Problemática antropológica detrás de la discriminación generada a partir de los algoritmos de la inteligencia artificial

Anthropological problem behind the discrimination generated from artificial intelligence algorithms

Gabriela Morales Ramírez*
Universidad Panamericana, México

https://doi.org/10.36105/mye.2023v34n2.04

Resumen

Actualmente la inteligencia artificial se encuentra en un punto de desarrollo nunca visto prometiendo grandes beneficios que trascienden en las distintas esferas sociales. Una problemática al respecto es la aparente neutralidad de los algoritmos utilizados en su programación y su impacto a gran escala en relación con la discriminación generada a partir de los sesgos inmersos en ellos, provenientes de sus diseñadores. Esto como resultado de una mirada parcial a la realidad y la persona misma. La solución a la segregación es posible hallarla no solo en las llamadas paridades, que son una respuesta que pretende

* Licenciada en Filosofía por la Universidad Panamericana. Maestra en Filosofía de la Ciencia en el área de Estudios Filosóficos y Sociales de la Ciencia y la Tecnología. Correo electrónico: gmoralesr@up.edu.mx https://orcid.org/0000-0003-1297-2977
Recepción: 22/11/22 Aceptación: 10/01/2023

compensar errores en la programación y trae como consecuencia desigualdades en oportunidades y privilegios para ciertos grupos, sino en una mirada a la totalidad de la persona.

Palabras clave: sesgos, pensamiento automático, equidad algorítmica, neutralidad.

1. Introducción

El desarrollo y la aplicación de la inteligencia artificial (IA) puede representar un progreso considerable para la ciencia y la tecnología, pero también aparecer como una amenaza para las personas y su existencia en el planeta.

El uso de la IA hasta hace algunos años era un tema propio de los libros de ciencia ficción en historias que parecían demasiado lejanas e ilusorias para el tiempo en el que fueron escritas como *Yo, robot* de Isaac Asimov o *¿Sueñan los androides con ovejas eléctricas?* de Philiph K. Dick o, incluso, películas de culto como *Matrix* de las hermanas Wachowski. Hoy en día, los avances tecnológicos y los problemas éticos planteados en estos relatos nos han alcanzado e incluso traspasado.

Existe una gran ignorancia de la población en general con respecto a qué es propiamente la IA, cuáles son sus usos, cuáles son las consecuencias que podría traer consigo su desarrollo, por qué es moralmente correcto, o no, destinar tanto presupuesto a su elaboración, etcétera.

A la par, hay un gran desinterés por parte de los gobiernos y grandes compañías dedicadas a la IA por hacer evaluaciones de las posibles repercusiones éticas, políticas, sociales, ambientales y económicas de sus acciones dedicadas a ella, así como una falta de interés por hacer una declaración transparente de sus efectos en la vida diaria y futura de la sociedad.

Ante esta situación, es indispensable que se aborden estos temas a la luz de distintas disciplinas, pero sobre todo, que se consideren

https://doi.org/10.36105/mye.2023v34n2.04

sus implicaciones antropológicas. Una de ellas atiende a los llamados sesgos algorítmicos que se traducen como errores de tipo estadístico, estructural, cognitivo y social que traen consigo desventajas que son éticamente objetables pues dan lugar a resultados discriminatorios o bien producen beneficios sistemáticamente a un grupo de individuos frente a otros (11).

Por ello, en el presente artículo se trabajará sobre la hipótesis siguiente: si los algoritmos de la IA contienen sesgos cognitivos provenientes de su diseñador, entonces se perpetúan modelos discriminatorios que se centran sólo en aspectos accidentales de la persona.

Para lograr demostrar esta hipótesis, se partirá de algunas consideraciones sobre la persona y su dignidad. Después se atenderá a los sesgos cognitivos y su vínculo con la discriminación hacia las personas. Más adelante se argumentará cómo éstos son trasladados al campo de la ciencia y la tecnología, especialmente al ámbito de la IA a partir del entrenamiento de algoritmos. Se mostrarán algunos ejemplos de sus consecuencias y se planteará una base antropológica que permita esbozar una vía de solución.

2. Algunas consideraciones iniciales en torno a la persona y su dignidad

Si se pretende hacer un estudio sobre el ser humano se debe partir de su sustancialidad comprendida como "aquello que posee una totalidad en sí mismo." (9, p. 11) Es decir, que no depende de algo más para existir y que tiene características propias que lo distinguen de cualquier otra cosa, está dotado de "una densidad existencial tan fuerte que permanece en sí misma a través de los cambios." (5, p. 29)

Por otro lado, conviene referir también a los llamados accidentes que atiende a características que pueden o no estar presentes pero que no modifican el ser. "Que el ser humano sea sustancia significa entonces que de él se pueden predicar todas sus cualidades: tamaño, peso, color, edad, sexo, etc. y que, a su vez, éstas serán accidentales, es decir, si están o no, no afectarán a la sustancia que ya es" (9, p. 11).

https://doi.org/10.36105/mye.2023v34n2.04

La palabra "persona", desde el punto de vista etimológico, remite al término *prósopon* que alude a las máscaras de los personajes del teatro griego antiguo. En el derecho romano *personare* apunta al papel del individuo en la sociedad. Más adelante, el cristianismo recoge este término, pero enfatiza en el orden social y humano afirmando que persona se predica absolutamente de todos los seres humanos, designa la singularidad y carácter irrepetible de cada uno además de la igualdad de todos ellos ante Dios para rechazar cualquier posible discriminación. San Agustín de Hipona apunta a la idea de la persona como un ser que participa del Dios creador, todos participamos de la misma manera y es el origen de la igualdad. Ya santo Tomás de Aquino, retomando la definición establecida por Boecio, dice:

> En general persona indica la sustancia individual de naturaleza racional. Individuo es lo indistinto en sí mismo, pero distinto de los demás. Por lo tanto, en cualquier naturaleza, persona significa lo que es distinto en aquella naturaleza, como en la naturaleza humana indica esta carne, estos huesos y esta alma, que son los principios que individualizan al hombre. Estos principios, aun cuando no significan persona, sin embargo, sí entran en el significado de persona humana (8).

Disminuir a la persona a su dimensión racional es un reduccionismo que se olvida de la dimensión volitiva o afectiva. De la misma manera centrarnos en la parte intelectiva del ser humano dejaría de lado su psique y corporalidad. El ser humano es un ser individual y único a la par que es un ser espiritual que es capaz de autotrascenderse, salir de sí. Como señala Burgos (5, p. 29), tanto hombres como mujeres "son seres especialísimos por la perfección intrínseca que poseen y que les coloca por encima y en otro plano del resto de los seres de la naturaleza".

En la modernidad, Kant aludiría a la noción de dignidad del ser humano como el valor que tiene en sí mismo y que, por ende, elimina cualquier posibilidad de ser comprado, sustituido o instrumentalizado. A diferencia de los objetos que tienen precio, la persona tiene un valor incalculable por el mero hecho de existir, es ella quien dota de valor a las cosas y el universo mismo.

https://doi.org/10.36105/mye.2023v34n2.04

En el siglo xx, en *Populorum Progressio*, encíclica dedicada a promover la cooperación entre las naciones, Pablo VI enfatiza en el carácter social del ser humano:

> Y no es solamente este o aquel hombre, sino que todos los hombres están llamados a este desarrollo pleno (…) estamos obligados para con todos y no podemos desinteresarnos de los que vendrán a aumentar todavía más el círculo de la familia humana. La solidaridad universal, que es un hecho y un beneficio para todos, es también un deber (20).

Es un requerimiento para todos que nos reconozcamos como parte del género humano y busquemos el desarrollo no sólo de algunos sino de todos los que pertenezcan a él. Eso dará pauta para hablar también del término "adultez humana" que refiere a que todas las personas deben poder acceder a la posibilidad de construir su ser a partir de un tener que resulte suficiente. Como condiciones de posibilidad para decidir ser y alcanzar la adultez humana se presenta la necesidad del tener, se requiere de ciertas condiciones mínimas para su desarrollo y responder a su vocación (28). Por ello, Pablo VI enfatiza:

> (Las personas deben) verse libres de la miseria, hallar con más seguridad la propia subsistencia, la salud, una ocupación estable; participar todavía más en las responsabilidades, fuera de toda opresión y al abrigo de situaciones que ofenden su dignidad de hombres; ser más instruidos; en una palabra, hacer, conocer y tener más para ser más: tal es la aspiración de los hombres de hoy, mientras que un gran número de ellos se vean condenados a vivir en condiciones que hacen ilusorio este legítimo deseo (20).

La aspiración de Pablo VI es que todas las personas puedan acceder a un trabajo fijo, estar libres de enfrentar coyunturas que atenten contra su dignidad, conseguir su estabilidad y adultez humana. En su lugar, encontramos que a pesar de los progresos que ha hecho la humanidad para muchos seres humanos llegar a estas metas se presenta como un mero sueño inalcanzable.

https://doi.org/10.36105/mye.2023v34n2.04

El trato que debe tener una persona, un ser que no depende de otros y que es única frente al resto de cosas existentes, que tiene una vocación y que está llamada a lograrla en conjunto con los otros, es el del respeto y el reconocimiento más allá de sus cualidades accidentales. La pregunta ante estas reflexiones iniciales es ¿por qué si todos los seres humanos somos dignos, irremplazables, no debemos ser instrumentalizados, tenemos como fin a adquirir la adultez humana, etc. existen prácticas discriminatorias que promueven la plenitud de unos a costa de otros? A continuación, se pretende plantear una breve reflexión al respecto.

3. Discriminación y sesgos cognitivos

Daniel Kahneman y Amos Tversky (14) en su texto *Prospect Theory: An Analysis of Decision under Risk* fueron pioneros en señalar que las decisiones de los seres humanos no son absolutamente objetivas ni informadas. Esto puso sobre la mesa que la información parcial sumada a creencias, experiencias, prejuicios y conocimientos anteriores intervienen en la conducta y deliberación de los individuos.

> Las personas interpretan la realidad y con base en ello juzgan y actúan influenciadas por la información que perciben sus sentidos y la que reciben y acumulan de su entorno, además de mecanismos que no siempre son conscientes, pero que les permiten tomar decisiones inmediatas y reaccionar a los desafíos y cuestionamientos que se les presentan. Esta respuesta, que es variable en cada persona y puede estar o no apegada a una deliberación racional, es producto de mecanismos mentales llamados sesgos cognitivos que utilizamos para simplificar y facilitar nuestros juicios y actos cotidianos (12, p. 9).

Si bien los seres humanos están dotados de inteligencia es poco admisible pensar que todas sus decisiones son acompañadas únicamente por la razón y logran llegar siempre a conclusiones dotadas de objetividad. Contrario a ello, las evaluaciones que se hacen sobre la realidad en muchas ocasiones son parciales y las decisiones que se

https://doi.org/10.36105/mye.2023v34n2.04

toman a partir de ellas vienen cargadas de ideas previas, opiniones, convicciones que no necesariamente han sido demostradas o justificadas racionalmente.

El funcionamiento de nuestra inteligencia por momentos se nos presenta como desconcertante ¿por qué podemos ser brillantes en algunas cosas e inoperantes en otras? ¿por qué algunas tareas las realizamos con especial destreza y otras no? Para responder se han propuesto dos tipos de pensamiento: uno intuitivo y automático que tiene por características no ser controlado, no implicar esfuerzo, es asociativo, rápido y otro que es más bien reflexivo y racional que en contraposición es controlado, laborioso, deductivo, lento, sigue normas y es autoconsciente (27).

Se ocupa un sistema u otro de acuerdo con la situación que se enfrenta. Si un balón viene a toda velocidad hacia nosotros intentaremos esquivarlo sin mayor reflexión de por medio. Si alguien pregunta cuánto es 15,345 por 23 la mayoría de las personas utilizarán el sistema reflexivo. El sistema automático puede ser muy útil, pero fiarse completamente de él puede resultar en un error pues muchas de sus conclusiones se obtienen de manera inmediata sin que haya análisis o una amplia comprensión del problema detrás y se toman como si fueran correctas, aunque no necesariamente lo sean.

Las personas generalmente tienen vidas ocupadas lo que les impide reflexionar a cada instante. Cuando tienen que emitir juicios, por la necesidad de llegar a respuestas inmediatas, lo hacen a través del uso de reglas básicas y automáticas. Por supuesto son muy prácticas, pero también pueden traer consigo sesgos sistemáticos conocidos como cognitivos.

Un sesgo cognitivo, entonces, refiere a "la tendencia a decantarse por una vía específica de pensamiento, condicionada por la intuición más que por el discernimiento" (29, p. 59). Estos sesgos se entienden como atajos heurísticos que permiten al ser humano dar una respuesta rápida ante ciertas situaciones particulares del entorno. Esto conlleva imponer a la realidad un filtro selectivo y subjetivo de información que conducirá al sujeto a tomar decisiones o llevar a cabo conductas equivocadas bajo determinados contextos.

https://doi.org/10.36105/mye.2023v34n2.04

Los sesgos cognitivos han abierto la discusión sobre cómo pensamos y decidimos, la autonomía con la que elegimos. La manera en la que nuestra mente maneja actitudes y reacciones hacia los demás que pueden estar cargadas de heurísticas y afirmaciones en las que no medie una reflexión y, por ende, obtenga soluciones que se concentren sólo en una parte de la realidad, pero no contemplen aspectos relevantes para dictar un juicio válido y verdadero.

Una idea generalizada como sociedad, por ejemplo, es que el reconocimiento de la dignidad común que hace que reconozcamos al otro como una persona con igualdad de derechos y valor se ve como un aspecto teórico sin gran relevancia en el día a día. Esto conduce a generar prácticas de violencia, intolerancia y marginación y abre la puerta a tener una mirada incompleta sobre las personas o sólo ver algunas de sus dimensiones o características accidentales que no intervienen en que una persona sea tal o tenga un nivel distinto o superior.

La discriminación refiere a la diferenciación que se hacen entre unas cosas y otras. En sí misma, no se presenta como un problema, al menos no en todos los casos pues puede servir para distinguir características o determinar el trato que se debe dar, por ejemplo, a una persona y a un objeto. Sin embargo, existe una discriminación peyorativa que atiende al trato diferente que se realiza hacia algunos grupos de seres humanos a razón de su género, color, orientación sexual, entre otras con el objetivo de "mantener o establecer una relación opresiva entre grupos o mantenerlos en una posición de desventaja" (24, p. 46).

Bajo la discriminación peyorativa, es decir, cuando se diferencia entre seres que comparten una naturaleza ontológica, es indiscutible la demanda de crear políticas públicas y buscar medios para erradicar estas distinciones que se han realizado hacia grupos oprimidos o excluidos de manera activa a lo largo de la historia. De esta forma se garantizarán los derechos humanos que apuntan a la pretensión de que se reconozca a todas las personas como libres e iguales en lo que dignidad y derechos refiere sin que intervenga distinción alguna por cuestiones contingentes al ser humano como las dichas con anterioridad (24).

https://doi.org/10.36105/mye.2023v34n2.04

Podría tenerse la falsa creencia de que el ejercicio del pensamiento automático sólo se emplea en las actividades poco trascendentes del día a día o en los encuentros inmediatos que tenemos con los demás seres humanos. Sin embargo, nos encontramos con que los sesgos y las respuestas inmediatas también están presentes en áreas, como la ciencia y la tecnología, que integran en principio la reflexión, pero en las que se infiltra la automatización precisamente por su capacidad de ofrecernos resoluciones que no suponen gran empeño y son dinámicas de acuerdo a la situación en la que estamos parados.

4. Sesgos cognitivos en la ciencia y el desarrollo de tecnología

En el campo científico existe una lucha muy competitiva por obtener el monopolio de la autoridad científica pues esto supone tener legitimidad. Es preciso hacer notar que son los seres humanos quienes dotan de sentido a las prácticas científicas y su labor. Derivado de esto, se reconocerá la influencia que tiene la psique de quienes realizan la investigación con los axiomas de la ciencia.

El conocimiento científico y el desarrollo tecnológico son el resultado de la manera en que los científicos y tecnólogos ejecutan la ciencia pero, sobre todo, de cómo la aprenden y conciben para transmitirla a los demás. Autores como Popper (21) aluden a este punto cuando señalan que la elección de una finalidad de este tipo debe ser objeto de una decisión que trascienda la argumentación racional lo que atañe a la individualidad del sujeto que trabaja desde convenciones y acuerdos previamente interiorizados, alejados de la racionalidad que después dan lugar a la ciencia. En otras palabras, ni siquiera los científicos y tecnólogos escapan a tener una mirada parcial de la realidad.

Si se sitúan estos sesgos, previamente mencionados, en el campo de la investigación científica es posible hablar de ilusiones inferenciales. Esto a causa de que nuestra razón trabaja con premisas que no

https://doi.org/10.36105/mye.2023v34n2.04

son otra cosa que inferencias. Ante esto, nos encontramos con que muchas de las tesis y clasificaciones científicas que han sido aceptadas durante un largo tiempo ahora se estudian como un producto de los sesgos cognitivos. Algunos de ellos son los siguientes:

a) Sesgo de confirmación: supone aceptar las pruebas que apoyan las propias ideas mientras que se adopta una actitud escéptica respecto a las tesis contrarias asumiéndolas como parciales. En el campo científico, es común que las personas alineen los resultados obtenidos a sus propias certidumbres (29).
b) Efecto halo: se presenta cuando un rasgo positivo de la persona se transfiere a su investigación o a toda su persona. Por ejemplo, cuando se asume que un científico destacado siempre tiene la razón y sus observaciones y conclusiones siempre son correctas. A su vez, esto se vincula con terceros que los citan como fuente indiscutible para sustentar sus argumentos dando pauta al llamado sesgo de autoridad (29).
c) Efecto de encuadre: se cae en este sesgo cuando el investigador ya tiene una conclusión en mente y busca enmarcarla con los resultados.
d) Ilusión de control: refiere a la tendencia de que es posible, a través del control y la manipulación, gobernar o al menos influir en los hechos sobre los que no se puede actuar totalmente. Supone que se podría observar sin error o sin fallo alguno (29).
e) La adhesión a las ideas: los científicos analizan los argumentos que se les oponen en afán de descubrir fallas de tal modo que no admiten que se cuestionen sus resultados fácilmente (30).

La ciencia se ocupa de conocer y comprender las causas de los fenómenos mientras la tecnología inventa productos que aún no existen pero que se presentan como una solución para los problemas actuales. El sector tecnológico incorpora conocimientos obtenidos gracias a la investigación científica aunado a la información del mercado, los precios de la competencia, etc. Si el avance científico trabaja

https://doi.org/10.36105/mye.2023v34n2.04

de la mano con el ámbito tecnológico, no se excluye a este último de contener los sesgos de los que se hablaba anteriormente.

La situación se plantea como problemática porque los resultados de las diversas investigaciones no se quedan encerrados en un laboratorio o en un escrito académico, ejemplo de ello es la IA que es usada para resolver múltiples eventos de corte práctico. Tienen repercusiones en la vida de las personas, los mercados e incluso me atrevería a decir, en la visión del mundo que hemos construido de la mano con el progreso de la ciencia y la tecnología.

5. Consideraciones en torno a la IA

La inteligencia es definida de múltiples maneras. Sin embargo, se retoma la del filósofo Burgos (5) porque enfatiza algunos de los aspectos que evidencian la diferencia entre las inteligencias artificial y humana: "(es) la capacidad que tiene la persona de salir de sí misma, trascendiéndose, acceder al mundo que la rodea, comprenderlo y poseerlo de modo inmaterial" (p. 65). Es decir, esta concepción supone que la inteligencia permite al ser humano comprender, conocer y acceder a la realidad y en ese sentido, poseerla haciendo especial énfasis en la abstracción y la inmaterialidad del conocimiento.

Mientras tanto, la IA propiamente es "una rama de la informática (que) se ocupa de métodos que permiten a un ordenador resolver tareas que, cuando resuelto por los seres humanos, requieren inteligencia." (3, p. 5) Asimismo, a la IA la caracteriza, como a otras nuevas tecnologías, la posibilidad de trabajar con incertidumbre, inexactitud, borrosidad y probabilidades (4).

Aunada a esta definición es posible distinguir entre tipos de IA: la débil es "aquella en la que las máquinas simulan un comportamiento de inteligencia utilizando las matemáticas y la informática en un área específica de aplicación y poseen la capacidad de aprender" (3, p. 5). La IA general es "una capacidad de aprendizaje en general, incluyendo la capacidad de desarrollarse de manera autónoma"

https://doi.org/10.36105/mye.2023v34n2.04

(3, p. 5). La superinteligencia o IA fuerte refiere a un desarrollo superior al del cerebro humano en muchas áreas (3).

La IA ha alcanzado una etapa de desarrollo en que tiene la posibilidad de modificar considerablemente la vida en el planeta a través de su aplicación. Dada la potencial peligrosidad del avance de IA, en 2017 se postularon los *Principios de Asilomar* para regular sus límites. Entre otras cosas, se apuesta por el progreso de "inteligencia beneficiosa", un vínculo entre ciencia y política, trasparencia, responsabilidad, seguridad, servicio al bien común y específicamente:

> 20. Capacidad de precaución: al no haber consenso, deberíamos evitar las asunciones sobre los límites superiores de las futuras capacidades de la IA.
> 22. Riesgos: los riesgos asociados a los sistemas de IA, especialmente los catastróficos o existenciales, deben estar sujetos a la planificación y esfuerzos de mitigación equiparables a su impacto esperado (22).

Ante estos principios, sin duda surgen algunos cuestionamientos en relación, por ejemplo, a la significación de la "inteligencia beneficiosa" y quienes serán aquellos que reciban ese beneficio, toda la humanidad o sólo unos pocos.

6. Los sesgos en la IA

La IA se presenta como una tecnología nueva introducida al mercado apenas hace alrededor de sesenta años. Si bien su desarrollo es bastante temprano se ha considerado que podría ser una opción viable para la toma de decisiones en asuntos, por ejemplo, sociales y económicos.

En principio se le observa como una herramienta para neutralizar la subjetividad que ha sido asociada a la decisión humana eliminando el trato discriminatorio y los sesgos destinados a ciertas personas o grupos. Pese a ello, los sistemas que utilizan IA pueden tener efectos mucho más amplios y perjudicar a muchas más personas sin

https://doi.org/10.36105/mye.2023v34n2.04

que existan los mecanismos de control social y autolimitación que sí están presentes en el comportamiento humano (26, p. 2).

Los sistemas de IA pertenecen al ámbito de IA débil, permiten ejecutar tareas y brindar soluciones en ámbitos particulares del conocimiento humano. El *machine learning* o aprendizaje automatizado, también perteneciente a la IA débil, refiere a un conjunto de técnicas y métodos que permiten a los algoritmos extraer correlaciones de los datos, que se constituyen como la materia prima a partir de la cual se pueden automatizar procesos de aprendizaje y realizar predicciones, sin supervisión (11).

Encontramos que existen distintas formas de aprendizaje con respecto a la IA. Uno de ellos es el llamado aprendizaje supervisado. En este caso, los sistemas son sometidos a un proceso de entrenamiento dirigido que pretende asociar determinadas características propias de los datos con las etiquetas que les corresponden. Dicho de otro modo, se analizan los datos de tal forma que se encuentren elementos que permitan distinguir a una categoría o etiqueta de otra. Por ejemplo, si deseamos entrenar un modelo para identificar rostros en fotografías tendríamos que ingresar una base de datos con fotografías de personas y etiquetas que, a la par, señalen en qué parte de la imagen aparece el rostro de cada una de ellas.

En un principio las asociaciones que haga la IA serán incorrectas, pero se irán corrigiendo hasta que, incluso, a partir de los datos, sea capaz de llegar a resultados nuevos con datos nunca antes vistos y establecer si son o no correctas sus conclusiones. Una premisa fundamental para considerar es que los datos con los que trabajará el modelo en el futuro serán de algún modo similares, aunque no iguales, a aquellos con los que el modelo ha sido entrenado (11).

En lugar de programar un ordenador para saber reconocer una imagen, recibe muchas, comienza a establecer conexiones entre ellas, y es él mismo quien pondera sus características para más adelante emplearlas en nuevas imágenes. Por ejemplo, se pasa la foto de un perro a la IA y después la de un Golden Retriever. Se informa al algoritmo que ambos son perros y éste será capaz de identificar cualquier

https://doi.org/10.36105/mye.2023v34n2.04

perro a pesar de que no cuente con las mismas características que los ejemplos dados inicialmente (18).

Hasta este punto sería factible pensar que sólo son cuestiones relacionadas a la mera programación de un sistema sin más. No obstante, existen algunas variaciones en la introducción de datos, provenientes de lo que nombraremos a partir de ahora sesgos algorítmicos, que pueden interferir de manera devastadora en la calidad de las predicciones. Como se señalaba en la introducción, estos sesgos aluden a los errores de tipo estadístico, estructural, cognitivo y social que traen consigo desventajas que son éticamente objetables pues dan lugar a resultados discriminatorios hacia personas o grupos o bien producen beneficios sistemáticamente a unos frente a otros (18). Es decir, refieren a una disparidad probabilística y estadística que proviene de un algoritmo generado por una computadora que sigue reglas muy específicas que le permiten tomar decisiones establecidas a través de distintos códigos (16).

Las estadísticas siempre tienen errores por lo que más que detenernos ante este punto se presentan dos cuestionamientos. El primero responde a la necesidad de saber si esos errores están equilibrados entre las distintas poblaciones que conforman la comunidad y el segundo a comprender de dónde ha surgido la inequidad en las reglas estadísticas.

La solución a lo anterior remite a que las reglas estadísticas no son aprendidas por los sistemas automatizados de la nada, sino que tienen la posibilidad de contener sesgos presentes en su diseñador:

> Los datos rara vez son neutros, están ligados a experiencias e historias de personas, por lo que reducirlos a modelos matemáticos sin tener en consideración las circunstancias que los rodean con el objeto de darle una aparente neutralidad, lleva ineludiblemente a resultados incompletos y equivocados (15, p. 279).

Por ello es fundamental entender cómo funcionan, evidenciarlos y controlarlos para lograr su erradicación y eliminar la discriminación que pueden traer consigo. (26, p. 5). A continuación, algunos ejemplos:

https://doi.org/10.36105/mye.2023v34n2.04

a) Sesgo de interacción: ocurre cuando el programador introduce en el modelo un sesgo, por ejemplo, al definir "éxito". Cuando se hace una selección de postulantes a una universidad, si el programador ha definido una preferencia que aplica sólo a quienes provienen de determinadas instituciones educativas por considerarlas académicamente superiores, entonces habrá un sesgo de interacción pues aquellos estudiantes que no hayan formado parte de estas instituciones serán rechazados más allá de cualquier aspecto.
b) Sesgo latente: refiere a cuando la IA realiza correlaciones inapropiadas entre los datos creando nexos falsos. Por ejemplo, un gerente no ha contratado a cierto grupo étnico y piensa que esas personas suelen vivir en ciertas zonas de la ciudad. Cuando se realiza el entrenamiento para la IA, basado en decisiones anteriores de ese mismo gerente, el sistema aprendería a no seleccionar a personas que vivan en esas zonas automatizando el descarte a las solicitudes que provengan de dicho grupo de individuos.
c) Sesgo de selección: cuando no se tienen los suficientes datos representativos de la diversidad existente en un medio social, es decir, hay una disparidad en el tamaño de la muestra. (26, p. 5). Si se entrenara a una IA para predecir las habilidades de la población de la universidad para las humanidades, el algoritmo utilizado resultaría inútil para realizar esa predicción en cualquier otra universidad dada la baja representatividad de esa población. Otro caso es el de:

> Joy Buolamwini, una científica informática, (quien) descubrió que su cara no era reconocida por un sistema de reconocimiento facial mientras desarrollaba aplicaciones en un laboratorio del departamento de ciencia de la computación de su universidad. Buolamwini descubrió que los datos (caras) con los que entrenaron aquel tipo de sistemas eran principalmente de hombres blancos. Esto explicaba por qué el sistema no reconocía su cara afroamericana (1).

https://doi.org/10.36105/mye.2023v34n2.04

Se ha creído que los resultados que ofrece la IA son más objetivos y neutros que aquellos a los que llegaría una persona pues excluyen, por ejemplo, sentimientos y emociones logrando mejores resultados que atiendan a las necesidades del grupo al que están dirigidos.

Pese a esto, los sistemas algorítmicos en ocasiones no son más que "opiniones escritas en código", según Cathy O'Neil matemática y experta en datos. Por eso habrá que considerar que no se trata solamente de algoritmos o modelos matemáticos, sino que tienen una repercusión en la vida de las personas. La autora señala: "I worried about the separation between technical models and real people, and about the moral repercussions of that separation" (17, p. 42).

Olvidamos que son los seres humanos quienes desarrollan y diseñan esta tecnología, lo que implica que los sesgos que ellas posean podrían ser transferidos a la IA de manera consciente o inconsciente. Al respecto Coeckelbergh (7) dice: "a menudo el sesgo no es intencionado: es habitual que lo desarrolladores, usuarios y otros involucrados, como podría ser la dirección de una empresa, no prevean los efectos discriminatorios contra ciertos grupos o individuos" (p. 117).

Esto nos conduce a que, si las variables y datos iniciales con los que la IA ha sido entrenada están sesgados con prejuicios sus resultados, por muy bueno que sea el algoritmo que utilice la IA, estarán viciados. Si estos algoritmos son utilizados en un programa social, en analizar si un megaproyecto es viable en un territorio donde reside cierto grupo, si una persona merece o no ser sujeta de crédito o contratada en una empresa, etc. entonces la aprobación no depende de meros datos, sino que tiene detrás todo un marco contextual que habrá que identificar, analizar y que se constituye como parte irremplazable del desarrollo de los algoritmos de IA.

Los sesgos aprendidos por la IA no son casos aislados, sino que han sido identificados en diferentes ambientes. Por ejemplo, la empresa *Clearview AI* prometía predecir dónde se iba a cometer un delito e identificar al perpetrador. Dejó de utilizarse en muchos países como Canadá cuando dieron cuenta de "la tendencia a identificar como

https://doi.org/10.36105/mye.2023v34n2.04

delincuentes a personas con rasgos no caucásicos" (6). Es decir, por el hecho de tener rasgos latinos o afroamericanos, minorías en muchos de los territorios donde se empleaba este sistema, se presuponía una mayor disposición a cometer actos criminales.

Otro caso es el de *Amazon* cuando intentó emplear un sistema de reclutamiento basado en IA. No obstante, el sistema tenía un sesgo contra las mujeres pues si aparecían las palabras "mujer" o "mujeres" en el currículum cuando solicitaban roles técnicos automáticamente recibían bajas calificaciones (25). El criterio de *Amazon* fue entrenar a su herramienta de reclutamiento a partir de la identificación de las palabras clave más utilizadas en la currícula de los mejores empleados, pero sin contar con la capacidad de comprender el contexto social.

De acuerdo con estos ejemplos encontramos que los sesgos algorítmicos traen consigo repercusiones que se acentúan cada vez más. Es decir, no afectan sólo a las diez o quinientas personas que no fueron aceptadas en un empleo, sino que generan un descarte general hacia ciertos grupos que se ven negados a obtener oportunidades por algo tan irrelevante como su origen étnico o su género más allá de sus capacidades.

Al final se tendrán en las empresas siempre altos directivos que cumplan con los estereotipos o en las cárceles a personas con cierto color de piel bajo la creencia común de que eso es lo correcto y lo normal. No podemos olvidar que ni "…la política, la ciencia, el arte, las formas religiosas…, son éticamente neutras o inhumanas o antisociales por naturaleza. Es una actividad del hombre y, precisamente porque es humana, debe responder a criterios humanizantes" (19, p. 77). No deberíamos confiar ciegamente en que un algoritmo tome decisiones sin previamente asegurarnos que ha sido analizada y que tiene criterios admisibles cuando se trata de evaluar personas cuyas vidas tienen la posibilidad de ser trastocadas en gran medida por miradas parciales o irreflexivas.

Ante esto, se procede a establecer algunas pautas que nos conducen a una solución al menos momentánea.

https://doi.org/10.36105/mye.2023v34n2.04

7. ¿Cómo integrar la equidad en el diseño de algoritmos?

Eliminar la discriminación, las desigualdades y promover el respeto a la dignidad no son temas nuevos, sino que han sido tratados desde múltiples perspectivas. Al día de hoy, el reto no cambia cuando la IA juega un rol, pero supone ciertos matices.

Hasta ahora hemos dicho que la persona es un ser único, que no depende de otros para existir, irremplazable, etcétera. Se ha señalado también que es un ser racional pero que su toma de decisiones y su visión del mundo no necesariamente es guiada por ella únicamente. Más bien intervienen otros factores como las creencias y prejuicios que abren la puerta a los sesgos cognitivos que, trasladados al desarrollo de tecnología, pueden filtrarse en el diseño de algoritmos de IA a partir de datos y reglas estadísticas.

Lamentablemente resulta imposible alcanzar el error cero tanto en el ser humano como en aquello que produce o proyecta. Sería deseable lograr la excelencia en los algoritmos y la estadística evitando cualquier falla, pero ante su inaccesibilidad se ha propuesto la aplicación de las llamadas "paridades" para mitigarlas:

a) La paridad demográfica: "refiere a una distribución demográfica en la que se busca que las personas que son parte de un grupo de interés estén representadas igualmente en una población demográfica" (16, p. 141). Es decir, que exista una cuota que permita introducir un equilibrio en los datos que se introducen al algoritmo según sea el caso, números similares entre hombres y mujeres, personas caucásicas y afroamericanas, entre otros.
b) La paridad de umbrales: establece si una decisión es admisible como justa midiendo a las personas de acuerdo con los mismos criterios sin estimar su origen étnico (16). Más allá de las diferencias que implica pertenecer a un país, tener la piel de cierto color, género, tendría que utilizarse la misma evaluación con unas personas que con otras. Si se van a administrar

https://doi.org/10.36105/mye.2023v34n2.04

pruebas usando IA, para la obtención de un empleo o el acceso a educación superior, los requisitos de puntaje y dificultad habrían de ser los mismos si se aplican a estadounidenses que a salvadoreños.

c) La paridad de errores: alude a la posibilidad de que, cuando se toma una decisión a partir de una regla estadística, pueda existir una equivocación que únicamente se pueda verificar a posteriori. Si un algoritmo es equitativo se emplearía en distintos grupos poblacionales y tendería a equivocarse con la misma frecuencia de tal modo que se generen tanto falsos positivos como falsos negativos (16). Es decir, se garantiza que cualquier categoría de personas divididas por un criterio cualesquiera integre la falla como un posible resultado.

Se presenta como viable aplicar estas paridades porque la responsabilidad y el impacto cambia cuando una nueva tecnología, en este caso la IA, trasciende la relación directa de persona a persona y tiene la capacidad de normalizar e institucionalizar sesgos en una sociedad y no sólo en el presente sino a largo plazo. No es la pretensión ahondar aquí en el principio de responsabilidad de Jonas, pero abona a la discusión un punto central de su propuesta:

> El bien y el mal por los cuales había de preocuparse la acción residían en las cercanías del acto, bien en la praxis misma, bien en su alcance inmediato; no eran asunto de una planificación lejana. Esta proximidad de los fines rige tanto para el tiempo como para el espacio. El alcance efectivo de la acción era escaso. El lapso para la previsión, la determinación del fin y la posible atribución de responsabilidades, corto. Y el control sobre las circunstancias, limitado. La conducta recta tenía criterios inmediatos y un casi inmediato cumplimiento (13, p. 29-30).

Es decir, anteriormente las preocupaciones éticas remitían a una cercanía en el actuar entre los individuos, la responsabilidad y las consecuencias no superaban el corto plazo. En el caso de la IA y los algoritmos utilizados, según lo dicho en el apartado previo, tienen consecuencias no sólo en la vida presente de los individuos sino

https://doi.org/10.36105/mye.2023v34n2.04

incluso en generaciones futuras que se verán afectadas por la desigualdad y exclusión que se desprendan de sus resultados.

Sumado a las paridades, un camino que contribuiría a disminuir estas dificultades también serían las auditorías algorítmicas que, entre otras cosas, solicitan la información de los responsables tanto del diseño, desarrollo e implementación del algoritmo; la metodología usada para crearlo; los datos sobre el proceso de aprendizaje y funcionamiento del sistema; las bases de datos utilizadas durante el entrenamiento y una definición clara de los posibles grupos vulnerables afectados por la puesta en práctica del algoritmo (10).

Por sí misma, la auditoría permitiría acceder a un análisis externo que compruebe que el algoritmo está libre de sesgos y que, si los tuviera, hay manera de mitigarlos. Por otro lado, saber quiénes están detrás de ellos da pauta a establecer responsabilidades y comprender intereses e incluso contextos detrás. Tener claridad con las bases de datos implica transparencia y apertura a admitir que es imposible incluir todas las variables y que por ello es importante la mirada del otro.

Antes de salir al mercado, todos los algoritmos deberían haber ser auditados y superar evaluaciones dirigidas a comprobar que la segregación hacia ciertas personas no es la respuesta común.

8. De vuelta al punto de partida

La problemática central de los sesgos algorítmicos se pensaría, reside en los datos que se introducen a los sistemas, la baja representatividad de algunos grupos, etc. de tal forma que una primera solución sería introducir paridades. Paradójicamente, éstas traen consigo la opción de que el error aparezca con la misma constancia en unos grupos que en otros. Es decir, que no sólo sean unos cuantos grupos los que se vean afectados, sino que exista la posibilidad de que cualquiera pueda serlo. Esto nos lleva a la pregunta ¿es esto deseable?

La respuesta no recae en perfeccionar el reconocimiento facial o determinar una cuota de paridad que asegure que el número de datos

https://doi.org/10.36105/mye.2023v34n2.04

ingresados al sistema es admisible. Al final esto no garantiza que se elimine la discriminación (2,14). La verdadera solución es volver la mirada hacia la persona, en su totalidad. Hay que reconocer que ante nosotros tenemos un ser valioso, digno, merecedor de alcanzar la adultez humana, de lograr la mejor versión de sí mismo y responder a su llamado.

Los sesgos cognitivos no dejan de estar presentes en la manera en la que observamos el mundo y decidimos. Aun así, es posible disminuir su impacto o incluso erradicarlo si antes de decidir quién vale más o es mejor reflexionamos sobre todo aquello que hemos aprendido, las creencias que hemos adquirido y vemos, en su lugar, a la persona.

Los sesgos algorítmicos no son otra cosa que el reflejo de una sociedad que está dividida por prejuicios injustificados, que a lo largo de la historia se ha ocupado por diferenciar más que construir puentes y que ha puesto primero intereses de otro tipo que a la persona.

> Explicar cómo se produce la discriminación a través de sistemas de IA exclusivamente desde el punto de vista técnico sería una limitación (…), la IA es un concepto sociotécnico que sólo se explica teniendo en consideración propósitos, motivos y relaciones sociales que influyen en su desarrollo e implementación (15, p. 281).

A pesar de esto, no es una razón para dejar de utilizar la IA en el presente ni mucho menos parar su desarrollo pues como señala Idoia Salazar, presidenta de Odise IA, el Observatorio del Impacto Social y Ético de la Inteligencia Artificial:

> La IA es un software con capacidad para analizar datos, extraer conclusiones, tomar decisiones de forma autónoma y aprender. Es una tecnología con enormes posibilidades para ayudarnos a tener una vida mejor si se usa para el bien (23).

Es una oportunidad para replantearnos el trato hacia nuestros congéneres buscar las medidas para mitigar posibles daños a la sociedad

https://doi.org/10.36105/mye.2023v34n2.04

ya sean intencionales o un mero accidente de la irreflexión pues en ello hay responsabilidad.

9. Comentarios finales

La discriminación proviene de una diferenciación a las personas proveniente de centrarnos en características accidentales como color de piel, preferencia sexual, edad, estatura como si éstas definieran su ser y dependiera de estas particularidades el ser más o menos humano y por ende tener más o menos valor.

La discriminación a los otros atiende a una falta de atención a lo que verdaderamente significa ser persona y responder al respeto que cada una nos merece por ser dignas ontológicamente hablando.

De acuerdo con la hipótesis planteada, como se señalaba, los sesgos cognitivos están presentes de manera constante en nuestra manera de razonar y actuar frente al mundo y las personas lo que nos conduce a tener una visión parcial de la realidad que nos permitiría, por ejemplo, atenernos a mirar en los otros sólo sus accidentes y no su sustancialidad en tanto personas. Esto aplica a todas las personas, incluso a los científicos, los tecnólogos, los diseñadores de algoritmos, etc. Como se mencionaba anteriormente, muchas veces los sesgos son introducidos de manera involuntaria. Volcar culpas sobre ellos sería de nueva cuenta pretender trabajar con autómatas exentos de las facultades propiamente humanas como son la razón, la afectividad, la voluntad, la libertad y su propia biografía.

Por otro lado, no podemos olvidar que si estos últimos actores utilizan también el pensamiento automático y rápido para obtener respuestas y a partir de ahí programan se pueden filtrar fácilmente opiniones, prejuicios y creencias injustificadas. Frente a ello, entonces, se nos presentan como humanidad al menos tres opciones: reflexionar sobre los sesgos, determinar su impacto negativo y buscar medios para erradicarlos o simplemente ignorarlos o negar su existencia.

https://doi.org/10.36105/mye.2023v34n2.04

Llevados estos sesgos al ámbito de los algoritmos y el aprendizaje automatizado de la IA representa problemáticas importantes pues su impacto trasciende al ámbito práctico en la vida de las personas y su entorno. No se pretende edificar un muro que impida el progreso científico y tecnológico, sino que habría que establecer las directrices mínimas que aseguren que la IA ha de ser utilizada en pro del ser humano que logre un avance inclusivo y equitativo para todos erradicando la ilusión de neutralidad y que sea, sobre todo, capaz de responder a las exigencias de la sociedad.

Detrás de estos sesgos y la segregación misma está detrás un rechazo a lo diferente o a lo que se muestra como separado de un "yo" o un "nosotros" que se perpetúa de unos seres humanos hacia otros. Cuando estas miradas seccionadas de la realidad se trasladan a algoritmos, que tendrán impactos en grupos de personas, la discriminación por un aspecto u otro aumenta exponencialmente e incluso se normaliza.

Es imposible encerrar a las personas en categorías pues eso implicaría caer en una observancia de ellas pobre y cortada. Paradójicamente, la inteligencia humana y los productos desarrollados por ésta como la ciencia y la tecnología operan de esa manera. Nos vemos obligados a dividir la realidad, generar modelos, incluir y excluir variables porque no podemos conocerlo todo en un mismo instante. Si eso ocurre con el mundo, es una ilusión pretender conocer a un ser multidimensional e inefable como la persona de manera absoluta y a partir de ello construir lo demás.

Conocer las limitantes, a su vez, es lo que abre la posibilidad de evitar pensar que nuestra mirada, la de científicos o tecnólogos, es única y omniabarcante. Es prácticamente imposible mantenernos en el pensamiento reflexivo y analizar cada uno de los pasos que damos. Sin embargo, si conseguimos contrastar los algoritmos, no con otros de ellos ni reducir su estudio a su efectividad según criterios de nueva cuenta establecidos por algunos, sino con el examen de más personas y todo lo que hay detrás de ellas como sus contextos, historias, modos de comprender el mundo, se crearán de manera paulatina algoritmos que respondan mejor a lo conlleva ver a la persona.

https://doi.org/10.36105/mye.2023v34n2.04

Al día de hoy no se tiene una regla que permita construir algoritmos libres de sesgos discriminatorios, pero se tienen personas que son capaces de aprehender un poco de quien está frente a sí. Al modo de un rompecabezas, uno coloca la pieza que tal vez otro no ha visto o intentaba poner en el lugar incorrecto. Por ejemplo, cuando se construye un algoritmo alguien ha considerado que incluirá por igual a hombres y mujeres en edad productiva para generar un seguro de asistencia médica y considera únicamente a aquellos con un trabajo remunerado. Alguien más da cuenta de que ha olvidado a las y los trabajadores del hogar que también realizan un trabajo indispensable para que la sociedad se mantenga en movimiento y que podrían hacerse acreedores al seguro, aunque su labor no sea pagada monetariamente. Es decir, es el otro quien nos ayuda a ver este y muchos otros matices y circunstancias de la persona fuera del radar, a contrastar con nosotros mismos, hasta construir de manera gradual el todo.

Un paso más será reconocer que los algoritmos no pueden ser considerados como universales ni permanentes. Deben estar en constante revisión de acuerdo con cimientos firmes como la dignidad, el respeto a la diferencia que nos permitan navegar en el nuevo horizonte de la IA. Una vez que se identifique alguna falla en ellos habrá que, en primera instancia, apoyarse en las paridades y las auditorías algorítmicas, pero ante múltiples inconsistencias habrá que desecharlos y construir nuevos. Será la disparidad del otro la que rompa la fragilidad de los sesgos y prejuicios que hasta ahora hemos admitido como inamovibles.

La IA es un sistema computacional sometido a la decisión entre diferentes opciones programada de acuerdo con un algoritmo, al menos hasta ahora, imposibilitado a salir de sí. La persona es capaz de elegir, crear, dar cuenta de fallos e inventar nuevos escenarios más allá de su situación individual. Expandir el análisis en el diseño de algoritmos es reconocer la complejidad de la persona y dar cuenta de que está en nuestras manos la importante decisión de enfrentar nuestras ideas preconcebidas que hemos adoptado sin justificación e involucrarnos en deslumbrarnos ante la inmensidad del otro.

https://doi.org/10.36105/mye.2023v34n2.04

Es una realidad que el error cero tanto en la estadística como en nuestro pensamiento no existe, pero genera un cambio importante que se ponga sobre la mesa que, dada la situación en la que estamos parados hoy en día tecnológicamente hablando, no podemos dejar el estudio y denuncia sobre los algoritmos que perpetúan modelos de injusticia y discriminación hacia las personas.

La perspectiva antropológica en el tema aquí tratado es indispensable pues cualquier progreso que se tenga en el ámbito del saber o de la técnica deben estar fundamentados en una mirada correcta hacia la persona. Si no se conoce exactamente qué o quién es ella, sus facultades, por qué se distingue del resto de las criaturas, por qué es digna entonces los algoritmos simplemente obtendrían respuestas guiadas por datos y reglas estadísticas, pero no habrá mayor finalidad.

Es cierto que una ambición es hacer más eficientes los procesos, pero el punto de partida y el final es la persona. Es ella quien da sentido a nuestros quehaceres como humanidad, así como la exigencia que hacemos para que todas sean reconocidas como valiosas y tengan las mismas oportunidades. Esto bajo la aspiración de construir en conjunto una sociedad conformada por personas que puedan alcanzar su plenitud y desarrollarse con excelencia.

Referencias

1. Adetunji J. Los sesgos en inteligencia artificial, el reflejo de una sociedad injusta: The Conversation [Internet] 17 de mayo de 2021 [consultado 20 de abril de 2022]. Disponible en: https://theconversation.com/los-sesgos-en-inteligencia-artificial-el-reflejo-de-una-sociedad-injusta-160820
2. Baeza R, Muñoz C. Académicos viendo Netflix: sesgos codificados. CIPER Académico [Internet]. 8 de mayo de 2021 [consultado 2 de noviembre de 2022]. Disponible en: https://www.ciperchile.cl/2021/05/08/academicos-viendo-netflix-sesgos-codificados/
3. Schmiedchen F, Bartosch U, Bauberger S, Stefan S, von Damm T, Engels R, Rehbein M, Stapf-Finé H, Sülzen A. Informe sobre los principios Asilomar en Inteligencia Artificial. Berlín: Grupo de Estudio Evaluación de la tecnología de la digitalización de la Federación de Científicos Alemanes; 2018. https://vdw-ev.de/wp-content/uploads/2019/05/Informe-sobre-los-principios-Asilomar-en-Inteligencia-Artificial_final.pdf

https://doi.org/10.36105/mye.2023v34n2.04

4. BIKTOM. Künstliche Intelligenz verstehen als Automation des Entscheidens. Berlín: Leitfaden; 2018. https://www.bitkom.org/sites/default/files/file/import/Bitkom-Leitfaden-KI-verstehen-als-Automation-des-Entscheidens-2-Mai-2017.pdf
5. Burgos J. Antropología Breve. España: Palabra; 2010.
6. Charte F. Qué peligro implican los sesgos en los modelos de inteligencia artificial: Campus MVP [Internet]. 17 de mayo de 2021 [consultado 25 de abril de 2022]. Disponible en: https://www.campusmvp.es/recursos/post/que-peligro-implican-los-sesgos-en-los-modelos-de-inteligencia-artificial.aspx
7. Coeckelbergh M. Ética de la inteligencia artificial. España: Cátedra; 2021.
8. De Aquino T. S Th.: HJG [Internet]. septiembre 2012 [consultado 2 de noviembre de 2022]. Disponible en: https://hjg.com.ar/sumat/. I, q. 29, a. 4.
9. De los Ríos M. ¿Quién es el ser humano? Bioética. Aporte para un debate necesario. México: Fundación Rafael Preciado Hernández; 2018. p. 11-27.
10. Eticas Research and Consulting SL. Guía de Auditoría Algorítmica [Internet]. 2021 [consultado 2 de enero de 2023]. Disponible en: https://www.eticasconsulting.com/wp-content/uploads/2021/01/Eticas-consulting.pdf
11. Ferrante E. Inteligencia artificial y sesgos algorítmicos ¿Por qué deberían importarnos? Nueva Sociedad: Fundación Friedrich Ebert, 2021; (294):27-36.
12. González L. Discriminación, discriminación peyorativa y la Declaración Universal de los Derechos Humanos. Aguilar A. Discriminación, sesgos cognitivos y derechos humanos: perspectivas y debates transdisciplinarios. México: UNAM; 2022. p. 9-12.
13. Jonas H. El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica. Barcelona: Herder; 1995.
14. Kanheman D, Tversky A. Prospect Theory: An Analysis of Decision under Risk. Econometrica. 1979; 47(2):263-291. https://doi.org/10.2307/1914185
15. Muñoz C. La discriminación en una sociedad automatizada: Contribuciones desde América Latina. Rev. chil. derecho tecnol. (en línea) [Internet]. 30 de junio de 2021 [citado 4 de febrero de 2023]; 10(1):271-307. Disponible en: https://rchdt.uchile.cl/index.php/RCHDT/article/view/58793
16. Noriega A. Discriminación algorítmica y costo de equidad. Aguilar A. Discriminación, sesgos cognitivos y derechos humanos: perspectivas y debates transdisciplinarios. México: UNAM; 2022. p. 139-144. https://biblio.juridicas.unam.mx/bjv/detalle-libro/7065-discriminacion-sesgos-cognitivos-y-derechos-humanos-perspectivas-y-debates-transdisciplinarios-coleccion-pudh
17. O´Neil K. Weapons ot Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Nueva York: Crown; 2016.
18. Ortega A. La imparable marcha de los robots. España: Alianza; 2016.
19. Osorio B. Antropología de la donación: el don como principio de la acción humana. Escritos. 2015; 23(50):67-82.
20 Pablo VI. Carta Encíclica Populorum Progressio del Papa Pablo VI a los obispos, sacerdotes, religiosos y fieles de todo el mundo y a todos los hombres de buena voluntad sobre la necesidad de promover el desarrollo de los pueblos [Internet]. 26 de marzo de 1967 [consultado 3 de noviembre de 2022]. Disponi-

https://doi.org/10.36105/mye.2023v34n2.04

ble en: http://w2. vatican. va/content/paul-vi/es/encyclicals/documents/hf_p-vi_enc_26031967_populorum. Html
21. Popper K. La lógica de la investigación científica. Madrid: Tecnos; 1977.
22. ROBOTechnics. Principios de Asilomar de la Inteligencia Artificial [Internet]. 11 de noviembre de 2017 [consultado 22 de abril de 2022]. Disponible en: https://www.robotechnics.es/asilomar/
23. De los sesgos a la manipulación, la cuestión ética es ineludible en el desarrollo de la inteligencia artificial: Nektiu [Internet]. 24 de junio de 2021 [consultado 24 de abril de 2022]. Disponible en: https://nektiu.com/de-los-sesgos-a-la-manipulacion-la-cuestion-etica-es-ineludible-en-el-desarrollo-de-la-inteligencia-artificial/
24. Risse M. Sobre los sesgos cognitivos y los derechos humanos. Aguilar A. Discriminación, sesgos cognitivos y derechos humanos: perspectivas y debates transdisciplinarios. México: UNAM; 2022. p. 46-57.
25. Sabán A. Amazon desecha una IA de reclutamiento por su sesgo contra las mujeres: Genbeta [Internet]. 10 de octubre de 2018 [consultado 23 de abril de 2022]. Disponible en: https://www.genbeta.com/actualidad/amazon-desecha-ia-reclutamiento-su-sesgo-mujeres
26. Sánchez M. Prevenir y controlar la discriminación algorítmica. RC D [Internet]. 2021 [consultado 18 de abril de 2022]; (427). Disponible en: https://www.researchgate.net/publication/358207305_Prevenir_y_controlar_la_discriminacion_algoritmica
27. Sunstein C, Thaler R. Un pequeño empujón. El impulso que necesitas para tomar mejores decisiones sobre salud, dinero y felicidad. Estados Unidos: Taurus; 2008.
28. Verdoy A. El concepto de progreso en la doctrina de Montini. Sols J. La humanidad en camino. Medio siglo de la Encíclica Populorum Progessio. Barcelona: Herder; 2019. p. 12-83.
29. Villarruel-Fuentes M. El quehacer del científico: una perspectiva crítica desde referentes psicológicos. Revista Ensayos Pedagógicos, 2019; 14(1):55-68.
30. Vinck D. Ciencias y sociedad: Sociología del trabajo científico. Barcelona: Gedisa; 2014.