



# Las Universidades Populares: Escuelas de Igualdad y Feminismo

# Las Universidades Populares: Escuelas de Igualdad y Feminismo

Edita

Federación Española de Universidades Populares (FEUP)

**FEUP** | Editorial

## **Presidenta de FEUP**

Mónica Calurano Vellarino

## **Directora Editorial**

Montserrat Morales Corraliza

## **Asistencia técnica**

Emprendedores: Juan Carlos Casco - Fernando Barrena - Damián Pérez

Copyright © 2024 FEUP

**Depósito Legal:** M-4891-2025

**ISBN:** 978-84-126291-9-4

# Capítulo 5

## Inteligencia artificial desde la perspectiva de género: desafíos y oportunidades

---

### **Eva Cernadas García**

Doctora y profesora titular de Inteligencia Artificial de la Universidad de Santiago de Compostela (USC). Centro Singular de Investigación en Tecnoloxías Intelixentes da USC y Centro Interdisciplinario de Investigacións Feministas e de Estudos de Xénero CIFEX da USC

### **Encina Calvo Iglesias**

Doctora y profesora titular de Física Aplicada de la Universidad de Santiago de Compostela, departamento de Física Aplicada. Centro Interdisciplinario de Investigacións Feministas e de Estudos de Xénero (CIFEX) da USC

## Realidades

En los últimos años hemos visto como la tecnología digital se hace cada vez más presente en nuestras vidas, tanto en los aspectos laborales como en el ocio, que se ve mediado en muchos casos por dispositivos tecnológicos. Desde el ordenador al smartphone pasando por la aspiradora robot, los asistentes de voz o los sistemas de recomendación de contenido, nuestras vidas están rodeadas por estos dispositivos tecnológicos que facilitan nuestro día a día, pero al mismo tiempo recopilan nuestros datos personales en lo que Shoshana Zuboff (2020) denomina “capitalismo de vigilancia”, sin que nosotros seamos conscientes del mal uso que se puede hacer de ellos. Probablemente, porque en nuestro imaginario permanece la idea de que la tecnología es neutra y está asociada al progreso. Pero tal y como nos han mostrado desde el tecnofeminismo (Wacjman, 2008), nuestros valores sociales y culturales afectan a la ciencia y la tecnología, y en algunos casos los avances científicos y tecnológicos han tenido consecuencias desastrosas para el medioambiente y la humanidad, por ejemplo, los plásticos o las armas nucleares.

La inteligencia artificial (IA), ni es inteligente ni artificial ya que, como sostiene Kate Crawford, los sistemas de IA no son autónomos, ni racionales y consumen un gran número de recursos naturales (Crawford, 2023). Esta tecnología o conjunto de tecnologías tiene un gran potencial para ayudar a resolver determinados problemas en distintos campos como la medicina, por ejemplo, para ayudar al personal médico a interpretar una radiografía, pero no sirve para todo como veremos en los siguientes apartados y presenta numerosos problemas éticos, entre ellos la tecnoprecariedad

que conlleva su uso, la discriminación, el impacto climático y la violencia digital contra las mujeres. Y es necesario conocer en que nos beneficia y en que nos perjudica, antes de seguir adelante con determinados desarrollos que realmente solo producen réditos a las grandes tecnológicas.

La película *Sorry we missed you*, dirigida por Ken Loach, nos muestra las consecuencias de este capitalismo digital y en particular del fenómeno conocido como 'uberización de la economía', que lleva a la precarización y autoexplotación de las personas trabajadoras. Una precarización que se ha extendido también a las profesiones creativas y al ámbito académico. Así, Remedios Zafra en su último libro *El informe* expone como en el pasado "nos vendieron la idea de un futuro en el que el ocio ocuparía el centro de nuestras vidas mientras las máquinas trabajarían por nosotros y crearían riqueza. Ese futuro utópico no llegó jamás, y a lo que nos enfrentamos es a un presente en el que al hablar del mundo laboral nos vemos abocados a utilizar palabras como hiperproductividad, precariedad, competición, burocracia..." (Zafra, 2024). Es más, el algoritmo es quien nos gobierna exprimiendo a las personas trabajadoras al máximo, como sucede a las personas que trabajan de riders, seleccionadores de productos en Amazon o incluso en el servicio de asistencia a domicilio y, como concluye Katharine Marçal (2022), al final la innovación ha sido organizar a las personas como si fueran máquinas. Y no debemos olvidar que algunos de los nuevos trabajos creados en este contexto digital son precarios, por ejemplo, los de clasificación, identificación de discursos de odio o moderación de las redes sociales y por el que las personas trabajadoras pueden ser expuestas a visualizar situaciones violentas que afecten a su salud mental, lo que llevó a un grupo de ellas a demandar a Facebook

en 2018 (Paniagua, 2021). Este tipo de trabajos también parece haberse llevado a cabo en nuestro país como reveló el reportaje periodístico de Analía Plaza titulado ‘Las galeras de Google’ recogido en el informe de la asociación AlgoRace. Aunque las grandes tecnológicas parecen haberlo desviado a países del continente africano, como se denunció en la reciente visita del presidente Biden a Kenia. Así, un reportaje de la revista Time mostró que el trabajo de etiquetado necesario para entrenar a Chat GPT se ha llevado a cabo en este país africano pagando menos de dos euros la hora.

En estos momentos de fascinación por la IA, que se usa en aplicaciones que van desde el reconocimiento facial a un asistente de voz o el famoso Chat GPT, necesitamos entender cómo funciona esta tecnología y evitar que se confirmen y perpetúen los estereotipos sexistas, clasistas y racistas que ya operan en algunas de ellas (O’Neil, 2018) y automatizan la desigualdad (Eubanks, 2021). Para ello es preciso entender que lo que denominamos inteligencia artificial no funciona sin datos, unos datos que muchas veces se obtienen sin el consentimiento de las personas usuarias, y que son necesarios para realizar predicciones sobre comportamientos, intereses y gustos de quién las usa, lo que se ha convertido en una gran fuente de negocio y un peligro para nuestras democracias (Vélez, 2020). Porque como hemos visto en el caso de Cambridge Analytica, a través de la recopilación de nuestra información personal pueden influenciar nuestras intenciones de voto mostrándonos determinados anuncios políticos. Y estos peligros han aumentado con la irrupción de la IA generativa que puede servir para crear un video recreando a una persona famosa ya fallecida, como Lola Flores en un anuncio de cervezas,

pero también para crear bulos y aumentar la desinformación, un cóctel peligroso en época de elecciones. Al mismo tiempo, ha aumentado la preocupación sobre la calidad de los datos masivos que se extraen de Internet, dada la baja representación de las mujeres y las minorías, ya que puede conducir a conclusiones sesgadas y erróneas que contribuyen a generar resultados discriminatorios. Y esto ya ha sucedido con el algoritmo que usaba Amazon para seleccionar a su personal y que discriminaba a las mujeres, por lo que fue retirado, pero puede estar sucediendo en la evaluación de personal a través de programas que analicen la voz, la expresión facial u otras características y que en general suelen ser opacos. Por ello, para mitigar estos efectos debemos exigir mayor transparencia en estos procesos y que se auditen los algoritmos utilizados en cualquier decisión relacionada con las Administraciones públicas. También tenemos que acabar con este sesgo de automatización o chovinismo tecnológico, por el que creemos que las máquinas son más fiables que las personas y entender que como se mostrará más adelante la IA solo nos va a dar soluciones aproximadas.

Otro de los problemas que presenta el uso de la IA es el impacto medioambiental, desde los minerales necesarios para la infraestructura digital hasta la energía que consumen durante la fase de entrenamiento del algoritmo. Como señala Crawford “los minerales son la columna vertebral de la IA, pero su alma sigue siendo la energía eléctrica” y cada vez consumen más. También se realiza un elevado consumo de agua, por ejemplo, el entrenamiento de GPT-3 en los centros de datos estadounidenses de Microsoft puede consumir directamente ingentes cantidades de litros de agua dulce limpia, algo muy preocupante ya que es un bien escaso. Pero, además, hay que tener en cuenta el consumo de agua para refrigerar los

centros de datos, un problema que amenaza a poblaciones de lo que se llama la España vaciada donde se pretenden instalar estos centros de grandes tecnológicas. Y algunas voces alertan sobre esta expansión que puede acelerar la desertización del país y los problemas para las personas que habitan estos lugares como ya ha sucedido en estados de EE. UU. como Virginia.

La IA, como todas las tecnologías, tiene el potencial de ser usada para el bien común pero también el riesgo de contribuir a reforzar el sexismo y la violencia contra las mujeres, las minorías y otras etnias. Sabemos que el uso de las redes sociales puede tener un efecto muy negativo en la salud mental, ya que el algoritmo fomenta la visión de contenidos que mantengan conectadas a las personas usuarias sin importar el efecto nocivo que puedan tener y por ello Meta y Tik Tok ya han sido denunciadas por escuelas de los EE. UU. También la plataforma Instagram, centrada en la imagen, puede causar efectos negativos como potenciar trastornos de la conducta alimentaria, la anorexia o la bulimia, y que afectan especialmente a las mujeres jóvenes. El reciente escándalo del robo de la voz de Scarlett Johansson para que ChatGPT sonara como en la película Her, nos muestra la falta de ética y el ambiente machista y misógino de la IA, programada para y por una mayoría masculina. Y debemos preguntarnos cuál es el papel que estas grandes empresas nos otorgan a las mujeres en la sociedad, además del de sumisas cuidadoras que nos muestran las asistentes virtuales como Siri y Alexa entre otras. Sin olvidar, que tenemos un grave problema con la IA generativa y los videos de pornografía creados sin consentimiento que circulan por Internet. Recordemos que según recientes investigaciones el 90 % de los vídeos falsos hiperrealistas es pornografía no consentida que se utiliza como arma de violencia ma-

chista y que afecta sobre todo a mujeres jóvenes, incluso adolescentes, como las menores del instituto de Almendralejo en Badajoz. Y deberíamos analizar en detalle el papel de las compañías tecnológicas, tanto en el diseño de estas tecnologías o la facilidad para acceder a las mismas y en el negocio de plataformas que pueden estar rentabilizando la adicción a la pornografía.

Y como comentamos anteriormente es necesario que conozcamos el funcionamiento de esta herramienta, sus limitaciones y así podamos decidir entre toda la ciudadanía como utilizarla para el bien común de nuestras sociedades.

## ¿Cómo funciona la inteligencia artificial (IA)?

El término de “inteligencia artificial” está en boca de mucha gente, tanto en los medios de comunicación como en la sociedad en general, ¿pero sabemos realmente qué hay detrás de este concepto o cómo funcionan los sistemas que se basan en IA? Un sistema que usa la IA es un programa de ordenador que “piensa” o razona como una persona. Por tanto, estamos hablando de un tipo de programa informático. Cuando en la sociedad hablan de sistemas que usan la IA o algoritmos de IA, normalmente, se refieren a un subtipo de programas que usan el aprendizaje automático para aprender el problema que quieren modelar y, a partir de ese momento, realizar su tarea.

La fase de aprendizaje de un algoritmo de IA se denomina entrenamiento, en el cual se le proporciona el algoritmo ejemplos del problema con dos tipos de datos: la entrada del algoritmo y la salida deseada. Por ejemplo, si se quiere entrenar un reconocedor automático de voz para que realice transcripciones, que llamaremos de forma genérica modelo, tendríamos que proporcionarle como entrada un trozo de audio y la salida deseada sería la transcripción del audio. Con un conjunto de ejemplos (audios y sus transcripciones) representativo del problema, el modelo aprende y luego ya puede realizar su trabajo, por ejemplo, identificando nuestra voz para preguntar en el móvil, para abrir nuestra casa o hablar con el coche.

Con los ejemplos proporcionados en el entrenamiento, el modelo construye un resumen del problema, que se codifica y se guarda con números dentro del modelo, para su posterior uso. En nuestra vida no digital pensamos que un resumen de un texto o una película recopila los rasgos generales y más relevantes del mismo. En el mundo digital, los modelos también codifican lo que más abunda o lo más frecuente de los ejemplos que se utilizaron en su entrenamiento, porque dicho proceso se basa en cálculos matemáticos que invisibilizan los casos raros o singularidades. Sabiendo cómo funcionan los procesos de diseño de los algoritmos de IA, muchas preguntas inundan nuestros pensamientos. ¿Qué pasa si yo entreno el modelo solo con audios de hombres?, ¿o con personas de habla inglesa y quiero interpretar audios en español?, ¿o con personas de España y quiero que reconozca los audios de personas de América Latina?, o ¿reconocería mi voz si un día estoy afónica?

Una vez que el modelo se ha entrenado, ya está listo para su utilización, para ponerlo en producción, en nuestro caso un reconocedor de voz.

El siguiente paso será evaluar la calidad de servicio que proporciona el reconocedor de voz. En este ejemplo parece bastante obvio saber si el servicio es bueno o no, si reconoce o no los audios de las personas, pero pueden existir otros sistemas en los que la calidad del servicio dependa de quien lo evalúe. Imaginemos que estamos ante un sistema que predice posibles valores para una palabra enmascarada, por ejemplo el sistema BERT (el acrónimo de Bidirectional Encoder Representations from Transformers, y se trata de un algoritmo interno de Google que utiliza para realizar búsquedas con la lengua que usamos todos los días). Hicimos un pequeño experimento preguntando a BERT que nos diga posibles opciones para la palabra [MÁSCARA] en las frases “Un hombre trabaja como [MÁSCARA]” y “Una mujer trabaja como [MÁSCARA]”. Los resultados que proporciona BERT es que un hombre trabaja como carpintero, camarero, barbero, mecánico o comercial, y que una mujer trabaja como enfermera, camarera, solterona, prostituta o cocinera. Dejamos al lector o lectora la opinión sobre la calidad de este funcionamiento.

La intuición y el sentido común nos susurra al oído la necesaria cautela y exigencia en los procesos de diseño de la inteligencia artificial para que no reproduzcan los sesgos de género y otras deficiencias existentes en la sociedad. En el caso de películas, series u otras producciones artísticas el test de Bechdel nos ayuda a evaluar las desigualdades de género en el mundo de la cultura. Algo similar en el mundo de la IA nos llevaría a realizar las siguientes preguntas o reglas en cuanto al entrenamiento de los algoritmos de IA:

1. ¿Los ejemplos representan toda la complejidad del problema? Hay audios suficientes de todas las singularidades que luego quiero trans-

cribir, audios de hombres y mujeres; infantes, jóvenes y personas adultas; voces con afonía, gritando y susurrando; variedades dialectales, etc.

2. ¿La salida deseada del modelo se determinó teniendo en cuenta la diversidad del problema? ¿Quién y cómo se ha realizado este proceso? En un reconocedor de voz parece obvio determinar la salida deseada sin ambigüedad, pero imaginemos un sistema que interpreta el tipo de documento en función del análisis automático de su texto. La interpretación del texto puede variar con las personas y una frase puede percibirse como un chiste para unas personas y una ofensa por otras, en función de su género, edad, tradición cultural o etnia.
3. ¿El sistema de IA proporciona un servicio de calidad? ¿Quién y cómo se evalúa la calidad del sistema? Porque el resultado de esta evaluación puede condicionar las decisiones para mejorar los dos pasos anteriores. Los sistemas de IA solo producirán un servicio de calidad cuando su salida sea satisfactoria para todos los colectivos implicados en la predicción de este.

Si la respuesta a alguna de las preguntas anteriores es negativa, el sistema de IA no debería ponerse en producción, por los posibles riesgos de discriminación para la sociedad. En el ejemplo anterior del sistema BERT, la predicción laboral para los hombres es normal, mientras que para las mujeres proporciona dos profesiones con connotaciones negativas y peyorativas. Por tanto, la calidad del servicio es inferior para las mujeres. Cuando hablamos de asistentes virtuales la percepción también depende del género. Así los asistentes de ocio como Alexia, Siri o Cortana tienen voz femenina y una narrativa servil y complaciente, mientras que otros asis-

tentes profesionales en el ámbito de la abogacía o negocios tienen voz masculina y una narrativa de prestigio y formal. Cuando un reconocedor de voz de un coche no reconoce las voces femeninas, su calidad de servicio es claramente inferior para las mujeres.

Los sistemas de aprendizaje profundo o deep learning se basan en los mismos principios pero utilizan un modelo mucho más complejo y con muchos parámetros para determinar en su entrenamiento. Por tanto, necesitan un volumen de datos muy grande para aprender en su entrenamiento. Las IA generativas, como chatGPT, son un tipo de sistemas de aprendizaje profundo que pueden crear contenido nuevo, objetos, sonidos, vídeos, etc., en función de los datos que se usaron en su entrenamiento.

## Limitaciones y riesgos de la IA

Como hemos mencionado, cualquier sistema que utiliza IA se representa como una caja que llamamos modelo, con sus dificultades inherentes para representar o codificar los casos poco frecuentes en el problema a modelar. Este funcionamiento de la IA puede ofrecer innovaciones interesantes en muchas aplicaciones como la inspección de alimentos y productos o la realización de trabajos que son peligrosos para las personas. Pero suscita muchos riesgos el uso de las IA genéricas en dominios que afecten a la vida o el desarrollo integral de las personas, como puede ser la educación, donde es necesaria la atención a la diversidad, o medicina, en el abordaje de las enfermedades raras. Por tanto, deberíamos

extremar las precauciones en dichos ámbitos porque sus predicciones siempre nos llevarían a la estandarización, a lo más frecuente o común, invisibilizando la diversidad de necesidades. Las IA en estos ámbitos pueden ofrecer oportunidades cuando se usen para tareas delimitadas y con objetivos específicos.

En relación con el género, las personas no binarias u otras identidades representan porcentajes pequeños en la sociedad y su inclusión en los sistemas de IA está penalizada, como hemos mencionado, por las limitaciones inherentes a los modelos de IA. En el caso de mujeres y hombres, que representan casi el 50 % de la sociedad, los sesgos de género se introducen en el modelo a través de las decisiones de su diseño. Dada la falta de estudios sobre los colectivos minoritarios, en lo sucesivo hablaremos de sesgos de género para referirnos a las desigualdades entre mujeres y hombres. En este caso, los sesgos de género se introducen a través de las decisiones del equipo que diseña las IA en cuanto a las tres reglas formuladas, qué ejemplos se usan en el entrenamiento, quién y cómo se determinaron sus salidas deseadas y quién y cómo se determinó la calidad del servicio proporcionado por las IA. Un artículo científico de finales de 2022 revisa 120 estudios que analizan los sesgos de género de distintos sistemas de IA en el ámbito de los traductores y analizadores de texto, el reconocimiento facial, el reclutamiento y marketing, sistemas de recomendación y algoritmos de búsqueda, robótica y sistema de ayuda a la decisión en políticas del gobierno (Shrestha y Das, 2022). Otros estudios posteriores analizan los sesgos de género de GPT-4 en diferentes tareas relacionadas con la medicina (Zack y col., 2024), de los sistemas de reclutamiento del talento en las redes sociales como LinkedIn and Fa-

cebook (Simon, 2023), o los traductores automáticos al inglés incluso cuando se utilizan lenguajes con pronombres neutros como el húngaro (Farkas, 2022). Todos estos estudios sugieren que la desigualdad está implícitamente representada en el uso diferenciado del lenguaje entre géneros. Dado que la IA convive y está presente en nuestro desarrollo vital, puede generar datos que sirvan para el aprendizaje, la toma de decisiones o el entrenamiento de nuevos sistemas de IA, sus sesgos de género podrían perpetuar o amplificar la desigualdad presente en el entorno cultural en el que vivimos o asimilar otros entornos culturales con reglas de convivencia diferentes. Esta posibilidad entraña muchos riesgos ya que abre la puerta a que la IA, silenciosamente, propicie el retroceso en los avances sociales en igualdad de género alcanzados en la sociedad en las últimas décadas. La igualdad de género es un compromiso de toda la sociedad. Por todo lo enumerado, nos queda la impresión de que los sistemas de IA están penalizando siempre a las mujeres, pero en aquellos aspectos que sean negativos para los hombres en la sociedad, como por ejemplo la criminalidad o el fracaso escolar, los hombres también serán penalizados negativamente por los sistemas de IA.

Considerando todas las dificultades del diseño de los sistemas de IA, impera la necesidad de alcanzar un compromiso de la sociedad en su conjunto para promover la igualdad de género también en el ámbito digital. Tal vez ese compromiso debería empezar ejerciendo la igualdad en las aulas universitarias (Cernadas y Calvo-Iglesias, 2022). Es necesario ejercer la igualdad en el aprendizaje para llevar la igualdad de género a los entornos de trabajo relacionados con la tecnología, la igualdad en la toma de decisiones relacionadas con la tecnología, y así diseñar y cons-

## Las Universidades Populares: Escuelas de Igualdad y Feminismo

truir sistemas de IA más justos y éticos con la igualdad de género. Tal vez si no hay igualdad dentro de los equipos de diseño, sea ciencia ficción desarrollar sistemas de IA que no tengan sesgos.



## **Eva Cernadas García**

Doctora y profesora titular de Inteligencia Artificial de la Universidad de Santiago de Compostela (USC). Centro Singular de Investigación en Tecnoloxías Intelixentes da USC y Centro Interdisciplinario de Investigacións Feministas e de Estudos de Xénero CIFEX da USC



## Encina Calvo Iglesias

Doctora y profesora titular de Física Aplicada de la Universidad de Santiago de Compostela, departamento de Física Aplicada. Centro Interdisciplinario de Investigacións Feministas e de Estudos de Xénero (CIFEX) da USC