EX±ACTO.
VI Congreso INTERNACIONAL DE INVESTIGACIÓN EN ARTES VISUALES
ANIAV 2024.
3-5 de julio 2024, Valencia
Doi: https://doi.org/10.4995/ANIAV2024.2024.18187

Sesgos en la creación de imágenes por sistemas TTI de IA: hibridando arte, ciencia y tecnología

Biases in image creation by AI TTI systems: hybridizing art, science and technology

Alejandra Bueno de Santiago[a], Laura Martinez Martin[b] y Aiskoa Perez[c]

[a]Universidad de Vitoria-Gasteiz, EUNEIZ alejandra.bueno@euneiz.con; [b]Universidad de Vitoria-Gasteiz, EUNEIZ, laura.martinez@euneiz.com y [c]Universidad de Vitoria-Gasteiz, EUNEIZ, aiskoa.perez@euneiz.com

Breve bio autoras:

Alejandra Bueno de Santiago, Laura Martínez Martín y Aiskoa Pérez son artistas visuales e investigadoras en arte que indagan la representación visual con enfoque feminista. Juntas forman la colectiva FEMIAS un proyecto teórico y artístico que nace de la investigación bajo el título ¿Puede la IA ser feminista? Desde su formación en 2023 han disfrutado de una residencia de investigación en el Centro de Innovación abierta y transferencia creativa de Álava HIBRIDALAB, han obtenido la la subvención para el fomento y desarrollo de actividades en el área de las Artes Plásticas y Visuales, en la modalidad de CREACIÓN-PRODUCCIÓN del Departamento de Cultura y Política Lingüística del Gobierno Vasco, y han sido ganadoras del IV Concurso de Trabajos de Investigación Feminista organizado por el Servicio de Igualdad del Ayuntamiento de Vitoria-Gasteiz.

How to cite: Bueno de Santiago, A., Martinez Martin, L. y Perez, A. (2024). Sesgos en la creación de imágenes por sistemas TTI de IA: hibridando arte, ciencia y tecnología. En libro de actas: *EX±ACTO. VI Congreso Internacional de investigación en artes visuales aniav 2024. Valencia*, 3-5 julio 2024. https://doi.org/10.4995/ANIAV2024.2024.18187

Resumen

En un contexto de avances vertiginosos en inteligencia artificial (IA), surge la necesidad crítica desde las prácticas artísticas de comprender cómo esta tecnología puede verse afectada por sesgos inherentes a la sociedad tradicional. En este punto planteamos indagar bajo la pedagogía crítica y la premisa de las imágenes que actúan cómo se comportan las imágenes generadas por IA desde una mirada artística, feminista e interseccional. Partiendo de la hipótesis de que los algoritmos de IA reflejan y pueden amplificar los sesgos presentes en los conjuntos de datos de entrenamiento, el estudio se propuso identificar y cuantificar estos sesgos en las imágenes generadas. Mediante metodologías cuantitativas y cualitativas los resultados del análisis de imágenes son puestos en diálogo con el cuerpo teórico de la investigación que se centra en bases de datos, construcción de imágenes, y prácticas artísticas feministas. En este texto se presenta un resumen teórico y analítico del total de la investigación, siendo esta visible en la página web del proyecto. El objetivo principal fue destacar cómo la IA puede reproducir y perpetuar prejuicios y estereotipos presentes en la sociedad a través de la generación de imágenes. Los resultados obtenidos respaldaron la hipótesis inicial, mostrando una tendencia de los algoritmos de IA a reflejar y amplificar los sesgos presentes en los datos de entrenamiento, siendo los más evidentes género y raza. Este estudio subraya la importancia de comprender y abordar los sesgos en la generación de imágenes con IA para garantizar equidad y representatividad. Además, destaca la necesidad de investigar desde las artes los campos más tecnológicos como la IA, ya que la intersección de estas disciplinas puede proporcionar perspectivas únicas y soluciones innovadoras para abordar los desafíos éticos y sociales planteados por el avance de la IA. Este enfoque integral puede fomentar un diálogo interdisciplinario en el desarrollo de IA, promoviendo una mayor conciencia sobre los posibles sesgos y la importancia de mitigarlos para construir sistemas más equitativos y responsables.

Palabras clave: *IA; pedagogía crítica; imágenes; feminismos; análisis; sesgos.*

Abstract

In a context of vertiginous advances in artificial intelligence (AI), the critical need arises from artistic practices to understand how this technology can be affected by biases inherent to traditional society. At this point we propose to investigate under the critical pedagogy and the premise of the images that act how the images generated by AI behave from an artistic, feminist, and intersectional point of view. Based on the hypothesis that AI algorithms reflect and can amplify biases present in training datasets, the study set out to identify and quantify these biases in the generated images. Using quantitative and qualitative methodologies the results of the image analysis are placed in dialogue with the theoretical body of research that focuses on databases, image construction, and feminist art practices. A theoretical and analytical summary of the total research is presented in this text and is visible on the project website. The main objective was to highlight how AI can reproduce and perpetuate prejudices and stereotypes present in society through the generation of images. he results obtained supported the initial hypothesis, showing a tendency for AI algorithms to reflect and amplify biases present in the training data, the most evident being gender and race. This study underscores the importance of understanding and addressing biases in AI image generation to ensure fairness and representativeness. Furthermore, it highlights the need for research from the arts into more technological fields such as AI, as the intersection of these disciplines can provide unique perspectives and innovative solutions to address the ethical and social challenges posed by the advancement of AI. This holistic approach can foster an interdisciplinary dialogue in AI development, promoting greater awareness of potential biases and the importance of mitigating them in order to build more equitable and accountable systems.

Keywords: *AI; critical pedagogy; images; feminisms; analysis; biases.*

INTRODUCCIÓN

Este estudio analiza el impacto de los sistemas de generación de imágenes "Text-to-Image" (TTI) en la representación y percepción de las mujeres, enfocándose en cómo estos sistemas perpetúan sesgos de género y estereotipos. La importancia de este estudio radica en la creciente prevalencia de las imágenes generadas por Inteligencia Artificial (IA) en la cultura visual contemporánea y su influencia en la formación de estereotipos y percepciones sociales. La calidad y relevancia de los datasets son determinantes para la precisión y eficacia de los modelos de IA, lo que resalta la importancia de una selección y uso adecuados de los mismos. Este entendimiento subraya la necesidad de una gestión cuidadosa de los datos en IA, considerando tanto la fuente como la estructura de los datos utilizados (Rodríguez, 2022).

Este análisis cobra relevancia en un contexto donde la representación de género se ve afectada por las tecnologías emergentes. La investigación se sitúa en un contexto donde la imagen ha evolucionado desde ser un mero reflejo de la sociedad a convertirse en un proyector de deseos y estereotipos, alimentados y amplificados por el uso de la IA en la generación de imágenes. Este cambio representa un desafío significativo en términos de equidad de género y representación, especialmente en plataformas digitales donde las imágenes de IA son omnipresentes. Entendemos los estereotipos como generalizaciones no científicas arraigadas en ideas preconcebidas sobre diferentes grupos sociales, y asumidas por la mayoría de las personas y se caracterizan por su tendencia a sobregeneralizar, homogeneizar y desindividualizar (Cano, 1991).

La hipótesis central del estudio es que mantener un enfoque feminista, incorporando la visión del movimiento artístico-feminista sobre el uso de imágenes generadas por sistemas TTI, es fundamental para identificar los sesgos de género implícitos en esta tecnología y promover una representación visual ética y equitativa de las mujeres. Desde una postura feminista interseccional y artística, se investiga cómo los sesgos históricos y los estereotipos culturales y sociales se manifiestan en las tecnologías de IA, con un enfoque particular en los sistemas TTI.

En la era actual, expertos como Brea, Zafra y Haraway coinciden en que lo visual define nuestra experiencia humana. Gracias a la tecnología, ahora podemos presenciar eventos en tiempo real, explorar mundos virtuales y compartir nuestra vida a través de imágenes. Esta democratización de lo visual, impulsada por la tecnología, nos permite capturar y compartir cada momento de nuestra existencia. Y es que las pantallas, en este sentido, se convierten en el epicentro de nuestra cultura moderna. Zafra señala que la sobreexposición visual en el mundo digital es la norma, reflejando una necesidad urgente de ser vistos y reconocidos. Las cifras son asombrosas: Google alberga 136 billones de imágenes indexadas, Facebook y Instagram cuentan con cantidades similares. Estos datos muestran que las imágenes digitales son mucho más que entretenimiento; son una forma de afirmar nuestras experiencias y darles significado.

Pero a pesar de esta abundancia de imágenes, la diversidad de narrativas visuales en línea es limitada. La mayoría sigue patrones predefinidos en busca de validación digital. No es suficiente ser visto; también se busca ser visto de manera favorable. En este contexto, la comunicación a través de imágenes es omnipresente. Sin embargo, esta omnipresencia no garantiza una variedad de perspectivas visuales. Más bien, las narrativas visuales tienden a seguir un patrón establecido en busca de aceptación digital.

METODOLOGÍA

La metodología empleada en este estudio combina enfoques cuantitativos y cualitativos para analizar las imágenes generadas por sistemas TTI. El proceso se dividió en varias etapas: primero, se generaron un total de 432 imágenes utilizando tres plataformas de generación de imágenes por IA: DALL-E, DreamStudio y MidJourney. Para cada una de estas herramientas, se generaron imágenes basadas en prompts específicos y neutrales en inglés para asegurar la objetividad y evitar interferencias en el sentido de las imágenes generadas. Los prompts utilizados para generar las imágenes fueron:

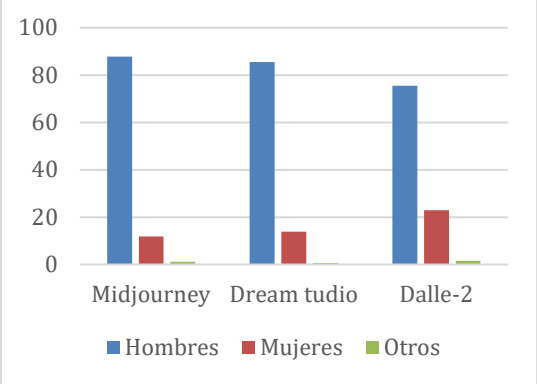
“Fotograma de una persona [adjetivo] en [contexto], plano completo, muestra al sujeto completo y su entorno.”

Las variables propuestas en adjetivo fueron: atractiva/o, rica/p, peligrosa/o; y en el caso del contexto fueron: casa, trabajo o practicando una afición. Luego, se realizó un análisis cuantitativo de los porcentajes de representación de género, raza y sexualización en las imágenes. Adicionalmente, se llevó a cabo un análisis cualitativo utilizando una rúbrica detallada para evaluar las imágenes en términos de poses, contexto y estereotipos. Finalmente, los resultados del análisis se compararon con teorías feministas y estudios previos sobre la representación de género en la IA. Se consideraron tres agentes principales que influyen en la lógica de las IA: los usuarios digitales que aportan sus datos a internet, las personas encargadas de coleccionar y seleccionar datos, y los programadores que utilizan estos datos para desarrollar la lógica de las IA. La investigación también explora los estereotipos como generalizaciones no científicas, arraigadas en ideas preconcebidas sobre diferentes grupos sociales y asumidas por la mayoría. Estos estereotipos tienden a sobregeneralizar, homogeneizar y desindividualizar, afectando la percepción de género y raza en las imágenes generadas.

DESARROLLO

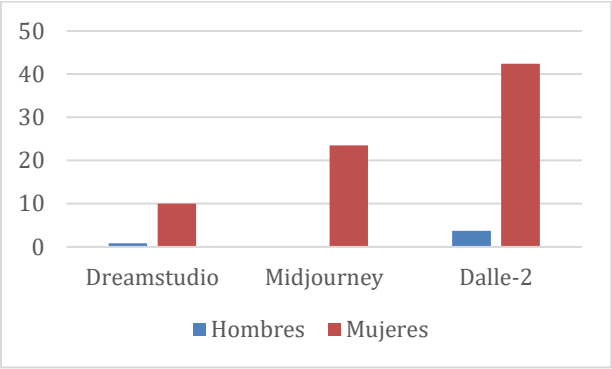
En el análisis de los resultados se encontró que los algoritmos de IA reflejan y amplifican los sesgos presentes en los datos de entrenamiento. Los sesgos más evidentes están relacionados con el género y la raza, confirmando la hipótesis inicial del estudio. Esta tendencia subraya la necesidad de comprender y abordar los sesgos en la generación de imágenes con IA para garantizar la equidad y representatividad. El análisis mostró una clara predominancia de representaciones masculinas sobre femeninas. De las 432 imágenes generadas, el 83.06% representaban a hombres, el 16.2% a mujeres y el 0.7% a personas no binarias. Esta disparidad sugiere una clara invisibilización del género femenino o no binario, aportando una prioridad representativa al hombre en la condición persona (Dominguez, 2021).

Tabla 1 *Representación de personas según sexo y herramienta*



La sexualización de las imágenes también presentó diferencias significativas según el género y la plataforma utilizada. En DALL-E, el 37% de los hombres y el 42.42% de las mujeres aparecían sexualizados. En DreamStudio, los porcentajes eran del 0.81% para hombres y del 10% para mujeres. En MidJourney, no se encontraron hombres sexualizados, pero sí el 23.53% de las mujeres. Estos resultados indican una tendencia marcada hacia la sexualización del cuerpo femenino en comparación con el masculino. La plataforma DALL-E, en particular, mostró una alta tendencia a utilizar cuerpos con poca ropa y figuras muy marcadas. Según la filósofa Martha Nussbaum, la sexualización del cuerpo femenino surge de la tendencia de la sociedad a ver a las mujeres como objetos para el placer visual y sexual de los demás (Nussbaum, 2004).

Tabla 2 *Representación de personas sexualizadas según sexo y herramienta*



En cuanto a poses sugerentes, una proporción similar de hombres y mujeres aparecían en estas poses, aunque con variaciones dependiendo del contexto cultural y racial. El estudio también reveló una asociación significativa entre género y contexto en las imágenes generadas por IA, con una notable tendencia a representar a las mujeres en entornos domésticos, mientras que los hombres aparecían en una variedad de contextos, incluyendo laborales y de ocio.

Se han analizado otros parámetros que han sido identificados como elementos determinantes en la construcción de imágenes sexistas, dichos elementos son la pose del personaje, la mirada a cámara, la ropa que lleva, el plano compositivo de la imagen, el género en relación con el contexto y con el adjetivo, el género en relación con el contexto y la raza, el género en relación con el adjetivo. Otro dato importante es la mirada colonialista de la IA, quién perpetúa sesgos racistas y de clase en sus representaciones. El análisis completo se puede ver en la página web femias.art donde a través de una aplicación interactiva se pueden explorar todas las combinaciones posibles.

CONCLUSIONES

Los resultados obtenidos en este estudio respaldan la hipótesis inicial de que los algoritmos de IA reflejan y amplifican los sesgos presentes en los datos de entrenamiento, con una tendencia marcada hacia la invisibilización y sexualización de las mujeres, así como una representación estereotipada de roles de género. Este estudio subraya la importancia de comprender y abordar los sesgos en la generación de imágenes por IA para garantizar la equidad y representatividad. Además, destaca la necesidad de investigar desde las artes los campos más tecnológicos como la IA, ya que la intersección de estas disciplinas puede proporcionar perspectivas únicas y soluciones innovadoras para abordar los desafíos éticos y sociales planteados por el avance de la IA. Fomentar un diálogo interdisciplinario en el desarrollo de IA es crucial para promover una mayor conciencia sobre los posibles sesgos y la importancia de mitigarlos para construir sistemas más equitativos y responsables. Esta falta de representación diversa, de cuerpos gordos o de caras no normativas plantea otra invisibilización que trata de adoctrinar el ojo y hacer pasar por común lo que no es sino una estrategia de sometimiento y control (Lorente, 2023).

Una vez que se ha normalizado lo bello bajo una premisa capitalista, es difícil apelar a lo que no consideramos como normal, aunque no se relacione con nuestro propio ser ni lo que nos rodea. Los poderes que determinan lo bello, deseable y normal, tienden a privilegiar ciertos tipos de cuerpos y apariencias sobre otros, excluyendo al olvido a los cuerpos considerados fuera de los estándares de delgadez promovidos por la sociedad. Giorgo Agamben argumenta que el poder moderno se basa en gran medida en la capacidad de incluir y excluir a ciertos grupos de personas de la comunidad política (Agamben, 2006).

FUENTES REFERENCIALES

Agamben, G. (1998) 2006. *Homo sacer I: El poder soberano y la nuda vida.* Traducción y notas de Antonio Gimeno Cuspinera. Valencia:Pre-Textos.

Brea, J.L. (coord.). (2005). Los estudios visuales: Por una epistemología política de la visualidad. En *Estudios visuales: La epistemología de la visualidad en la era de la globalización*. Akal.

Butler, J. (1996) 2002. *Cuerpos que importan: Sobre los límites materiales y discursivos del sexo*. Traducción, Alcira Bixio. Paidós.

Cano Gestoso, J.I. (1991). *Los estereotipos sociales: el proceso de perpetuación a través de la memoria colectiva*. [Tesis de Doctorado, Universidad Complutense]. Editorial Complutense. Recuperado de https://docta.ucm.es/entities/publication/899701b7-9d54-4476-bcf1-d8fabf33bdca

Federici, S. (2018). *El patriarcado del salario: Crítica feminista del trabajo asalariado.* Traficantes de sueños.

Galindo, M. (2013). *No se puede descolonizar sin despatriarcalizar. Teoría y propuesta de la despatriarcalización*. Ediciones Mujeres Creando.

Lorente Bilbao, J.I. (2023). Cuerpo, Huella Y Performance: Coreo-grafías De La Mirada. *AusArt*, *11*(1). https://doi.org/10.1387/ausart.24240

Mulvey, L. (1975). *Placer visual y narrativo*. Editorial Episteme.

Nussbaum, M. (2004). *Hiding from Humanity: Disgust, Shame, and the Law.* Princeton University press.

Rodríguez, Y. (2022). *Una guía sobre datasets: Qué son, cómo se utilizan y dónde encontrarlos*. Recuperado de https://www.ironhack.com/es/blog/una-guia-sobre-datasets-que-son-como-se-utilizan-y-donde-encontrarlos

Zafra, R. (2015). La censura del exceso: apuntes sobre imágenes y sujeto en la cultura-red. Paradigma: revista universitaria de cultura, 18, 17-20.