mye.2025v36n4.02

los equipos de desarrollo, así como la implementación de auditorías éticas rigurosas a partir de las normativas y principios éticos deben estar en primer plano. Tal y como Diakopoulos (10) y Barocas, Hardt y Narayanan (7) mencionan, es vital priorizar estas estrategias para racionalizar y combatir los prejuicios sociales e individuales de los programadores, que están inherentes en los procesos de creación, desarrollo y programación de la IA.

En ese sentido, la colaboración interdisciplinaria en los procesos de desarrollo, aunados al diálogo interreligioso y de diversas perspectivas es esencial, ya que esto garantiza que la amplia gama de experiencias, puntos de vista y consideraciones éticas contribuya a la creación de algoritmos más justos y equitativos. Sólo mediante la pluralidad de enfoque podemos confrontar, abiertamente, y superar los prejuicios arraigados en la sociedad y, por ende, en la IA, que amenazan perpetuar las desigualdades y discriminaciones que conllevan dichos sistemas.

También hemos señalado la importancia de hacer accesible la IA paraque sea utilizada como medio de mejora de la sociedad y la condición humana siempre respetando la dignidad humana y los principios de la bioética (26). La tecnología, como se dice mucho, debe estar al servicio de la humanidad y, por ende, debe buscar activamente eliminar las desigualdades sociales y combatir la discriminación; objetivo que va más allá del simple uso técnico de la IA para obtener ganancias, sino que, de fondo es un imperativo ético y bioético que debe guiar el desarrollo y uso de esta.

Finalmente, concluimos con un urgente llamado a la acción de todos los involucrados en los sistemas de la IA (desde la creación hasta el usuario final): es necesario adoptar un enfoque interdisciplinario e integral que abarque desde la transparencia en el diseño y el funcionamiento de los algoritmos, hasta un firme enfoque en la equidad y la justicia en todos los aspectos de su implementación ya que sólo así podremos asegurar que el progreso tecnológico beneficie a la humanidad alineándose con los más altos estándares éticos y bioéticos.

https://doi.org/10.36105/mye.2025v36n4.02

Referencias

1. Mateescu A, Elish M. AI in Context: The Labor of Integrating New Technologies. Chicago: Data & Society Research Institute; 2015.
2. Meissner P, Narita Y. Así es como la inteligencia artificial transformará la toma de decisiones. Tecnologías emergentes – WEF; 2023. Disponible en: https://es.weforum.org/agenda/2023/10/la-inteligencia-artificial-transformara-la-toma-de-decisiones-asi-es-como/
3. Lanzagorta-Ortega D, Carrillo-Pérez DL, Carrillo-Esper R. Inteligencia artificial en medicina: presente y futuro. Gac Med Méx. 2021; 158(1):17-21. Disponible en: https://doi.org/10.24875/gmm.m22000688
4. Chui M, Manyika J, Miremadi M. Where machines could replace humans and where they can't (yet). McKinsey Quarterly. 2019:1-12. Disponible en: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet
5. Ferrante E. Inteligencia artificial y sesgos algorítmicos. Nueva Sociedad. 2021; (294):27-36. Disponible en: https://nuso.org/articulo/inteligencia-artificial-y-sesgos-algoritmicos/
6. Simon J, Wong PH, Rieder G. El sesgo algorítmico y el enfoque del diseño sensible al valor. Rev. Latinoam. Econ. Soc. Digit. 2022; (Número Especial 1):1-18. Disponible en: https://doi.org/10.53857/tzvn9229
7. Barocas S, Hardt M, Narayanan A. Fairness and Machine Learning: Limitations and Opportunities. 2019. Disponible en: fairmlbook.org
8. Asís RD. Una mirada a la robótica desde los derechos humanos. Dykinson; 2014.
9. Terrones AL. Inteligencia Artificial fiable y vulnerabilidad: una mirada ética sobre los sesgos algorítmicos. En Suárez-Álvarez R, Martín-Cárdaba MÁ, Fernández-Martínez LM. Vulnerabilidad digital: desafíos y amenazas de la Socieda hiperconectada. Madrid: Dykinson; 2023.
10. Diakopoulos N. Accountability in Algorithmic Decision Making. Commun ACM. 2016; 59(2):56-62. Disponible en: https://doi.org/10.1145/2844110
11. Velázquez H. Transhumanismo, libertad e identidad humana. Thémata. 2009; (41):577-590. Disponible en: https://revistascientificas.us.es/index.php/themata/article/view/594
12. Olarte C, Plegrín J, Reinares E. Implantes para aumentar las capacidades innatas: integrados vs apocalípticos. ¿Existe un nuevo mercado? Universia Bus Rev. 2015:86-101. Disponible en: https://www.redalyc.org/articulo.oa?id=43343050003
13. Abràmoff M, Tarver M, Loyo-Berrios N, Trujillo S, Char D, Obermeyer Z, Maisel, W. Considerations for addressing bias in artificial intelligence for health equity.npj Digital Med. 2013; 6(170):1-7. Disponible en: https://doi.org/10.1038/s41746-023-00913-9
14. Straw I, Rees G, Nachev P. Sex-Based Performance Disparities in Machine Learning Algorithms for Cardiac Disease Prediction: Exploratory Study. J Med Internet Res; 2024; 26:e46936. Disponible en: https://doi.org/10.2196/46936

15. Akter S, Dwivedi YK, Sajib S, Biswas K, Bandara RJ, Michael K. Algorithmic bias in machine learning-based marketing models. J. Bus. Res. 2022; 144:201-216. Disponible en: https://doi.org/10.1016/j.jbusres.2022.01.083
16. U.S. Department of Education, Office of Educational Technology. Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations. Washington; 2023.
17. Bostrom N. A history of transhumanist thought. Journal of Evolution and Technology. 2005; 14(1):1-25. Disponible en: https://nickbostrom.com/papers/a-history-of-transhumanist-thought/
18. Diéguez A. Transhumanismo: la búsqueda tecnológica del mejoramiento humano. Barcelona: Herder; 2017.
19. Harvard University. Justice with Michael Sandel - CCCB: Bioethics: Designer children. [video]; 2011. Disponible en: https://youtu.be/aFcfygkMM0I
20. Tobin H. Transhumanismo SuperPoderes Humanos – Castellano. [video]; 2016. Disponible en: https://www.youtube.com/watch?v=ClxzKdfQAkY
21. Fuentes, MÁ. Principios fundamentales de bioética. Roma: Instituto del Verbo Encarnado; 2006.
22. Berti B. Los principios de la bioética. Prudent Iuris. 2015; (79):269-280. Disponible en: https://erevistas.uca.edu.ar/index.php/PRUDENTIA/article/view/4030
23. Thompson J. (2017). Transhumanism: How Far Is Too Far? The new bioethics. 2017; 23(2):165–182. Disponible en: https://doi.org/10.1080/20502877.2017.1345092
24. Floridi L. The Logic of Information: A Theory of Philosophy as Conceptual Design. Oxford: Oxford University Press; 2019.
25. Beauchamp T, Childress, JF. Principles of biomedical ethics. Oxford: Oxford University Press; 2019.
26. Sgreccia E. Manual de bioética: fundamentos y ética biomédica. Navarra: EUNSA; 2016.
27. Latour B. Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. En Bijker W, Law J. Shaping Technology-Building Society. Studies in Sociotechnical Change. MIT Press, Cambridge Mass; 1992.
28. Future of Life Institute. Asilomar AI Principles. De Future of Life Institute; 2017. Disponible en: https://futureoflife.org/ai-principles/
29. Paglia VS. Rome Call for AI Ethic; 2023. Disponible en: https://www.romecall.org/wp-content/uploads/2022/03/RomeCall_Paper_web.pdf
30. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The Ethics of Algorithms: Mapping the Debate. Big Data Soc. 2016; 3(2). Disponible en: https://doi.org/10.1177/2053951716679679
31. Attiah M, Farah M. Minds, motherboards, and money: futurism and realism in the neuroethics of BCI technologies. Front Syst Neurosci. 2014; 86(8):1-3. https://doi.org/10.3389/fnsys.2014.00086
32. Evans J. Faith in Science in Global Perspective: Implications for Transhumanism. Public Underst Sci. 2014; 23(7):1-33. Disponible en: https://doi.org/10.1177/0963662514523712

33. Schaefer O, Savulescu J. Better Minds, Better Morals: A Procedural Guide to Better Judgement. J. Posthuman Stud. 2017; 1(1):26-43. Disponible en: http://dx.doi.org/10.5325/jpoststud.1.1.0026
34. Zehr P. Future think: cautiously optimistic about brain augmentation using tissue engineering and machine interface. Front Syst Neurosci. 2015; 9(72):1-5. Disponible en: https://doi.org/10.3389/fnsys.2015.00072
35. Huxley A. Un mundo feliz. México: Éxodo; 2011.
36. Doshi-Velez F, Narayanan M, Chen E, He J, Kim B, Gershman S. How do Humans Understand Explanations from Machine Learning Systems? An Evaluation of the Human-Interpretability of Explanation. New York: Cornell University; 2018. Disponible en: https://doi.org/10.48550/arXiv.1802.00682
37. Felzmann H, Fosch-Villaronga E, Lutz C, Tamò-Larrieux A. Towards Transparency by Design for Artificial Intelligence. Sci Eng Ethics. 2020; 26(6):3333–3361. Disponible en: https://doi.org/10.1007/s11948-020-00276-4
38. Kirat T, Tambou O, Do V, Tsoukiàs A. Fairness and explainability in automatic decision-making systems. A challenge for computer science and law. EURO J. Decis. Process. 2023; 11:1-19. Disponible en: https://doi.org/10.1016/j.ejdp.2023.100036
39. Gavilán I. Ideas y buenas prácticas para evitar el sesgo algorítmico. En: Ignacio G.R. Gavilán; 2022. Disponible en: https://ignaciogavilan.com/ideas-y-buenas-practicas-para-evitar-el-sesgo-algoritmico/
40. Sandu E. Sesgo en los algoritmos de inteligencia artificial: causas y soluciones. metaverso.pro - Metaverso para Profesionales; 2023. Disponible en: https://metaverso.pro/blog/sesgo-en-los-algoritmos-de-inteligencia-artificial-causas-y-soluciones/
41. UNESCO. Inteligencia artificial y educación: guía para las personas a cargo de formular políticas. Francia: UNESCO; 2021.
42. Comisión Europea. Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican diversos actos legislativos de la Unión. Bruselas; 2024.
43. Gebru T, Morgenstern J, Vecchione B, Wortman J, Wallach H, Daume HI, Crawford K. Datasheets for Datasets. Commun ACM. 2021; 64(12):86-92. Disponible en: https://doi.org/10.1145/3458723