

Galatea o el problema de género en la inteligencia artificial

La inteligencia artificial y la robótica presentan importantes sesgos de género



por Pablo Rodríguez Canfranc



El desarrollo actual de la inteligencia artificial conlleva un elevado volumen de sesgos de género, que proceden tanto de la influencia social, que traen consigo los datos con los que se alimenta el *machine learning*, como de los los prejuicios inconscientes que vierten en el diseño los propios creadores de algoritmos.

De acuerdo con el mito que nos transmite Ovidio, Pigmalión, rey de Chipre y escultor, modeló la estatua de una mujer perfecta de la que se enamoró perdidamente. La diosa Afrodita, conmovida por la pasión del monarca, le otorgó el don de convertir a Galatea –pues ese era el nombre de la escultura- en humana, para que se convirtiese en su mujer. Se trata de una leyenda más, dentro de la cultura de Occidente, de una larga tradición que parte de la idea de que la mujer es creada o formada a partir del varón, y que por tanto es inferior a él y carece de voz. Al igual que en la historia de Adán y Eva que aparece en el Génesis, el varón desempeña un papel de transmisor de la sabiduría, que ella recibe de forma pasiva y sumisa.

En palabras de la profesora María del Carmen Rodríguez, este mito ha servido históricamente para transmitir un mensaje: “aquél que concede capacidad de actuación a un ser masculino sobre una representación femenina,

estática y sin voz, ya sea escultórica o literaria”. Es decir, que la leyenda se convierte en vehículo para difundir los valores patriarcales que han regido en la sociedad a lo largo de las distintas épocas.

Nuestro mundo tecnológico en plena transformación digital no ha conseguido todavía esquivar el mensaje machista que lleva aparejado el mito de Pigmalión y Galatea. La inteligencia artificial y la robótica son dos campos en los que los sesgos de género se hacen especialmente evidentes. Por un lado, la supuesta neutralidad de la tecnología se ve deformada al tener que aprender los sistemas inteligentes de los humanos –el *machine learning*–, puesto que adquieren todos los prejuicios y comportamientos sexistas que siguen bien presentes en la sociedad. La inteligencia artificial se vuelve más peligrosamente sexista a medida que los sesgos de género se filtran en los algoritmos, programas y diseños robóticos.

Un buen ejemplo de esto fue la desastrosa experiencia de Microsoft con su robot de Twitter Tay. Se trataba de un chatbot femenino con inteligencia artificial que debía aprender a través de sus conversaciones con los usuarios de Twitter. Entró en funcionamiento en marzo de 2016 y en unas pocas horas comenzó a desarrollar un comportamiento sexista y racista tras sus interacciones con un público no siempre bienintencionado, de forma que la empresa decidió apagarla a las veinticuatro horas de su debut en las redes. Tay había aprendido de los humanos, pero, por desgracia, del lado más oscuro de la naturaleza humana.

“ Algunos diseñadores de autómatas humanoides convierten el desarrollo de un robot femenino en un proyecto masculino de creación artificial de la mujer perfecta ”

Y luego está el “complejo de Pigmalión” de algunos diseñadores de autómatas humanoides que, como dice la profesora de antropología Lucy Suchman (citada por Erika Hayasaki), convierten el desarrollo de un robot femenino en un “*proyecto masculino de creación artificial de la mujer perfecta*”. A menudo estos diseños tienen rasgos de mujer estereotipada e hipersexualizada, como el robot Sophia de Hanson Robotics, características innecesarias para las funciones que deben desempeñar, o el prototipo chino Jia Jia, que es una recreación hiperrealista de una bellísima mujer oriental.

Los sesgos de género en la inteligencia artificial

Aparte de la asignación de roles de género en la robótica, lo que implica, por ejemplo, que los robots militares sean por norma masculinos y los sociales o relacionales femeninos, existen otros sesgos de género que afectan a la inteligencia artificial, más allá del mero diseño de andróides. Un reciente informe de la Fundación Alternativas los clasifica en dos grupos: sesgos de datos y de diseño.

El primer aspecto son los datos, el combustible de la inteligencia artificial actual basada en *machine learning* o aprendizaje automático. Estos algoritmos aprenden a base de consumir ingentes cantidades de datos –cuanto más volumen dispongan de ellos, más precisos serán sus diagnósticos–, sin embargo, nadie puede garantizar que la información que reciben los algoritmos sea realmente representativa. De hecho, se trata de datos que tienen los mismos sesgos que existen en las sociedades. Por ejemplo, el primer concurso de belleza en contar con un jurado no humano, Beauty.AI 2.0, demostró que el algoritmo había desarrollado un sesgo racista que le llevó a puntuar más bajo a las mujeres asiáticas o de color. Probablemente, el sesgo se produjo porque el sistema fue alimentado a base de modelos de belleza más occidentales.

En el aspecto de género, los datos pueden no ser lo suficientemente representativos de lo que ocurre en la

sociedad y ocultar la presencia y el papel de las mujeres en la misma, convirtiéndolas en invisibles. Otro ejemplo que ilustra esto es la experiencia de Amazon, que empezó a utilizar la inteligencia artificial para llevar a cabo las contrataciones, entrenando al algoritmo sobre el ideal de empleado utilizando datos de las contrataciones llevadas a cabo en los diez años precedentes. El problema es que los puestos técnicos durante ese periodo habían sido ocupados en su mayoría por hombres, por lo que el sistema determinó que las candidatas no se ajustaban al perfil requerido por la empresa. Los datos que aparentemente son neutros esconden poderosos sesgos.

La segunda categoría corresponde a los sesgos de diseño, es decir, los prejuicios que vierten en la inteligencia artificial los propios diseñadores de la misma. Por ejemplo, es muy común que los chatbots de atención al público tengan una voz femenina, llevando al plano tecnológico el rol estereotipado de la mujer como asistente, como recepcionista, o como empleada de atención al cliente. En España nos encontramos numerosos ejemplos de esto, como Lola de la Universidad de Murcia, Eva de Vueling, o Rosi, que contesta preguntas sobre información básica de Murcia.

“ Es muy común que los chatbots de atención al público tengan una voz femenina, llevando al plano tecnológico el rol estereotipado de la mujer como asistente, como recepcionista, o como empleada de atención al cliente ”

También los asistentes de voz suelen tener voces femeninas -Siri de Apple, Alexa de Amazon, Cortana de Microsoft, y Google Assistant-, y sus creadores les asignan una personalidad específica de mujer. Por ejemplo, James Giangola, uno de los desarrolladores de Google Assistant, reconoció ante un medio de comunicación que habían concebido al asistente como una mujer joven del estado de Colorado, hija pequeña de una pareja de investigadores. El estereotipo de género en los asistentes inteligentes tiene su paradigma en la respuesta que da Siri ante la expresión “Eres una zorra”, que es “Me sonrojaría si pudiese”, frase que dio título a un conocido informe de Unesco sobre brechas de género tecnológicas.

Una de las causas de esto es la escasa presencia de mujeres en los equipos que desarrollan sistemas basados en inteligencia artificial. De acuerdo con un reciente informe de la Fundación Alternativas, en la Unión Europea la brecha digital de género en actividades de programación era de 8 puntos en 2014; las mujeres solo representan el 12% de los autores de artículos en las principales conferencias sobre *machine learning* y el 13,83% de los miles que se escriben en general sobre inteligencia artificial. Por otro lado, un estudio realizado en 2017 por Emerj demostró que sólo el 13% de las altas posiciones ejecutivas en empresas tecnológicas dedicadas a la inteligencia artificial son ocupadas por mujeres, y, si nos centramos en el subámbito del lenguaje natural, el porcentaje cae hasta el 5%.

Suecia, inteligencia artificial por la igualdad

La Unión Europea lleva ya tiempo poniendo mucho énfasis en el desarrollo de una inteligencia artificial europea robusta, transparente y centrada en las personas. El objetivo es que la tecnología contribuya a mejorar el bienestar y la libertad humana y, aunque en sus estrategias no hace una alusión directa y específica a la igualdad entre hombres y mujeres, sí que aboga por facilitar el cumplimiento de los Objetivos de Desarrollo Sostenible establecidos por Naciones Unidas, entre los que se encuentra la promoción de la igualdad de género.

De todo el continente, Suecia es el país que más ha avanzado en uso de la inteligencia artificial para acabar con la discriminación por razones de género. En este sentido, la tecnología se plantea como una herramienta para poder identificar la discriminación que sufren las mujeres, y también como una palanca para promover la inclusión.

En primer lugar, en el marco del proyecto Ceretai de la Agencia Sueca de Innovación (Vinnova) se desarrollan herramientas automatizadas y diseñadas para la detección de normas, patrones y estereotipos discriminatorios presentes en la cultura popular, con el fin de cuantificar y visibilizar comportamientos discriminatorios que sufren las mujeres, y que a menudo pasan inadvertidos. En la misma línea, la iniciativa NoBias utiliza el *machine learning* para eliminar los prejuicios y las normas a través de explicaciones que muestran usos no inclusivos y discriminatorios del lenguaje, con el fin de difundir el lenguaje inclusivo en el mundo de la empresa.

En el ámbito de la educación, Suecia plantea utilizar la inteligencia artificial para evaluar tareas y exámenes de modo que se eviten sesgos por género (proyecto Gradecam), y también para detectar casos de *bullying* desde momentos tempranos y para evitar sesgos de género en los procesos de admisión en las universidades.

Con el fin de promover la igualdad económica entre hombres y mujeres en el ámbito laboral, se persigue analizar las bases de datos de las empresas para identificar desigualdades en el salario entre hombres y mujeres. Por su parte, el proyecto Rikare II busca reducir los sesgos de género en la concesión de financiación para proyectos emprendedores de áreas tradicionalmente masculinas, a través de un algoritmo.

Dentro de la lucha contra la violencia de género, la inteligencia artificial se utiliza para detectar comportamientos agresivos o violentos hacia las mujeres, analizando llamadas de emergencia, imágenes de vídeo o publicaciones en redes sociales.

Foto de [Magda Ehlers](#) en [Pexels](#)

Bibliografía

Hayasaki, E. (2017) "Is AI Sexist?" en FP. Disponible en: <https://foreignpolicy.com/2017/01/16/women-vs-the-machine/>

Huertas Sánchez, M. A. (2021) "¿Por qué es necesario incorporar visión de género a la robótica y la inteligencia artificial?" en *The Conversation*. Disponible en: <https://theconversation.com/por-que-es-necesario-incorporar-vision-de-genero-a-la-robotica-y-la-inteligencia-artificial-160655>

Motyka, J. (2017) "Conoce a Jia Jia, la robot humanoide china que habla" en *Computer Hoy*. Disponible en: <https://computerhoy.com/noticias/life/conoce-jia-jia-robot-humanoide-china-que-habla-56776>

Ortiz de Zárate Alcarazo, L. y Guevara Gómez, A. (2021) "Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España". Fundación Alternativas

Rodríguez, M. del C. (2010) "El mito de Pigmalión en textos literarios y fílmicos" en *ARBOR Ciencia, Pensamiento y Cultura* CLXXXVI 741 enero-febrero (2010).

Unesco (2020) "Artificial intelligence and gender equality". Disponible en: https://en.unesco.org/system/files/artificial_intelligence_and_gender_equality.pdf

Unesco (2019) "I'd blush if I could : closing gender divides in digital skills through education". Disponible en: <http://repositorio.minedu.gob.pe/bitstream/handle/20.500.12799/6598/1%27d%20blush%20if%20i%20could%20closing%20gender%20divides%20in%20digital%20skills%20through%20education.pdf>