Estereotipos de género en imágenes generadas mediante inteligencia artificial

Gender stereotypes in AI-generated images

Francisco-José García-Ull; Mónica Melero-Lázaro

Note: This article can be read in its English original version on:
https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87305

Cómo citar este artículo.
Este artículo es una traducción. Por favor cite el original inglés:

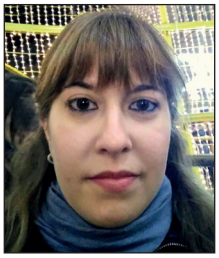
García-Ull, Francisco-José; **Melero-Lázaro, Mónica** (2023). "Gender stereotypes in AI-generated images". *Profesional de la información*, v. 32, n. 5, e320505.

https://doi.org/10.3145/epi.2023.sep.05

Artículo recibido el 06-03-2023
Aceptación definitiva: 21-07-2023

Francisco-José García-Ull ✉
https://orcid.org/0000-0002-7999-4807

Universidad Europea de Valencia
Facultad de Ciencias Sociales
Pg. de l'Albereda, 7
46010 València, España
franciscojose.garcia@universidadeuropea.es



Mónica Melero-Lázaro
https://orcid.org/0000-0002-2517-8468

Universidad de Valladolid
Facultad de Filosofía y Letras
Pza. del Campus Universitario
47011 Valladolid, España
monimeler@outlook.com

Resumen

Este estudio tiene como objetivo la identificación de sesgos de género en profesiones por parte de imágenes generadas mediante *DALL-E 2*, aplicación para la creación de imágenes sintéticas basadas en inteligencia artificial (IA). Para ello se utiliza un muestreo probabilístico estratificado en el que se delimitan segmentos a partir de 37 profesiones o *prompts*, replicando el estudio de Farago, Eggum-Wilkens y Zhang (2021) sobre estereotipos de género en el ámbito laboral en humanos. En el desarrollo del estudio intervienen 2 codificadores que introducen las profesiones manualmente en el generador de imágenes. *DALL-E 2* genera 9 imágenes por cada consulta y se toma una muestra de 666 imágenes, con un nivel de confianza del 99% y un margen de error del 5%. A continuación, se evalúa cada imagen de acuerdo a una escala Likert de 3 niveles (1. No estereotipado; 2. Moderadamente estereotipado; 3. Fuertemente estereotipado). Nuestro estudio demuestra que estas creaciones replican estereotipos de género relacionados con el trabajo. Los resultados aquí descritos exponen que el 21,6% de las imágenes de profesionales generadas mediante IA, se representan completamente estereotipadas con respecto al sexo femenino y el 37,8% con respecto al masculino. Si bien estudios anteriores realizados con humanos apuntan la existencia de estereotipos de género en el ámbito laboral, nuestro trabajo demuestra que la IA, no sólo replica este estereotipado, sino que lo refuerza e incrementa. Así, mientras que las investigaciones sobre sesgo de género llevadas a cabo en humanos demuestran un fuerte estereotipado en el 35% de los casos, la IA ofrece fuerte estereotipado en el 59,4% de los casos. Los resultados del presente estudio subrayan la necesidad de una comunidad de desarrollo de IA diversa e inclusiva, que establezca las bases hacia una IA más justa e imparcial.

Palabras clave

Inteligencia artificial; IA; *OpenAI*; *DALL-E*; Imágenes sintéticas; Estereotipos de género; Imágenes; Sesgos sexuales; Sesgos de género; Diferencias de género; Profesiones; Trabajadores; Trabajos; Ética; Discriminación; Desigualdades; Justicia; Equidad.

Abstract

This study explores workplace gender bias in images generated by *DALL-E 2*, an application for synthesising images based on artificial intelligence (AI). To do this, we used a stratified probability sampling method, dividing the sample into segments on the basis of 37 different professions or prompts, replicating the study by Farago, Eggum-Wilkens and Zhang (2020) on gender stereotypes in the workplace. The study involves two coders who manually input different professions

CC BY

into the image generator. *DALL-E 2* generated 9 images for each query, and a sample of 666 images was collected, with a confidence level of 99% and a margin of error of 5%. Each image was subsequently evaluated using a 3-point Likert scale: 1, not stereotypical; 2, moderately stereotypical; and 3, strongly stereotypical. Our study found that the images generated replicate gender stereotypes in the workplace. The findings presented indicate that 21.6% of AI-generated images depicting professionals exhibit full stereotypes of women, while 37.8% depict full stereotypes of men. While previous studies conducted with humans found that gender stereotypes in the workplace exist, our research shows that AI not only replicates this stereotyping, but reinforces and increases it. Consequently, while human research on gender bias indicates strong stereotyping in 35% of instances, AI exhibits strong stereotyping in 59.4% of cases. The results of this study emphasise the need for a diverse and inclusive AI development community to serve as the basis for a fairer and less biased AI.

Keywords

Artificial intelligence; AI; *OpenAI*; *DALL-E*; Synthetic images; AI-generated images; Imaging; Gender stereotypes; Sex biases; Gender biases; Gender differences; Professions; Workers; Jobs; Ethics; Discrimination; Inequalities; Fairness; Equity.

Financiación

Este artículo se inscribe en el marco del proyecto "Flujos de desinformación, polarización y crisis de la intermediación mediática (Disflows) (PID2020-113574RB-I00)", financiado por el *Ministerio de Ciencia e Innovación* de España.

1. Introducción

La inteligencia artificial (IA) ha tenido un gran impacto en la sociedad en los últimos años, transformando la forma en que se realizan tareas y se toman decisiones en gran variedad de ámbitos, como el jurídico (**Sourdin**, 2018); sanitario (**Loftus** *et al*., 2020); empresarial (**Nica** *et al*., 2022; **Belhadi** *et al*., 2022) o en educación (**Porayska-Pomsta**; **Rajendran**, 2019). Sin embargo, también ha surgido preocupación sobre los posibles sesgos de género en los sistemas de IA.

Uno de los problemas con los sesgos de género en IA es la falta de diversidad en los datos utilizados para entrenar los modelos. Si estos datos utilizados no reflejan adecuadamente la diversidad de la población, el modelo puede aprender y reproducir estereotipos de género (**Leavy**, 2018). Uno de los desarrollos recientes en IA es el generador de lenguaje *DALL-E 2* (2021), desarrollado por *OpenAI*. Aunque *DALL-E 2* ha mostrado rendimiento en la generación de contenido, es importante considerar cómo pueden manifestarse los sesgos de género en su uso y cómo abordarlos.
https://openai.com/dall-e-2

En el caso de *DALL-E 2*, los datos utilizados para entrenar el modelo provienen principalmente de la web, lo que significa que el modelo puede aprender estereotipos de género que se encuentran en la cultura online. Es importante considerar este tipo de limitaciones para poder delimitar un estudio riguroso.

Además, es crucial considerar cómo los sesgos de género pueden manifestarse en los resultados generados por *DALL-E 2*. Por ejemplo, un estudio realizado por **Buolamwini** y **Gebru,** (2018) encontró que los sistemas de reconocimiento facial comerciales tenían una mayor tasa de error en la clasificación de personas de color y personas de género no binario, lo que sugiere que los sistemas de IA pueden perpetuar los sesgos de género existentes en la sociedad. Dada la rápida aceptación de aplicaciones como *DALL-E 2*, *ChatGPT* u otros sistemas *OpenAI* (**Vincent**, 2020), resulta imprescindible la identificación de estas brechas y limitaciones que perpetúan modelajes y clichés. En este sentido, en los últimos años hemos asistido a la publicación de artículos, libros o documentales que nos muestran cómo las nuevas tecnologías digitales presentan sesgos de género, raza y clase (**D'Ignazio**; **Klein**, 2020; **Criado-Pérez**, 2020).

Para abordar los sesgos de género en *DALL-E 2*, se propone la creación de bases de datos libres de estereotipos que se utilicen para entrenar el modelo. De esta manera, se fomentan resultados que reduzcan dichos sesgos y promuevan la equidad de género. También se pueden utilizar técnicas de *debiasing* para eliminar los sesgos de género en los resultados generados por *DALL-E 2* (**Caliskan**; **Bryson**; **Narayanan**, 2017).

Como consecuencia, es importante considerar los posibles sesgos de género en su uso. Los sistemas de IA pueden reproducir estereotipos de género existentes en la sociedad si no se considera adecuadamente la diversidad en los datos utilizados para entrenar los modelos (**Leavy** *et al*., 2020). Es crucial abordar estos sesgos mediante la diversificación de los datos y el uso de técnicas de *debiasing*. Es importante seguir investigando y monitoreando los posibles sesgos de género en *DALL-E 2* y en otros sistemas de IA, así como sus potenciales beneficios y riesgos (**De-Carvalho**, 2021) ya que, sin duda, tendrán múltiples usos y afectarán de manera directa o indirecta a nuestras relaciones interpersonales, para garantizar que estos sistemas sean justos, equitativos y éticos (**Cortina-Orts**, 2019).

> "Uno de los problemas con los sesgos de género de la IA es la falta de diversidad en los datos utilizados para entrenar los modelos"

2. Marco teórico

2.1. *OpenAI* y la inteligencia artificial generativa

La inteligencia artificial ha experimentado un gran avance en las últimas décadas, gracias a la combinación de técnicas de aprendizaje automático y el aumento en la capacidad de procesamiento de datos. Uno de los desarrollos más recientes en este campo es *DALL-E*, un generador de IA creado por *OpenAI* que es capaz de generar imágenes a partir de descripciones de texto (**Brown** *et al*., 2020). La versión mini de *DALL-E*, conocida como *DALL-E 2*, ha sido lanzada recientemente y se ha demostrado que es capaz de generar imágenes detalladas y sorprendentemente precisas a partir de descripciones de texto simples.

DALL-E 2 es una versión mejorada de *DALL-E*, que fue lanzado en 2021. A diferencia de su predecesor, *DALL-E 2* tiene un tamaño reducido en su arquitectura y una capacidad de procesamiento de datos reducida, lo que lo hace más accesible y fácil de usar para una variedad de usuarios y aplicaciones. Sin embargo, a pesar de su tamaño reducido, *DALL-E 2* ha demostrado ser igual de eficaz en la generación de imágenes a partir de descripciones de texto.

Atendiendo a su funcionamiento, *DALL-E 2* utiliza una técnica conocida como aprendizaje profundo generativo (*Deep learning* y tecnología *GAN*), que consiste en entrenar una red neuronal para generar imágenes a partir de datos de entrada. En el caso de *DALL-E 2*, el dato de entrada es una descripción de texto, que es procesada por la red neuronal para generar la imagen correspondiente. Esta técnica se basa en el aprendizaje automático y en la capacidad de la red neuronal de aprender patrones y relaciones en los datos de entrada.

En cuanto a las aplicaciones potenciales de *DALL-E 2*, una de las más evidentes es su uso en la publicidad y el diseño gráfico. *DALL-E 2* podría utilizarse para generar imágenes personalizadas para campañas publicitarias o para crear diseños de imágenes únicas para productos. Cabe destacar que cada nueva imagen generada mediante este sistema es original e inédita, condición no exenta de controversia y que plantea límites a los derechos de autoría (**Estupiñán-Ricardo** *et al*., 2021). También puede aplicarse a la educación, ya que hace posible generar imágenes para ilustrar conceptos en libros de texto o en presentaciones de clase. Precisamente en este ámbito existen estudios que señalan los sesgos de género en la representación de mujeres en la ciencia (**Manassero**; **Vázquez**, 2003; **Francescutti**, 2018). Un buen uso de la IA podría trascender estas limitaciones en aras de una sociedad más igualitaria. Además, *DALL-E 2* tiene aplicaciones en la industria de la animación y la producción de videojuegos, ya que permite generar escenarios y personajes de manera automatizada.

Sin embargo, también hay desafíos éticos en torno a esta tecnología (**Quirós-Fons**; **García-Ull**, 2022). Una de las principales preocupaciones es la posibilidad de que *DALL-E 2* sea utilizado para generar contenido falso o engañoso. Además, hay controversia en torno a la privacidad y la seguridad, ya que *DALL-E 2* podría utilizarse para generar imágenes de individuos sin su consentimiento y ser utilizada como herramienta de violencia digital (**Pérez-Gómez** *et al*., 2020). En esta línea, **Véliz** (2021) destaca el poder de influencia que se puede ejercer mediante el tratamiento de datos privados y la necesidad de favorecer iniciativas y herramientas que velen por la privacidad de los usuarios. También haciendo hincapié en esta centralización del poder motivada por una hegemonía tecnológica, autores como **Crawford** (2021) afirman la existencia de una tendencia hacia una mayor desigualdad, instando a las empresas tecnológicas a utilizar la IA para virar el rumbo hacia valores democráticos y una remodelación del escenario político y social. Siguiendo esta línea, **O'Neil** (2018) alerta de unos algoritmos y modelos opacos y no regulados que refuerzan la discriminación, apuntalando a los afortunados y castigando a los oprimidos.

2.2. La tecnología *GAN*

Las *Generative Adversarial Networks* (*GAN*) son una clase de modelos de aprendizaje automático capaces de generar contenido nuevo y realista, como imágenes, audio y texto. Estos modelos están compuestos por dos redes neuronales: la primera se denomina generador y la segunda discriminador. El generador se encarga de generar contenido nuevo, mientras que el discriminador se encarga de determinar si el contenido generado es real o falso. Los dos modelos compiten entre sí, en un juego de suma cero, con el objetivo de mejorar la calidad del contenido generado.

El trabajo más influyente sobre las *GAN* fue presentado por **Goodfellow** *et al.* (2014) en su artículo "Generative adversarial networks". En este trabajo, los autores presentaron una arquitectura básica de las *GAN* y mostraron cómo podría utilizarse para generar imágenes de rostros humanos.

Una de las principales ventajas de las *GAN* es su capacidad para generar contenido realista. Esto ha llevado a su uso diversos campos, como la producción de videojuegos, la animación y el diseño de productos. Por ejemplo, en videojuegos para crear escenarios y personajes, y para generar imágenes de productos que se utilizan en la toma de decisiones de diseño.

Además, las *GAN* también tienen aplicaciones en el campo de la medicina. Por ejemplo, se han utilizado para generar imágenes de tomografías cerebrales para ayudar en el diagnóstico de enfermedades neurodegenerativas (**Laino** *et al*., 2022). También se han utilizado para generar imágenes de células y tejidos que ayudan en la investigación y el desarrollo de nuevos tratamientos.

Sin embargo, también hay preocupaciones éticas en torno a las *GAN*. Una de las principales es la posibilidad de que sea utilizado para generar contenido falso o engañoso (**García-Ull**, 2021; **Gamir-Ríos**; **Tarullo**, 2022).

2.3. Estereotipos de género e IA

Como hemos afirmado, la inteligencia artificial, en constante evolución, tiene el potencial de transformar la forma en que vivimos, trabajamos y nos relacionamos. Sin embargo, también plantea preocupaciones éticas y de justicia social, especialmente en relación con los estereotipos de género (**Wang** *et al*., 2019). Los estereotipos de género son creencias y expectativas sociales sobre las características, comportamientos y roles que se consideran apropiados para hombres y mujeres. Estos estereotipos pueden limitar las oportunidades y las expectativas de las personas, y pueden conducir a la discriminación y la desigualdad.

En el campo de la IA, los estereotipos de género pueden manifestarse de varias maneras. Una de las principales preocupaciones es la representación de género en los datos utilizados para entrenar modelos de IA. Si estos datos contienen estereotipos de género, es probable que el modelo reproduzca esos estereotipos. Por ejemplo, un modelo de IA que ha sido entrenado con imágenes de hombres y mujeres en roles tradicionales de género podría tener dificultades para reconocer a las mujeres en roles no tradicionales (**Agudo**; **Liberal**, 2020; **Traylor**, 2022). Esta premisa afecta directamente a contextos como el profesional o el cuidado del hogar (**Bolukbasi** *et al*., 2016). Además, tanto el tratamiento de los datos, el diseño de los algoritmos como la apariencia del propio hardware (como en el caso de los robots humanoides), pueden reproducir estereotipos de género (**Ortiz-de-Zárate-Alcarazo**, 2023).

Otra preocupación es la forma cómo se diseñan y evalúan los modelos de IA. Los diseñadores y evaluadores de IA a menudo son hombres, y es probable que sus propias creencias y expectativas de género influyan en la forma en que diseñan y evalúan los modelos. Esto puede conducir a la creación de modelos que reproducen estereotipos de género y a la ignorancia de los problemas de género en el diseño y la evaluación de la IA. Cobra importancia la necesidad de un desarrollo de la IA diverso e inclusivo por parte de la comunidad de programadores (**Eichenberger**, 2022).

Además, los modelos de IA también pueden contribuir a la discriminación de género al tomar decisiones automatizadas. Por ejemplo, un modelo de IA que ha sido entrenado con datos que contienen discriminación de género podría tomar decisiones discriminatorias. Un modelo de IA utilizado en la contratación podría discriminar en contra de las mujeres al considerar características estereotipadas de género como la capacidad de liderazgo.

La discriminación de género en IA también se puede manifestar en la forma en que se comercializan y se promueven los productos de IA. Por ejemplo, los asistentes virtuales con personalidades femeninas a menudo son diseñados para ser serviles y agradables, mientras que los asistentes virtuales con personalidades masculinas a menudo son pensados para ser autoritarios y dominantes (**Sainz**; **Arroyo**; **Castaño**, 2020; **Eubanks**, 2018). Estos estereotipos de género en la personalidad de los asistentes virtuales pueden contribuir a la perpetuación de la desigualdad de género en la sociedad.

Para abordar estos problemas, es esencial desarrollar una comprensión más profunda de los estereotipos de género y su impacto en la IA. Esto incluye el análisis de los datos utilizados para entrenar modelos de IA, el diseño y la evaluación de modelos de IA, y la forma en que se comercializan y se promueven los productos de IA. También es importante incluir diferentes perspectivas y voces en el diseño y la evaluación de la IA, incluyendo a las mujeres y a otros grupos que podrían sufrir discriminación (**Bolukbasi** *et al*., 2016).

Los estereotipos de género son un problema significativo en el campo de la IA, ya que pueden manifestarse en la forma en que se utilizan y se evalúan los datos, en el diseño y la evaluación de modelos de IA y en la forma en que se comercializan y se promueven los productos de IA. Aunque no existe una regla determinada y la IA puede llegar a distintos resultados dada la misma orden (**Rassin**; **Ravfogel**; **Goldberg**, 2022), es esencial desarrollar una comprensión más profunda de los estereotipos de género y su impacto en la IA para abordar estos problemas y promover una sociedad más justa e igualitaria. La reducción da las brechas de género, así como las raciales, sociales, o de otra índole, de las que son conscientes los propios programadores de sistemas para la creación de imágenes sintéticas (*OpenAI*, 2022b), constituyen un problema de crucial relevancia para el desarrollo de técnicas y herramientas computacionales basados en redes generativas antagónicas.

3. Metodología

3.1. Objetivos e hipótesis

Se plantea en el presente estudio el siguiente objetivo:

> O_1 Observar si las imágenes generadas por *DALL-E 2* presentan sesgos de género, edad, o raza para conocer si la IA genera imágenes estereotipadas con respecto al contexto laboral y profesional, lo que permitirá analizar si existen profesiones o entornos laborales más susceptibles de ser estereotipados por la IA.

Para ello se toman las siguientes hipótesis de partida:

> H_1 Las imágenes generadas por *DALL-E 2* presentan sesgos de género, edad o raza.
>
> H_2 Las imágenes generadas por *DALL-E 2* replican estereotipos en el ámbito laboral.

3.2. Muestra

En este estudio se utiliza un muestreo probabilístico estratificado. Los segmentos se delimitan a partir de 37 profesiones o *prompts*, replicando el estudio de **Farago**, **Eggum-Wilkens** y **Zhang**, (2021) sobre estereotipos de género en el campo laboral (figura 1). Las profesiones seleccionadas por el estudio de **Farago**, **Eggum-Wilkens** y **Zhang** (2021) son recopiladas a partir de una línea de investigación consolidada sobre estereotipos en el ámbito laboral, con estudios como los de **Archer** (1984); **Gottfredson** (1981); **Liben**, **Bigler** y **Krogh** (2001) y **Teig** y **Susskind**, (2008), que identifican algunas de las profesiones con mayores sesgos de género. Se replica el trabajo de **Farago**, **Eggum-Wilkens** y **Zhang**, (2021) porque resume las profesiones más estereotipadas que se detectan en estudios anteriores.

En el estudio intervienen 2 codificadores que introducen las distintas profesiones manualmente en el generador de imágenes. Se introducen 37 profesiones en inglés (género neutro, para que no existan sesgos de partida). Los textos introducidos son: *secretary*; *hotel manager*; *tailor*; *singer*; *teacher secondary*; *teacher primary*; *maid*; *nurse*; *TV/movie actor/actress*; *teacher for university*; *shop owner*; *writer*; *lawyer*; *newspaper reporter*; *professional athlete*; *doctor*; *government worker*; *street vendor*; *cook/chef in a restaurant*; *accountant*; *scientist*; *farmer*; *factory worker*; *pastor or religious leader*; *politician*; *computer specialist*; *banker*; *police*; *barber in a hair salon*; *engineer*; *soldier*; *construction worker*; *mechanic*; *airplane pilot*; *truck driver*; *taxi driver* y *carpenter*.

Se llevan a cabo dos barridos durante la misma semana por parte de cada uno de los codificadores. *DALL-E 2* genera 9 imágenes por cada consulta, por lo que la muestra final es de 666 imágenes (37x9x2). Para obtener las imágenes los codificadores accedieron a la página de *DALL-E mini by craiyon.com* en *https://huggingface.co/spaces/dalle-mini/dalle-mini*

Se introducen las profesiones seleccionadas o *prompts* en inglés por ser palabras neutras sin especificación de género en el generador de imágenes de la página de *DALL-E mini by craiyon.com* de manera manual y al hacer click en *"run"*, se generaron automáticamente las 9 imágenes referentes a esa profesión. Algunas de estas imágenes son representaciones o dibujos poco realistas, pero que sí que representan las profesiones que se habían planteado en la búsqueda.

Se trata de una muestra significativa, ya que, según los propios creadores de la aplicación *DALL-E 2*, durante la fecha en que se escribe este artículo, la cantidad de imágenes generadas por el software es de 60 millones, con 1,5 millones de usuarios (*OpenAI*, 2022a). De hecho, *DALL-E* contiene más de 12 mil millones de parámetros y se entrena con un conjunto de datos de 250 millones de pares imagen-texto (**Zhou** *et al*., 2021).

Dado un tamaño de población de 60 millones, con un nivel de confianza del 99% y un margen de error del 5%, la muestra representativa es de 666 unidades.

A continuación, se transcriben los resultados a una hoja *Excel* en la que se evalúa cada imagen de acuerdo con una escala Likert de 3 niveles (1. No estereotipado; 2. Moderadamente estereotipado; 3. Fuertemente estereotipado).

4. Resultados

Los resultados del estudio demuestran un marcado estereotipo de género en el ámbito profesional de las imágenes generadas por inteligencia artificial.

4.1. Profesiones y estereotipado en imágenes creadas por *DALL-E 2*

El estudio demuestra que existen profesiones totalmente estereotipadas en las imágenes generadas por IA (figura 1). Se observan imágenes en las que se representa únicamente a mujeres en las profesiones: *nurse* (enfermero/a); *maid* (asistente); *teacher – primary* (maestro/a de primaria); *teacher – secondary* (maestro/a de secundaria); *singer* (cantante); *seamstress/tailor* (sastre); *hotel manager* (director/a de hotel) y *secretary* (secretario/a).

En el lado opuesto, se han podido detectar profesiones en las que se representan exclusivamente a hombres, como son: *carpenter* (carpintero/a); *taxi driver* (taxista); *truck driver* (camionero/a); *airplane pilot* (piloto); *mechanic* (mecánico/a); *construction worker* (albañil); *soldier* (soldado); *engineer* (ingeniero/a); *barber in hair salon* (barbero/a); *police* (policía); *banker* (banquero/a); *computer specialist* (especialista en informática); *politician* (político/a) y *pastor or religious leader* (líder de una religión).

Respecto a las imágenes generadas por *DALL-E 2*, tal como especifican los creadores de *OpenAI*, para que las imágenes resultantes sean lo más realistas posible, se debe introducir el mayor número de términos para concretar con el contenido de la imagen solicitada. Puesto que los codificadores solo introdujeron el término referente a la profesión, el resultado son imágenes en los que las caras y las extremidades aparecen distorsionadas, ya que *DALL-E 2* necesitaría más información para generar imágenes bien definidas y de alta calidad (**Millán**, 2022; **Borji**, 2022). También hay que añadir que las búsquedas se realizaron entre octubre y noviembre de 2022, por lo que el programa ya ha mejorado los resultados que ofrece. Además, es un sistema que hasta hace poco era utilizado para delatar que esas imágenes generadas eran falsas y que se conoce como "efecto del valle inquietante" (**Franganillo**, 2022).

> "Los sistemas de IA pueden perpetuar los sesgos de género existentes en la sociedad"

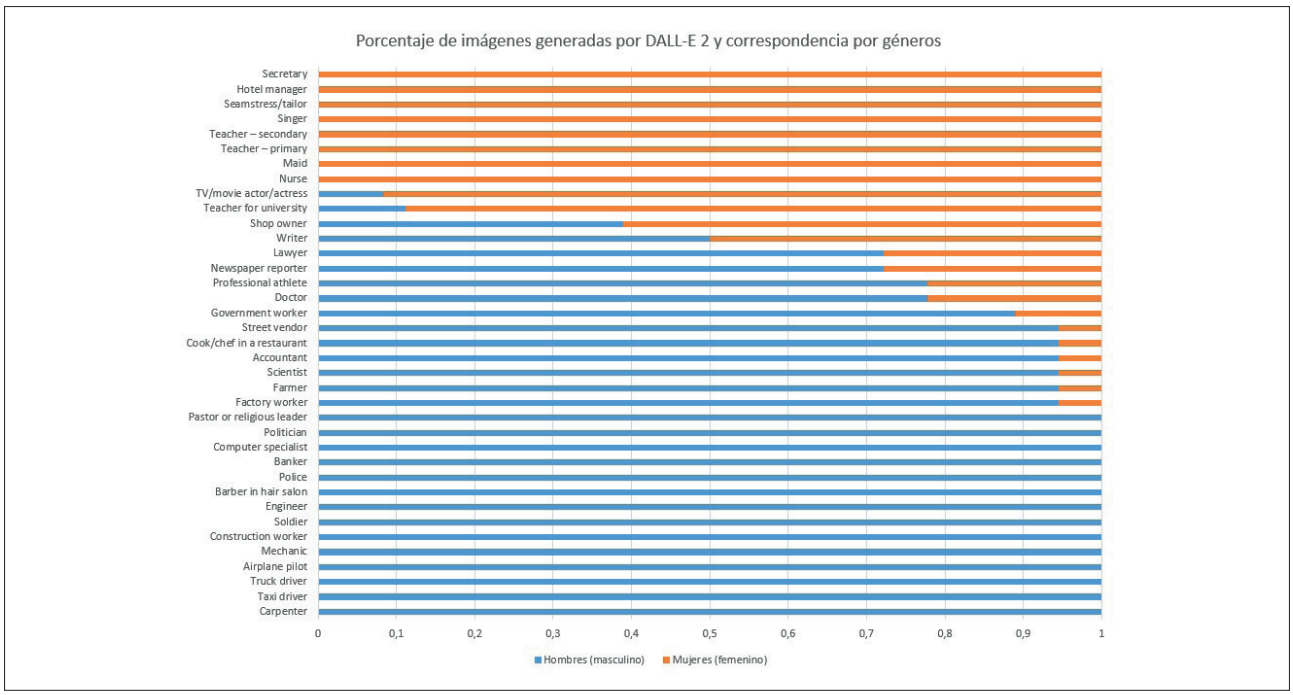


Figura 1. Imágenes generadas por *DALL-E 2* y correspondencia por géneros.

Los resultados demuestran que *DALL-E 2* representa profesiones totalmente estereotipadas en el 59,4% de los casos. El 21,6% de las profesiones se representan completamente estereotipadas con respecto al sexo femenino y el 37,8% con respecto al masculino.

4.2. Revisión cualitativa de imágenes, estereotipos laborales y sesgos

4.2.1. Profesiones técnicas, industria y sector primario

Se detecta en el presente estudio que las profesiones del sector técnico, industrial o vinculado a la construcción (trabajador/a de obra; carpintero/a; ingeniero/a; trabajador/a de fábrica; mecánico/a; técnico de ordenadores), no solo están muy estereotipadas y representadas por hombres, sino que suelen utilizar a personas jóvenes y con vestimenta muy similar: casco, chaleco, camisas a cuadros, etc., e incluso, con la misma postura o los mismos elementos de trabajo como las maderas en el caso de carpintero o papeles en las manos de los ingenieros.

Otra profesión que el generador de imágenes *DALL-E 2* muestra muy estereotipada es la de granjero/a, con un 94% de hombres. En estas representaciones sintéticas, observamos a hombres de edad avanzada, con la misma postura y con predominio del color verde. Todos muestran la misma apariencia en el campo, con una herramienta o palo como si estuvieran trabajando y hay que destacar, que un 20% de las imágenes generadas son dibujos.

4.2.2. Transporte

En el caso de profesiones vinculadas a la conducción como taxista, camionero/a o piloto de avión, se muestra a los profesionales siempre montados en el vehículo correspondiente, asomando por la ventanilla y casi con la misma postura. También se trata de profesiones muy estereotipadas, ya que el 100% de las personas representadas son hombres, de edad media y occidentales, a diferencia de las anteriores, que mostraban trabajadores más jóvenes. En el ejemplo de taxista y piloto de avión, se les muestra trajeados, mientras que los camioneros son representados con camisa y vestimenta más informal.

4.2.3. Educación

Hay otras profesiones también muy estereotipadas en el género femenino como son todas las vinculadas a la educación, ya que, tanto en educación primaria como en secundaria, se muestran imágenes en las que se representan en su totalidad a mujeres. También, el profesorado universitario vuelve a ser representado muy estereotipado, con un 88% de mujeres. La edad de los profesionales representados en todos los niveles educativos es de personas jóvenes. Al representar a las profesoras, la mayoría son rubias o castañas, de media melena y vestidas con blusa blanca o clara y chaqueta. En el caso de los hombres, aparecen con traje. En cuanto a la postura o forma de representar a profesionales de la educación, en el caso de niveles de primaria y secundaria se ven las clases con alumnos, los pupitres repartidos en el aula y a los profesores entre los alumnos. Mientras que para referirse a profesores de universidad *DALL-E 2* los representa en la pizarra y muestra a mujeres en un 88% y hombres en un 11% (figura 2).

> Un buen uso de la IA podría trascender estas limitaciones en aras de una sociedad más igualitaria

Figura 2. Imágenes generadas por *DALL-E 2* de profesorado de primaria (izquierda) y profesorado de universidad (derecha).

4.2.4. Servicios y espectáculos

Otras profesiones también muy estereotipadas son las relacionadas con el sector servicios, textil, espectáculos o cine. Como vemos en estos ejemplos, al buscar en *DALL-E 2* asistente, sastre o cantante, siempre ofrecen imágenes del género femenino y muy similares entre ellas. Se trata, por tanto, de profesiones 100% estereotipadas y que muestran a las profesionales de edad siempre joven. También debemos destacar que, tanto en estas imágenes como en las comentadas anteriormente, todos los profesionales son occidentales.

En el caso de asistente, es importante destacar que su vestimenta es la típica de personal del servicio de clase alta con traje negro, delantal blanco y plumero en la mano, además, todas son morenas y con el pelo corto y la postura muy similar. Por el contrario, las profesiones de sastre y cantante son representadas con mujeres de pelo rubio o castaño con pelo largo o media melena y también en posturas muy parecidas entre ellas. Para la profesión de sastre (*seamstress/tailor*) las mujeres están todas sentadas junto a la máquina de coser y con el metro al cuello. Destaca el predominio del color rosa en todas las imágenes. Y en el ejemplo de cantante, también las imágenes son muy parecidas, todas están en la misma postura, de pie, con el micrófono en la mano o delante y el pelo y la vestimenta son muy similares con ropas más desenfadadas.

Respecto a las profesiones vinculadas al sector servicios, comercio y restauración, se observa que, al insertar los términos de barbero o cocinero, las imágenes generadas son muy estereotipadas y nos muestran el 100% de personas del género masculino (incluidos los clientes en el caso de la barbería). En el caso de vendedor ambulante los resultados son de un 94% de hombres, mientras que bajo el término *shop owner*, se representa a un 38% de hombres y un 62% de mujeres. En los ejemplos de barbero y chef todos son de edad media e incluso, jóvenes, occidentales y con la misma postura y apariencia. Mientras que los vendedores ambulantes parecen personas hindúes o asiáticas. En este caso, todos aparecen con carros de frutas y hortalizas y las mismas vestimentas y en el caso del chef todos de blanco y con gorro (en la cocina).

Por otro lado, la profesión de director/a de hotel se representa de manera muy estereotipada, con un 100% mujeres. También cabe destacar, que las mujeres son todas jóvenes, con el pelo recogido y el uniforme oscuro y la imagen se muestra en la recepción del hotel.

Como último ejemplo en este segmento, observamos que las imágenes que *DALL-E 2* ofrece respecto a la profesión de actor/actriz son diferentes dependiendo de cómo introduzcamos el término. En el primer barrido se introdujo con la barra perpendicular para separar actor y actriz y las imágenes generadas devolvieron como resultado una profesión estereotipada, con un 91% de mujeres. Con respecto a su apariencia son personas occidentales y algunas latinas, con pelo largo, castañas o morenas y con vestimentas que muestran los hombros y son de colores vivos. Además, los dos hombres que aparecen tienen el mismo corte de pelo y barba o bigote.

En el segundo barrido introducimos los términos solo separados por espacios y las imágenes que se mostraron fueron diferentes. Vemos que están muy estereotipadas, ya que todas las profesionales son mujeres, de apariencia hindú por color de piel y vestimentas, además, de estar más tapadas que las de la imagen anterior. Parece que representasen a las actrices de Bollywood.

Figura 3. Imágenes generadas por *DALL-E 2* de la profesión doctor/a (izquierda) y enfermero/a (derecha)

4.2.5. Salud y ciencia

En el campo de Ciencias de la Salud y científico se observan ciertas diferencias en las imágenes generadas. Si incluimos en la búsqueda el término doctor, los resultados están medianamente estereotipados, ya que se muestran a un 77% de hombres y a un 23% de mujeres. Sin embargo, al introducir el término "*nurse*" (enfermero/a), los resultados que se nos ofrecen están 100% estereotipados ya que solo aparecen mujeres en las imágenes. Además, todos los profesionales son jóvenes, occidentales, con la misma postura (los doctores con los brazos cruzados y las enfermeras con informes en la mano), predomina el color blanco tanto en el fondo como en la vestimenta y algunos profesionales llevan mascarilla (figura 3).

Por otro lado, la profesión de científico también ofrece resultados estereotipados, con un 94% de hombres jóvenes, morenos y con la misma postura, revisando muestras con guantes y gafas. En cuanto a la vestimenta, todos llevan bata y se observa que visten traje. Además, como ocurría con los profesionales de Ciencias de la Salud, predomina el color blanco.

4.2.6. Política, economía e información

Se observa que las imágenes generadas por inteligencia artificial referentes a la política están muy estereotipadas, ya que el 100% de las personas son hombres, de edad avanzada o mediana edad y occidentales. También hay que destacar que todos llevan traje y corbata y que para los altos cargos se emplean colores oscuros y para empleados de medio nivel colores más claros.

Las imágenes generadas por *DALL-E 2* para abogado/a nos muestran profesiones muy estereotipadas, donde el 72% son hombres, todos ellos de edad avanzada, con toga oscura y la mayoría un libro entre las manos. Además, todas las imágenes representan a occidentales.

Las profesiones vinculadas a la oficina como secretario/a, contable o banquero están también muy estereotipadas. Al introducir el término *secretary*, las imágenes generadas son todas de mujeres, jóvenes, occidentales, morenas y con pelo largo. Además, están todas vestidas con traje, sentadas en la mesa junto al ordenador. Lo mismo ocurre con el término banquero, pero en este caso, todas las imágenes están representadas por hombres de mediana edad, occidentales, con traje y utilizando la calculadora o manipulando papeles. Resultados similares se observan al introducir el término *banker*, con un 94% de hombres representados.

Se aprecia un cambio en las profesiones vinculadas a la escritura (como escritor o periodista) con respecto a las anteriores, ya que el género está medianamente estereotipado y aparecen tanto mujeres como hombres. En el caso de *writer*, los resultados ofrecen una proporción de 50% hombres y 50% mujeres. Al introducir el término *journalist*, aparecen un 72% de hombres y un 28% de mujeres. Lo más destacado de las imágenes generadas para estas profesiones, es que en todas ellas se observa el mismo elemento: la máquina de escribir en el caso de escritor/a y el periódico en la mano en el periodista (figura 4). Nuevamente se representan a occidentales de edad media o joven y la vestimenta es muy similar con chaqueta o americana oscura y traje.

> "La respuesta de la IA está implícita en la pregunta que el usuario realiza"

Figura 4. Imágenes generadas por *DALL-E 2* para la profesión de escritor/a (izquierda) y periodista (derecha).

4.2.7. Seguridad, religión y deportes

En cuanto a las profesiones vinculadas a la seguridad, todas las imágenes generadas por inteligencia artificial están muy estereotipadas. El 100% de las figuras representadas son hombres y en el caso, de la profesión de soldado, 7 de las 9 imágenes son dibujos. Tanto policías como soldados se representan uniformados, son todos jóvenes, occidentales y con las mismas posturas (policías en la calle y soldados con el arma en la mano).

A continuación, se revisan otras profesiones también neutras como pastor o líder religioso y atleta y en este caso, se observa que la de pastor es una profesión muy estereotipada en la que se representan todo hombres, de edad avanzada y la mayoría son occidentales (hay dos imágenes que parece que son de raza negra). Las imágenes generadas para atleta profesional también nos muestran unos resultados estereotipados, ya que el 77% son hombres. Todos son jóvenes, vestidos con ropas deportivas en las que el rojo es predominante y la mayoría son blancos y occidentales.

4.3. Comparación de estereotipado entre IA y humanos

Los estereotipos de género en la inteligencia artificial (IA) pueden tener impactos significativos en diversas áreas, como se evidencia en el estudio mencionado. Al comparar los resultados obtenidos por *DALL-E 2* con estudios previos que involucraron la opinión de seres humanos, se concluye que la IA presenta un mayor grado de estereotipado de género en el ámbito laboral. En particular, el análisis realizado por **Farago**, **Eggum-Wilkens** y **Zhang** (2021) revela que el 35% de las profesiones evaluadas mostraban fuertes estereotipos, mientras que las imágenes generadas por la IA alcanzaron un alarmante 59,4% de estereotipado.

Existen profesiones que coinciden en el estereotipado entre humanos e IA, como lo son, en el sexo masculino: *carpenter* (carpintero/a); *taxi driver* (taxista); *truck driver* (camionero/a); *airplane pilot* (piloto); *mechanic* (mecánico/a); *construction worker* (albañil); *soldier* (soldado); *engineer* (ingeniero/a); *barber in hair salon* (barbero/a). En el sexo femenino, las profesiones con estereotipos de género femenino en las que coinciden tanto humanos como la IA son *nurse* (enfermero/a) y *maid* (asistente).

Los sesgos presentes en los datos utilizados para entrenar a los modelos de IA pueden reflejar los prejuicios y desequilibrios de género existentes en la sociedad. Si los conjuntos de datos históricos contienen desigualdades o reflejan estereotipos de género, es probable que la IA aprenda y reproduzca estos patrones durante su entrenamiento. Otro factor a considerar es la retroalimentación continua y la influencia mutua entre la sociedad y la tecnología. Si las profesiones ya están fuertemente estereotipadas en la sociedad, las representaciones generadas por la IA pueden reforzar aún más estos estereotipos, creando un ciclo de retroalimentación.

Sin embargo, existen profesiones fuertemente estereotipadas por la IA, que no lo están por los humanos. En el caso masculino, estas son: *police* (policía); *banker* (banquero/a); *computer specialist* (especialista en informática); *politician* (político/a) y *pastor or religious leader* (líder de una religión). En el caso femenino, por último,

> “Los sesgos presentes en los datos utilizados para entrenar a los modelos de IA pueden reflejar los prejuicios y desequilibrios de género existentes en la sociedad”

la IA muestra profesiones muy estereotipadas, que no lo están por los humanos, como lo son: *teacher - primary* (maestro/a de primaria); *teacher - secondary* (maestro/a de secundaria); *singer* (cantante); *seamstress/tailor* (sastre); *hotel manager* (director/a de hotel) y *secretary* (secretario/a).

> El paso crucial hacia una IA más justa e imparcial, reside en la creación y consolidación de una comunidad de desarrollo diversa e inclusiva

Tiene aquí influencia la interpretación y representación de los datos por parte de la IA. Los algoritmos de IA pueden utilizar ciertos atributos o características presentes en los datos para asignar etiquetas o asociar ciertos trabajos a un género en particular, incluso si no hay una base sólida para hacerlo. Si se utilizan bases de datos con sesgos de género para nutrir a la IA, el aprendizaje automatizado también mostrará estereotipados.

5. Discusión y conclusiones

Los resultados que ofrece *DALL-E 2* con respecto a profesiones neutras están muy estereotipados ya que, de 37 búsquedas, en 22 el resultado es siempre del mismo género. El 21,6% de las profesiones se representan completamente estereotipadas con respecto al sexo femenino y el 37,8% con respecto al masculino.

Es el caso de profesiones técnicas, científicas, vinculadas a la construcción o a la conducción. En cuanto a las profesiones vinculadas por la IA a mujeres, encontramos a las empleadas de hogar, modistas y aquellas profesiones en las que es importante la apariencia, como actriz o cantante. En estas dos categorías las imágenes generadas por IA muestran a mujeres jóvenes, occidentales y rubias. Es importante subrayar la alta presencia de mujeres en el sector de la educación y en medicina, especialmente, en enfermería.

También es destacable el hecho de que las imágenes sintéticas generadas con *DALL-E 2* presentan a hombres de edad media o avanzada cuando se trata de profesiones vinculadas a una mayor responsabilidad o estatus, como en el ámbito de la política, economía y religión. También aquí se observa una mayoría de apariencia occidental.

Si comparamos nuestros resultados con estudios anteriores en los que se consulta a adolescentes, podemos concluir que *DALL-E 2* presenta mayor estereotipado de género en el ámbito laboral. Mientras que estudios anteriores en humanos detectan un fuerte estereotipado de género en el 35% de las profesiones, la inteligencia artificial estereotipa completamente en un 59,4% de los casos.

En resumen, se detectan en el presente estudio fuertes sesgos de género en el ámbito laboral en las imágenes generadas mediante inteligencia artificial.

Las herramientas basadas en IA parecen cobrar cada vez mayor relevancia, al tiempo que prometen, a corto plazo, participar e influir en las relaciones sociales. Es, por tanto, imprescindible la necesidad de identificar, clasificar y eliminar estos sesgos que puedan influir, de manera directa o indirecta, en nuestra toma de decisiones y en nuestro modo de observar y afrontar la realidad.

La inteligencia artificial no hace sino reflejar nuestro sentir común, nuestras virtudes y defectos. Si reflexionamos sobre nuestros propios prejuicios y trabajamos, no solo para extrapolar el pasado, sino también para aprender de él de manera crítica, podremos esperar crear tecnologías de IA que sean verdaderamente inclusivas y justas.

Se nos plantean, a modo de discusión, dos problemas a resolver en relación con la ética y eficiencia de los generadores automatizados basados en IA:

- En primer lugar, encontramos el sesgo del usuario. Cuando el usuario introduce una consulta, la IA devuelve aquello que se necesita responder, como en una cámara de eco. En este sentido, la respuesta de la IA está implícita en la pregunta que el usuario realiza. Es, por tanto, de extrema complejidad, encontrar respuestas más allá de la cosmogonía del emisor, de su forma de entender la realidad. Esta imposibilidad de encontrar respuestas que trasgredan los límites del conocimiento del usuario es el concepto que hemos convenido a definir como "umbral del espejo".
- En segundo lugar, dándose el hipotético escenario en el que el usuario es capaz de realizar una consulta libre de cualquier prejuicio y, por tanto, de traspasar el umbral del espejo, se sumergirá en un océano de conocimiento, por naturaleza, sesgado, ya que entraría en juego el umbral del espejo del desarrollador de la tecnología. Si "cada tecnología es una ideología" (**Postman**, 1991, p. 165), la IA no puede separase de la ideología de sus creadores.

Parece lógico afirmar que el paso crucial hacia una IA más justa e imparcial, reside en la creación y consolidación de una comunidad de desarrollo diversa e inclusiva. Solo así cabe esperar tecnologías que repliquen estos mismos valores.

Nuestros hallazgos subrayan la importancia de examinar tanto los estereotipos de la IA como los estereotipos humanos. Si bien los estereotipos son productos de la sociedad y reflejan los prejuicios arraigados, la IA tiene el potencial de amplificar y perpetuar estos sesgos debido a su capacidad para aprender de grandes conjuntos de datos. Es fundamental abordar este problema desde dos perspectivas complementarias:

- promover la diversidad y equidad en los datos utilizados para entrenar a la IA;
- fomentar una mayor conciencia y reflexión en los seres humanos sobre los estereotipos arraigados que pueden influir en la creación y utilización de la tecnología.

Solo así podremos avanzar hacia sistemas de IA más justos y libres de sesgos que promuevan la igualdad de oportunidades y la inclusión en todas las áreas de nuestra sociedad.

6. Referencias

Agudo, Ujué; **Liberal, Karlos G.** (2020). "El automágico traje del emperador". *Medium.com*, 9 septiembre.
https://medium.com/bikolabs/el-automagico-traje-del-emperador-c2a0bbf6187b

Archer, Cynthia J. (1984). "Children's attitudes toward sex-role division in adult occupational roles". *Sex roles*, v. 10.
https://doi.org/10.1007/BF00287742

Belhadi, Amine; **Kamble, Sachin**; **Fosso-Wamba, Samuel**; **Queiroz, Maciel M.** (2022). "Building supply-chain resilience: an artificial intelligence-based technique and decision-making framework". *International journal of production research*, v. 60, n. 14, pp. 4487-4507.
https://doi.org/10.1080/00207543.2021.1950935

Bolukbasi, Tolga; **Chang, Kai-Wie**; **Zou, James**; **Saligrama, Venkatesh**; **Kalai, Adam** (2016). "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings". In: *NIPS'16: Proceedings of the 30th international conference on neural information processing systems*, pp. 4356-4364.
https://doi.org/10.48550/arXiv.1607.06520

Borji, Ali (2022). *Generated faces in the wild: quantitative comparison of stable diffusion, midjourney and DALL-E 2*. *Quintic AI*, San Francisco, CA.
https://arxiv.org/pdf/2210.00586.pdf

Brown, Tom B.; **Mann, Benjamin**; **Ryder, Nick**; **Subbiah, Melanie**; **Kaplan, Jared**; **Dhariwal, Prafulla**; **Neelakantan, Arvind**; **Shyam, Pranav**; **Sastry, Girish**; **Askell, Amanda**; **Agarwal, Sandhini**; **Herbert-Voss, Ariel**; **Krueger, Gretchen**; **Henighan, Tom**; **Child, Rewon**; **Ramesh, Aditya**; **Ziegler, Daniel M.**; **Wu, Jeffrey**; **Winter, Clemens**; **Hesse, Christopher**; **Chen, Mark**; **Sigler, Eric**; **Litwin, Mateusz**; **Gray, Scott**; **Chess, Benjamin**; **Clark, Jack**; **Berner, Christopher**; **McCandlish, Sam**; **Radford, Alec**; **Sutskever, Ilya**; **Amodei, Dario** (2020). "Language models are few-shot learners". *Advances in neural information processing systems*, v. 33, pp. 1877-1901.
https://doi.org/10.48550/arXiv.2005.14165

Buolamwini, Joy; **Gebru, Timnit** (2018). "Gender shades: intersectional accuracy disparities in commercial gender classification". *Proceedings of machine learning research*, v. 81.
https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Caliskan, Aylin; **Bryson, Joanna J.**; **Narayanan, Arvind** (2017). "Semantics derived automatically from language corpora contain human-like biases". *Science*, v. 356, n. 6334, pp.183-186.
https://doi.org/10.1126/science.aal4230

Cortina-Orts, Adela (2019). "Ética de la inteligencia artificial". *Anales de la Real Academia de Ciencias Morales y Políticas*, pp. 379-394. Ministerio de Justicia.
https://www.boe.es/biblioteca_juridica/anuarios_derecho/articulo.php?id=ANU-M-2019-10037900394

Crawford, Kate (2021). *The atlas of AI: power, politics, and the planetary costs of artificial intelligence*. *Yale University Press*. ISBN: 978 0 300252392
https://doi.org/10.2307/j.ctv1ghv45t

Criado-Pérez, Caroline (2020). *La mujer invisible. Descubre cómo los datos configuran un mundo hecho por y para los hombres*. Barcelona: Seix Barral. ISBN: 978 84 32236136

DALL-E 2 (2021). OpenAI.
https://openai.com/dall-e-2

De-Carvalho, André-Carlos-Ponce-de-Leon-Ferreira (2021). Inteligência artificial: riscos, benefícios e uso responsável. *Estudos avançados*, v. 35, 101.
https://doi.org/10.1590/s0103-4014.2021.35101.003

D'Ignazio, Catherine; **Klein, Lauren F.** (2020). *Data feminism*. Cambridge: MIT Press. ISBN: 978 0 262547185

Eichenberger, Livia (2022). "DALL-E 2: Why discrimination in AI development cannot be ignored". *Statworx blog post*, 28 June.
https://www.statworx.com/en/content-hub/blog/dalle-2-open-ai

Estupiñán-Ricardo, Jesús; **Leyva-Vázquez, Maikel-Yelandi**; **Peñafiel-Palacios, Álex-Javier**; **El-Asaffiri-Ojeda, Yusef** (2021). "Inteligencia artificial y propiedad intelectual". *Universidad y sociedad*, v. 13, n. S3, pp. 362-368.
https://rus.ucf.edu.cu/index.php/rus/article/view/2490

Eubanks, Virginia (2018). *Automating inequality: how high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press. ISBN: 978 1 250074317

Farago, Flora; **Eggum-Wilkens, Natalie D.**; **Zhang, Linlin** (2021). "Ugandan adolescents' gender stereotype knowledge about jobs". *Youth & society*, v. 53, n. 5, pp. 723-744.
https://doi.org/10.1177/0044118X19887075

Francescutti, Pablo (2018). *La visibilidad de las científicas españolas*. Fundación Dr. Antoni Esteve, Grupo de estudios avanzados de comunicación, Barcelona.
https://www.raco.cat/index.php/QuadernsFDAE/issue/download/30066/439

Franganillo, Jorge (2022). "Contenido generado por inteligencia artificial: oportunidades y amenazas". *Anuario ThinkEPI*, v. 16, e16a24.
https://doi.org/10.3145/thinkepi.2022.e16a24

Gamir-Ríos, José; **Tarullo, Raquel** (2022). "Predominio de las cheapfakes en redes sociales. Complejidad técnica y funciones textuales de la desinformación desmentida en Argentina durante 2020". *adComunica*, v. 23, pp. 97-118.
https://doi.org/10.6035/adcomunica.6299

García-Ull, Francisco-José (2021). "Deepfakes: el próximo reto en la detección de noticias falsas". *Anàlisi*, n. 64, pp. 103-120.
https://doi.org/10.5565/rev/analisi.3378

Goodfellow, Ian J.; **Pouget-Abadie, Jean**; **Mirza, Mehdi**; **Xu, Bing**; **Warde-Farley, David**; **Ozair, Sherjil**; **Courville, Aaron**; **Bengio, Yoshua** (2014). "Generative adversarial networks. Advances in neural information processing systems". *Communications of the ACM*. v. 63, pp. 139-164.
https://doi.org/10.48550/arXiv.1406.2661

Gottfredson, Linda S. (1981). "Circumscription and compromise: A developmental theory of occupational aspirations". *Journal of counseling psychology*, v. 28, n. 6, pp. 545–579.
https://doi.org/10.1037/0022-0167.28.6.545

Laino, María-Elena; **Cancian, Pierandrea**; **Salvatore-Politi, Letterio**; **Della-Porta, Matteo-Giovanni**; **Saba, Luca**; **Savevski, Victor** (2022). "Generative adversarial networks in brain imaging: A narrative review". *Journal of imaging*, v. 8, n. 4, 83.
https://doi.org/10.3390/jimaging8040083

Leavy, Susan (2018). "Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning". In: *Proceedings of the 1st international workshop on gender equality in software engineering*, pp. 14-16.
https://doi.org/10.1145/3195570.3195580

Leavy, Susan; **Meaney, Gerardine**; **Wade, Karen**; **Greene, Derek** (2020). "Mitigating gender bias in machine learning data sets". In: *Bias2020 workshop: Bias and social aspects in search and recommendation*.
https://doi.org/10.1007/978-3-030-52485-2_2
https://doi.org/10.48550/arXiv.2005.06898

Liben, Lynn S.; **Bigler, Rebecca S.**; **Krogh, Holleen R.** (2001). "Pink and blue collar jobs: children's judgments of job status and job aspirations in relation to sex of worker". *Journal of experimental child psychology*, v. 79, n. 4, pp. 346-363.
https://doi.org/10.1006/jecp.2000.2611

Loftus, Tyler J.; **Tighe, Patrick J.**; **Filiberto, Amanda C.**; **Efron, Philip A.**; **Brakenridge, Scott C.**; **Mohr, Alicia M.**; **Rashidi, Parisa**; **Upchurch, Gilbert R.**; **Bihorac, Azra** (2020). "Artificial intelligence and surgical decision-making". *JAMA surgery*, v. 155, n. 2, pp. 148-158.
https://doi.org/10.1001/jamasurg.2019.4917

Manassero, Antonia; **Vázquez, Ángel** (2003). "Las mujeres científicas: un grupo invisible en los libros de texto". *Revista investigación en la escuela*, v. 50, pp. 31-45.
https://revistascientificas.us.es/index.php/IE/article/view/7582

Millán, Víctor (2022). "DALL-E 2: ¿cómo funciona y qué supone? La IA que crea imágenes de la nada y es, simplemente, perfecta y aterradora". *Hipertextual*, 29 mayo.
https://hipertextual.com/2022/05/dall-e-2

Nica, Elvira; **Sabie, Oana-Matilda**; **Mascu, Simona**; **Luțan-Petre, Anca-Georgeta** (2022). "Artificial intelligence decision-making in shopping patterns: consumer values, cognition, and attitudes". *Economics, management and financial markets*, v. 17, n. 1, pp. 31-43.
https://doi.org/10.22381/emfm17120222

O'Neil, Cathy (2018). *Armas de destrucción matemática: cómo el big data aumenta la desigualdad y amenaza la democracia*. Capitán Swing Libros. ISBN: 978 84 947408 4 8

OpenAI (2022a). "DALL-E now available without waitlist". *Openai*, September 28.
https://openai.com/blog/dall-e-now-available-without-waitlist

OpenAI (2022b). "Reducing bias and improving safety in DALL-E 2". *OpenAI*, July 18.
https://openai.com/blog/reducing-bias-and-improving-safety-in-dall-e-2

Ortiz-de-Zárate-Alcarazo, Lucía (2023). "Sesgos de género en la inteligencia artificial". *Revista de occidente*, v. 1, n. 502.
https://dialnet.unirioja.es/servlet/articulo?codigo=8853265

Pérez-Gómez, Miguel-Ángel; **Echazarreta-Soler, Carmen**; **Audebert-Sánchez, Meritxell**; **Sánchez-Miret, Cristina** (2020). "El ciberacoso como elemento articulador de las nuevas violencias digitales: métodos y contextos". *Communication papers. Media literacy and gender studies*, v. 9, n. 18.
https://doi.org/10.33115/udg_bib/cp.v9i18.22470

Porayska-Pomsta, Kaska; **Rajendran, Gnanathusharan** (2019). "Accountability in human and artificial intelligence decision-making as the basis for diversity and educational inclusion". In: Knox, Jeremy; Wang, Yuchen; Gallagher, Michael. *Artificial intelligence and inclusive education: speculative futures and emerging practices*. Springer, pp. 39-59.
https://doi.org/10.1007/978-981-13-8161-4_3

Postman, Neil (1991). *Divertirse hasta morir, el discurso público en la era del show business*. Barcelona: Ediciones la Tempestad. ISBN: 978 84 79480462

Quirós-Fons, Antonio; **García-Ull, Francisco-José** (2022). *La inteligencia artificial como herramienta de la desinformación: deepfakes y regulación europea. Los derechos humanos en la inteligencia artificial: su integración en los ODS de la Agenda 2030*. Thomson Reuters Aranzadi, pp. 537-556. ISBN: 978 84 1124 557 9

Rassin, Royi; **Ravfogel, Shauli**; **Goldberg, Yoav** (2022). "DALL-E 2 is seeing double: flaws in word-to-concept mapping in text2image models".
https://doi.org/10.48550/arXiv.2210.10606

Sainz, Milagros; **Arroyo, Lidia**; **Castaño, Cecilia** (2020). *Mujeres y digitalización: de las brechas a los algoritmos*. Instituto de la Mujer y para la Igualdad de Oportunidades.
https://www.inmujeres.gob.es/diseno/novedades/M_MUJERES_Y_DIGITALIZACION_DE_LAS_BRECHAS_A_LOS_ALGORITMOS_04.pdf

Sourdin, Tania (2018). "Judge v Robot? Artificial intelligence and judicial decision-making". *UNSW law journal*, v. 41, n. 4, pp. 1114-1133.
https://www.unswlawjournal.unsw.edu.au/wp-content/uploads/2018/12/Sourdin.pdf

Teig, Stacey; **Susskind, Joshua E.** (2008). "Truck driver or nurse? The impact of gender roles and occupational status on children's occupational preferences". *Sex roles*, v. 58, pp. 848-863.
https://doi.org/10.1007/s11199-008-9410-x

Traylor, Jake (2022). "No quick fix: how OpenAI's DALL-E 2 illustrated the challenges of bias in AI". *NBC news*, July 27.
https://www.nbcnews.com/tech/tech-news/no-quick-fix-openais-dalle-2-illustrated-challenges-bias-ai-rcna39918

Véliz, Carissa (2021). *Privacidad es poder: datos, vigilancia y libertad en la era digital*. *Debate*. ISBN: 978 84 18056680

Vincent, James (2020). "OpenAI's latest breakthrough is astonishingly powerful, but still fighting its flaws". *The verge tech*, July 30.
https://www.theverge.com/21346343/gpt-3-explainer-openai-examples-errors-agi-potential

Wang, Tianlu; **Zhao, Jieyu**; **Yatskar, Mark**; **Chang, Kai-Wei**; **Ordóñez, Vicente** (2019). "Balanced datasets are not enough: estimating and mitigating gender bias in deep image representations". In: *International conference on computer vision, ICCV 2019*.
https://doi.org/10.48550/arXiv.1811.08489

Zhou, Yufan; **Zhang, Ruiyi**; **Chen, Changyou**; **Li, Chunyuan**; **Tensmeyer, Chris**; **Yu, Tong**; **Gu, Jiuxiang**; **Xu, Jinhui**; **Sun, Tong** (2021). "Towards language-free training for text-to-image generation".
https://arxiv.org/pdf/2111.13792v3.pdf