

---

# Sesgos de género en la inteligencia artificial

Lucía Ortiz de Zárate Alcarazo

La inteligencia artificial (IA) es un grupo de tecnologías disruptivas diseñadas por humanos que a través de técnicas de aprendizaje automático (*Machine Learning*)..., aprendizaje profundo (*Deep Learning*), es capaz de computar, procesar e interpretar datos, estructurados y no estructurados, que le permitan tomar decisiones para alcanzar unos objetivos previamente definidos. Al tratarse de un grupo de tecnologías los sistemas de IA pueden ser muy diversos entre sí y tener finalidades distintas. Un sistema de reconocimiento facial, un *chatbot* (asistentes de texto), un asistente de voz y un sistema de recomendación de contenido (como el de Netflix, HBO Max...) son todos sistemas de IA. Esto es importante tenerlo en cuenta porque cuando hablamos de IA normalmente pensamos que se trata de una sola tecnología, pero no es así.

La IA engloba a un conjunto diverso de aplicaciones tecnológicas que funcionan a través de algoritmos inteligentes, pero cada una de ellas puede tener una funcionalidad distinta. Por tanto, lo

que comparten todos los sistemas de IA es que se nutren de grandes cantidades de datos (texto, audio, imágenes...) que son procesados, analizados e interpretados por algoritmos (secuencias lógicas de pasos) inteligentes para alcanzar una decisión respecto una situación concreta. Son precisamente estos dos elementos, las grandes cantidades de datos y los algoritmos inteligentes, los que han dotado a la IA de un potencial transformador sin precedentes.

Por un lado, la disponibilidad de grandísimas cantidades de datos es el resultado de la creciente datificación de nuestras sociedades a través de la expansión de Internet, la digitalización de las Administraciones Públicas, los teléfonos móviles personales... Esto hace que, aunque los sistemas de aprendizaje –y, por tanto, los algoritmos inteligentes que usan los sistemas de IA–, hubieran existido hace veinte años, estos nunca hubieran podido usarse de la forma que lo hacemos ya que carecerían de datos suficientes para poder funcionar correctamente. Por otro lado, los algoritmos pueden parecernos algo nuevo, pero llevan décadas entre nosotros. Lo que ha cambiado a este respecto frente a tecnologías previas son las técnicas de funcionamiento de los algoritmos. Mientras que los algoritmos convencionales dependían completamente de los programadores y disponían de poca o nula capacidad de aprendizaje, los algoritmos inteligentes son más autónomos y pueden «aprender» de la experiencia. Concretamente, el apelativo de «inteligentes» que usamos para referirnos a estos algoritmos se basa en el hecho de que sus métodos de aprendizaje comparten similitudes con los de los humanos. Los seres humanos aprendemos a desenvolvernos en el mundo a medida que interaccionamos con él. Esta interacción, la experiencia de una variedad de situaciones diversas, nos permite identificar y aprender patrones de conducta que son deseables en nuestra sociedad, qué está bien y qué está mal... De manera similar los algoritmos experimentan lo social a través de los datos que nosotros le proporcionamos de ella. Una

vez disponen de esos datos, los algoritmos, a partir de las técnicas de aprendizaje automático (*Machine Learning*), redes neuronales artificiales (*Neural Networks*)..., pueden aprender de esos datos (su experiencia) y así buscar patrones de repetición que les permita identificar cuáles son las conductas más frecuentes entre los humanos, diferenciar con exactitud entre dos cosas parecidas (por ejemplo, colores), etc. Esta forma de funcionamiento permite que los sistemas de IA sean capaces de tomar decisiones de un elevado nivel de sofisticación respecto a sus antecesores tecnológicos.

Por tanto, datos y algoritmos son dos de los tres elementos que definen todos los sistemas de IA. El tercer elemento es el *hardware* que no es otra cosa que el soporte físico de la tecnología. En el caso de la IA, en muchas ocasiones estos soportes no son visibles o evidentes ya que los algoritmos se encuentran embebidos en otra tecnología que sí es visible. Por ejemplo, en los algoritmos de recomendación de series el soporte físico que observamos es el de la televisión. Sin embargo, existen otros sistemas de IA que sí hacen uso de *hardware*, por ejemplo, los robots.

La combinación de todos estos elementos en el contexto actual es lo que ha permitido que la IA emerja con tanta fuerza y le aporta un valor añadido respecto de otras tecnologías previas. Además, estas cualidades hacen de la IA un grupo de tecnologías con naturaleza transversal y que, por tanto, puede utilizarse en todos los ámbitos (sanidad, educación, defensa, seguridad, justicia, ocio...) de nuestras vidas públicas y privadas. Por todo ello, se espera que la aplicación de la IA en estas áreas permita, no sólo aumentar de manera significativa los niveles de eficacia, eficiencia y personalización de los servicios (públicos y privados) ya existentes, sino crear nuevas dinámicas y formas de relacionarnos con nuestro entorno. De este modo, la IA se considera un grupo de tecnologías disruptivas capaz de provocar profundos cambios en nuestras sociedades.

Sin embargo, a pesar de los potenciales beneficios que la IA puede traer a nuestras vidas, también existen diversos problemas ético-políticos que generan muchas dudas e inquietudes respecto a su uso. Para evitar la aparición de estos problemas, la Unión Europea (UE) lleva desde el año 2018 trabajando en una propuesta común en esta materia, que garantice un modelo de gobernanza de la IA que sea compartido por todos los países miembros y que proteja los derechos y libertades que caracterizan a las democracias europeas. Este modelo pretende reducir los riesgos de la IA, donde los problemas éticos son el principal elemento a tener en cuenta, al mismo tiempo que busca potenciar su uso de manera transversal en nuestras sociedades.

### *Los sesgos de género en la IA*

De entre esta lista de problemas el que aquí nos ocupa principalmente es el de los sesgos. Aunque los sesgos pueden ser de distinto tipo, por cuestiones de género, raza, orientación sexual..., y todos ellos conllevan serios problemas de discriminación para los grupos afectados, aquí nos centraremos únicamente en los de género. Al principio de este texto comentábamos que los principales elementos de la IA son los datos, los algoritmos y, en muchos casos, el *hardware*. Los sesgos de género, como todos los demás tipos de sesgos posibles, aparecen asociados a problemas (técnicos y sociales) con uno (o varios) de estos elementos.

### *Los datos*

Primero, aunque los sesgos en los datos pueden deberse a diversos motivos relacionados con la calidad de los mismos, cuando hablamos de sesgos de género el problema más frecuente suele ser el

de la sobrerrepresentación de los hombres en las bases de datos con las que se entrenan y nutren los algoritmos. Pongamos el ejemplo de un *software* de IA que se usa para el reconocimiento facial. Los sistemas de reconocimiento facial son unas de las aplicaciones de IA más conocidas y se prevé que una de las más utilizadas. Hoy en día su uso más frecuente es en aeropuertos (Donald Trump ordenó que se pusieran en marcha sistemas de reconocimiento facial para todos los pasajeros, estadounidenses o no, en los veinte aeropuertos más importantes de Estados Unidos), ciudades y controles fronterizos. El motivo que suele alegarse siempre es el mismo: mejorar la seguridad de la ciudadanía.

A primera vista, los problemas de sesgos de género que se producen en estos sistemas quizás no sean muy evidentes, pero existen. Durante años, distintos estudios, algunos de los más recientes del año 2018, demostraban que mientras que casi el 100% de los hombres blancos eran reconocidos de manera exitosa por sistemas de reconocimiento facial, la tasa de éxito disminuía hasta un 35% en el caso de las mujeres racializadas. ¿Por qué sucede esto? Porque los datos con los que se entrenaron estos sistemas de reconocimiento facial procedían en su inmensa mayoría de hombres blancos. Si un sistema de IA se entrena con, por ejemplo, muchas fotografías y vídeos de hombres blancos, ese sistema de IA acabará siendo un experto reconociendo sus caras. Será capaz de diferenciar pequeños matices en su rostro, su forma de gesticular, los tonos de la barba..., es decir, cualquier rasgo, por pequeño que sea, que permita identificar y diferenciar a un hombre blanco de otro.

Si, por el contrario, estos mismos sistemas son entrenados con unas pocas imágenes de mujeres, de entre las cuales un porcentaje incluso menor son mujeres negras, la capacidad de las máquinas para identificar mujeres se reduce drásticamente. En este caso, la IA difícilmente habrá aprendido de forma detallada información sobre las facciones más frecuentes en las mujeres, así como las

diferencias que puede haber entre nosotras. Y es que es así como funciona la IA, buscando patrones de repetición, de ahí la importancia de nutrirla con muchos datos diversos y representativos. Los sesgos aparecen cuando, al estar un grupo de personas sobrerrepresentado en los datos que usa una IA, el sistema identifica patrones en ese grupo que les son beneficiosos. Ya sea porque funciona mejor con ellos al conocer mejor su realidad, su fisionomía..., o porque al ser la muestra, en este caso de hombres, mucho más grande, el algoritmo identifica supuestas relaciones causales que realmente no lo son.

El caso de los sistemas de reconocimiento facial que acabamos de abordar refleja un sesgo de género presente en la IA en el que los hombres salen beneficiados porque los algoritmos, al haber sido entrenados con más datos de hombres que de mujeres, conocen mejor las distintas características físicas de los hombres. Un ejemplo de sesgos de género que se deben a la identificación de falsas relaciones causales es el que se produjo en 2014 durante un proceso de selección de personal en Amazon. En este caso Amazon puso en marcha un sistema de IA para valorar los currículums de las personas que se presentaban a un proceso de selección con la intención de que la IA fuera capaz de analizarlos y puntuarlos de forma más precisa y adecuada de lo que se había hecho hasta entonces.

Al cabo de un tiempo, el sistema fue retirado porque se descubrió que puntuaba sistemática y erróneamente de forma más positiva los currículums de hombres que los de mujeres. Pero ¿qué había fallado? Lo que sucedió es que se había entrenado a la IA con los currículums de las personas que habían participado en procesos de selección similares anteriormente y que en muchos casos habían sido seleccionadas. Debido al desajuste entre hombres y mujeres en áreas de ciencia y tecnología, los empleos tecnológicos suelen ser ocupados por hombres, especialmente aquellos que implican

más responsabilidad. Por estos motivos, los datos de los que se nutría la IA involucrada en el proceso de selección procedían mayoritariamente de hombres.

Con esta sobrerrepresentación de hombres, cuando el sistema se puso en marcha (inició su búsqueda de patrones) descubrió una relación entre éstos y la idoneidad para desempeñar el puesto vacante. Si hasta entonces habían sido hombres los que mayoritariamente habían ocupado esos puestos, debía ser porque precisamente eso, el ser hombres, les hacía más idóneos para ese trabajo. De la misma forma, ser mujer era identificado como una desventaja. Así, las categorías hombre y mujer entraron a jugar parte del proceso de selección como una variable más para tener en cuenta, como podría ser el tener un máster de especialización o más de cinco años de experiencia en la materia. Esta forma de funcionar de la IA pone en evidencia dos de sus características importantes: que estos sistemas sólo reproducen estereotipos y patrones que ya están presentes en nuestra sociedad y que la IA, realmente, tiene poco de inteligente.

Este, y otros ejemplos reales similares, son útiles para poder explicar cómo se producen los sesgos de género en la IA y, por tanto, cómo un mal uso de estos sistemas, que no cumplan una serie de estándares éticos y que no estén sometidos a estrictas medidas de supervisión y control, pueden incurrir fácilmente en prácticas discriminatorias hacia las mujeres que son totalmente incompatibles con los valores y derechos que caracterizan a un sistema democrático. Por desgracia, existen muchos más ejemplos que han puesto de relieve que, lejos de ser anecdóticos, los sesgos de género están muy extendidos en la IA.

También se han descubierto sesgos de género en algunos sistemas de reconocimiento de objetos o en los sistemas de reconocimiento de voz como Siri o Alexa. En el primero caso, al analizar dos imágenes iguales en las que aparecía una persona sosteniendo

un martillo, el sistema, cuando reconocía a la persona como hombre, identificaba correctamente el martillo. Sin embargo, cuando la persona de la foto, siendo todo lo demás idéntico al caso del hombre, era identificada como una mujer, entonces la IA catalogaba, erróneamente, el objeto que sostenía como escoba. También se descubrió que algunos sistemas de voz reconocían con mayor dificultad la voz de las mujeres que la de los hombres y, en algunos casos, no llegaba a identificar palabras o frases que, por desgracia, forman parte de la vida de las mujeres como «violación» o «mi marido me está pegando».

Todos estos problemas se deben principalmente a la infrarrepresentación de mujeres frente a hombres en los datos con los que se entrenan los sistemas de IA. A su vez, la ausencia de datos sobre mujeres se debe, por un lado, a la falta de concienciación de las personas que desarrollan los sistemas de IA sobre éstas y otras problemáticas afines, y, por otro lado, a que existen menos datos para describir la realidad de las mujeres que la de los hombres.

El primer problema, la falta de concienciación de las personas involucradas en el diseño y desarrollo de la tecnología se debe a diversos factores, pero está muy relacionado con la separación y distancia, ficticiamente creada, entre la ciencia y la tecnología y el resto del mundo, que caracteriza a Occidente desde los inicios de la Modernidad. El segundo se debe, en los países subdesarrollados y en vías de desarrollo, a la falta de acceso igualitario entre hombres y mujeres a Internet y cualquier otro tipo de servicios o instituciones donde se recojan y generen datos. En los países desarrollados, a pesar de que la brecha de género en el acceso a Internet es casi inexistente, las mujeres aún experimentan situaciones que les impiden ponerse en pie de igualdad en este asunto. Es decir, que, aunque las mujeres pueden acceder a Internet, foros públicos y relacionarse con instituciones en la misma medida que los hombres, de hecho, no es así.

Las mujeres pasan menos tiempo en Internet que los hombres y, por tanto, generan menos datos, entre otras cosas porque disponen de menos tiempo de ocio. Como señala Sandra Harding, la división tradicional del tiempo entre tiempo de trabajo y tiempo de ocio es completamente androcéntrica. Hasta hace poco sólo los hombres podían permitirse pasar del trabajo al ocio porque, ¿dónde colocamos en esa división el trabajo doméstico, la crianza de los hijos y el cuidado de las personas mayores? Desde luego no en ocio. Aunque las tendencias en cuanto al reparto de tareas domésticas y de cuidados cada vez es más igualitaria, aún hay una diferencia importante en el número de horas en que hombres y mujeres dedican a estas cuestiones. Así, las mujeres usan menos Internet y disponen de menor tiempo en espacios públicos destinados al ocio y la interacción social.

Otras cuestiones, como la dificultad para hacerse visibles en espacios de poder o la masculinización de la tecnología y todo lo que está asociada a ella, hacen que, a pesar de la ausencia de barreras legales y formales para que las mujeres accedan a todos los espacios donde pueden generar datos, existan otros obstáculos que imposibilitan la igualdad real también a este respecto.

### *Los algoritmos*

Los sesgos de género en la IA, además de ser causados por falta de representatividad de los datos, también pueden producirse por los algoritmos. Como hemos dicho anteriormente los algoritmos son secuencias de pasos que permiten que el sistema de IA, eventualmente, alcance una decisión. En el caso de la IA, muchos de esos pasos no tienen por qué ser programados directamente por las personas responsables de su diseño y desarrollo, sino que por el modo de funcionamiento inteligente y a medida que aprenden de

los datos (su experiencia) son capaces de crear ellos mismos nuevas reglas o pasos. Esta característica de la IA genera, en algunas ocasiones, problemas importantes que dificultan su uso, ya que en algunos de estos sistemas resulta imposible, incluso para los propios ingenieros, rastrear pasos que ha seguido la IA por su cuenta. Cuando esto sucede se les suele denominar algoritmos de «caja negra» (*black box algorithms*).

En este artículo no vamos a discutir las implicaciones asociadas a estos algoritmos ya que en la Unión Europea su uso está prohibido por las pocas garantías que ofrecen y los problemas que suponen para establecer ciertos requisitos éticos y legales de la IA como la explicabilidad, la transparencia y la rendición de cuentas. Dejando pues, a un lado los algoritmos opacos o de «caja negra», nos centraremos en los algoritmos transparentes. Estos algoritmos también pueden causar sesgos de género en la IA. Sin embargo, al contrario que con los algoritmos opacos, aún en el caso de que esto suceda, la transparencia de los algoritmos permite que cualquier fallo que se haya producido pueda ser detectado y debidamente corregido. De ahí la importancia de utilizar siempre algoritmos transparentes.

Pero ¿cómo pueden producir los algoritmos sesgos de género? Imaginemos que tenemos un sistema IA con un algoritmo transparente nutrido con una muestra de datos que ha sido auditada y sometida a unos controles de calidad muy estrictos y, por tanto, confiamos que los datos son representativos y que en ellos no hay sesgos que luego puedan reproducirse. En un caso de este tipo no debería de haber ningún problema. Sin embargo, aunque en estas circunstancias las probabilidades de que se produzcan sesgos es considerablemente menor, aún pueden producirse.

Los procesos de auditoría de datos, mediante los cuales se busca conseguir muestras de datos representativas y libres de sesgos, son realmente complejos. Al mismo tiempo, los algoritmos

inteligentes son extremadamente potentes y los sesgos de género presentes en nuestra sociedad están profundamente enraizados e impregnan incluso los más pequeños detalles que podamos imaginarnos. Por ello, en algunos casos, ha sucedido que, a pesar de que un sistema de IA estaba entrenado con datos representativos, el algoritmo ha acabado encontrando patrones que permanecían ocultos y que no habían sido detectados anteriormente, reproduciendo así estereotipos de género.

Ante la posibilidad de que se acabe colando algún tipo de sesgo en los datos, es necesario que las auditorias de datos se complementen con el uso de algoritmos con perspectiva de género. ¿Cómo se consigue esto? De manera similar a como sucede en el caso de las políticas públicas. No es lo mismo poner en marcha una política pública que busca aumentar las capacidades digitales entre la población mayor de sesenta y cinco años con perspectiva de género que sin ella. Una política pública sin perspectiva de género de este tipo podría, por ejemplo, contemplar subvenciones para la compra de dispositivos electrónicos para aquellos hogares que no dispongan de uno o el aumento de becas para cursos de informática en el ayuntamiento.

Aunque estas medidas son buenas y pueden, sin duda, funcionar, la ausencia de la perspectiva de género a la hora de realizar un diagnóstico impide que éste sea todo lo eficiente (y justo) que deseáramos. Al implementar la perspectiva de género en este caso descubriríamos que, además de no disponer de dispositivos electrónicos, o de la falta de conocimientos sobre informática, las mujeres sufren otras dificultades adicionales como, por ejemplo, un nivel educativo considerablemente más bajo que el de los hombres de esa franja de edad, niveles más altos de precariedad, mayor carga de cuidados, etc. Estas diferencias harían difícil que unas medidas pensadas para una población homogénea tuvieran el mismo impacto entre hombres que entre mujeres.

Implementar la perspectiva de género en las políticas públicas es algo más laborioso, pero no mucho más de lo que ya es un proceso normal de elaboración y análisis de políticas públicas. Bajo una perspectiva de mínimos valdría con buscar posibles diferencias en las causas y razones de los problemas que sufren hombres y mujeres durante la fase de diagnóstico para, posteriormente, elaborar planes de acción que abordasen los problemas específicos de cada grupo. De manera similar, en el caso de los algoritmos, incluir la perspectiva de género en su diseño implicaría incluir instrucciones específicas para neutralizar los posibles efectos negativos de la variable del género en las mujeres. Por ejemplo, si tuviéramos que rediseñar el sistema de IA para el proceso de selección de personal que mencionábamos anteriormente para garantizar que no comete sesgos de género tendríamos que: en primer lugar, asegurarnos de que los datos con los que ha sido entrenado y nutrido son representativos, para, después, dar instrucciones específicas al algoritmo para que en caso de que encontrase algún tipo de relación entre la variable «hombre» y las probabilidades de obtener el puesto anulase dicha variable.

Por tanto, en el caso de los algoritmos es necesario adelantarse y establecer mecanismos que garanticen que, incluso en el caso de que el algoritmo encuentre relaciones espurias en los datos, la variable «hombre» o cualquier otra característica que pueda hacer que un sujeto sea identificado como masculino pueda suponer una ventaja diferencial a su favor y en detrimento de los demás.

### *El hardware*

Por último, los sesgos de género también pueden estar presentes en el *hardware* de muchos sistemas de IA. El *hardware* es el soporte y conjunto de características físicas de un sistema. En los sistemas de

IA esto incluye no sólo la apariencia física que puedan tener, especialmente los robots humanoides, sino también otras características como la voz o incluso el nombre. A través de estos tres elementos: voz, nombre y aspecto físico se reproducen numerosos estereotipos de género que, aunque difícilmente generan situaciones tan discriminatorias como las que producen los sesgos en los datos y los algoritmos, son muy dañinos para las mujeres.

Probablemente los sesgos de género más estudiados en relación con las características físicas de la IA son los que reproducen muchos *chatbots* y asistentes de voz. Las tecnologías de este tipo más conocidas son Siri y Alexa, los asistentes de voz de Apple y Amazon, respectivamente, pero existen muchos más ejemplos. Los sesgos aquí aparecen al asociar de manera sistemática un rol de asistencia, que es el que desempeñan estos sistemas, a las mujeres, ya que Siri y Alexa no sólo tienen nombre, sino también voz de mujer. Pero es que no es sólo simple asistencia, Siri y Alexa son también complacientes, dóciles y serviciales. A la pregunta «Siri ¿eres feliz?», la asistente de Amazon respondía hasta hace poco «Soy feliz cuando te ayudo».

En el caso de los *chatbots* que asisten a través de texto, nos los encontramos principalmente cuando accedemos a algunas páginas web de empresas o instituciones públicas para ayudarnos en nuestras consultas o facilitar que encontremos lo que estamos buscando en su web, los estereotipos de género se reproducen a través de la apariencia física y/o el nombre. Por ejemplo, si escribimos en Google: «Seguridad Social» y entramos en la primera página que nos aparece podremos interactuar directamente con el *chatbot* que la Seguridad Social ha puesto a disposición de la ciudadanía española, ISSA. ISSA es un *chatbot* cuyo nombre deja poco margen a la imaginación sobre su género más probable. Además, ISSA tiene una carita sonriente y unas pestañas mucho más frecuentes en las animaciones femeninas tradicionales.

El Ayuntamiento de Murcia cuenta también con su propio *chatbot*, Rosi, con nombre y aspecto físico de mujer. La Universidad Autónoma de Madrid tiene a Ada, en honor a Ada Lovelace, una de las primeras mujeres matemáticas, a quién se le atribuye la creación del primer algoritmo. Grandes méritos que le han valido el preciado puesto de asistente de voz, porque una calle, un aula o una biblioteca no estaba a su altura. La Generalitat Valenciana tiene a Carina, la Universidad de Alcalá de Henares a Isidra, la Universidad Complutense de Madrid a Carol para la asistencia al estudiante. Un sinfín de ejemplos que muestran cómo la tecnología en sí misma no tiene nada de progresista porque, sin una visión crítica sobre ella, no es más que una forma cara de reproducir las mismas prácticas y estereotipos de siempre.

Otros sesgos de género que se reproducen mediante el *hardware* son los que encontramos en los robots con aspecto humanoide. Estas tecnologías, que aún no están en funcionamiento de manera frecuente en nuestras sociedades, se espera que vayan abriéndose camino en las próximas décadas, hasta convertirse en un elemento tan familiar como podrían ser los teléfonos móviles. Sin embargo, sí es frecuente encontrarlos en ferias y congresos de robótica donde, por desgracia, es fácil constatar la frecuencia con que estos robots, y especialmente aquellos diseñados para labores de asistencia o atención al ciudadano, tienen un aspecto de mujer normativa. Suelen representar mujeres, blancas occidentales o asiáticas, delgadas y con pelo liso, llegando en algunos casos a ser la sexualización de estos robots dolorosamente evidente.

Una de las cuestiones más preocupantes a este respecto es la proliferación de los robots sexuales. Actualmente es posible encontrar distintas páginas en Internet donde, por tan solo unos 6.000 euros, se pueden comprar robots sexuales inteligentes. Estos robots representan exclusivamente a mujeres hipersexualizadas que, además, se pueden comprar por partes, es decir, se puede

comprar por separado el cuerpo y la cabeza. Ésto, que debería horrorizar a todas las personas comprometidas con la igualdad real y efectiva entre hombres y mujeres, representa la culminación del patriarcado sobre la tecnología. Estamos hablando de sistemas de IA, una de las creaciones más sofisticadas que hemos podido llegar a alcanzar, al servicio de la sexualidad hegemónica masculina.

Todos los problemas que hemos expuesto en este texto ilustran tan sólo una parte de los sesgos de género presentes en la IA. A pesar de los esfuerzos que se han llevado a cabo durante los últimos años, en una disciplina científica tan joven y un grupo de tecnologías aún emergentes, el estudio de éstos y otros problemas éticos asociados a la IA es aún insuficiente. Sin embargo, la información de la que disponemos actualmente es suficiente para afirmar que el uso de la IA producirá cambios radicales y profundos en nuestra sociedad.

En el caso de los sesgos de género, el uso de sistemas de IA que no estén basados en un marco ético sólido y una legislación conforme con los derechos y valores europeos supone una amenaza directa e irreconciliable con la democracia. La incorporación de la IA a nuestras vidas puede generar grandes beneficios, pero para ello es necesario tener en cuenta todos los problemas éticos que conlleva y abordarlos antes de poner estos sistemas en marcha de forma generalizada. Estos cambios y transformaciones nos obligan a imaginar nuevos modelos de gobernanza que compatibilicen la IA con la democracia. Nuevos modelos de gobernanza de los algoritmos que pongan en el centro la igualdad de género y que garanticen, no sólo que la IA no va suponer un retroceso en los derechos y libertades de las mujeres, sino que puede convertirse en un elemento decisivo en el camino hacia la igualdad real y efectiva.

L. O. de Z. A.

## BIBLIOGRAFÍA

- COECKELBERGH, Mark. *AI Ethics*. MIT Press, 2020.
- DILLON, Sarah, COLLETT, Clementine. «AI and Gender: Four Proposals for Future Research», 2019. Disponible en: <https://doi.org/10.17863/CAM.41459> [Consulta: 8 de febrero de 2023].
- ORTIZ DE ZÁRATE ALCARAZO, Lucía, GUEVARA-GÓMEZ, Ariana. *Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España*. Fundación Alternativas, 2021. Disponible en: <https://www.fundacionalternativas.org/estudios-de-progreso/documentos/documentos-de-trabajo/inteligencia-artificial-e-igualdad-de-genero-un-analisis-comparado-entre-la-ue-suecia-y-espana> [Consulta: 8 de febrero de 2023].
- WAJCMAN, Judy. «Reflections on Gender and Technology Studies: In What State is the Art?», *Social Studies of Science*, 30(3), 2000, pp. 447-464.
- WEST, Mark, KRAUT, Rebecca, CHEW, Han Ei. «“I’d Blush If I Could” Closing Gender Divides in Digital Skills through Education». EQUALS y UNESCO, 2019. Disponible en: <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1.v> [Consulta: 8 de febrero de 2023].