

Arte y engaño en la red: IA como recurso creativo

Art and deception on the web: AI as a creative resource

Helena Hernández Acuaviva ^a, Rocío García Robles ^b

^aUniversidad de Sevilla, acuaviva@us.es; ^b Universidad de Sevilla, rociogarcia@us.es.

Breve bio autor/es:

Helena Hernández Acuaviva, artista y Personal Investigador en Formación (PIF), Departamento de Dibujo, Facultad de Bellas Artes, Universidad de Sevilla. Grupos de Investigación: HUM-822 Gráfica y Creación Digital (US, España) y GIIP - Grupo Internacional e Interinstitucional de Pesquisa em Arte, Ciência e Tecnologia (UNESP, Brasil).

Dra. Rocío García Robles, Ingeniera Informática, Lda. Bellas Artes, Artista, profesora e investigadora. Universidad de Sevilla. Responsable del grupo de Investigación ASTERISM: Art, Science, Technology, Engineering Research: Innovation, Synergies and Methodologies.

How to cite: Hernández Acuaviva, H. y García Robles, R. (2024). Arte y engaño en la red: IA como recurso creativo. En libro de actas: EX±ACTO. VI Congreso Internacional de investigación en artes visuales aniaav 2024. Valencia, 3-5 julio 2024. <https://doi.org/10.4995/ANIAV2024.2024.17920>

Resumen

En los últimos años, con el auge exponencial del uso de internet y las redes sociales, hemos aprendido rápidamente a manejar nuestros dispositivos digitales y, con ello, a entender las posibilidades que nos brindan. Además, poco a poco nos hemos ido acostumbrando a discernir nombres de usuarios falsos, identificando posibles phishing, textos spam y otros tipos de fraudes similares. Sin embargo, con las imágenes que circulan en el espacio virtual no sucede lo mismo. A día de hoy es extremadamente fácil manipular la imagen de una persona mediante filtros y aplicaciones que alteren sus rasgos. Con suma sencillez asumimos que un rostro desconocido puede ser verdadero. Actualmente, cuando entramos en la red nos sumergimos en un maremágnum invisible de bots y algoritmos de aprendizaje inteligentes capaces, entre otras cosas, de generar imágenes sintéticas de personas que no existen; imágenes de personas que, por otra parte, resultan ser cada vez más perfectas y verosímiles, lo que hace que el fake sea difícil de detectar. Nuestro objetivo es reflexionar, desde una perspectiva creativa, sobre la sencillez con la que se pueden crear identidades falsas a través de la inteligencia artificial (IA). Con este fin, discutiremos acerca de varios artistas que emplean estas herramientas como recurso artístico y sobre proyectos que, desde un posicionamiento crítico, abordan esta problemática.

Palabras clave: algoritmos; creación artística; identidad; inteligencia artificial; internet.

Abstract

In recent years, with the exponential rise in the use of the internet and social media, we have quickly learned to operate our digital devices and, with it, to understand the possibilities they offer us. In addition, we have gradually become accustomed to discerning fake usernames, identifying phishing, spam texts and other similar types of fraud. However, the same is not true of images circulating in virtual space. Nowadays it is extremely easy to manipulate a person's image through filters and applications that alter their features. It is very easy to assume that an unknown face can be real. Today, when we go online, we are immersed in an invisible sea of bots and intelligent learning algorithms capable, among other things, of generating synthetic images of people who do

not exist; images of people who, on the other hand, are increasingly perfect and plausible, which makes the fake difficult to detect. Our goal is to reflect, from a creative perspective, on the ease with which fake identities can be created through artificial intelligence (AI). To this end, we will discuss several artists who use these tools as an artistic resource and projects that, from a critical stance, address this problem.

Keywords: *algorithms; artificial intelligence; artistic creation; identity; internet.*

INTRODUCCIÓN

El arte, sin quedarse atrás y siempre abierto a nuevos cambios, ha ido evolucionando de la mano de las tecnologías pertinentes según cada época. Desde hace unos años, el uso de dispositivos digitales conectados a la red ha aumentado exponencialmente por la fascinación de la sociedad hacia la rapidez del desarrollo tecnológico y su facilidad funcional, hasta convertirse en una parte esencial en nuestro día a día. De hecho, no nos resulta inusual que a diario miremos, por ejemplo, las notificaciones de nuestros *smartphones* y abramos nuestros chats virtuales, como WhatsApp o Telegram, para mandar mensajes rápidos -muchas veces automatizados por el corrector de texto- a amigos, familiares o compañeros de trabajo. De igual forma, la implementación acelerada de la inteligencia artificial (IA), no cesa en su avance y cada vez tiene una mayor sofisticación desde su aparición en el año 1956¹; ésta, se encuentra latente en nuestro entorno, como por ejemplo en coches automatizados o aparatos electrónicos con conexión *IoT*², y nos ayuda a ser más productivos. Eso sí, como es natural, estas tecnologías digitales, aunque nos hagan la vida más amena y tengan infinidad de ventajas positivas, tienen diversos contras o desventajas que suelen ser evidentes ante nuestra capacidad crítica. Pongamos como ejemplo el caso de la suplantación de identidad de una empresa por, por ejemplo, contener símbolos extraños en sus nombres de usuario, que es algo recurrente; o, recibir mensajes *spam*, aunque muchas veces son automáticamente bloqueados por nuestros dispositivos. Por lo contrario a este tipo de *fakes* con gran obviedad en su detección, las imágenes que habitan la red no suelen ser tan claras al respecto, dado que la verosimilitud de las imágenes manipuladas no hace juicio a las imágenes reales, haciendo altamente difícil distinguir una de otras. Con esto, podemos declarar que el ámbito artístico no ha estado exento de contar con las influencias que brinda la red y las derivadas del pensamiento automático suponiendo un reseñable giro discursivo en el arte que ha afectado a la representación, creación y visualización de imágenes. Con ello, a lo largo de las siguientes páginas reflexionaremos desde una visión artística la problemática en torno a la sencillez de crear identidades falsas con la IA, meditaremos sobre el impacto social de la tecnología y observaremos cómo lo abordan diversos artistas ofreciendo alternativas creativas.

METODOLOGÍA

La aproximación de esta comunicación, metodológicamente, será siempre llevada a cabo desde el posicionamiento de las Bellas Artes. Nos centraremos en el estudio procedente de la reflexión y la referencia

¹Tanto el término como el campo de estudio de la inteligencia artificial, tuvo lugar en el evento *The Dartmouth College Artificial Intelligence Conference*, organizado por John McCarthy, Marvin Minsky, Claude Shannon y Nathaniel Rochester en el año 1956. <https://doi.org/10.1609/aimag.v27i4.1911>

² El concepto *IoT*, o Internet de las cosas, alude a la tecnología y a la red colectiva de dispositivos conectados en que facilita la comunicación tiempo real entre distintos dispositivos y el almacenamiento en la nube -almacenamiento de datos que está basado en redes de computadoras-, así como entre los propios dispositivos.

artística -de pensamiento y creación-, además de otros autores de distintos campos adyacentes del entendimiento.

DESARROLLO

Desde comienzos de la expansión del uso de internet en los años 90 (Telefónica, 2023), los net.artistas aprovecharon las peculiaridades de la red, su inmediatez e inmaterialidad, para tomarla como un lugar ideal para la redefinición del concepto de identidad. Estos artistas en red, creaban imágenes generadas que, como da por supuesto el profesor Juan Martín Prada, serán “siempre blandas, como “cuerpos de datos” para poetizar sobre lo múltiple y lo colectivo” (MARTÍN PRADA, 2015, p.155). La identidad siempre ha sido un tema importante en lo tocante al arte y ha estado presente a lo largo de la historia. No obstante, en la actualidad, en el contexto de las redes sociales en línea y la IA, se ha visto inevitablemente afectada por la sobreexposición y saturación del yo (GERGEN, 2006) que, aunque no tengamos la intención ni queramos ser expuestos en los espacios virtuales de compartición, donde nos proyectamos a través de imágenes, queramos o no, vamos a estar forzosamente bajo la mirada de ojos robotizados e inteligentes que nos rastrean y nos observan: siempre estamos “en escena” (MARTIN PRADA, 2023, p.64). Las tecnologías digitales, internet y los algoritmos permiten extender la identidad más allá del cuerpo, y con ello, posibilitar libremente la elección de representación de cada usuario ante los demás, de la forma que cada uno desee, como puede ser a través de nuestras mejores fotografías para nuestro perfil de una red social.

Según la RAE, la identidad se podría definir como el “conjunto de rasgos propios de un individuo o de una colectividad que los caracterizan frente a los demás” (RAE, 2023) y se hace eco tanto por el lado físico como el mental. Creemos oportuno mencionar el punto de vista de los profesores James E. Côté y Charles G. Levine, que sostienen en su libro *Identity, Formation, Agency, and Culture. A Social Psychological Synthesis* (CÔTÉ, LEVINE, 2002), sobre un modelo de diferenciación identitaria en tres niveles de actualización: 1. Identidad social (posición de cada individuo en la estructura social); 2. Identidad personal (experiencias individuales concretas basadas en interacciones); 3. Identidad del ego (sentido subjetivo más fundamental de la continuidad de la personalidad) (González et al., 2005). Así, teniendo en cuenta la forma que tenemos de relacionarnos y habitar las plataformas digitales, indudablemente, podemos decir que están emanando nuevas identidades. Es por ello, que la cuestión identitaria se concibe como un objeto de reflexión para una multitud de artistas hoy día.

Así, para pensar en torno a la identidad del ser humano y la máquina, podemos señalar la IA como una herramienta que proporciona recursos artísticos esenciales. Aunque el empleo de esta tecnología, en muchas ocasiones, implica tener conocimientos informáticos esenciales, cada día existen más opciones de softwares y algoritmos preentrenados que no requieren preparación por parte del usuario y que son fáciles de usar. Por ejemplo, podríamos mencionar programas de IA que crean imágenes a partir de descripciones textuales o *prompts*, como DALL-E, Firefly, Stable Diffusion o Midjourney. Asimismo, nos parece oportuno mencionar el sitio web viral *This Person Does Not Exist*³, del ingeniero de software Phillip Wang, que genera rostros diferentes completamente convincentes e hiperrealistas de personas que, como alude su título, no existen (PÁEZ, 2019). Cada imagen se crea cada vez que se refresca la página gracias al aprendizaje automático llamado ‘red generativa adversativa’ (RGA) o ‘generative adversarial network’ (GAN)⁴. Este modelo de aprendizaje automático ha servido

³ *This Person Does Not Exist* parte del proyecto *This X Does Not Exist*, el cual utiliza redes generativas adversarias para crear versiones falsas e inexistentes de casi cualquier cosa con aspecto realista -como, entre otros ejemplos, gatos, currículums, sustancias químicas, diputados, cielos nocturnos, aeropuertos o caballos-. Link de acceso: <https://thisxdoesnotexist.com/>.

⁴ La red generativa adversativa (RGAs), también conocida como GAN en inglés, es una clase de algoritmos de IA usados en el aprendizaje no supervisado para la generación de imágenes, especialmente de rostros. Originalmente fue codificada y publicada por Nvidia en 2019. <https://github.com/NVLabs/stylegan>

de base para la realización del proyecto *DoppelGANger.agency* (Fig. 1) creado por Mitra Azar en 2019. Este proyecto tiene como objetivo hacer que los temas relacionados con nuestra privacidad y el control por los aparatos electrónicos se expongan desde una perspectiva artística y política a través de la performance y la ingeniería inversa, basadas en la ficción y las narrativas que, aun mostrándose como utópicas, no se alejan mucho de la realidad (The Influencers, 2019). El proyecto parte de la idea que faltan humanos con capacidad de relacionarse emocionalmente entre ellos y el mundo tecnocrático que les rodea, a causa del acelerado flujo de información, cada vez más acelerado, y por los conocidos ‘filtros burbujas’ online, que, como remarca el autor en su web, “generan comunidades cerradas en las que las personas rebotan continuamente las mismas ideas y estéticas como si estuvieran atrapadas entre espejos” (DoppelGANger.agency, s.f.). Así como también, fomentan la intolerancia a lo distinto, personalizando los resultados de las búsquedas en internet y redundando en el mismo tipo de información (Pariser, 2017). El proyecto pretende que todos los seres humanos tengamos nuestro doble algorítmico único, con la esperanza de dar un primer paso hacia una nueva idea de privacidad relacionada con las tecnologías biométricas. Por ahora, existen 7000 millones de estos rostros y se comprometen a emparejar a todo ser humano de la Tierra con su *DoppelGANger*⁵. Por otra parte, la creación de nuevas identidades inexistentes le sirvió al artista como inspiración para la exposición que tuvo lugar en el Fotomuseum de Winterthur titulada con *DoppelGANger*⁶, entre 2019-2020. Ésta consistió en realizar distintas intervenciones en las que Pariser difundió carteles de personas desaparecidas, creadas mediante el algoritmo, en las calles de Cuba, París y Milán.



Fig.1. Captura de pantallas de página web. “DoppelGANger.agency”. Mitra Azar, 2019 (Fuente: Página web del proyecto. <http://doppelganger.agency/>)

Por otro lado, un trabajo que combatiría con esa obsesiva persecución por apariencia verosímil de los rostros que suelen habitar las redes es el del artista mediático, hacker e investigador británico Jake Elwes, quien desarrolló el proyecto denominado *Zizi - Queering the Dataset* (Fig. 2) en 2019 y que actualmente está en curso. Un videoarte a través del cual se combinan la identidad *drag*⁷ sintetizada y *deepfake*⁸ creada mediante

⁵ Doppelgänger es un vocablo alemán que se podría traducir como "doble andante" ('doppel' sería doble y 'gänger' andante). Este término fue acuñado por el escritor Jean Paul en 1796. https://www.elconfidencial.com/alma-corazon-vida/2022-02-09/doppelganger-donde-viene-razones-gemelo_3371014/

⁶ Acceso a más información sobre la exposición: <https://www.fotomuseum.ch/en/situations-post/doppelganger/>

⁷ El concepto *drag* se refiere a la performance de expresión de individuos que, muchas veces, exageran los aspectos de género y sus roles con fines de entretenimiento. Por lo general, encontraríamos tres tipos de *drag*: *drag queen*, que representa la feminidad; *drag king*, que escenifica la masculinidad; y, *drag queer*, que interpreta una androginia o neutralidad de género. <https://journals.sagepub.com/doi/10.1177/1363460709352725>

⁸ El término *deepfake* es un acrónimo inglés formado por las palabras *fake* (falsificación) y *deep learning* (aprendizaje profundo). Fue acuñado por primera vez en 2017 por un usuario, con el mismo nombre, de la plataforma de noticias sociales Reddit, que compartió videos pornográficos que utilizaban tecnología de intercambio de rostros de código abierto. <https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained>

aprendizaje automático. El proyecto expone cómo el *drag* desafía el género y explora la alteridad, el ‘ser otro’, y, cómo la IA, responde con imágenes que se alejan de las identidades normativas; con ello, se evidencia cómo los bancos de imágenes más empleados para el entrenamiento de sistemas de reconocimiento con IA carecen de cierto tipo de identidades. Esta obra, según consta en la web del autor, se realizó alterando los sistemas StyleGAN⁹ de 2019 entrenado sobre FFHQ¹⁰ 2018, que contiene 70.000 fotografías de rostros, y los volvió a entrenar añadiendo 1000 imágenes de rostros travestidos y de género fluido halladas online. Con esto, la red neuronal se aleja del entrenamiento original y adquiere un carácter para la creación de nuevas identidades *queer*, evidenciando lo que ha aprendido y lo que no (ELWES, 2019). Y con ello, abriría camino a una libre interpretación a raíz de las deformaciones y expresiones desproporcionadas de los diferentes sujetos.



Fig.2. Serie de 24 impresiones con pigmentos de archivo sobre papel Platine Fibre Rag 21x21cm. “Zizi - Queering the Dataset”. Jake Elwes, 2019
(Fuente: Página web del artista. <https://www.jakeelwes.com/project-zizi-2019.html>)

Otros artistas contemporáneos que utilizan el *deepfake* en su proceso creativo son Bill Posters y Daniel Howe, quienes desarrollaron la obra *Spectre* (Fig. 3) en 2019. Se trata de una instalación interactiva e inmersiva que expone una serie de vídeos falsos hiperrealistas mostrando a figuras públicas -tales como Mark Zuckerberg, Donald Trump o Kim Kardashian- mediante el arte sintético generados con IA que expresan comentarios disparatados o provocadores sobre temas relacionados con el control informativo, el libre albedrío, la publicidad en línea, la privacidad y la influencia de las redes sociales. *Spectre* toma su título del nombre del personaje en línea del Dr. Aleksandr Kogan, el investigador de datos que vendió 87 millones de perfiles de Facebook a Cambridge Analytica. Su objetivo es provocar una reflexión crítica sobre la facilidad con la que se puede manipular la información digital y cómo puede afectar a la percepción pública y al constructo social de lo que suponemos como verdad. Hace pensar sobre la necesidad de desarrollar nuevas formas de alfabetización

⁹ Las GANs tradicionales, mencionadas anteriormente, y las StyleGAN se diferencian en la arquitectura del generador. En StyleGAN se introduce una arquitectura nueva y distintiva, que mejora la calidad y la variedad de las imágenes generadas. <https://www.toolify.ai/es/ai-news-es/style-gan-una-nueva-forma-de-generar-imagenes-2109567>

¹⁰ Flickr-Faces-HQ (FFHQ) es un conjunto de datos de imágenes de alta calidad de rostros humanos, creado originalmente como referencia para las redes generativas adversariales (GAN). Se compone de 70.000 imágenes PNG de alta calidad con una resolución de 1024 × 1024 y contiene una variación considerable en términos de edad, origen étnico, fondo de la imagen y accesorios. Las imágenes, bajo licencia previa, fueron rastreadas desde la plataforma Flickr y alineadas y recortadas automáticamente por la biblioteca de software multiplataforma dlib. <https://github.com/NVLabs/ffhq-dataset>

mediática y ética tecnológica para enfrentar estos desafíos. Para contribuir a esta finalidad, la obra, compuesta por seis monolitos de 2,2 metros de altura, incluye elementos interactivos que permiten a los visitantes alimentarla con sus datos, generando una experiencia personalizada, para entender mejor cómo funcionan estas tecnologías y despojarse mentalmente de la supuesta verosimilitud basada en su hiperrealismo. Evidentemente este proyecto se difundió rápidamente por redes sociales, además de tener una cobertura inesperada en la prensa global, que causó gran revuelo y que llegó, incluso, a provocar respuestas oficiales de Facebook, Instagram y Youtube.



Fig. 3. Imagen digital. "Spectre". Bill Posters y Daniel Howe, 2019.

(Fuente: Página web de Bill Posters. <https://billposters.ch/the-zuckerberg-deepfake-heard-around-the-world/>)

Hoy, como hemos observado, la imagen digital es fácil de crear por las máquinas y, a menudo, se presenta en la red con cambios casi imperceptibles para el ojo humano. Precisamente por ello, los artistas nos exponen el peligro que puede llegar a ocasionar la inestabilidad de la información que consumimos a través de las imágenes y que podemos llegar a aceptar como verdades.

CONCLUSIONES

Como hemos podido vislumbrar a través de distintos proyectos artísticos que experimentan y reflexionan en torno al uso del aprendizaje automático y el *deepfake*, hemos procurado evidenciar cómo la identidad se ha visto indudablemente afectada por el uso exponencial de las redes sociales en línea.

Por un lado, observamos con el proyecto *DoppelGANger.agency* de Mitra Azar cómo los algoritmos nos permiten extender la identidad más allá del cuerpo, y con ello, nos propone como solución un doble algorítmico único para cada persona del planeta Tierra como una nueva idea de privacidad que resguarde nuestra intimidad.

Por otro lado, trabajos como *Zizi - Queering the Dataset* combatiría las apariencias verosímiles de los rostros que suelen habitar las redes y presentarían nuevas propuestas de identidades que no pretenden ser hiperrealistas y que se presentan como alternativas nuevas, creadas mediante programación de algoritmos basados en grandes cantidades de datos.

Por último, la obra instalativa *Spectre* es un claro ejemplo de cuán fácil puede ser manipular y editar cualquier imagen digital y, en este caso, se hace evidente a través de videos de celebridades alimentados por los propios datos de los visitantes.

En suma, a lo largo de las páginas anteriores hemos mostrado, de la mano de los artistas mencionados, que la problemática en torno a la sencillez de crear identidades falsas con la IA es una cuestión vigente en nuestra sociedad. Además, hemos visto cómo el arte, sin poder desligarse de los acontecimientos sociales de cada época, hoy está marcado por las influencias que brinda la red y las derivadas del pensamiento automático, suponiendo un relevante giro discursivo que ha afectado a la representación, manipulación, creación y visualización de imágenes.

AGRADECIMIENTOS Y FINANCIACIÓN

La investigación presentada cuenta con el apoyo financiero del VII Plan Propio de Investigación y Transferencia de la Universidad de Sevilla (VII PPIT-US), al cual agradecemos la oportunidad de hacer posible esta producción científica.

FUENTES REFERENCIALES

- Côté, J. y Levine, C. (2002). *Identity formation, agency and culture: a social psychological synthesis*. Lawrence Erlbaum Associates.
- DoppelGANger.agency. (s.f.). *DoppelGANger.agency*. <http://doppelganger.agency:3000/>
- Elwes, J. (2019). *Zizi - Queering the Dataset*. <https://www.jakeelwes.com/project-zizi-2019.html>
- Gergen, K. J. (2006). *El yo saturado: Dilemas de identidad en el mundo contemporáneo*. Paídos.
- González, S. et al. (2005). Revisión del constructo de Identidad en la Psicología Cultural. *Revista de Psicología de la Universidad de Chile*, 14(2), 9-25. <https://www.redalyc.org/pdf/264/26414202.pdf>
- Paez, D. (2019). 'This Person Does Not Exist' Creator Reveals His Site's Creepy Origin Story. *Inverse*. <https://www.inverse.com/article/53414-this-person-does-not-exist-creator-interview>
- Mcloughlin, M. (16 de octubre de 2018). El algoritmo entrenado con porno y clásicos que pinta fascinantes retratos en directo. *El Confidencial*. https://www.elconfidencial.com/tecnologia/2018-10-16/algoritmo-porno-cuadros-maestros-clasicos-madrid_1630658/
- Martín Prada, J. (2015). *Prácticas artísticas e Internet en la época de las redes sociales*. Ediciones Akal.
- Pariser, E. (2017). *Filtro Burbuja. Cómo la red decide lo que leemos y lo que pensamos*. Taurus.
- Real Academia Española. (s.f.). Identidad. En *Diccionario de la lengua española*. Recuperado el 19 de mayo de 2023, de <https://dle.rae.es/identidad>
- Telefónica. (06 de septiembre de 2023) Historia de Internet: ¿cómo nació y cuál ha sido su evolución? *Telefónica* <https://www.telefonica.com/es/sala-comunicacion/blog/historia-internet-como-nacio-evolucion/>
- The Influencers. (2019). *Mitra Azar* <http://doppelganger.agency/>. <https://theinfluencers.org/mitra-azar>