Nure Inv. 22 (138)
https://doi.org/10.58722/nure.v22i138.2747
nure investigación
Editorial
Open Access

Los sesgos de género en la inteligencia artificial: por qué ocurren y cómo corregirlos

Autora: *Ana Belén Salamanca Castro*

*** Dirección de contacto:** *nureinvestigacion@fuden.es*

Diplomado y Grado en Enfermería. Máster en Cuidados Perinatales y la Infancia. Máster en Salud y Género. Experto en Metodología de la Investigación en Ciencias de la Salud. Directora de la revista NURE Investigación.

Desde hace ya décadas los científicos consideraban la posibilidad de crear una máquina que pudiera pensar como un ser humano. De hecho, algunas tecnologías con inteligencia existen desde hace más de 50 años y el término "inteligencia artificial" *(artificial intelligence)* fue acuñado por John McCarthy en 1956 (1). No obstante, la irrupción de la inteligencia artificial (IA) en situaciones cotidianas de nuestras vidas es aún reciente y el hecho de que su utilización se haya extendido a la población general nos abre un mundo de posibilidades, pero también nos sitúa ante nuevos retos y problemas que deben ser abordados, habida cuenta de los sesgos que la IA puede perpetuar.

Actualmente no existe una definición única de IA. En algunos casos, se definen como máquinas que se comportan como humanos o que son capaces de realizar acciones que requieren inteligencia (2) (lo que incluye la realización de procesos de percepción, análisis, razonamiento y aprendizaje). La Comisión Europea adopta la definición de IA del *High Level Expert Group on Artificial Intelligence* (grupo de expertos de alto nivel sobre inteligencia artificial), quienes la definen como "sistemas de software (y posiblemente también de hardware) diseñados por humanos que, ante un objetivo complejo, actúan en la dimensión física o digital percibiendo su entorno, a través de la adquisición e interpretación de datos estructurados o no estructurados, razonando sobre el conocimiento, procesando la información derivada de estos datos y decidiendo las mejores acciones para lograr el objetivo dado" e indican que "los sistemas de IA pueden usar tanto reglas simbólicas como aprender modelos numéricos, y pueden también adaptar su comportamiento analizando cómo el entorno resulta afectado por sus acciones previas" (2).

Las inteligencias artificiales basan su funcionamiento en el uso de algoritmos y modelos matemáticos para procesar grandes cantidades de datos y tomar decisiones basadas en patrones y reglas establecidas a través del aprendizaje automático (dado que pueden aprender de forma autónoma a partir de datos, sin necesidad de ser programada específicamente para hacerlo). Esos algoritmos inicialmente son una creación humana (puesto que son las personas las que elaboran las reglas que el algoritmo utilizará), pero tienen la capacidad de aprender de la experiencia, desarrollando nuevas reglas o directrices que van más allá de las inicialmente programadas y por eso se denominan algoritmos inteligentes (3). Es esa capacidad de aprender la responsable de que podamos tener recomendaciones personalizadas basadas en la música que escuchamos o las series que vemos, por ejemplo.

Russell y Norvig clasifican la IA en cuatro tipos: los sistemas que piensan como humanos, los que razonan como humanos, los que piensan racionalmente y los que actúan racionalmente. La actuación de forma racional supone que la IA no considera necesariamente el comportamiento humano, sino la información disponible dado (1) dado que, para estos autores: un sistema es racional si hace "lo correcto" dado lo que sabe (2), y ahí radica el primer reto: alimentar adecuadamente al sistema, proporcionarle información insesgada y fidedigna. Si "lo que sabe" el sistema, los datos a los que accede o los algoritmos que utiliza se encuentran sesgados, lógicamente la decisión que tome también lo estará.

El sesgo algorítmico se produce cuando los errores sistemáticos en los algoritmos (la forma en que el equipo de ciencia de datos recopila y codifica los datos de

Nure Investigación está comprometida con el movimiento Open Access. Como revista que publica en acceso abierto ofrece acceso gratuito e inmediato a todos sus contenidos. Todos los trabajos se publican bajo licencia Creative Commons, según la modalidad Reconocimiento-NoComercial (CC-BY-NC), que permite el uso de la obra siempre que no sea con fines comerciales y se reconozca la autoría de su creador. https://creativecommons.org/licenses/by-nc/4.0/

entrenamiento) originan resultados injustos o discriminatorios, que a menudo reflejan o refuerzan sesgos socioeconómicos, raciales y de género (4).

En el caso de los sesgos de género en la IA, este ocurre cuando la IA establece una diferencia en el trato según el género de la persona bien porque es lo que aprende de los datos sesgados con que se alimenta o bien porque utiliza un algoritmo cuya formulación es sesgada.

En el caso de los datos que la nutren la IA, cuando un determinado grupo se encuentra sobrerrepresentado en las bases de datos con las que se entrenan y nutren los algoritmos (por ejemplo, hombres blancos) la IA, que funciona buscando patrones, se especializará en ese determinado grupo en detrimento del grupo infrarrepresentado (y, por ejemplo, será capaz de realizar exitosamente reconocimientos faciales en el caso de hombres, pero no tanto en mujeres, siguiendo este ejemplo) (5).

En relación a los algoritmos, como ya se ha indicado, la IA aprende de la experiencia y puede desarrollar algoritmos inteligentes que, en algunas ocasiones ni los propios ingenieros pueden rastrear los pasos que ha seguido la IA para generarlos. En estos casos, se habla de "algoritmos de caja negra" *(black box algorithms)*. El uso de este tipo de algoritmos está prohibido en la Unión Europea (5). Por contra, los sistemas de IA transparentes documentan y explican claramente la metodología del algoritmo que subyace, así como quién lo entrenó (4).

Consecuentemente, los datos que nutran a la IA y los algoritmos que esta utilice pueden hacer que la IA replique y perpetúe los estereotipos y sesgos de género que observamos en la sociedad actual. Son ejemplos de la existencia de este tipo de sesgos el hecho de que la mayoría de asistentes virtuales tengan, por defecto, voz femenina (reforzando el rol de servicio que se atribuye al sexo femenino); los estereotipos y sesgos en el lenguaje que se observan con los programas de traducción online (donde se asigna el femenino a la palabra *nurse* y el masculino a *doctor* o *lawyer*, por citar algún ejemplo); o la discriminación de aquellos algoritmos que favorecen a hombres frente a mujeres para la contratación laboral o que los algoritmos que deciden las líneas de crédito proporcionen importes menores a las mujeres, pese a que se hagan declaraciones conjuntas (3,4). La existencia de este tipo de sesgos de género en una tecnología que cada vez se encuentra más universalizada no cabe duda que menoscaba las posibilidades de las mujeres, que ya son víctimas de desigualdades estructurales. Por otro lado, estos sesgos hacen que la IA participe en el afianzamiento y la normalización de prejuicios y situaciones discriminatorias entre hombres y mujeres y, en una suerte de ciclo que se retroalimenta, sus resultados seguirán alimentando algoritmos que, lógicamente, serán cada vez más sesgados.

Pero además la infrarrepresentación de determinados grupos en los algoritmos predictivos de sistemas de diagnóstico asistido por ordenador puede hacer puede hacer que los resultados sean menos precisos en el caso de personas negras o mujeres, por ejemplo.

Es preciso, por ello, emprender acciones que ayuden a romper ese círculo que se retroalimenta, proporcionando a la IA datos e información inclusiva y diversa (incorporando también a los grupos minoritarios o infrarrepresentados en el desarrollo de sistemas de IA) y revisando de forma periódica si el sistema utiliza algún algoritmo sesgado. No cabe duda que una IA es tan buena como los datos que la entrenan y por ello los datos deben representar a todos los grupos de personas y reflejar la demografía real de la sociedad. Además, la monitorización y las pruebas continuas pueden ayudar a detectar y corregir algoritmos sesgados y, por tanto, se deben utilizar algoritmos transparentes, que permitan comprender cómo se entrenan y ajustan los sistemas de IA y cómo toman sus decisiones (5,6). Del Villar incluye la necesidad de contar con marcos éticos sólidos para los sistemas de IA e integrar políticas sensibles al género en el desarrollo de estos sistemas. En cualquier caso, como esta experta en IA indica, la concientización y educación de la población es fundamental ya que entender cómo funciona la IA y sus posibles sesgos. Este conocimiento ayudará a reconocer y prevenir sistemas sesgados, y a conservar la supervisión humana en los procesos de toma de decisión (6).

Referencias Bibliográficas

1. Gobierno de España. Qué es la inteligencia artificial. Noticia (14 sep 2023). [Citado 1 sep 2025]. Disponible en: *https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr*
2. European Commission. AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence. JRC Technical Reports. Luxemburgo: Publications Office of the European Union; 2020.
3. Flores Anarte L. Sesgos de género en la inteligencia artificial: el estado de derecho frente a la discriminación algorítmica por razón de sexo. Revista Internacional de Pensamiento Político. 2023;18:97-122
4. Jonker A, Roberts J. Qué es el sesgo algorítmico. IBM [Internet]. [Citado 15 sep 2025]. Disponible en: *https://www.ibm.com/es-es/think/topics/algorithmic-bias*
5. Ortiz de Zárate Alcarazo L. Sesgos de género en la inteligencia artificial. [Citado 24 sep 2025]. Disponible en: *https://ortegaygasset.edu/wp-content/uploads/2023/03/RevistadeOccidente_Marzo2023_L.Ortiz_de_Zarate.pdf*
6. ONU Mujeres. Cómo la inteligencia artificial refuerza los sesgos de género y qué podemos hacer al respecto. (Actualizado 5 feb 2025). [Citado 25 sep 2025]. Disponible en: *https://www.unwomen.org/es/noticias/entrevista/2025/02/como-la-inteligencia-artificial-refuerza-los-sesgos-de-genero-y-que-podemos-hacer-al-respecto*