



Trabajo Fin de Grado

Grado en Filosofía. Facultad de Humanidades
Universidad de La Laguna. 2023-2024.

Impactos de la Inteligencia Artificial:
Sesgos de género y brechas sociales.

Autor: Nuzet Gutiérrez Hernández.

Tutora: Inmaculada Perdomo Reyes

1. Introducción: Inteligencia Artificial y el dilema de los sesgos.	2
2. Antecedentes. Vistazo a la historia de la IA y los sesgos de género.	4
2. 1. Inteligencia artificial.	4
2. 1. 1. Definición del concepto.	4
2. 1. 2. El origen de la IA. Los comienzos del futuro.	4
2. 1. 3. Evolución de los sistemas inteligentes hasta el comienzo del siglo XXI.	6
2. 2. Sesgos.	7
2. 2. 1. Definición contextualizada, la relación entre desigualdad y tecnología.	7
2. 2. 2. Cronología de los sesgos de género. La brecha entre los sexos llega hasta los algoritmos.	7
3. Estado actual. El nuevo mundo: automatización acelerada y sus consecuencias.	12
3. 1. Desequilibrios en la Inteligencia Artificial. Vulnerabilidad en la sociedad.	13
3. 2. La aparición de los sesgos de género en la IA.	14
4. Discusión y posicionamiento. La IA como herramienta cotidiana y la estrecha línea entre utopía y distopía.	17
4. 1. El paraíso en desarrollo. Mejora de las necesidades del ser humano.	17
4. 2. Una sociedad corrompida. El sesgo como círculo vicioso.	17
4. 3. Brechas en la sociedad, camino a la desgracia.	20
5. Conclusión y vías abiertas. Debates y preguntas sin respuesta sobre los sistemas inteligentes.	24
5. 1. Reacción ante la inestabilidad.	24
5. 2. Ética de la IA e inconvenientes en los procedimientos.	25
6. Bibliografía utilizada.	28

1. Introducción: Inteligencia Artificial y el dilema de los sesgos.

La Inteligencia Artificial¹ puede ser reconocida con facilidad como el foco principal de interés de las tecnologías durante el último lustro, e incluso podríamos extenderlo a la última década. Los medios de comunicación, los estamentos oficiales y las personas a pie de calle han mostrado un interés notable y creciente en este tipo de sistemas. Este interés se debe a que la IA no solamente facilita una variedad de tareas, sino que también puede considerarse como una herramienta revolucionaria en términos de creación. Estos algoritmos tienen la capacidad de realizar una generación de contenido propio, ya sea a través de imágenes, textos, audios u otros formatos, mostrando así su versatilidad y rango de utilidades.

Resulta innegable la gran cantidad de avances que la Inteligencia Artificial ha tenido a lo largo de este último periodo de años. Cada vez se muestra más sofisticada y precisa en la ejecución de sus comandos y tareas. Sin embargo, a pesar de estos impresionantes progresos, aún podemos considerar que nos encontramos en las etapas iniciales del desarrollo de una herramienta con un potencial incalculable. Esta situación es problemática, ya que todavía no se sabe hasta qué punto se puede llegar a controlar toda la masa de información que conforma la IA y las implicaciones de su uso. ¿Cuáles son los límites que debería alcanzar? ¿Es necesario poner un freno a este avance? Estas cuestiones se entrelazan con consideraciones éticas, ya que han comenzado a surgir casos a nivel global en los que la Inteligencia Artificial emite respuestas extremistas en una variedad de ámbitos, abarcando desde posiciones políticas hasta cuestiones de racismo.

A lo largo de este texto pondremos nuestro interés en aquellos casos donde la Inteligencia Artificial acentúa y amplifica la brecha existente entre los géneros. Aunque es evidente que el inicio de estas cuestiones proviene de mucho más atrás en el tiempo y tiene raíces históricas profundas, esta problemática específica de los sesgos de género en la IA es una de las más actuales y pertinentes en los debates contemporáneos sobre asuntos tecnológicos. Pues, al resultar ser una herramienta con unas características tan potentes, las cuales pueden ampliarse por su aprendizaje automatizado, puede llegar con facilidad a prácticamente cualquier rincón del planeta. Si la información que lleva consigo está directamente sesgada, lo que implicará es una mayor y más marcada división entre géneros a nivel global. Tal vez esta situación sea algo que temer y un gran motivo por el que preocuparse, pero también resulta horrífico y alarmante el hecho de que aún no se haya podido encontrar una solución eficiente y efectiva que lo frene, haciendo real y palpable la

¹ También nos referiremos a ella como IA.

posibilidad de que este problema aumente y se expanda exponencialmente con el paso del tiempo.

En este análisis repasaremos la historia de la IA, al igual que de los sesgos de género. Determinaremos el punto de conexión entre ambos y observaremos cuáles son las consecuencias de esta unión. Por último, después de exponer y entender los problemas que suscita esta cuestión, intentaremos plantear alguna solución que ejecutar en el futuro.

2. Antecedentes. Vistazo a la historia de la IA y los sesgos de género.

Para poder realizar una investigación más profunda, es necesario entonces definir varios conceptos en un contexto tecnológico adecuado. Para alcanzar la comprensión de la temática expuesta definiremos las dos ideas claves que encontramos en el propio título: Inteligencia Artificial y sesgos.

2. 1. Inteligencia artificial.

2. 1. 1. Definición del concepto.

Mark Coeckelbergh² (2021) determina que “la IA se puede definir como una inteligencia tecnológica desplegada o simulada por un código -el cual está compuesto por algoritmos- o por máquinas” (p. 61), la cual “siempre se basa en otras tecnologías y está integrada en prácticas y procedimientos científicos y tecnológicos más amplios” (p. 72). También, para darle una mayor apertura al concepto, podría señalarse que son aquellas máquinas capaces de realizar un trabajo que, si fuera realizado por humanos, requeriría de inteligencia. De esa manera se diferenciaría de la definición de autómeta³ por tener un grado de complejidad superior en la ejecución de sus tareas.

2. 1. 2. El origen de la IA. Los comienzos del futuro.

A pesar de que pudiéramos considerar que la Inteligencia Artificial es un invento de nuestra historia reciente, en realidad tiene su origen en el anterior siglo como gran parte de los avances tecnológicos relacionados con la computación. Si seguimos la segunda definición ofrecida anteriormente, nos daremos cuenta de que algunas problemáticas que rodean actualmente a la Inteligencia Artificial surgían ya en la fantasía de la literatura del siglo XIX a raíz los vertiginosos avances tecnológicos que se habían sufrido en esa época, donde comenzaba a plantearse la cuestión de la ética en la ciencia. Algunas de las obras donde podríamos cruzarnos con estas temáticas son consideradas verdaderos clásicos tales como *El hombre de arena* de Hoffman, *Fausto* de Goethe y *Frankenstein* de Mary Shelley.

En la primera mencionada, Pamela McCorduck, autora de *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*, sugiere que la

² Filósofo belga especializado en tecnología, Inteligencia Artificial y robótica.

³ Dispositivos que realizan labores predefinidas de manera automática sin intervención humana.

búsqueda desenfrenada de conocimiento y el deseo de poder de Fausto pueden verse como precursores de la preocupación sobre los peligros de crear una Inteligencia Artificial que supere al control humano. Por el lado de la obra de Hoffman, encontramos cuestiones sobre la identidad, la realidad y la naturaleza de lo humano, temas que se relacionan con los debates sobre la IA y la posibilidad de crear entidades artificiales que puedan llegar a ser indistinguibles de los humanos, tal y como se entiende al hombre de arena desde la visión del protagonista. Por último, Frankenstein, una de las novelas más célebres en la historia. En ella se abordan las preocupaciones éticas y morales relacionadas con la creación de vida artificial, además de plantear preguntas sobre la responsabilidad del creador hacia su creación, tal y como lo son los límites y consecuencias de la búsqueda del conocimiento y el poder.

Estas obras literarias clásicas, según McCorduck, han proporcionado un marco cultural e intelectual para discutir los desafíos éticos, filosóficos y existenciales asociados con la Inteligencia Artificial y han alimentado debates sobre el papel de la humanidad en la creación de entidades inteligentes y conscientes (McCorduck, 2004, pp. 18-22).

Sin embargo, el comienzo de la IA de la forma en que la conocemos se remonta a 1956 durante la Conferencia de Dartmouth. En ella, un grupo de científicos conformado por matemáticos, psicólogos e ingenieros eléctricos, entre otros, llegaron a la conclusión de que “el pensar” podía darse fuera del cráneo y, además, también concluyeron que el instrumento no humano más óptimo para ello serían los ordenadores digitales. Es en este momento en el que John McCarthy, uno de los investigadores, acuñó el término de Inteligencia Artificial. Durante esta conferencia no se expuso únicamente a la IA como una teoría que alcanzar, sino que Herb Simon y Allen Newell presentaron el “Logic Theorist”, el cual podemos considerar como el primer sistema funcional de Inteligencia Artificial. Este programa tenía como objetivo demostrar teoremas de la lógica proposicional utilizando el método de demostración de la resolución. Utilizaba una estrategia heurística para buscar pruebas eficientes de teoremas y fue notable por ser capaz de encontrar soluciones a problemas que requerían creatividad y pensamiento abstracto. El Logic Theorist marcó un hito en el desarrollo de la IA y sentó las bases para futuras investigaciones en el campo (McCorduck, 2004, p. 170). Aún habiendo conseguido este logro, siguieron surgiendo trabajos acerca de todas aquellas tareas que podrían llegar a realizar los algoritmos de las Inteligencias Artificiales, aunque durante estos primeros años de investigación los avances fueron más bien teóricos.

2. 1. 3. Evolución de los sistemas inteligentes hasta el comienzo del siglo XXI.

Surge en 1963 una idea que ayudará a procesar con mayor precisión a la IA. Es decir, se encuentra una forma para que las máquinas puedan tener un entrenamiento que refleje de forma real los avances que realice: los juegos. Estos procesadores inteligentes fueron desarrollados para poder jugar partidas de ajedrez y, así, entrenarse. La economización de la búsqueda por las mejores respuestas entre todas las posibilidades se trataba de un tema central para el correcto funcionamiento de la Inteligencia Artificial (McCorduck, 2004, p. 185).

La década de los 70 resultó ser muy fructífera en los avances de IA, pues surgieron los llamados “sistemas expertos” que son aquellos programas de computación con la capacidad de imitar el razonamiento humano en un campo específico de conocimiento. Uno de los primeros funcionales y de vital éxito, siendo tomado como referencia para continuar con el desarrollo de las Inteligencias Artificiales fue DENDRAL.⁴ Fue desarrollado por Edward Feigenbaum, a quien se le reconoce como el padre de los sistemas expertos, y Joshua Lederberg, premio Nobel en biología. DENDRAL fue utilizado para interpretar datos de espectrometría de masas y para ayudar en la identificación de estructuras moleculares desconocidas. Esta tarea podía ser realizada por humanos, pero debían ser expertos y además se tardaba bastante tiempo en conseguir resultados, por lo que la creación de este algoritmo demostró el gran potencial que tenía la Inteligencia Artificial para abordar temas complejos en campos específicos (McCorduck, 2004, pp. 325-335).

Hans Moravec, diseñador del famoso carro de Stanford, siguiendo los pasos de otros científicos especializados en la IA, quiso llevar la autonomía más allá de los programas, por lo que realizó el primer vehículo autónomo controlado por algoritmos (McCorduck, 2004, pp. 460-462). A raíz de este punto comienza una globalización de las Inteligencias Artificiales pues los avances son inmensos y se despierta el interés del mundo por todos estos sistemas. Científicos tecnológicos de todo el planeta comenzaron el desarrollo de distintos algoritmos con diferentes objetivos que abarcaban todas las posibilidades que se tuvieran al alcance. Al comienzo de la década de los 80, el gobierno de Japón comienza el Proyecto de la Quinta Generación cuyos objetivos principales eran el desarrollo de una IA aún superior a lo que ya se había conseguido y, con ello, llevar estos sistemas inteligentes a las masas (p. 417). Sin embargo, la prematura mercantilización de la Inteligencia Artificial hizo que su avance se

⁴ Acrónimo de *Dendritic Algorithm*.

frenara un poco más de lo esperado. Aún así, ya al inicio del siglo XXI son lanzados al mercado general juguetes inteligentes dotados con IA, marcando así el paso definitivo de la globalización de la Inteligencia Artificial, tanto para los científicos que la desarrollan, como para la ciudadanía en general.

2. 2. Sesgos.

2. 2. 1. Definición contextualizada, la relación entre desigualdad y tecnología.

Según el diccionario de la Real Academia Española el término “sesgo” se define como “error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras”. Sin embargo, esta definición no resulta del todo satisfactoria a la hora de hablar de los sesgos en el campo tecnológico de la Inteligencia Artificial. El sesgo estadístico se refiere a cuando un algoritmo no representa con precisión los datos; en este sentido, un algoritmo debe incluir cualquier sesgo que esté presente en los datos para no estar sesgado estadísticamente (Bui, 2021, p. 4).

En este contexto en el que hablamos de una Inteligencia Artificial a la que se le otorga la capacidad de llevar a cabo una toma de decisiones, definimos el sesgo como la obvia tendencia que existe en esa resolución establecida por la IA. Esta tendencia resulta relevante porque, aunque cumple con las expectativas y el resultado deseado de una gran parte de individuos, también debe ser entendida como el desplazamiento, desprecio, rechazo o exclusión de ciertos grupos o minorías. Por lo tanto, a pesar de que nuestro foco principal estará dirigido hacia los sesgos de género, debe mencionarse obligatoriamente que hay muchos tipos distintos. Algunos ejemplos de sesgos son: los sesgos de raza, de edad, de discapacidad, de sexualidad, etc.

2. 2. 2. Cronología de los sesgos de género. La brecha entre los sexos llega hasta los algoritmos.

Antes de poder hablar específicamente del tema que nos atañe, debemos primero exponer la existencia de los sesgos de género desde sus orígenes para entender que esta cuestión proviene desde la antigüedad. Para ello debemos primero poner nuestra atención en tiempos anteriores a los inventos tecnológicos actuales tales como el Internet, los ordenadores y demás dispositivos.

Los inicios de la historia han estado repletos de estos sesgos que afectan al género. Simone de Beauvoir, una de las pensadoras más célebres y reconocidas de la historia, en su famosa obra *El segundo sexo* define que la aparición de los sesgos de género surge con la expansión de la agricultura, hecho que aconteció en la Revolución Neolítica, entre el 10000 a.C. y 4000 a.C. El motivo por el que esta autora señala que este fue ser el punto de origen de la brecha de géneros es debido al aumento en los dominios de la agricultura, pues eran los hombres quienes asumieron las responsabilidades relacionadas con la misma. Esto, de forma lógica, generó que los trabajos domésticos de la mujer se vieran relegados a un segundo plano respecto a los trabajos y esfuerzos de los hombres en los bosques y los campos (De Beauvoir, 2017, p. 86). Es precisamente esta división del trabajo la que sentó los cimientos de la subordinación de las mujeres en muchas sociedades.

Ya en estas comunidades había conciencia de la superioridad que los hombres habían establecido contra las mujeres, pues podremos observar cómo en el Código de Hammurabi⁵ la cuestión de la división entre los sexos hace un fuerte acto de presencia. Algunas leyes que encontramos en esta recopilación traducidas al español por Gabriel Franco ordenan lo siguiente:

Si alguien habiendo contraído una deuda vende a su mujer, hijo, hija y los entrega a trabajo forzado y a la sujeción, el comprador coaccionador les empleará tres años, y al tercero los libertará. (Franco, 1962, p. 344. Ley 117).

En este caso observamos cómo es el hombre quien tiene el dominio absoluto sobre la mujer -además del control sobre la familia al completo- llegando al punto donde las mismas eran vendidas para saldar las deudas cometidas por sus padres o maridos. Otra ley del código dice:

Si una mujer ha sido maldecida por su marido, aun cuando este no la hubiere sorprendido durmiendo con otro, la mujer jurará por el nombre de Dios y retornará a su casa, al hogar de su padre. (Franco, 1962, p. 345. Ley 131).

De esta manera, observamos que las mujeres que habían contraído matrimonio eran obligadas a regresar a la casa de su padre en el caso de que sus maridos las acusaran,

⁵ El Código de Hammurabi es la primera gran recopilación de leyes de las que se tiene registro.

independientemente de si eran culpables o no lo eran, quedando así en absoluto control de las decisiones que los varones tomaran sobre ellas.

Durante el avance de los siglos y de los milenios seguimos encontrando que la brecha entre géneros es una constante. Si marchamos a la Antigua Grecia, veremos que la mujer se encontraba en una situación donde no podía tomar decisiones políticas y únicamente se ocupaba de las labores del hogar. Aristóteles, uno de los mayores filósofos de la historia, cuyos escritos y aportaciones siguen siendo trascendentes hasta el día de hoy, decía esto acerca de la posición de la mujer:

Las partes de la administración doméstica eran tres: una, la del dominio del amo, de la que antes se ha hablado; otra, la paterna; la tercera, la conyugal. Pues también hay que gobernar a la mujer y a los hijos, como a seres libres en ambos casos, pero no con el mismo tipo de gobierno, sino a la mujer como a un ciudadano, y a los hijos monárquicamente. En efecto, el hombre es por naturaleza más apto para mandar que la mujer. (Aristóteles, 1998, pp. 78-79).

Este tipo de comportamientos y razonamientos formaban parte de la sociedad. Además, eran considerados como habituales y comunes. Incluso, como acabamos de observar, algunas de las mayores eminencias filosóficas tenían una posición extrema acerca de la relación que debía mantenerse entre los géneros.

Por otro lado, y con el avance del tiempo, encontramos que el aumento exponencial del cristianismo a nivel mundial fue también uno de los grandes causantes de los sesgos de género, pues la representación principalmente patriarcal del contenido de la Biblia hizo que todos los fieles y seguidores de esta religión tomaran algunas posturas que aumentaron la brecha entre ambos sexos. El mayor de estos ejemplos es el que se puede encontrar en el libro del Génesis en la Biblia:

Luego el Señor Dios dijo: «No es bueno que el hombre esté solo. Voy a hacerle una ayuda adecuada.»

Entonces el Señor Dios hizo que el hombre cayera en un sueño profundo, y mientras éste dormía, le sacó una de las costillas y cerró el lugar con carne.

Con la costilla que había sacado del hombre, el Señor Dios hizo una mujer y se la presentó al hombre.

El hombre dijo: «Esta sí que es hueso de mis huesos y carne de mi carne. Se llamará “mujer” porque del hombre fue sacada.»

Por eso el hombre deja a su padre y a su madre, y se une a su mujer, y los dos se funden en un solo ser. (Génesis 2:18-22 Reina-Valera 1960).

Es precisamente en estos versículos donde San Pablo, uno de los considerados Padres de la Iglesia, trata de justificar la subordinación a la que las mujeres debieron someterse, pues señala que es esta quien ha sido creada para el hombre y no al revés, por lo que la mujer debe someterse a su marido como la Iglesia a Cristo (De Beauvoir, 2017, pp. 161-162). Muchas escrituras revelan a la hembra como la representación encarnada de la tentación, tal como sucede con el pecado original donde es Eva quien primero prueba la manzana para después dársela a Adán. De esta forma, no solamente el papel de la mujer quedó relegado a un segundo plano, sino que se le estableció como antagonista respecto al deber que el hombre debía realizar para seguir el camino de Dios.

La opresión hacia las mujeres fue inmensa durante este periodo, aunque con el tiempo aparecieron personas de vital importancia que mostraban su apoyo a las mismas argumentando a su favor durante sus trabajos, tales como Diderot, Montaigne o Stuart Mill. A pesar de que muchas de sus luchas e ideas por la igualdad se habían visto reflejadas mucho más atrás en la historia, el feminismo como movimiento social y político moderno tiene sus raíces en el siglo XVII. Esta fecha coincide con los inicios de la llamada Revolución Industrial. Resulta completamente lógico que así sea, pues una de las consecuencias de este proceso es la participación de la mujer en el trabajo productivo. Este hecho causó la reacción del sector masculino de trabajadores, pues veían en ellas una amenaza de perder su trabajo porque realizaban las mismas labores, pero por un salario mucho menor (De Beauvoir, 2017, pp. 57-58). A pesar de las condiciones precarias y de la gran presión establecida por el sexo contrario, queda claro que este periodo fue clave para la emancipación de la mujer.

Podemos destacar que el siglo XIX también tuvo algunos avances respecto a la posición social de la misma, a pesar de que la opresión siguiera siendo la tónica habitual. En 1837 abrió sus puertas Mount Holyoke, el reconocido como el primer centro educativo que ofreció una educación igual para mujeres y hombres. Con el paso de los años se abrieron algunas escuelas femeninas y comenzaron a contratarse a un mayor número de maestras -aunque el motivo volvió a ser que sus salarios eran mucho menores que el de su contraparte masculina-. Este último hecho sumado a la agitación feminista que se venía gestando, dio

como resultado un mayor impulso a la igualdad entre sexos, pues se había conseguido, en gran medida, una educación igual para ambos (Millet, 1995, pp. 149-151).

Durante el siglo XX aún podemos encontrar todos aquellos aspectos que formaban parte de los anteriores siglos: una sociedad patriarcal donde las relaciones de poder siguen quedando a cargo de los hombres, los cuales mantienen la opresión hacia las mujeres (Millet, 1995, pp. 68-69). Se utilizaron métodos de publicidad y mensajes subliminales para establecer el llamado “carácter femenino”, el cual es sumamente artificial y viene moldeado por el interés de los propios varones poderosos (pp. 183-184). La llegada de las Guerras Mundiales también ocasionó grandes consecuencias para las mujeres, pues estas eran utilizadas como mano de obra o se les imponía el puesto de amas de casa con la responsabilidad de hacer caso a sus respectivos maridos (pp. 297-300). A pesar de todo ello, el Movimiento Feminista se vuelve un ámbito cultural más amplio y consigue algunos hitos reseñables durante este periodo. En el año 1964, después de muchas décadas -que en realidad son siglos, incluso milenios- de lucha, 104 países habían otorgado a las mujeres algún tipo de derecho civil. Podría señalarse que el gran hito que se buscó y se consiguió fue el sufragio universal, donde la mujer por fin obtuvo derecho a voto político (pp. 160-167), consiguiendo así que las opiniones de la ciudadanía femenina fueran por fin tomadas en cuenta.

Una vez habiendo entendido los antecedentes de la Inteligencia Artificial y, algo superficialmente, los de los sesgos de género, veamos entonces la relación entre ambos. En el contexto de la IA, el sesgo de género se refiere al trato desigual, desfavorable e injusto de un género sobre otro. En otras palabras, este hecho ocurre cuando los resultados obtenidos por los algoritmos tienen una tendencia claramente masculina.⁶

Guiados por la obra de Pamela McCorduck anteriormente mencionada, observamos que en los nombres de los desarrolladores e investigadores que trabajaron durante el desarrollo y creación de las primeras Inteligencias Artificiales, se escatima la presencia femenina. Este hecho ocurre porque, a pesar de los grandes avances que habían ocurrido respecto a la educación durante esas décadas, aún era difícil la admisión de mujeres en temas de especialización científica más allá de la enfermería o la medicina. Por ese motivo, el comienzo de las IAs ya estuvo marcado por los sesgos de género, lo que, de forma irremediable, termina plasmándose en el entrenamiento sesgado de los algoritmos por la falta de un enfoque femenino y feminista.

⁶ También existen casos inversos, donde la respuesta es una tendencia femenina, pero en una balanza entre ambos, los casos habituales siguen esta fórmula expuesta.

3. Estado actual. El nuevo mundo: automatización acelerada y sus consecuencias.

La evolución de la Inteligencia Artificial desde el comienzo del siglo hasta la actualidad resulta innegablemente impresionante, especialmente cuando hablamos de los últimos años en los cuales se ha desarrollado y perfeccionado una gran cantidad de algoritmos con distintos objetivos. Algunos de los más célebres son Chat GPT⁷, una IA capaz de responder preguntas y de generar contenido escrito -actualmente se encuentran en un mayor desarrollo para generación de imágenes-; Wix ADI, que sirve para crear una página web que se ajuste a las necesidades exigidas por el usuario; Suno, para la creación de música por medio de texto, etc. Casi sería mejor resumir este aspecto señalando que existe una IA para cada necesidad o problema que le surja a un humano.

Los mayores avances se están mostrando en el ámbito de las ciencias, especialmente el de la medicina. Según los datos de IBM⁸ (2021), a través de los desafíos que se establecieron con la pandemia del COVID-19 se comenzaron a desarrollar y probar nuevas tecnologías en este campo, como algoritmos que monitoreaban a los pacientes y otras herramientas de evaluación. Topol (2019) asegura que la digitalización de la medicina resulta conveniente y necesaria, pues la IA tiene la capacidad de hacer un seguimiento constante y personalizado para cada paciente en específico (pp. 12-18). Expuesto de esta manera es fácilmente reseñable que estos algoritmos son una forma de ayuda vital para los pacientes y también para los doctores, ya que podrían recomendar medicamentos, hacer un seguimiento específico y dar más rápido con ciertos resultados.

Sin embargo, paralelamente al crecimiento y desarrollo de estos sistemas digitales, también han crecido la cantidad de problemas al respecto en sus respuestas. Por lo que en la actualidad nos encontramos lejos de un algoritmo que pueda considerarse perfecto, ya sea en la medicina o en el resto de ámbitos.

3. 1. Desequilibrios en la Inteligencia Artificial. Vulnerabilidad en la sociedad.

Cabe destacar que una innumerable cantidad de estos poderosos algoritmos están disponibles para cualquiera que goce de un dispositivo con conexión. El potencial y la exposición de estos ha sido denotado como incontrolable, pues el desarrollo que la IA realiza

⁷ *Generative Pre-trained Transformer.*

⁸ *International Business Machines Corporation.*

en su aprendizaje automático es mucho más rápido que las propias mentes humanas que la crean. Virginia Eubanks señala que gran parte del poder de la toma de decisiones acerca de oportunidades de empleo, seguros o servicios gubernamentales recaen en manos de “máquinas sofisticadas” (Eubanks, 2018, p. 3). Es por ello por lo que han comenzado a viralizarse todos aquellos “errores” o respuestas desafortunadas que han tenido estos sistemas.

El ejemplo perfecto de ello lo encontramos en el sistema de nombre COMPAS,⁹ el cual ha sido utilizado en los juzgados de Estados Unidos para predecir cuáles serán las personas que tienen una mayor probabilidad de reincidir en un delito (Coeckelbergh, 2021, p. 17). Como su uso había sido tan controversial, se le hizo una investigación para evaluar su capacidad real de acierto. Los resultados de esta mostraban que los falsos positivos del algoritmo, es decir, aquellas personas que predijo que iban a volver a delinquir, pero no lo hicieron; se daban desproporcionadamente entre gente de piel oscura. Y, por el contrario, sus errores donde predijo que alguien no volvería a delinquir y acabó haciéndolo, eran mayoritariamente personas con la piel clara (pp. 18-19).

Con este caso específico podemos determinar que se trata, objetivamente, de la aparición de un sesgo racial, pues el sistema estaba inclinado favorablemente hacia las personas de tez clara. Queda entonces demostrado que la IA de esta manera puede reforzar prejuicios y discriminaciones injustas. Joy Boulamwini, investigadora de inteligencia artificial, realizó otra investigación en distintos sistemas de reconocimiento facial, dando como resultado una mayor tasa de error en las personas que tenían la piel oscura y, también, en rostros femeninos. Además, señala que gran parte de la información con la cual habían sido entrenados estos algoritmos era por medio de hombres de piel blanca (Kantayya, 2020, 2:53).

A este hecho se le debe sumar que la mayoría de los informáticos e ingenieros que desarrollan la IA son hombres blancos de países occidentales que rondan entre los 20 y los 40 años. Por lo que, como señala Boulamwini en el documental *Sesgo codificado*, lo tecnológico no puede escapar de lo social. Por ende, todas aquellas experiencias personales, opiniones y, por supuesto, prejuicios, influyen en el trabajo de estos ingenieros. Este hecho termina afectando potencialmente de manera negativa a las personas que no encajan con esta descripción -hombre blanco occidental en su adultez-, como mujeres, discapacitados, gente mayor, gente de color y gente de países en vías de desarrollo (Coeckelbergh, 2021, p.112).

⁹ *Correctional Offender Management Profiling for Alternative Sanctions.*

Podemos volver a definir sin ningún tipo de duda que los sesgos aparecen en la Inteligencia Artificial desde la propia creación de la IA, ya que su información proviene directamente sesgada por culpa de los ingenieros que formulan el código de entrenamiento de este algoritmo. Este factor propicia que la sociedad se encuentre actualmente en un punto de vulnerabilidad, puesto que la automatización está siendo, tal vez, prematura y con errores. Queda así demostrado que los algoritmos sesgados generan un mayor distanciamiento entre los ciudadanos.

3. 2. La aparición de los sesgos de género en la IA.

“Al no incluir la perspectiva de las mujeres se impulsa un sesgo masculino no intencionado que (a menudo de buena fe) pasa por «neutro» desde la perspectiva de género”. (Criado Perez, 2020, p. 11). Esta frase engloba a la historia de la sociedad en general. El mundo ha estado dirigido por hombres y en la actualidad seguimos sufriendo las diferencias entre sexos por este hecho. Cathy O’Neil en su clarificador libro *Armas de destrucción matemática: cómo el Big data aumenta la desigualdad* cuestiona “si hemos eliminado el sesgo o si simplemente lo hemos camuflado con tecnología” (O’Neil, 2019, p.23). La realidad es que este se ha integrado en los algoritmos inteligentes a raíz de la falta de opiniones femeninas en su creación, de los juicios sesgados de los ingenieros y de los datos que habitan en la nube, los cuales recoge el sistema para aprender de ellos.

Algunos casos donde se ha demostrado este hecho se encuentran en los sistemas de detección facial. Estos pasaron por grandes polémicas por los casos que hemos mencionado anteriormente: su funcionamiento tenía una mayor tasa de error en los rostros femeninos. De esta manera se reveló al mundo un suceso que ya venía establecido desde los cimientos de las Inteligencias Artificiales: la presencia de sesgos de género en los resultados que aportaba, al igual que en la información con la que fue entrenada.

Estos sesgos también han sido detectados en otros sistemas que han sido analizados. En este caso hablaríamos de la herramienta de reclutamiento desarrollada por Amazon, la cual estaba siendo utilizada para recomendaciones de contratación de personal. Sin embargo, durante el estudio que se le realizó, los resultados mostraron una clara tendencia hacia los hombres, incluso cuando existía algún currículum mejor por parte de una mujer (Dastin, 2018).

Sin necesidad de salir del ámbito del trabajo y la Inteligencia Artificial, volvemos a cruzarnos con un estudio que demostró la existencia de sesgos de género en los algoritmos. En este caso se realizó un anuncio que promovía oportunidades laborales en campos de

STEM¹⁰. Este algoritmo había sido estrictamente entrenado para mostrar el anuncio de forma absolutamente neutral, pero después al realizar un sondeo de los usuarios a los que había aparecido, quedó demostrado que la cantidad de mujeres había sido bastante menor que la de hombres. La justificación para ello es que el grupo de mujeres jóvenes es considerado como un grupo demográficopreciado, lo que implicaría que mostrar publicidad al mismo es más caro. Es entonces cuando el algoritmo determinó, por medio de su propia inteligencia de toma de decisiones basada en la optimización de efectividad-precio, no mostrar la misma cantidad de anuncios a los usuarios femeninos. De esta manera queda al descubierto que todos aquellos algoritmos publicitarios que pretenden ser neutrales en la cuestión de género terminan siendo una nueva forma de discriminación (Lambrech y Tucker, 2019).

Cuando pensamos en ayudantes virtuales controlados por inteligencia artificial, estos, a pesar de definirse como entes neutrales, en su gran mayoría suelen estar representados con una voz femenina. Lo que señalan los desarrolladores es la búsqueda de una voz que sea calmada y agradable para el oído, pero encontramos que mayormente coincide con que se trata del tono de una mujer. Siri o Alexa son algunos -y los más famosos- casos de estos ayudantes virtuales, los cuales cumplen con lo anteriormente señalado. Esta es una forma de establecer unos roles predeterminados en los datos informáticos, llevando a las tecnologías la imagen de “secretaria obediente” (Adams y Ni Lodeian, 2020), la cual está sumida en el patriarcado y demuestra la brecha de género entre hombres -quienes tienen el poder y deben ser obedecidos- y las mujeres -quienes tienen el rol de sumisión, por lo que están “obligadas a obedecer” a sus “superiores”-.

Los generadores de imágenes también han causado revuelo por la constante diferenciación estereotipada entre hombres y mujeres. Esto se debe a que los algoritmos encargados de la generación de estas imágenes, dependiendo del oficio que les sea encargado de representar, hacen siempre una selección predeterminada que muestra un rasgo estereotipado. Por ejemplo, cuando a esta IA se le ordena generar la imagen de doctores, el resultado se tratará de hombres, mientras que si utilizamos la terminología neutra anglosajona y lo que exigimos es una imagen donde pongamos el oficio *nurse*, los resultados que obtendremos serán mujeres (Kay et al., 2015). *Google Translate* también ha estado en el foco de atención por motivos similares, ya que a la hora de traducir ciertos oficios -escritos intencionalmente de forma neutra- tiende a asumir el género de los mismos, volviendo a caer en los errores que se han mostrado en el caso anterior (Bryson et al., 2017).

¹⁰ *Science, Technology, Engineering, Mathematics.*

Este tipo de resultados vuelve a mostrarnos la existencia de los sesgos de género en la Inteligencia Artificial. A pesar de que estas cuestiones ya son lo suficientemente problemáticas como para generar preocupación por las posibles consecuencias, algo que también hace que salten las alarmas en los gobiernos del mundo es el hecho del gran alcance de estos sistemas al público general, lo cual implica una exposición máxima de toda esta información sesgada a todo tipo de usuario. Algunos de estos usuarios no cuentan con la suficiente formación en el conocimiento acerca de la posibilidad del sesgo, por lo que se vuelven vulnerables y se encuentran más expuestos al riesgo. A menudo, son susceptibles de manipulaciones y es precisamente la IA la que brinda una mayor oportunidad para que estos actos ocurran (Coeckelbergh, 2021, p. 88).

Durante el último lustro, la UE¹¹ y otros países han aprobado y puesto en vigor leyes para regular la utilización de las IAs. En el caso de la ley europea, destacamos que se ha establecido una normativa que señala que todas aquellas Inteligencias Artificiales que sean destinadas a la administración de justicia y los procesos democráticos, o aquellas que puedan causar daños a las personas físicas, deben ser evaluadas como de “alto riesgo”, por lo que tendrán un mayor nivel de atención para ser reguladas y utilizadas. También exige que los datos utilizados para el entrenamiento de la IA sean estadísticamente equilibrados y contrastados, además señalan que tienen conciencia de que los sesgos inherentes tienden a aumentar gradualmente, por lo que deben realizarse revisiones constantes en las mismas.

Muchas otras leyes han sido propuestas y actualmente están siendo revisadas y modificadas a la espera de entrar en vigor, pero aún es una etapa temprana del comienzo de este tipo de normativas, aunque habrá un aumento de las mismas en los años venideros. Dicho de otra manera, los gobiernos se han puesto en marcha para solucionar la problemática expuesta con algunos de estos ejemplos, pero aún se mantiene la incertidumbre de cómo puede llegar a afectar el uso de las IAs en la vida cotidiana.

¹¹ Unión Europea.

4. Discusión y posicionamiento. La IA como herramienta cotidiana y la estrecha línea entre utopía y distopía.

4. 1. El paraíso en desarrollo. Mejora de las necesidades del ser humano.

El Libro Blanco sobre la Inteligencia Artificial de la Comisión Europea (2020) señala en su primer párrafo que “la inteligencia artificial se está desarrollando rápido. Cambiará nuestras vidas, pues mejorará la atención sanitaria (por ejemplo, incrementando la precisión de los diagnósticos y permitiendo una mejor prevención de las enfermedades), aumentará la eficiencia de la agricultura, contribuirá a la mitigación del cambio climático y en la correspondiente adaptación, mejorará la eficiencia de los sistemas de producción a través de un mantenimiento predictivo, aumentará la seguridad de los europeos y nos aportará otros muchos cambios que de momento solo podemos intuir”.

Las previsiones acerca del futuro cercano de la IA son sumamente positivas y la esperanza de la evolución tecnológica está depositada en su desarrollo. Comprobamos anteriormente que los avances en la medicina son una de las propuestas más interesantes, por lo que debemos empezar a plantearnos sistemas capaces de detectar enfermedades a gran velocidad a la vez que, al mismo ritmo, puedan recomendar el diagnóstico médico para combatirlos con mayor precisión al hacer un estudio personalizado de los datos del paciente y sus síntomas. También cabría pensar en la facilitación que existiría al automatizar las tareas administrativas que las personas del ámbito de la medicina deben realizar cada día, liberando tiempo para que los expertos puedan prestar más atención a sus respectivos pacientes. Sin embargo, ese es el esquema más simple, las posibilidades que ofrece la IA en el futuro son inciertas, pero hay una expectativa muy alta, por lo que no sonaría extraño que fueran capaces de detectar y predecir nuevas enfermedades, al igual que ser capaces de generar nuevos medicamentos de forma efectiva.

En realidad todos los sistemas expertos están resultando de gran ayuda en todos los ámbitos que podamos imaginar. La sociedad se encuentra en un punto donde parece que se aproximan evoluciones exponenciales en las tecnologías, por lo que el comienzo de un nuevo mundo está desarrollándose.

4. 2. Una sociedad corrompida. El sesgo como círculo vicioso.

Sin embargo, seguidamente de aquel optimismo en el Libro Blanco sobre la Inteligencia Artificial de la Comisión Europea (2020), se señala que “al mismo tiempo, la

inteligencia artificial conlleva una serie de riesgos potenciales, como la opacidad en la toma de decisiones, la discriminación de género o de otro tipo, la intromisión en nuestras vidas privadas o su uso con fines delictivos”. El debate instantáneo que surge es un cruce de opiniones entre aquellos que consideran que la sociedad debe arriesgarse para avanzar, y aquellos que consideran que el uso de un algoritmo que cuenta con tantos riesgos solamente nos llevará a una sociedad aún más corrompida.

Es por ello que tras exponer distintos casos que muestran sólidamente la existencia y la actualidad de los sesgos de género en la Inteligencia Artificial, cabe volver a cuestionarse el origen de estos. Para ello nos guiaremos por lo afirmado por Coeckelbergh:

El sesgo puede surgir de diversas maneras en todas las fases de diseño, prueba y aplicación. Puede surgir en la selección del conjunto de datos de entrenamiento; en el propio conjunto de datos de entrenamiento, por no ser representativo o por estar incompleto; en el algoritmo; en el conjunto de datos que se le da al algoritmo una vez que se lo ha entrenado; en decisiones basadas en correlaciones espurias; en el grupo que crea el algoritmo; o en la sociedad en general. (Coeckelbergh, 2021, p. 111).

Queda entonces claro que el sesgo puede surgir en cualquier punto de la generación de la Inteligencia Artificial. Comenzaremos por el hecho de lo que ya habíamos señalado anteriormente: los datos que son utilizados para el entrenamiento de los algoritmos son, en su mayoría, aportados por hombres blancos. Si estos ingenieros tienen una opinión sesgada, acabarán aportándola a la base de datos, por lo que formará parte de la información que el sistema tratará como verdadera, y esta información será la que sea expuesta como tal a aquellos usuarios que hagan uso de la IA.

En este caso no estaríamos hablando únicamente de sesgos de género, todo tipo de minorías sufrirían de sesgo al no encontrarse representados en el aprendizaje de los algoritmos. La falta de diversidad en el ámbito propicia que se brinde una mayor inexactitud en los resultados, pues la IA solamente estaría contando con un único punto de vista. Dicho de otro modo, la información proviene de hombres blancos, por lo que el abanico de respuestas se encuentra limitado a lo que puede experimentar un hombre blanco. Si esta es la información con la que es alimentada la IA, entonces es la que el algoritmo reconocerá como ideal. Por ende, la preferencia de búsqueda que ejecutará para responder al comando que le sea ordenado tendrá como referencia y objetivo a este público específico, excluyendo así a gran parte de la sociedad.

Imaginemos un supuesto donde una IA es creada y entrenada desde todas las posiciones posibles existentes, sin dejar fuera a ninguna de las minorías que habitualmente son sesgadas. ¿Valdría esto para conseguir un algoritmo absolutamente neutro que nos asegure la inclusividad? Posiblemente esa inclusividad se esfumará en las primeras interacciones que tuviera. El motivo es la constante posición de aprendizaje que tienen estos sistemas automatizados, por lo que memorizará todos los datos que encuentre de forma común en Internet, además del contenido de las propias interacciones a las que esté expuesta. Entonces, como si se tratara de un ciclo repetitivo, nos daremos cuenta de que gran parte de la información que se encuentra en la nube también está sesgada. Todos los comentarios de redes sociales, todas las noticias sensacionalistas, todo archivo que ha sido subido, etc. El total de esta información puede quedar en manos de una Inteligencia Artificial y, a pesar de que su propio algoritmo haya sido entrenado para clasificar la calidad de la información que termina utilizando, el filtro no tiene la calidad suficiente como para poder identificar dónde se encuentra el sesgo.

La propia sociedad está sesgada, es más, cada individuo está sesgado. Crecemos y nos criamos en un ambiente en el cual desarrollaremos nuestros gustos, nuestras opiniones, incluso nuestra forma de pensar y de ser. Inevitablemente nuestro cerebro creará tendencias y preferencias. Sin ser conscientes de ellos, nuestras acciones o pensamientos terminan siendo también una forma de exclusión. Por ejemplo, al considerar que un adulto siempre va a tener más madurez que alguien joven o cuando se da por hecho que una madre es la que más cuidados brindará a su hijo.

La ignorancia ha sido, desde tiempos inmemorables, uno de los mayores potenciadores para la exclusión. Esta puede estar alimentada por desinterés -el individuo tiene constancia de su ignorancia, pero no trata de remediarla-, tradición -el individuo ha sido criado para tener un pensamiento o un ideal que le lleve a ella- o por falta de medios -como en el mito de la caverna¹², aquellos que no tienen acceso a los recursos necesarios para entender una idea más globalizada de la sociedad, únicamente podrán tener como experiencia aquello a lo que se les sea expuesto-. En realidad, no resulta relevante el motivo en este caso, porque el punto de coincidencia es que esta exclusión de las minorías o del resto de grupos o individuos se mantiene con el paso de los años.

¹² El mito de la caverna es prácticamente la alegoría más famosa de la filosofía. Platón describe a personas que han vivido toda su vida encadenadas en una caverna, viendo únicamente sombras proyectadas delante de ellos. Estas son todo lo que conocen, su realidad. Uno consigue escapar de la caverna y descubre la realidad fuera de la caverna. Cuando regresa y lo cuenta a los otros, estos lo tratan como un loco, quedándose aferrados a aquella realidad a la que habían sido forzados.

Es cierto que se han realizado grandes avances en la inclusividad dentro de la sociedad, los países son mucho más multiculturales, se han establecido leyes que dan un mayor grado de derecho a las personas del colectivo LGTB, se realizan obras adaptadas a personas con baja movilidad, etc. Sin embargo, aún queda un amplio camino por luchar, pues todavía existen demasiados casos donde la exclusión se hace presente en la sociedad, ya sea por opiniones individuales o por acciones colectivas.

La conexión con el resto del mundo permite interacciones instantáneas con otro tipo de pensamientos. Ahora bien, la existencia del Internet y de las redes sociales dan una de cal y una de arena cuando se trata de hablar de inclusividad. Al igual que personas que se han encontrado reprimidas por su condición, aspecto o vivencias pueden expresarse con libertad y recibir un *feedback* de otras personas en una situación similar, o el apoyo de personas concienciadas de su situación; también podemos encontrar discursos de odio que reciben mucho apoyo.

Es aquí donde el algoritmo de la IA puede tomar caminos equivocados. Incluso cuando tiene la capacidad de diferenciar palabras malsonantes o insultos, no es tan eficiente en detectar el discurso de odio si se encuentra bien estructurado y expuesto. De esa forma también puede nacer el sesgo, con una fórmula sumamente sencilla donde el sistema sigue su conducta predeterminada en la cual, al haber sido creado por hombres blancos, buscará la información expuesta por hombres blancos, si estos exponen un discurso que menosprecia a la mujer, entonces la IA tomará esa información y la hará parte de sí misma. Ya hemos observado anteriormente los ejemplos, no se trata de casos aislados, por momentos incluso parece tratarse de un denominador común.

4. 3. Brechas en la sociedad, camino a la desgracia.

Uno de los casos más preocupantes respecto a las tecnologías desde la aparición de los *smartphones* hasta la actualidad es el gran ascenso en la estadística de adicción infantil. Paralelamente han sufrido un aumento desproporcionado los casos de ansiedad y suicidio juvenil. Los más jóvenes son, en realidad, los más expuestos y vulnerables al contenido que se encuentra en la nube, pues la tendencia muestra una mayor cantidad de horas por parte de los infantes en la red que de otras edades (Pons, 2020). A pesar de la existencia del control parental, la cual es una herramienta muy útil para limitar el contenido, queda demostrado que una gran cantidad de padres o tutores legales no mantienen un control sobre ese contenido al que sus hijos están expuestos. El resultado de ello es que, como cómputo global, los jóvenes

tienen, además de una dependencia emocional o “mono por el móvil”, una pérdida de inocencia prematura. El acceso a pornografía, a vídeos de accidentes o asesinatos, a videochats online, etc. Son todos lugares a “solo un click” a los que un niño no debería tener acceso.

La exposición resulta ser máxima, pues es a través del acceso a Internet que consiguen los estímulos o el conocimiento acerca de temas que requieren de una mayor madurez y profundidad. Por desgracia, su primera toma de contacto muchas veces es por medio de las redes sociales. La etapa de la adolescencia es clave para la construcción de una personalidad, es una época de cambios y descubrimientos, pero también uno de los momentos donde más se puede influenciar a alguien. Aunque se tenga un conocimiento básico sobre las temáticas que se abordan, aún la gran mayoría no tiene una capacidad de discernimiento hacia aquellas opiniones sesgadas, por lo que, si las ven repetidas de forma constante, acabarán tomándolas como propias y verdaderas, haciendo una perfecta similitud con lo que ocurriría con el aprendizaje de la IA.

La globalización y el mercado de las Inteligencias Artificiales está sufriendo un auge en todo el mundo, cada vez tiene una mayor exposición y acercamiento al público general y, por lo tanto, estos jóvenes vulnerables también la tienen. Como hemos mencionado antes, los usuarios de edad joven o adolescente son altamente manipulables, sin embargo este factor se ve incrementado con el uso de unos algoritmos que se retroalimentan de información sesgada.

Imaginemos a un adolescente que acaba de leer una opinión por X donde alguien expone que si su casa estuviera en llamas preferiría que fuera en su ayuda un hombre bombero porque confía en que biológicamente esté más preparado para ayudarlo. Este adolescente considera que tiene lógica aquello que se está exponiendo, pero para confirmar aún más esa opinión decide hacerle una consulta a alguna inteligencia artificial. Si la respuesta de esta misma confirmara el hecho de que biológicamente el hombre es más fuerte que la mujer, entonces el sesgo quedaría definitivamente sellado en la opinión de este adolescente, cuando la realidad de este caso es muy distinta. Cualquier bombero, sin depender del género, está preparado y entrenado para estas situaciones de emergencia, además de que la fuerza de un humano depende de distintos factores como peso, altura, preparación, masa muscular, etc. Y aún así, estos no determinarían la fuerza exacta de los individuos, pues algunos pueden estar más preparados para levantar peso, mientras que otros tendrán más fuerza para correr a una gran velocidad. Sin embargo, la perspectiva de este joven terminaría estando sesgada si no se le informa de manera correcta. A raíz de este caso, debemos señalar que se causaría un efecto de bola nieve, porque el joven asumirá está

información y la propagará en las redes sociales, y la IA se retroalimentará de su interacción y de la información que este adolescente aporte a la nube.

Por lo tanto, la primera consecuencia de los sesgos de género en la IA es clara: la impregnación de estos mismos sesgos en los jóvenes, que serán los adultos que guíen el volante de las sociedades en el futuro. De esta manera, se perpetuarían los estereotipos de género al asignar de forma predeterminada los roles de las mujeres, lo que nos conduciría de vuelta a una comunidad aún menos inclusiva e igualitaria.

Por otro lado, cabe señalar la tendencia a la automatización que están teniendo las grandes empresas para realizar algunas labores repetitivas, como habíamos observado que había hecho Amazon. Pero ha quedado demostrado, precisamente con este caso, que los algoritmos se encuentran sesgados y sufren de una inclinación hacia los perfiles masculinos para el sistema laboral. Si esta automatización prospera de la misma manera en la que lo está haciendo actualmente, no sonaría tan extraño que en cualquier oficio se hiciera uso de algoritmos que gestionen las contrataciones de la empresa. Si dentro de algunos de estos algoritmos surgiera el sesgo -como ya lo ha hecho en alguna ocasión-, entonces el índice de contratos laborales estaría completamente desequilibrado haciendo que aumentara la densidad de mujeres en el paro.

También hablaríamos del mismo caso si se automatiza una labor de recursos humanos que se dedique a despedir a personal con la intención de reemplazar a esos trabajadores por otros que se acerquen más al perfil ideal que le ha sido mostrado. Si ocurriera lo mismo que en el caso de Amazon, sabríamos que los principales afectados serían las personas de piel oscura y las mujeres. Determinamos entonces que la falta de equidad debida a la discriminación algorítmica es una consecuencia que se está sufriendo en la actualidad.

Resulta esencial prestar atención a la digitalización que está aconteciendo en la sociedad. A pesar de los grandes avances que puedan aportar las Inteligencias Artificiales en nuestra vida cotidiana como son los robots de limpieza o los *smartwatches*, en realidad también representan una forma recopilatoria de información personal de cada usuario que habita en la sociedad. Byung-Chul Han, filósofo surcoreano, señala que la nueva forma de sociedad disciplinaria de la que hablaba Foucault habría evolucionado y cambiado radicalmente, pues ahora nuestra prisión es la propia "libertad". Gracias a todos los movimientos que ofrecemos sobre nosotros mismos a través de nuestra huella digital establecemos nuestra nueva prisión, pues ahora la vigilancia y el control se consiguen a través de los datos y la información. Las Inteligencias Artificiales también están destinadas a

este control sobre nosotros, debido a que la intencionalidad de estas es saber todo. Tanto lo que sucede, como lo que sucederá (Han, 2022, pp. 9-24).

A raíz de toda esta vigilancia e información ofrecida a la IA, esta será capaz de mostrarnos siempre algún dato que, con seguridad, vaya a satisfacernos, pero que no tiene que aferrarse con fuerza a la realidad. La intencionalidad de este hecho es ese sentimiento de “satisfacción” fácil con el que los humanos tienden a conformarse: una vez llegado a él no se indaga más en las cuestiones que nos atañen. Entonces, ¿por qué motivo se encuentra sesgada la información de una IA que pretende satisfacer? Porque también alcanza su objetivo al posicionar en desequilibrio a dos “bandos enemigos”¹³. Por un lado, consigue la satisfacción de aquel que, con sus respuestas, logra reafirmar su -falsa- superioridad, alimentando un ego que se sostiene por el patriarcado. Solo este hecho ya sería suficiente para que la brecha entre géneros se ampliara más. Sin embargo, cabe añadir que esto consigue que el bando no beneficiado quiera conseguir su objetivo de igualdad, por lo que utilizarán los medios de comunicación que, efectivamente, servirán para volver a llenar de datos personales precisos a la Inteligencia Artificial.

Es decir, la digitalización está creando una red de cadenas para mantener el control sobre la sociedad entera. La sociedad está empezando a ser controlada por estos sistemas digitales sin ser conscientes de ello, como ocurre con el caso Bosco¹⁴, en el cual una Inteligencia Artificial se encargaba de dictaminar quiénes recibirían la ayuda, aunque el uso de la misma no era notificada a los solicitantes. Es de esta manera que estas IAs logran tener manejo en nuestra toma de decisiones, deciden lo que debemos ver o escuchar, tal y como sucedería con el Gran Hermano.¹⁵ De esta manera conseguirían que nuestra “libertad” se encuentre encarcelada en nuestras pantallas.

Ahora imaginemos que en esta red donde se somete a la ciudadanía los datos de los algoritmos se mantuvieran sesgados en términos de género. Todo ciudadano acabaría creyendo firmemente que el orden natural de la humanidad ubicaría al hombre por encima de la mujer, por lo que esta distopía estaría aún más multiplicada para ellas. La conclusión a la que llegamos es la que resulta más obvia. Es absolutamente necesario que se solucione la cuestión del sesgo en las inteligencias artificiales, pues su existencia asegura más problemas a la sociedad que soluciones al individuo.

¹³ Entiéndase esto como una metáfora para dividir a los sexos en dos grupos contrarios en sus objetivos.

¹⁴ Bosco era el algoritmo encargado de revisar las solicitudes del Bono social de Electricidad. Se hizo una prueba de funcionamiento, demostrando que tenía errores que negaban la ayuda a personas que cumplían con las especificaciones para recibirla.

¹⁵ Personaje de George Orwell en su obra *1984*, el cual es un líder supremo y omnipresente que utiliza técnicas de vigilancia masiva y manipula la información para mantener su poder.

5. Conclusión y vías abiertas. Debates y preguntas sin respuesta sobre los sistemas inteligentes.

Habiéndose ya plasmado en este escrito el origen y el desarrollo de los sesgos de género en la IA, además de cómo afecta actualmente y cómo afectará en el futuro a la sociedad, nuestra responsabilidad, casi que por obligación ética, es intentar plasmar algunas ideas que permitan regular este problema. A pesar de ello, debemos ser conscientes de que las soluciones teóricas siempre se muestran con un grado de sencillez mayor de lo que realmente puede realizarse en la práctica. Sin embargo, trataremos, desde una perspectiva ética, de aportar algunas.

5. 1. Reacción ante la inestabilidad.

La solución más obvia para regular el sesgo de género es una mayor inclusión femenina en el equipo de ingenieros que las desarrollan. Su mera presencia equilibraría más la balanza de opiniones y tendencias. Por ello, también es necesaria una mayor cantidad de datos sobre mujeres -y también de otros grupos que no han sido tan incluidos en las bases de datos- en los entrenamientos que se realizan a los algoritmos. Al igual que debería especializarse a los algoritmos acerca del género para que así no cometa ninguna discriminación al mantener de forma consciente la inclusión. De esta manera, al menos habría una menor probabilidad de que se desarrolle el sesgo de género en una etapa temprana del sistema.

Cabría destacar también la necesidad de que las IAs muestren las fuentes y referencias de las que extraen sus contenidos. El lado positivo es que algunos sistemas como Perplexity ya han comenzado a implementarlo para aclarar al usuario el origen de las respuestas que les brindan. Este apartado únicamente aplica para aquellas IAs que son de tipo chat, por lo que no es una solución definitiva para esta problemática. Además, incluso si pudiera aplicarse a todos los sistemas, habría que contar con que el usuario fuera capaz de diferenciar la calidad de las fuentes o información citadas, y también deberían de ser ellos mismos quienes debieran juzgar si se trata de datos sesgados. Ese hecho es una incertidumbre que no nos podría asegurar que los sesgos dejaran de existir, pues únicamente veríamos de donde proviene, pero no se extirparían.

Una solución que debe alcanzarse de forma obligatoria en la inteligencia artificial es la revisión constante por parte de los ingenieros de los contenidos que la IA está utilizando para monitorear las posibles apariciones de sesgos de género -y de sesgos en general-, al igual

que deberían hacerlo para revisar posibles erratas o datos sin actualizar. Es un trabajo complejo y que requiere de bastante atención, pero resulta necesario si se pretende implementar la utilización de las Inteligencias Artificiales de forma global y para todos los ámbitos. Además, si esta labor requiere de un personal que pretenda acabar con los sesgos o las exclusiones, acabaría aportando un oficio inclusivo para muchas personas.

Como se ha mencionado anteriormente, algunos gobiernos se han puesto en marcha para regular la utilización y el alcance de la IA. De esta manera se están comenzando a desarrollar limitaciones que pueden ayudar a llegar a una solución de una forma más eficaz. Sin embargo, como bien dice la FRA¹⁶, estamos lejos de conseguir una solución definitiva y mucho menos existe una solución rápida del problema (Agencia de los Derechos Fundamentales de la Unión Europea, 2022, p. 7).

5. 2. Ética de la IA e inconvenientes en los procedimientos.

Aún no es claro el tipo de intervención que debe realizarse para poder calmar o minimizar el sesgo. Surgen inconvenientes por la propia naturaleza de la inteligencia artificial, la cual aprende e integra nueva información a un ritmo inhumano. Para poder regular la velocidad de esta información habría que incorporar ciertos códigos que bloquearan al algoritmo de aprender datos sesgados, pero aun así habría que hacer una revisión de 24 horas o un bloqueo semanal para repasar el contenido que la IA selecciona y exporta a los usuarios.

Como esta globalización de la Inteligencia Artificial es muy reciente, aún no se han establecido con exactitud los parámetros en los que debería intervenir la justicia, por lo que resulta incluso más complejo denunciar este tipo de cuestiones, ¿hasta qué punto se puede permitir la existencia de un sesgo en la IA? ¿cuándo es recomendable llevar un caso más allá de un reporte a la empresa encargada del algoritmo? Las respuestas a estos problemas aún se están moldeando, por lo que no existen unos estándares fijos por los que guiarse con exactitud.

Sumado a ello, hay otros tipos de cuestiones que necesitan responderse proponiendo políticas de actuación: no solamente basta con decir qué debe hacerse, sino también el por qué, el cuándo debe hacerse, determinar cuáles son la naturaleza, la extensión y la urgencia del problema y, además, el quién debe realizarlo. Coeckelbergh cuestiona “¿corre principalmente de parte de los gobiernos llevar a cabo la acción, o deberían las empresas y la

¹⁶ Agencia de los Derechos Fundamentales de la Unión Europea.

industria desarrollar sus propias líneas de acción para asegurar una IA ética?” (Coeckelbergh, 2021, p. 123). Estas son cuestiones que tienen que resolverse cuanto antes, pero hay que ser conscientes de que no debemos precipitarnos, pues poner en las manos equivocadas las soluciones que queremos aplicar para nuestro futuro, es asegurarnos de que los problemas acabarán multiplicándose.

Sin embargo, a pesar de la urgencia con la que debemos solucionar estos problemas, aún nos encontramos en el proceso de cómo deberíamos trazar una ética de la IA, ¿se le debe considerar un mero algoritmo al que poder modificar deliberadamente a pesar de haber demostrado su capacidad de inteligencia? ¿o tal vez haya que darle una serie de derechos que deban cumplirse? Además, imaginemos que por negligencia de alguna IA que realiza recomendaciones médicas un paciente termina siendo afectado, ¿quién debería ser juzgado? ¿el sistema por haber errado en su pronóstico? ¿la empresa que haya distribuido su utilización? ¿el propio equipo de ingenieros que la desarrolló y entrenó? Aunque alguna de estas cuestiones no sean las más lógicas en un inicio, también deben plantearse para establecer correctamente una ética de la IA con la que poder orientarnos para realizar correctamente los juicios hacia los algoritmos.

Dictaminar una ética de la Inteligencia Artificial resulta todavía un reto, pero la CE¹⁷ (2020) señala que un grupo de expertos acogió siete requisitos esenciales para unas directrices éticas de la IA, los cuales son: acción y supervisión humana; solidez técnica y seguridad; gestión de la privacidad de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y medioambiental; y redición de cuentas. Esta lista debería ser facilitada a todas aquellas empresas o grupos que realicen un algoritmo para exigirles el cumplimiento del mismo. Sin embargo, en la actualidad continúan incumpléndose gran parte de estos requisitos por parte de muchos desarrolladores de estos sistemas inteligentes.

Aún así, como bien señala Coeckelbergh: “lidiar con el sesgo en la Inteligencia Artificial no es solamente un asunto técnico, sino que también es una cuestión política y filosófica” (Coeckelbergh, 2021, p. 114). El *quid* de la cuestión reside en revelar qué tipo de mundo queremos y, tras eso, plantearnos si realmente debemos intentar cambiarlo, pensando entonces en qué tipos de cambio son aceptables y equitativos. La única afirmación que podemos tener respecto a las soluciones que deben establecerse ante los sesgos de género en la IA, es que será imposible eliminarlos en su totalidad si no somos capaces de extirparlos de nuestra sociedad. Si aplicáramos las soluciones necesarias en los algoritmos, obtendremos un

¹⁷ Comisión Europea.

Uróboro¹⁸, ya que la IA siempre estará actualizándose con los datos que le sean mostrados en la nube. Si estos están llenos de sesgos de género, entonces volverá a llenarse de los mismos. Porque, una vez más, como señaló Boulamwini, lo tecnológico no puede escapar de lo social. Por lo que, dicho de otra manera: mientras los sesgos de género permanezcan en la sociedad, la IA no podrá escapar de los sesgos.

¹⁸ La serpiente que se muerde la cola, símbolo de las civilizaciones antiguas que representa el proceso cíclico de la naturaleza.

6. Bibliografía utilizada.

- Adams, R. y Ni Loideain, N. (2020). From Alexa to Siri and the GDPR: The gendering of Virtual Personal Assistants and the role of Data Protection Impact Assessments. *Elsevier, ScienceDirect*.
- Agencia de los Derechos Fundamentales de la Unión Europea (FRA). (2022). *Bias in Algorithms: Artificial Intelligence and Discrimination*.
https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf
- Aristóteles. (1998). *Política*. Editorial Gredos. Madrid, España.
- Bryson, J. J., Caliskan, A., y Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*.
- Bui, C. K. T. (2021). *Exploring Bias Against Women in Artificial Intelligence*. Tesis de maestría, Departamento de Informática, Universidad de Oslo. Oslo, Noruega.
- Coeckelbergh, M. (2021). *Ética de la inteligencia artificial*. Cátedra. Madrid, España.
- Comisión Europea. (2020). *Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza*.
https://commission.europa.eu/document/download/d2ec4039-c5be-423a-81ef-b9e44e79825b_es?filename=commission-white-paper-artificial-intelligence-feb2020_es.pdf
- Criado Perez, C. (2020). *La mujer invisible. Descubre cómo los datos configuran un mundo hecho por y para los hombres*. Seix Barral. Barcelona, España.
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*.
- De Beauvoir, S. (2017). *El segundo sexo*. Cátedra. Madrid, España.
- Eubanks, V. (2018). *Automating Inequality: How high-tech tools profile, police and punish the poor*. St. Martin's Press. Nueva York, Estados Unidos.
- Fantham, E., Foley, H. P., Kampen, N. B., Pomeroy, S. B., y Shapiro, H. A. (1995). *Women in the Classical World: Image and Text*. Oxford University Press. Oxfordshire, Inglaterra.
- Foucault, M. (2023). *Vigilar y castigar: Nacimiento de la prisión*. Siglo XXI Editores. Madrid, España.
- Franco, G. (1962). Las leyes de Hammurabi. *Revista de Ciencias Sociales*, (3), 331-356. <https://revistas.upr.edu/index.php/rcs/article/view/9466>
- Goethe, J. W. (2004). *Fausto*. Alianza Editorial. Madrid, España.

- Han, B. C. (2022). *Infocracia. La digitalización y la crisis de la democracia*. Taurus. Madrid, España.
- Hoffmann, E. (2009). *El hombre de arena y otros relatos*. El País. Madrid, España.
- IBM. (2021). ¿Qué es la inteligencia artificial en la medicina? Recuperado de <https://www.ibm.com/es-es/topics/artificial-intelligence-medicine#:~:text=Adem%C3%A1s%20de%20ayudar%20a%20los,y%20presentarles%20las%20im%C3%A1genes%20relevantes.>
- Kantayya, S. (2020). Sesgo codificado [Documental]. Netflix.
- Kay, M., Matuszek, C., y Munson, S. A. (2015). Unequal representation and gender stereotypes in image search results of occupations. En *CHI 2015 - Proceedings of the 33rd Annual CHI Conference on Human Factors in Computing Systems: Crossings* (pp. 3819-3828).
- Lambrecht, A., y Tucker, C. (2019). Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Pubsonline, INFORMS*.
- Lerner, G. (1990). *La creación del patriarcado*. Editorial Crítica. Barcelona, España.
- McCorduck, P. (2004). *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. A K Peters, Ltd. Massachusetts, Estados Unidos.
- Millet, K. (1995). *La política sexual*. Cátedra. Madrid, España.
- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press. Nueva York, Estados Unidos.
- O'Neil, C. (2019). *Armas de destrucción matemática: cómo el Big Data aumenta la desigualdad*. Capitán Swing. Madrid, España.
- Orwell, G. (2013). *1984*. DeBolsillo. Barcelona, España.
- Parlamento Europeo. (2024). *Reglamento de Inteligencia Artificial*. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_ES.pdf
- Platón. (2020). *Diálogos IV. República*. Editorial Gredos. Madrid, España.
- Pons, È. (2020). Adicción a Internet en niños y adolescentes: consecuencias de una vida entre pantallas. *El Diario de la Educación*. <https://eldiariodelaeducacion.com/2020/07/06/adicion-a-internet-en-ninos-y-adolescentes-consecuencias-de-una-vida-entre-pantallas/>
- Shelley, M. (1995). *Frankenstein*. Ediciones Orbis. Barcelona, España.
- Topol, E. (2019). *Deep Medicine. How Artificial Intelligence can make healthcare human again*. Basic Books. Nueva York, Estados Unidos.