ISSN impreso 1996-6660 / ISSN online 2519-5719
DOI: https://doi.org/10.31381/perfilesingenieria. v20i21.7056

Implicancias de la equidad algorítmica en la Inteligencia artificial

Implications of algorithmic fairness in artificial intelligence

Augusto Cortez Vásquez, Maria Manyari Monteza, Gilberto Salinas Azaña, Jorge Chávez Soto

RECIBIDO: 01 de noviembre de 2024.
ACEPTADO: 15 de diciembre de 2024.

Resumen

El uso de algoritmos de Inteligencia artificial no se limita como a veces se supone a procedimientos efectivos, el uso de este vocablo plantea varias concepciones, interpretaciones y problemas. Con el fin de no distraernos en este laberinto lingüístico, hemos tomado posición por una caracterización muy común en sentido psicológico que consiste en concebirla como una capacidad poseída por ciertos organismos/mecanismos para adaptarse a situaciones nuevas utilizando para tal efecto el conocimiento adquirido en el curso de anteriores procesos de adaptación. La irrupción de la inteligencia artificial (IA) se integra cada vez más en la sociedad y generalmente se utiliza para la toma de decisiones oportunas que afectan a la sociedad y por ende a las personas, en diferentes ámbitos. En el desarrollo de algoritmos de la IA, pueden ocurrir errores sistémicos y repetibles en un sistema informático que crean resultados injustos, como privilegiar a un grupo arbitrario de usuarios frente a otros. Estos algoritmos denominados sesgados, se caracterizan generalmente por la existencia de prejuicios o distorsiones en los datos de entrenamiento. La comunidad científica e instituciones gubernamentales han lanzado propuestas para combatir estos riesgos que buscan reducir su impacto negativo en la sociedad. Es imperativo resolver aspectos que van contra la Ética, la justicia, la Transparencia y la Equidad de los datos, algoritmos y sus predicciones. El presente trabajo pretende sensibilizar que desarrollo de algoritmos para tomar decisiones deben cumplir tres requisitos: en primer lugar, garantizar el equilibrio entre el conjunto de datos utilizados y la programación del algoritmo con equidad que evite la discriminación y el sesgo, segundo, garantizar condiciones de transparencia en los resultados, es decir, el resultado obtenido debe ser explicable a cualquier usuario de forma clara y sencilla. No debe soslayarse la regulación de requisitos para el desarrollo y uso de la IA, debe alinearse a la no afectación de los derechos humanos fundamentales

Palabras claves: Algorítmica, equidad algorítmica, inteligencia artificial, equidad de datos, sesgos algorítmicos

Abstract

The use of Artificial Intelligence algorithms is not limited, as is sometimes assumed, to effective procedures; the use of this vocabulary raises several conceptions, interpretations and problems. In order not to get distracted in this linguistic labyrinth, we have taken a position on a very common characterization in the psychological sense that consists of conceiving it as a capacity possessed by certain organisms/mechanisms to adapt to new situations using for this purpose the knowledge acquired in the course. . from previous adaptation processes. The emergence of artificial intelligence (AI) is increasingly integrated into society and is generally used to make timely decisions that affect society and therefore people, in different areas. In the development of AI algorithms, systemic and repeatable errors can occur in a computer system that create unfair results, such as privileging an arbitrary group of users over others. These so-called biased algorithms are generally characterized by the existence of biases or distortions in the training data. The scientific community and government institutions have launched proposals to combat these risks that seek to reduce their negative impact on society. It is imperative to resolve aspects that go against Ethics, justice, Transparency and Equity of data, algorithms and their predictions. The present work aims to raise awareness that the development of algorithms to make decisions must meet three requirements: first, guarantee the balance between the set of data used and the programming of the algorithm with fairness that avoids discrimination and bias, second, guarantee conditions of transparency in the results, that is, the result obtained must be explainable to any user in a clear and simple way. The regulation of requirements for the development and use of AI should not be ignored, it must be aligned with the non-affecting of fundamental human rights

Keywords: Algorithmic, algorithmic fairness, artificial intelligence, data fairness, algorithmic biases

Problemática

La IA presenta un panorama que incluye un conjunto de problemas que son inherentes a su construcción, comportamiento y resultados. A esta problemática puede añadirse, que, los algoritmos lejos de eliminar las desigualdades, está reproduciendo, sistematizando y amplificando los sesgos. Esto se debe en gran medida a un desequilibrio entre los datos utilizados y la programación del algoritmo con atributos de sesgo y discriminación, al mismo tiempo se soslayan condiciones de transparencia, en la que no se explica a los usuarios los resultados obtenidos ni la forma como se obtuvieron. La profesora titular de **Esade Anna** Ginès i Fabrellas[1] , afirma que los problemas presentados al inicio de la IA no han desaparecido, por el contrario (Frabellas, 2023), los equipos encargados de desarrollar esta tecnología es responsable de decisiones discriminatorias, en la que se magnifican sesgos y estereotipos de género,

1 Anna Gines: Es directora del Instituto de Estudios Laborales y profesora titular del Departamento de Derecho en Esade. Se graduó en Derecho y Economía en la Universidad Pompeu Fabra, Sus áreas de investigación incluyen el mercado laboral y las plataformas digitales, las nuevas tecnologías y el futuro del trabajo, la salud y la seguridad, así como el derecho y la economía raza, orientación sexual y discapacidad a un ritmo alarmante que afectan a millones de vidas.

La CEPAL consciente de esta problemática y siendo uno de sus roles la promoción de un crecimiento económico equitativo de largo plazo y la generación y asignación eficiente de recursos financieros para apoyar el desarrollo y la igualdad en los países de América Latina y el Caribe, considera que se requiere redefinir los patrones de desarrollo de la región teniendo como columna vertebral, la equidad, vale decir, la reducción de la desigualdad social en sus diferentes formas, siendo este el instrumento para medir la calidad del desarrollo (CEPAL, 2024). Consciente de ello, la CEPAL lanzo en un foro latinoamericano el primer Índice Latinoamericano de Inteligencia Artificial (ILIA) en la CEPAL, en la que se señalaba que la IA puede contribuir de gran manera a la transformación de los modelos de desarrollo en América Latina y el Caribe haciéndolos mas más productivos, inclusivos y sostenibles, no obstantes, asisten también riesgos por lo que se hace imperativo provechar sus oportunidades y minimizar sus potenciales amenazas. La Unión Europea (EUROPEA, 2023), no se puede soslayar a esta problemática, señala que más del 30% de sus grandes empresas ya utilizan la IA, teniendo como horizonte la excelencia y la confianza, basta observar en nuestro alrededor cómo la IA se ha integrado en diversos aspectos de la vida, desde la educación, salud y comercio hasta la ciberseguridad. No obstante, la IA soslaya en sus algoritmos: los sesgos, y se plantea uno de los mayores retos**,** encontrar mecanismos para mitigarlos y alcanzar la equidad algorítmica. Esto no es un mero convencionalismo, sino una necesidad para construir modelos que impulsen la investigación, pero con unos criterios de respeto a los derechos humanos fundamentales.

Cundo se habla de equidad algorítmica se refiere claramente al diseño de soluciones basadas en IA incluyendo aquellos de aprendizaje automático (ML) sustentados en la justicia, transparencia, el respeto, la no discriminación y oportunidad equitativa, explica Marco Creatura, científico de datos en BBVA AI Factory:

ISSN (Digital): 2519-5719

> *"La preocupación principal es que estos sistemas no reproduzcan, refuercen ni amplifiquen los sesgos sociales existentes. La equidad en este contexto se define como la ausencia de sesgos, que en el campo de la toma de decisiones se define como cualquier prejuicio o favoritismo hacia un individuo o grupo basado en sus características inherentes o adquiridas"*.

La discriminación algorítmica, también denominada como sesgo algorítmico, hace referencia al hecho de que los algoritmos de IA y ML involucran un conjunto de prejuicios y sesgos humanos existentes en los datos utilizados en su entrenamiento. Estos prejuicios afectan una amplia gama de derechos humanos incluyendo el derecho a comunicar o recibir información libremente (Grigore, 2022)Este problema plantea serias preocupaciones porque puede conducir a la exclusión social, la marginalización y la perpetuación de estereotipos negativos. La complejidad de los algoritmos su opacidad, pueden dificultar la rendición de cuentas y la apelación de decisiones injustas. Es un requerimiento imperativo para conseguir la equidad implementar prácticas de desarrollo de algoritmos éticos, como la transparencia en el diseño y la recopilación de datos, la auditoría de algoritmos y la diversidad en los equipos de desarrollo. Estos mecanismos deben ir acompañados de una regulación efectiva que garantice la protección de los derechos humanos en el contexto de la discriminación algorítmica (Jose, 2023)

Los algoritmos tienen una dimensión política en la medida en que intervienen en el orden social y estructuran nuestras decisiones. **"La injusticia**, en cualquier parte, es una amenaza a la justicia en todas partes", señalaba Martin Luther King referente a los derechos civiles en los 70, NIKKEN[2], personaje histórico señalaba que la noción de derechos humanos se alinea con la afirmación de la dignidad de la persona frente al Estado. En ese sentido, el poder público se debe ejercer como un servicio dirigido a conseguir bienestar del ser humano, por lo tanto, no se puede ni debe ser empleado lícitamente como instrumento para ofender cualidades inherentes al ser humano, por el contrario debe ser un vehículo que permita vivir en sociedad en consonancia con la misma dignidad..

Paradójicamente, siendo de vital importancia sus usos y decisiones, por el impacto directo que tiene en la vida de las personas y en sus derechos, los organismos que los utilizan desarrollan e implementan sistemas de algoritmos con bajos niveles de transparencia, conocimiento público, y medidas de supervisión o responsabilidad (Innerarаty, 2023).

El sesgo algorítmico ocurre, cuando en un sistema se soslaya el equilibrio en los datos y la construcción de los algoritmos con criterios no transparentes y subjetivos: «*Sistemas y predicciones que benefician sistemáticamente a un grupo de individuos frente a otro, resultando así injustas o desiguales* «según señala (Seminario de Integracion II, 2024)…..

Estos algoritmos aprenden mediante distintos paradigmas de aprendizaje. Siendo uno de los más utilizados el del aprendizaje supervisado, en la que se someten a proceso de entrenamiento guiado (supervisado) por anotaciones o etiquetas, *buscando alinear* características o patrones propios de los datos con las correspondientes etiquetas. Se analizan los datos con la intención de encontrar patrones distintivos que permitan diferenciar las categorías. El propósito de este método es que, a través del entrenamiento,

los sistemas aprendan a encontrar patrones en estas características y asociarlas con las correspondientes categorías. Probablemente en un inicio las asociaciones encontradas serán incorrectas, no obstante, durante el proceso de entrenamiento, el modelo se irá ajustando y mejorando su desempeño.

Para resolver estos problemas se debe considerar dos aspectos: en primer lugar, debe existir un equilibrio entre el conjunto de datos utilizados y la programación del algoritmo que evite la discriminación y el sesgo, maximizando el nivel de equidad. En segundo lugar, el algoritmo utilizado debe cumplir condiciones de transparencia, que permita explicar a cualquier usuario de forma clara y sencilla. (ASLAM, 20213)

2 Pedro NIKKEN: anterior presidente del Instituto Interamericano de derechos humanos (HDH), juez y expresidente de la Corte Interamericana de derechos Humanos Implicancias de la falta de equidad algorítmica en la Inteligencia artificial

a) Sesgos en sistemas de reconocimiento facial

Los sistemas de reconocimiento facial presentan sesgos algorítmicos, se han presentado casos para reconocer adecuadamente a personas de diferentes razas y etnias, sobre todo en aquellos con tonos de piel más oscuros. La razón más común es que los algoritmos utilizados en estos sistemas suelen basarse en conjuntos de datos desequilibrados y poco representativos, derivándose de ello en un trato injusto y discriminación hacia ciertos grupos. IBM, presento un caso en la que hubo muchas críticas su sistema de reconocimiento facial mostró un sesgo racial al tener dificultades para reconocer adecuadamente a personas de piel más oscura. (https://hub.laboratoria.la/10-casos-donde-la-inteligencia-artificial-jugo-en-contra-de-la-diversidad, 2023)

b) **Discriminación laboral:** Cuando se utilizan sistemas para selección de personal, estos pueden basarse en algoritmos con tendencias a discriminación en el análisis de sus datos, pueden presentar sesgos favoreciendo ciertos perfiles, excluyendo que a personas de diferentes orígenes o con trayectorias no convencionales. Esto hace que se refuercen las desigualdades, se favorezcan preferencias y se dificulte la diversidad en los centros laborales.
c) **Sesgos en la publicidad en línea:** Muchas empresas utilizan soluciones que usan algoritmos que favorecen estereotipos de género y raza, excluyendo a ciertos grupos de oportunidades comerciales, laborales o educativas, lo que refuerza las desigualdades y limita las opciones disponibles. Los anuncios o contenido generados contribuyen a la segregación ocupacional y limitan la representación en diversas áreas, restringiendo así las oportunidades de ciertos grupos.
d) **Asistentes de voz y estereotipos de género:** Los asistentes de voz más populares, como Siri, Alexa y Google Assistant, a menudo tienen voces femeninas y están diseñados para responder con un tono servicial, lo que puede reforzar los estereotipos de género y perpetuar la desigualdad de género.
e) **Sesgos en los algoritmos de recomendación**: Las redes sociales y plataformas en línea utilizan algoritmos para personalizar la experiencia del usuario, pero es importante tener en cuenta que estos algoritmos pueden limitar la exposición a diferentes perspectivas y crear burbujas de filtro que reafirman los prejuicios y estereotipos existentes.

ISSN (Digital): 2519-5719

f) **Discriminación algorítmica en la concesión de créditos:** Los algoritmos que evalúan el crédito pueden orientarse hacia la exclusión financiera de ciertos grupos, impidiendo su acceso a préstamos y oportunidades económicas.
g) **Chatbots y atención a la cliente automatizada:** Los chatbots y sistemas de atención al cliente automatizados pueden perpetuar la discriminación si no se considera la diversidad y las necesidades de diferentes grupos de usuarios. Es posible que estos sistemas no tengan la capacidad de comprender acentos, empatía cultural o lidiar con situaciones que requieran sensibilidad cultural, lo que puede generar experiencias negativas para algunos usuarios.
h) **Tecnologías biométricas y privacidad:** Las tecnologías biométricas, como el reconocimiento facial y de huellas dactilares, pueden ser utilizadas de manera inadecuada, lo que plantea preocupaciones en cuanto a la privacidad y la discriminación. Si no se considera la diversidad y las necesidades de diferentes grupos de usuarios, estas tecnologías pueden llevar a la vigilancia masiva y la recolección de datos sensibles, lo que podría tener impactos negativos en comunidades minoritarias y grupos vulnerables.
i) **Aplicaciones de citas y sesgos raciales**: Las apps de citas en línea pueden ser propensas a sesgos raciales y reforzar estereotipos. Algunas de estas apps han sido criticadas por la discriminación racial presente en la elección de parejas.
j) **Automatización y pérdida de empleo**: La automatización puede ser una herramienta efectiva para aumentar la eficiencia y la productividad, pero también puede tener efectos negativos en la diversidad y la igualdad laboral. Según algunos estudios, los trabajadores de bajos ingresos y los mayores podrían ser los más afectados por la pérdida de empleos debido a la automatización, lo que podría empeorar las desigualdades existentes.

Es muy importante tomar en cuenta la diversidad en todos los aspectos, y especialmente en el desarrollo de tecnología como la IA, que puede reflejar y amplificar sesgos y prejuicios presentes en los datos utilizados para entrenarla. Por eso, es esencial que los equipos que crean tecnología estén formados por colaboradores con diferentes pensamientos, identidades y formaciones, para aportar diversas perspectivas al producto o servicio. De esta manera, podremos aprovechar todo el potencial de la tecnología para construir un futuro más justo y diverso.

La justicia algorítmica en modelos de aprendizaje se refiere a la aplicación de principios éticos y de equidad en el desarrollo y uso de algoritmos de aprendizaje automático. Estos algoritmos son utilizados para hacer predicciones o tomar decisiones con base en los datos, y pueden tener un impacto significativo en la vida de las personas, por ejemplo, en el ámbito de la selección de candidatos para empleos, la aprobación de préstamos, la evaluación del riesgo de incidir en delitos, entre otros. La justicia algorítmica busca evitar la discriminación y el sesgo en la toma de decisiones basadas en los datos. Sin embargo, existen múltiples definiciones y perspectivas sobre lo que significa la justicia algorítmica y cómo se puede lograr. Estas definiciones pueden ser tan heterogéneas que puede ser imposible obtener dos o más de éstas de forma simultánea en una misma estimación (Bernal, 2023).

Además, **los algoritmos de IA generativa presentan una complejidad mayor que los clásicos** de machine learning, cuya salida es generalmente una puntuación o una probabilidad. "Los grandes modelos de lenguaje se entrenan con inmensas cantidades de

datos en formato texto generalmente extraídos de internet, que no necesariamente han sido curados y que pueden contener estereotipos, representaciones sesgadas de la sociedad, lenguaje excluyente, denigrante o despectivo con respecto a ciertos grupos sociales y grupos vulnerables. La complejidad radica en que el lenguaje es en sí mismo una tecnología que captura normas sociales y culturales", subraya **Clara Higuera**, científica de datos en BBVA AI Factory. La detección de sesgos en IA generativa es un campo nuevo, todavía en fase de exploración, que ya se está utilizando aspectos como la construcción de guardarraíles o herramientas para detectar esos mismos sesgos (BBVA, 2024).

De hecho, según un estudio de la UNESCO, los modelos de lenguaje empleados por la IA generativa, pueden reproducir prejuicios de género, raciales y homófobos que colaboran a la desinformación.

Cómo se producen los sesgos, los grandes obstáculos de la equidad algorítmica

Los sesgos son diversos y pueden manifestarse en diferentes etapas, tal y como se apunta en un artículo de MIT Technology Review:

1. **La definición del problema**. Los desarrolladores comienzan estableciendo el objetivo del algoritmo que están desarrollando. Esto implica convertir en métricas aspectos tan difusos como la 'eficacia', un concepto difuso y subjetivo abierto a interpretaciones que no siempre son imparciales. Por ejemplo, si el algoritmo de una plataforma de contenidos de 'streaming' busca maximizar el tiempo de visualización de los espectadores, esto podría derivar en recomendaciones que refuercen sus intereses previos en lugar de diversificar sus experiencias con otros contenidos.

2. **Durante la recogida de datos**. Existen dos posibles razones para este fenómeno: o los datos recopilados no son representativos de la realidad o reflejan prejuicios ya existentes. Por ejemplo, si un algoritmo recibe más fotos de caras de piel clara que de piel oscura, el reconocimiento facial será menos preciso en el segundo caso. Otro ejemplo es lo que ha sucedido con algunas herramientas de contratación, que, por ejemplo, descartaban a las mujeres para puestos técnicos debido a que el algoritmo se había entrenado con decisiones de contratación históricamente sesgadas.

3. **En la preparación de los datos**. A menudo se seleccionan y preparan los atributos que el algoritmo utilizará para tomar decisiones, como la edad o el historial de sus acciones. Estos atributos pueden introducir sesgos en herramientas de evaluación, ya sean socioeconómicos o relacionados con el género. Estos sesgos también se pueden introducir cuando se etiquetan los datos que posteriormente utilizarán los algoritmos, especialmente cuando se llevan a cabo tareas de anotación de datos. Esto se debe a que los anotadores pueden traer sus propios sesgos; de ahí la importancia de escribir guías claras de anotación.

ISSN (Digital): 2519-5719

Qué tipos de sesgos debe combatir la equidad algorítmica

Los sesgos que pueden presentar los algoritmos pueden manifestarse en diversas formas y tipos. Para desarrollar sistemas de IA, debe tenerse en consideración aspectos esenciales como mitigar, con el propósito de que el tratamiento tanto en su generación como en su uso IA que sean justos y beneficiosos para todos los usuarios, asegurando decisiones equitativas y aumentando la confianza en las tecnologías emergentes", afirman Marco Creatura y Clara Higuera. Algunos de ellos, como señala google, son:

a) **Sesgo de selección**. Se manifiesta cuando el conjunto de datos de entrenamiento contiene instancias(ejemplos) con poca representatividad en el mundo real. No se puede entrenar a un algoritmo con todo el universo de los datos, por tanto se debe elegir una muestra sin soslayar el contexto. Una muestra no representativa del conjunto, o una desequilibrada hacia un colectivo, remitirá a resultados igualmente sesgados.

b) **Sesgo de automatización**. Se manifiesta cuando se cree ciegamente todo lo que resulta de los sistemas automatizados sin considerar la tasa de error que pueda existir, más aún cuando se toman decisiones en tiempos muy cortos, soslayando un análisis exhaustivo, dando por valida información que no ha sido formalmente contrastada. Se resalta el hecho de depender en exceso de los sistemas automatizados, incluso cuando estos cometen errores. "Cuando la gente tiene que tomar decisiones en plazos relativamente cortos, con poca información... ahí es cuando tiende a confiar en cualquier consejo que les proporcione un algoritmo", indica **Ryan Kennedy**[3], profesor de la Universidad de Houston especialista en automatización, en un artículo de investigación.

c) **Sesgo de correspondencia**. Se manifiesta cuando los algoritmos evalúan aspectos/criterios generalizándolas solo su pertenencia a un grupo en particular, soslayando sus características individuales. Por ejemplo, cuando se asume que todas las personas que han asistido a la misma universidad tienen el mismo nivel de conocimiento para un trabajo, en este caso se está calificando a la universidad y no a la persona.

d) **Sesgo implícito**. Se manifiesta cuando los desarrolladores de los algoritmos asumen posiciones subjetivas sobre situaciones y experiencias personales que no se aplican en forma general. Este sesgo podría afectar su modelización y entrenamiento, ya que podrían introducirse soterradamente o inconscientemente sus propios prejuicios.

3 Ryan Kennedy es profesor asociado de ciencias políticas en la Universidad de Houston, director fundador del Centro de Estudios Internacionales y Comparados (CICS) de la Universidad de Houston e investigador asociado del Hobby Center for Public Policy de la Universidad de Houston

Iniciativas y propuestas para alcanzar la equidad algorítmica

Frente a los diversos tipos de sesgos que se dan en desarrollo de los algoritmos de IA así como en el uso de las aplicaciones que usan estos algoritmos, se propician también iniciativas y normativas para promover la práctica de la equidad algorítmica y mitigar estas situaciones de injusticia. "Los gobiernos y organizaciones han comenzado a implementar directrices y normativas para asegurar que las tecnologías de IA sean justas y responsables. Esto incluye la creación de marcos éticos y legislación específica sobre el uso de IA para proteger contra la discriminación", afirma Marco Creatura, en referencia al Reglamento de Inteligencia Artificial (IA Act) de la Unión Europea.

De hecho, la Unión Europea también cuenta con un proyecto(Marco jurídico Europeo) que incluye propuestas para que los desarrolladores, implementadores y usuarios puedan intervenir solo en caso de que las leyes no cubran (Reglamento General de Protección de Datos (RGPD**)** y con diferentes requisitos de transparencia. Con lo cual, garantizan que disponen de las mejores prácticas de confianza y seguridad, asegurardo que la IA funciona de manera inclusiva.

En España, **Marco Creatura**[4] señala, que la Agencia Española de Protección de Datos (AEPD) ha difundido algunas normas con el propósito de auditar sistemas de IA con un apartado específico para el control del sesgo en las fuentes de datos utilizadas. Señala: "Las investigaciones actuales se centran en métodos para corregir sesgos en los datos y en los modelos. Esto incluye técnicas de recolección de datos más representativos, ajustes en los algoritmos y modificaciones en el posprocesamiento de los resultados para asegurar decisiones más justas".

Se realizan permanentemente métricas **de equidad** para de evaluar modelos de 'machine learning' con el propósito de garantizar que los modelos de IA sean más transparentes y multidisciplinares. Aunque es dificultoso por ejemplo cuando se define un problema convertir en métricas conceptos tan difusos y subjetivos como la eficacia que da la posibilidad de interpretaciones subjetivas que podrían ser imparciales.

Para conseguir equidad en los algoritmos no resulta suficiente involucrar a científicos de datos y desarrolladores, es también menester, incluir a expertos en ética, sociólogos y representantes de los grupos vulnerados. Vasta revisar los proyectos como la Liga de la Justicia Algorítmica, fundada por **Joy Buolamwini**[5], quien se dedica analizar y a divulgar las diferentes formas de discriminación. Señala que aunque las nuevas herramientas son prometedoras, es imperativo promover el desarrollo de una IA equitativa y responsable.

4 **Marco** es licenciado en Ingeniería Electrónica y tiene un máster en Bioingeniería. Ha trabajado en diferentes proyectos europeos centrando su investigación en escenarios de interfaz cerebro-ordenador (BCI) no invasivos. Ha aplicado la teoría de la probabilidad y los algoritmos de aprendizaje automático diseñando e implementando soluciones centradas en el cliente dentro de las industrias de la publicidad online y la banca

5 **Joy Buolamwini**, investigadora del Instituto de Tecnología de Massachusetts (MIT) y activista pionera en IA, conspicua revisora de las implicancias de la falta de equidad algorítmica,

ISSN (Digital): 2519-5719

Las empresas y los desarrolladores de algoritmos de IA generativa en su propósito de fomentar una IA sin prejuicios y responsable, vienen usando técnicas como los guardarraíles, como un elemento de seguridad: "Son directrices y herramientas para supervisar, controlar y guiar el comportamiento de los modelos de IA generativa. Pueden ser desde una simple instrucción en el 'prompt' (ejemplo: contesta educadamente y respetuosamente sin insultar a nadie) hasta la detección de una respuesta formulada en lenguaje inapropiado".

La equidad algorítmica sen ha constituido en un desafío crucial, la comunidad científica e instituciones gubernamentales vienen poniendo énfasis y concienciación en todos los niveles comprometidos en el ciclo de desarrollo de la IA, desde 'seniors leaderships' hasta 'data scientists', con el propósito de desarrollar sistemas realmente justos y transparentes. Como subraya Clara Higuera[6] (Higuera, 2024), "se debe seguir avanzando en investigación, regulación y colaboración para garantizar que los algoritmos tengan en cuenta a todas las personas de manera igualitaria ayudará a que se libren de prejuicios".

6 Higuera: Lead data scientist en BBVA AI Factory. Clara Higuera es Doctora en Inteligencia Artificial por la Universidad Complutense de Madrid. En sus más de 13años de experiencia profesional ha trabajado tanto en la academia como en industria aplicando IA en diferentes sectores como bioinformática y biomedicina, en la BBC en Londres y en los últimos años ha liderado equipos de científicos de datos en BBVA AI Factory.

Conclusión

La IA enfrenta serios desafíos, como crear explicaciones que sean completas e interpretables, aunque es difícil lograr la interpretabilidad y la integridad al mismo tiempo. Sin embargo, se permitirá solucionar la aplicabilidad y la integridad de cualquier decisión cuando no se consideren como cajas negras y la decisión venga acompañada con una argumentación y justificación de esta que incluirá la ausencia de discriminación y sesgo, sentándose las bases de la certificación ética

Aunque parezcan posiciones antagónicas: impulsar la investigación y la industria basada en la IA, y tomar conciencia de los enormes riesgos y desafíos que conlleva. Se pretende, y en eso están de acuerdo diversas personalidades investigadoras de la IA, también políticos, en conseguir un acuerdo equilibrado entre estos dos aspectos importantes. Buscando propiciar un lugar donde la IA prospera desde el laboratorio hacia el mercado, con criterios inclusivos, de un beneficio plural, con respeto a los derechos humanos fundamentales.

Los requisitos legales que se establezcan para el uso de la tecnología de IA deben estar acordes a los riesgos de afectación de los derechos humanos fundamentales, a mayor riesgo, mayores condiciones de legalidad.

Para mitigar los efectos de falta de equidad algorítmica minimizando así los sesgos inconscientes en la fase de desarrollo**,** se recomienda formar equipos de desarrollo inclusivos, de conformación diversa en cuanto a género, etnia y procedencia con el propósito de aportar perspectivas variadas**.**

Referencias

ASLAM. (20213). *Solución para medir la ética, equidad e integridad de los algoritmos de IA.* Obtenido de https://aslan.es/solucion-para-medir-la-etica-equidad-e-integridad-de-los-algoritmos-de-ia/

BBVA. (2024). *Equidad algorítmica, clave para crear una inteligencia artificial responsable.* Obtenido de https://www.bbva.com/es/innovacion/accesibilidad-neuromarketing-e-inteligencia-artificial-startups-que-ayudan-a-mejorar-la-experiencia-de-usuario/

Bernal, C. (2023). *Justicia algoritmica y sus limitaciones.* Obtenido de https://quantil.co/es/blog/justicia-algoritmica-y-sus-limitaciones-un-teorema-de-imposiblidad/

CEPAL. (2024). *https://repositorio.cepal.org/server/api/core/bitstreams/a5fcd682-bdec-4b63-9621-693d36c497f8/content.* Obtenido de https://repositorio.cepal.org/server/api/core/bitstreams/a5fcd682-bdec-4b63-9621-693d36c497f8/content

EUROPEA, U. (2023). *https://digital-strategy.ec.europa.eu/es/policies/european-approach-artificial-intelligence#:~:text=El%20enfoque%20de%20la%20UE,seguridad%20y%20los%20derechos%20fundamentales.* Obtenido de https://digital-strategy.ec.europa.eu/es/policies/europeanapproachartificialintelligence#:~:text=El%20enfoque%20de%20la%20UE,seguridad%20y%20los%20derechos%20fundamentales.

Frabellas, A. (2023). *Han desaparecido realmente los problemas de discriminación y sesgo que plagaban la IA temprana?* Obtenido de https://dobetter.esade.edu/es/algoritmos-desigualdad

Grigore, A. (2022). Derechos humanos e inteligencia artificial. Obtenido de https://revistascientificas.us.es/index.php/ies/article/view/19991/18602

Higuera, C. (2024). IA responsable, retos, avances y oportunidades.

https://hub.laboratoria.la/10-casos-donde-la-inteligencia-artificial-jugo-en-contra-de-la-diversidad. (2023).

Inneraraty, D. (2023). *Women evolution.* Obtenido de https://womenevolution.es/igualdad-algoritmica/

Jose, I. (2023). La discriminación algorítmica y su impacto en la dignidad de la persona y los derechos humanos. *12*. Obtenido de https://djhr.revistas.deusto.es/article/view/2910

Montesinos, A. (2023). INTELIGENCIA ARTIFICIAL EN LA JUSTICIA. Obtenido de https://revista-aji.com/wp-content/uploads/2024/07/AJI21_Art20.pdf

ISSN (Digital): 2519-5719

(2024). *Seminario de Integracion II.* Obtenido de https://seminarioiiuntref.wordpress.com/2022/04/12/hay-equidad-algoritmica/

Trayectoria académica

Augusto Parcemón Cortez Vásquez

Universidad Nacional Mayor de San Marcos, Lima, Perú.

Maestría en Ciencias de la computación en la Facultad de Ciencias Matemáticas, Estudios culminados de doctorado en filosofía en la Facultad de letras y ciencias humanas en la UNMSM. Segunda especialidad en Psicopedagogía en la Universidad Ricardo Palma.

Autora corresponsal: acortezv@unmsm.edu.pe

Orcid: https://orcid.org/0000-0002-5188-7962

Maria Manyari Monteza

Universidad Nacional Mayor de San Marcos, Lima, Perú.

Docente Universitario en Universidad Nacional Mayor de San Marcos.

mmanyarium@unmsm.edu.pe

Orcid:

Gilberto Salinas Azaña

Universidad Nacional Mayor de San Marcos, Lima, Perú.

Docente Universitario en Universidad Nacional Mayor de San Marcos.

gsalinasa@unmsm.edu.pe

Orcid: https://orcid.org/0000-0003-3591-1202

Jorge Luis Chávez Soto

Universidad Nacional Mayor de San Marcos, Lima, Perú.

Docente Universitario en Universidad Nacional Mayor de San Marcos.

jchavezs@unmsm.edu.pe

Orcid: https://orcid.org/0000-0002-9408-9266

Contribución de autoría

Augusto Parcemón Cortez Vásquez

Coordinación trabajos de campo y ensayos de laboratorio, análisis y discusión de resultados, redacción de artículo.

Maria Manyari Monteza

Coordinación y acompañamiento en los trabajos de campo, discusión en los ensayos de laboratorio e interpretación de resultados.

Gilberto Salinas Azaña

Coordinación de trabajos de campo, participación en los ensayos de campo y de laboratorio, procesamiento de información e registros de ensayos por muestras evaluadas.

Jorge Luis Chávez Soto

Coordinación de trabajos de campo, participación en los ensayos de campo y de laboratorio, procesamiento de información e registros de ensayos por muestras evaluadas.

Conflicto de intereses

Los autores declaran que no existen conflictos de intereses en el desarrollo de la presente investigación.

Responsabilidad ética y legal

El desarrollo de la investigación se realizó bajo la conformidad de los principios éticos del conocimiento, respetando la originalidad de la información y su autenticidad.

ISSN (Digital): 2519-5719

Declaración sobre el uso de LLM (Large Language Model)

Este artículo no ha utilizado para su redacción textos provenientes de LLM (ChatGPT u otros)

Financiamiento

La presente investigación ha sido realizada con el financiamiento con los recursos propios de los autores.

Correspondencia: acortezv@unmsm.edu.pe