



LA REVOLUCIÓN EN LA CREACIÓN VISUAL: La Inteligencia Artificial Generativa

MIGUEL CASAS ARIAS¹, ALEJANDRO PRIEGO DÍAZ¹, MARÍA LARA-MARTÍNEZ¹

¹ Universidad Complutense de Madrid, España

PALABRAS CLAVE

Inteligencia artificial
Fotografía
Midjourney
Visual
Redes Sociales

RESUMEN

La integración de la inteligencia artificial (IA) en la creación audiovisual está redefiniendo los límites entre la creatividad humana y el potencial tecnológico y su uso está muy extendido en redes sociales. Esta investigación revisará los antecedentes técnicos y se propone como objetivos analizar la aplicación de la inteligencia artificial en las diferentes etapas de la producción visual, donde se estudiará si el profesional de la comunicación puede aprovechar sus conocimientos para sacar un mayor rendimiento a estas herramientas. Las conclusiones determinan que la inteligencia artificial está involucrada en el surgimiento de nuevas formas de expresión artística y comunicativa.

Recibido: 29 / 04 / 2024

Aceptado: 16 / 06 / 2024

1. Introducción

A medida que el volumen y la velocidad de los datos masivos continúan expandiéndose, asumimos desde hace ya algunos años que la Inteligencia Artificial (IA) desempeñará un papel cada vez más importante en contextos relacionados con los medios (Forrester, 2017). Los productores de contenido multimedia están adoptando la IA para enriquecer la eficiencia y acelerar el rendimiento de la producción multimedia (Jayanthiladevi et al., 2020).

Sin embargo, resulta desafiante ya que el método de interacción entre hombre y máquina más extendido en tecnología con IA es el lenguaje natural o NLP, *natural language processing*, y tanto las imágenes como las descripciones del lenguaje en el mundo real son ruidosas, con variabilidad lingüística e inconsistencias al transmitir conceptos complejos (Zhang y Peng, 2018).

La tecnología audiovisual con IA ha desarrollado sistemas de generación de imágenes a partir de texto, que se han popularizado como herramientas para la creación de imágenes digitales y obras de arte (Crowson et al., 2023). Actualmente, tanto profesionales audiovisuales y artísticos como usuarios de redes sociales sin formación especializada pueden generar su contenido mediante estas aplicaciones. Es más, está al alcance de cualquier usuario de internet crear contenido mediante la simple introducción de textos en lenguaje natural. En la literatura científica analizada aparece referida con distintos nombres, pero es comúnmente conocida por el término utilizado para a los mensajes que se introducen en los distintos sistemas para generar instrucciones a la IA: *prompts*.

Estos sistemas de creación de contenido combinan algoritmos generativos con una interfaz fácil de entender para ingresar indicaciones en lenguaje natural y recibir imágenes generadas con los conceptos y características transmitidas en el texto. La accesibilidad de estos sistemas ha mejorado mucho recientemente, lo que ha aumentado el conocimiento de la población sobre su capacidad para producir imágenes de alta resolución. Los resultados son sorprendentes y originales, y aportan estilos estéticos que coinciden con las indicaciones de texto con una tasa de precisión alta (Brisco et al., 2023). Por ese motivo, en este estudio se plantea la cuestión de hasta qué nivel la IA generativa de texto a imagen podría llegar a influir o contribuir en el diseño y creación del contenido visual profesional, en concreto, de imágenes fijas. Además, la IA se está implementando en todos los procesos de producción audiovisual, sea cine, televisión o video, redes sociales y radiodifusión. Esta investigación analiza el camino que tomará la naturaleza de la creatividad, la productividad y la calidad de los resultados aplicados en fotografía.

Existen importantes preocupaciones éticas al incorporar la IA en el arte. Las cuestiones de autoría y derechos de autor todavía no están claras, y la destrucción de puestos de trabajo es un tema recurrente. Cuando una IA crea una pieza de arte, ¿quién es el artista: la persona que creó el algoritmo, la persona que lo puso en marcha o el propio sistema de IA? Además, en muchos campos, como el arte, la posibilidad de sesgo en los algoritmos de IA es una preocupación ética. El arte producido puede perpetuar estereotipos o reflejar puntos de vista parciales si los datos utilizados para entrenar estos sistemas contienen sesgos.

1.1. Objetivos de la investigación

Esta investigación tiene dos objetivos. Primero, ubicar la IA en la producción audiovisual de los medios y redes sociales, analizando aplicaciones con soluciones a través de IA y su funcionalidad. En segundo lugar, poner a prueba las herramientas de generación de imágenes a través de texto más extendidas. En este estudio serán examinados los desafíos y necesidades emergentes relacionadas con el uso de estas aplicaciones de *software* poniendo el foco en las nomenclaturas y herramientas utilizadas y en la necesidad de obtener resultados profesionales compatibles con las decisiones creativas humanas.

Este artículo se propone así lograr una comprensión de cómo la IA está cambiando el proceso de creación de contenido en el ámbito visual y cómo esto está afectando al surgimiento de nuevas formas de expresión artística y comunicativa.

1.2. Justificación

Hay abundante literatura que aborda las aplicaciones de la IA, pero muy poca se ha centrado en los medios audiovisuales. Y tampoco se ha llevado a cabo un estudio enfocado a las carencias presentes en la IA como medio para su explotación como generadora de contenido visual dentro de unos parámetros que alcancen un nivel profesional. Los avances son tan rápidos que cada semana las grandes compañías

dedicadas a esta tecnología implementan novedades técnicas más avanzadas, y se ha llegado a un momento de desarrollo en el que hay que clarificar el alcance y la posición de las aplicaciones de IA en el entorno visual.

2. Estado de la Cuestión

2.1. Contextualización de la IA en el entorno audiovisual

Desde finales de la década de 1950 han surgido varias prácticas artísticas novedosas vinculadas a la tecnología. Se ha encontrado una clara y sintética descripción de estas prácticas por parte de Margaret

A. Boden y Ernest A. Edmonds en su artículo «What is Generative Art» (Boden y Edmonds, 2009) que nos enmarcan, en el ámbito de la generación de contenido, dónde se desarrolla la actual tecnología de IA:

- Arte computarizado: arte en cuyo proceso productivo intervienen los ordenadores.
- Arte electrónico: Este concepto amplio abarca cualquier obra de arte cuya producción involucre ingeniería eléctrica y/o tecnología electrónica.
- Arte generativo: Obras generadas al menos en parte por algún proceso que no está bajo el control directo del artista.

Estas tecnologías han sido numerosas veces discutidas en cuanto a su estética frente a la teoría del arte. Concretamente y producto de su enorme crecimiento, ha sido en las redes sociales donde se ha dado un mayor auge en el uso de nuevas tecnologías de movimiento creativo, sobre todo tras la pandemia, muy por encima de cualquier otro ámbito (Arana Arrieta et al., 2020).

El actual crecimiento que ha tomado la IA comenzó con el desarrollo del Chat GPT (Generative Pre-trained Transformer) en 2022, una tecnología que se integra en la IA Generativa. Se ocupa de la creación de formas de comunicación entre humanos y máquinas utilizando el lenguaje cotidiano, tanto oral como escrito. Aunque la primera representación contemporánea del modelo de chat GPT apareció en 2018, su calidad y complejidad ha aumentado mucho.

Las imágenes creadas hace años a través de una IA tenían una resolución bastante baja, lo que las hacía oscuras y se eliminaban varios detalles (Zhang y Peng, 2018). La presente evolución creemos que marcará un hito en la tecnología. Hoy en día se puede presenciar un número cada vez mayor de casos en los que los cineastas emplean la IA (Momot, 2022).

2.2. Generación automática de contenido audiovisual con IA

Estos programas pueden crear imágenes completamente originales a partir de un texto con una descripción basada en un lenguaje de palabras. La información del texto se procesa para generar una imagen que se intenta ajustar a la descripción. Gracias al progreso en las técnicas de aprendizaje profundo de la IA, la existencia y la asequibilidad de conjuntos de datos y la capacidad informática, con un ordenador es posible generar imágenes a través de una descripción textual y a la inversa, descripciones textuales para una imagen (Brisco et al., 2023; Zhang y Peng, 2018).

Dentro de una producción audiovisual, es en esta área de la generación automática de contenidos, donde la IA ha avanzado enormemente en la automatización de tareas que anteriormente requerían la intervención humana. Gatys y otros investigadores discuten en «Image Style Transfer Using Convolutional Neural Networks», sobre el desarrollo de los algoritmos capaces de aprender representaciones visuales artísticas y aplicar estilos específicos a imágenes, mostrando el poder de la IA en la creación visual (Gatys et al., 2016). Aunque la IA generativa ya había demostrado su eficiencia en la generación de contenido, su implementación ha sido moderada hasta el año 2020 (Son et al., 2019). Su aplicación en industrias como el cine, las plataformas digitales y la radiodifusión es significativamente limitada. El potencial actual de la tecnología de IA generativa se sitúa dentro del dominio de las imágenes digitales y el video (Lee, 2023).

El lanzamiento de la comunidad Midjourney resultó ser un momento clave en este sentido. Los miembros pueden ver las imágenes creadas por otros miembros junto con las indicaciones (*prompts*) utilizadas. Por lo tanto, las indicaciones se comparten en la comunidad, lo que la convierte en recurso de aprendizaje social, desbloqueando la creatividad de lo que anteriormente era una práctica de un solo

creador (Oppenlaender, 2022). De esta manera se han descubierto muchos *prompts* para modificar el estilo y la calidad de las imágenes y se están aplicando regularmente en la comunidad Midjourney (Oppenlaender, 2023).

2.2.1. Generación de texto a imagen

En 2021, se lograron avances significativos en la generación de texto a imagen para la síntesis de contenido visual digital (Cobb, 2023). Estos sistemas ya han alcanzado un desarrollo que les permite generar imágenes de gran nivel estético, especialmente cuando se utilizan *prompts* para alterar el estilo o mejorar la calidad de la imagen (Brisco et al., 2023). Aunque en muchos casos, se necesitaban más pasos de procesamiento, como por ejemplo la corrección de ojos en sujetos humanos (Nightingale y Farid, 2022).

En la mayoría de los casos, especialmente en redes sociales donde dominan en número las imágenes de usuarios no profesionales, el arte de texto a imagen está impulsado por la exploración. Los usuarios sondan su idea o su modelo con indicaciones textuales para ver qué funciona (y qué no). Este método también podría extenderse a los profesionales. Por lo tanto, se trata de un desarrollo que es progresivo, y que vincula ideas entre indicaciones posteriores. A menudo se necesitan varias interacciones para llegar a un resultado subjetivamente satisfactorio (Oppenlaender, 2022). El proceso también se ve afectado por la propia interpretación de la IA y la ambigüedad de las palabras para representar ideas (Zhang y Peng, 2018). En este punto es voluntad del usuario permitir que la conversación con la IA se desvíe en direcciones imprevistas. Y lo mismo para los profesionales, donde igualmente pueden seguir por casualidad las oportunidades presentadas durante el desarrollo de su creación.

Los actuales usuarios de aplicaciones de texto a imagen no profesionales, en consecuencia, tienen como su única opción seleccionar cuidadosamente una muestra de sus mejores trabajos, mientras descartan (o deciden no publicar) otras imágenes. El arte de texto a imagen es similar al proceso de un fotógrafo de filtrar y eliminar una selección de imágenes de una colección más grande, lo que permite la exploración rápida de más alternativas, una ventaja competitiva significativa. La comunidad de Midjourney tiene este proceso de selección incorporado, en donde solo las imágenes que el miembro de Midjourney eligió mejorar se muestran de forma predeterminada. Esta práctica permite obtener unas métricas holísticas de filtrado que pueden ser aplicables para evaluar la creatividad del artífice de los *prompts* (Oppenlaender, 2022).

Liu y Chilton (2022) investigan las capacidades generativas de la generación de imagen a través de texto en «Design Guidelines for Prompt Engineering Text-to-Image Generative Models». En este artículo se lleva a cabo un análisis sobre los resultados de los *prompts* que han ganado popularidad entre los expertos. Su objetivo es establecer normas de diseño que faciliten la obtención de resultados superiores mediante el uso de esta tecnología.

2.2.2. Generación de texto a video

Esta rama de la IA generativa está en fase de experimentación y rápido crecimiento. Consideramos que será este un periodo de poco tiempo hasta que se consolide y ofrezca resultados óptimos y a gran escala.

Generar video requiere mucho más poder de cómputo y memoria que generar imágenes fijas, dado que se deben crear múltiples imágenes (cuadros) para recrear el movimiento de una imagen digital. Además, se debe mantener una coherencia estética y de contenido, y la fluidez del movimiento a lo largo de la secuencia. Esto implica no solo la generación de varias imágenes sino también la comprensión y aplicación de principios cinematográficos, como la continuidad y la narrativa visual (Fu et al., 2023).

En fecha de febrero de 2024, los resultados de calidad de imagen de los videos generados a través de IA son notables, aunque aún no han alcanzado el nivel de realismo que se puede alcanzar con imágenes fijas.

2.3. Herramientas de postproducción

El desarrollo de aplicaciones de inteligencia artificial (IA) en el ámbito de la postproducción audiovisual presenta con cada innovación, un avance notable. La singularidad de estas aplicaciones radica en su capacidad para especializarse en diferentes utilidades dentro del proceso creativo. Durante los últimos meses aparecen novedades cada semana con nuevas herramientas y versiones más evolucionadas. En la captura de imágenes y videos, los algoritmos pueden automatizar procesos como

el enfoque, la iluminación y la composición. Además, en la edición, la IA facilita la identificación de elementos clave, mejora de la calidad de imagen, e incluso puede sugerir opciones creativas para la narrativa visual. Entre las posibilidades de hacer postproducción se pueden encontrar herramientas para cambiar caras, crear arte original, quitar fondos a imágenes, mejorar la calidad de la imagen, incluir elementos en videos, adaptar los labios de un locutor a otro idioma, generar canciones y efectos de sonido, generar voces, doblar voces a otros idiomas, clonar voces, arreglar videos desenfocados, animar personajes y objetos 3D, crear contenido interactivo para videojuegos y VR, analizar y clasificar contenido visual y auditivo, generar de subtítulos, traducir audio en tiempo real, y reconocimiento facial y de objetos en videos.

Una de las aplicaciones más extendidas en postproducción es la edición de imágenes basada en *prompts*. A pesar de que las instrucciones humanas a veces son demasiado breves para que los métodos actuales las ejecuten adecuadamente (Fu et al., 2023), sin embargo, ya se puede automatizar la tarea de gestión de contenidos, la sincronización de audio/video, las operaciones multimedia y reducir los errores humanos (Jayanthiladevi et al., 2020).

Una de las herramientas con mayor repercusión, es la de poder completar partes internas de imágenes, lo que se denomina *image inpainting* (Zhang et al. 2020). El término *inpainting* se refiere específicamente a la técnica de reconstrucción de partes perdidas o dañadas de imágenes o videos, empleando algoritmos de inteligencia artificial para rellenar esos espacios de manera coherente con el resto de la imagen (Elharrouss et al., 2020). Estas novedosas prácticas creativas ofrecen un nivel de interactividad más allá de la mera generación de imágenes estáticas a partir de indicaciones de entrada de texto (Oppenlaender, 2023). Esta tecnología creemos que redefine la postproducción fotográfica y de video.

3. Diseño y método

3.1. Hipótesis

Los generadores de contenido y tomadores de decisiones ya sean usuarios de redes o profesionales de los medios necesitan no sólo generar, sino concluir en ideas creativas, pero también de calidad para que se produzca la innovación y el impacto deseado. En este ámbito tanto las redes como el mundo profesional se nutren de numeroso material fotográfico, sin ir más lejos, una de las redes más utilizadas para generar contenido es Instagram, especializada en fotografía y video.

Ante esa capacidad generativa de las IAs surge la siguiente pregunta, ¿son capaces las herramientas de generación de imágenes a partir de texto de proporcionar buenos resultados en la creación de fotografías que se asemejen a las realizadas por un fotógrafo profesional de una forma tradicional y puedan ser utilizadas en redes o en publicaciones profesionales sin que los usuarios sean capaces de diferenciarlas de fotografías reales?

Además, ¿qué tipo de *prompts* pueden ser más útiles para esa generación de imágenes fotográficas: los meramente artísticos y descriptivos o los que contienen parámetros profesionales fotográficos?

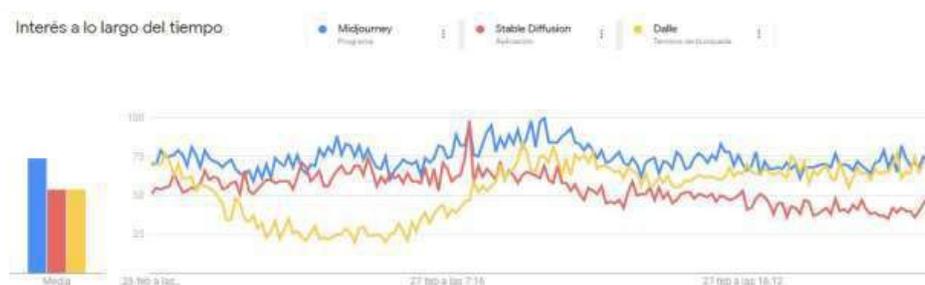
3.2. Metodología

Primeramente, se ha realizado una revisión documental de la literatura científica existente para poder situar objetivamente el lugar que ocupa la IA en la industria audiovisual. Un análisis en profundidad de la información disponible se ha enfocado sobre uno de sus usos con mayor crecimiento: los contenidos generados de texto a imagen, texto a video y sus aplicaciones concretas.

Nos ha llamado la atención la escasa literatura encontrada en Scopus y Web of Science. Sin embargo, Google Scholar tiene numerosas publicaciones, si bien entre estas predominan los estudios orientados a análisis particulares de las novedades tecnológicas de los desarrolladores sin mayor pretensión investigadora.

Por otro lado, se propone una metodología experimental utilizando las herramientas líderes del mercado en generación de imágenes: Midjourney, Stable Diffusion y Dall-e 3. Para llegar a este resultado se ha realizado una consulta previa a través de Google Trends sobre las 3 aplicaciones más buscadas en Google entre 20 de las más conocidas durante el último año.

Figura 1. Gráfico comparativo de Google Trends



Fuente: Google Trends, 2024.

Con este punto de partida, se realizará un análisis posterior del rendimiento de las tres seleccionadas en cuanto al realismo de las imágenes creadas basándonos en varios parámetros.

Procederemos a dar a estas aplicaciones *prompts* idénticos, con instrucciones con terminología técnica, para evaluar y comparar su capacidad de responder a comandos profesionales específicos. Este método permitirá una comparación imparcial de cómo cada sistema maneja y procesa los mismos *prompts*, proporcionando una base sólida para comprender las fortalezas y limitaciones de cada herramienta en el contexto de la generación de imágenes profesionales y poder así hacer determinar si el resultado se puede considerar profesional y sobre todo conforme en exactitud a las ideas preliminares del creador de la idea.

El experimento consistirá en una comparación de los métodos tradicionales de creación fotográfica y la generación de un trabajo similar mediante el uso de la IA generativa. Para ello se va a partir del trabajo de dos fotógrafos profesionales conocidos por su obra y se van a tratar de imitar algunos de sus trabajos a través de los modelos generativos profundos, introduciendo en ellos una serie de *prompts* tanto de nivel descriptivo como de nivel técnico que se puedan extraer de las propias fotografías de los autores.

La evaluación de los resultados se basará en un análisis cuantitativo y cualitativo de la experiencia y resultados obtenidos de primera mano utilizando ambas técnicas. Se compararán los resultados extraídos del uso de la IA con las fotografías reales que se toman como base de trabajo.

Para poder comprobar si esto es posible, se ha hecho una primera parte del experimento que ha consistido en proporcionar en estas tres IAs generativas de texto a imagen un *prompt* básico para la generación de una fotografía. En este caso se ha introducido la siguiente descripción textual para generar las fotografías requeridas: «a photography of Concha Beach in Donosti» («una imagen de la Playa de la Concha de Donosti»).

Una vez generadas las imágenes en cada uno de los modelos, se pueden observar grandes diferencias entre el resultado ofrecido por cada uno de los modelos. En DALL E 3 (Fig. 2) el resultado se asemeja a una ilustración. Se observa que parece una pintura, no hay elementos que destaquen por ser fieles a la realidad, simplemente se percibe una representación artística del lugar seleccionado. En Stable Diffusion (Fig. 3), aunque es un resultado más realista, los detalles no permiten que se pueda considerar que se acerca a un resultado fotográfico profesional, debido a que ni las personas ni las fachadas de los edificios que hay en la imagen están bien definidas.

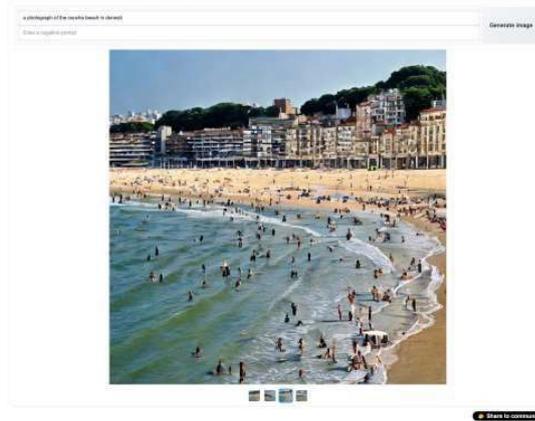
Mientras que Midjourney ha generado imágenes que son coherentes con el *prompt* propuesto, el resultado son cuatro imágenes que simulan una fotografía de la Playa de la Concha (Fig. 4). En ella se distinguen claramente la arena, el agua, las personas, los accidentes geográficos y los edificios y, además, se asemejan en un alto grado a lo que se percibe en una imagen fotográfica normal. Por tanto, nos centraremos en el uso del modelo de Midjourney para el resto del experimento.

Figura 2. Imagen de la playa de la Concha creada con DALL E-3



Fuente: DALL E-3, 2024.

Figura 3. Imagen de la playa de la Concha creada con Stable Diffusion



Fuente: Stable Diffusion, 2024.

Figura 4. Imágenes de la playa de la Concha creadas con Midjourney



Fuente: Midjourney, 2024.

Además, Midjourney ofrece una gran facilidad de manejo, lo que permite que el tiempo requerido para que su uso sea el adecuado es mucho menor que con otras, puesto que antes de estudiar unos *prompts* determinados se tienen que generar una gran cantidad de imágenes (Liu y Chilton, 2022).

El siguiente paso consiste en la generación de distintas imágenes con aspecto de fotografías a través de una serie de *prompts* que imiten la obra de fotógrafos profesionales con una obra muy emblemática. Para ello se ha decidido trabajar a partir de las obras de Martin Parr y Ansel Adams.

Martin Parr es un fotógrafo británico altamente conocido, miembro de la agencia Magnum y su obra se encamina principalmente a la fotografía documental, de un estilo costumbrista con un gran sentido

del humor. Los ejemplos de su obra desde los que se realiza el experimento son dos fotografías de la serie *Life's a beach* de mediados de los años 90 del siglo XX (Fig. 5 y 6).

Figura 5. Fotografía de Martin Parr de la serie *Life's a beach*



Fuente: www.martinparr.com, 1997.

Figura 6. Fotografía de Martin Parr de la serie *Life's a beach*



Fuente: www.martinparr.com, 1997.

Ansel Adams fue un fotógrafo americano que destacó por sus fotografías en blanco y negro de paisajes de parques naturales de Estados Unidos. Para el experimento se utilizan las siguientes obras: *In Glacier National Park*, Montana (Fig. 7) y *Long's Peak*, Rocky Mountain National Park, Colorado (Fig. 8).

Figura 7. Fotografía de Ansel Adams *In Glacier National Park*



Fuente: archives.com.gov, 1942.

Figura 8. Fotografía de Ansel Adams *Long's Peak*



Fuente: archives.com.gov, 1942.

Se efectúan tres experimentos diferentes introduciendo distintas descripciones textuales para generar las imágenes requeridas. En cada uno de ellos el objetivo principal es conseguir una imagen lo más cercana posible a la obra de ambos autores.

4. Trabajo de Campo y Análisis de Datos

Para llevar a cabo el experimento se realiza una sesión de generación de imágenes utilizando *prompts* en inglés para conseguir la imagen deseada en Midjourney. Se experimenta con una extensa diversidad de *prompts*, explorando el efecto de diferentes descripciones, estilos o número de palabras en las imágenes resultantes. Éstas se guardan en un formato PNG de alta resolución para su posterior análisis. En total se generan 125 imágenes originales en base a la obra de los dos autores de referencia entre los días 20 y 26 de febrero de 2024.

Las imágenes generadas se analizan de un modo cualitativo en base a distintos criterios entre los que se encuentran la cercanía con la realidad de los elementos generados, el grado de similitud con las obras de referencia o el grado de correspondencia con el *prompt* seleccionado.

Se escogen ejemplos representativos de distintos niveles de resultado para poder así mostrar en detalle las posibilidades que en la actualidad ofrece la IA generativa de texto a imagen y, en particular, el modelo Midjourney en la labor concreta de reproducción de imágenes fotográficas.

5.Resultados

Al comienzo del experimento el uso de *prompts* se limita a una descripción textual muy sencilla, se le pide Midjourney que genere una fotografía en el estilo de cada uno de los fotógrafos mencionados y otra en la que se especificará el contenido de la fotografía.

En primer lugar, se pide a la IA que genere fotografías al estilo de Martin Parr («a photography in Martin Parr's style»). Midjourney devuelve cuatro imágenes bastante realistas y que podrían considerarse casi como imágenes fotográficas (Fig. 9), éstas tienen con un estilo muy determinado, bastante cercano a la obra de Martin Parr y con unos colores e iluminación muy característicos.

Figura 9. Imágenes generadas en Midjourney al estilo de Martin Parr



Fuente: Midjourney, 2024.

En cuanto a la obra de Ansel Adams, para generar las imágenes se genera un *prompt* similar al usado anteriormente, en este caso se pide a Midjourney que genere una fotografía al estilo de Ansel Adams («a photography in Ansel Adams' style») (Fig. 10). En este caso el resultado son cuatro imágenes de paisajes muy cercanos a las fotografías realizadas por Ansel Adams, en las que se observan montañas nevadas, grandes pinos y elementos acuáticos, como ríos o lagos, en los cuales, hasta se pueden observar reflejos de otras partes de la imagen.

Figura 10. Imágenes generadas en Midjourney al estilo de Ansel Adams



Fuente: Midjourney, 2024.

En una segunda fase, se aprovecha una de las opciones que ofrece Midjourney para incluir a estas imágenes algún otro prompt más técnico. Se trata de una opción que permite modificar alguna de las cuatro imágenes generadas incluyendo nuevos textos. Esto se realiza con las imágenes generadas para ambos autores. El resultado que da la IA es otras cuatro imágenes basadas en la seleccionada de las creadas con anterioridad.

En el caso de Martin Parr, se ha seleccionado la segunda imagen, de una mujer con gafas y sombrero rosas en un barco (Fig. 11), y, en el caso de Ansel Adams, se ha elegido la simula las montañas con nieve con un lago en primer plano y el reflejo de las montañas en él (Fig. 12).

Figura 11. Imagen generada en Midjourney al estilo de Martin Parr



Fuente: Midjourney, 2024.

Figura 12. Imagen generada en Midjourney al estilo de Ansel Adams



Fuente: Midjourney, 2024.

Se le pide a Midjourney que genere desde la imagen de la mujer otras en las que se utilice una óptica de 18 mm y una menor profundidad de campo («a photography in Martin Parr style with a lens of 18 mm and less depth of field»). El resultado son cuatro imágenes (Fig. 13) parecidas a la anterior en las que se aprecia a una mujer con gafas y sombrero rosa en primer término y con los fondos totalmente desenfocados, también se observa que ninguna de las imágenes se corresponde con lo que sería una fotografía tomada con una lente de 18 mm. Por otro lado, las texturas de la ropa y de la piel guardan similitud con la realidad y salvo el detalle de una taza de café que al borde de una mesa de una forma extraña, no hay grandes errores en las imágenes.

Figura 13. Imágenes generadas en Midjourney al estilo de Martin Parr



Fuente: Midjourney, 2024.

Para el caso de la imagen del estilo de Ansel Adams, se pide a Midjourney que genere a partir de ella, una imagen que utilice una lente de 8 mm y una menor profundidad de campo («a photography in Ansel Adams style with a 8mm lens and less depth of field»), el resultado (Fig. 14) son cuatro imágenes muy similares a la anterior, donde se puede apreciar levemente que la óptica podría ser un angular, pero sin llegar a un 8mm y la profundidad de campo no es baja.

Figura 14. Imágenes generadas en Midjourney al estilo de Ansel Adams



Fuente: Midjourney, 2024.

En la segunda parte del experimento se trabaja de la siguiente manera: cogiendo como base las obras de ambos autores seleccionadas al comienzo del estudio se procede a describir esas fotografías para crear un *prompt* meramente descriptivo basado en ellas para que Midjourney genere nuevas imágenes.

Para que genere una imagen similar a las de la serie *Life's is beach* de Martin Parr (Fig. 5 y 6) se le introduce el siguiente *prompt*: «a photograph of people sitting on the beach with sun loungers and umbrellas, sunbathing. Some sun loungers are empty, others are people sitting. Other people are lying

on towels. All sun loungers and umbrellas are brightly colored. Not all the beach is full. You can see the water and some people in it. The photograph is from the eye level of a 1.75 meter person with an 18mm lens».

Figura 15. Imágenes generadas en Midjourney describiendo la figura. 6



Fuente: Midjourney, 2024.

Las imágenes generadas (Fig. 15) son bastante realistas, en todas ellas se cumplen los parámetros usados para su creación, las texturas de arena y agua son acertadas, las sombrillas, tumbonas y demás elementos playeros se acercan bastante a la realidad, pero se puede ver cómo hay carencias en la representación de algunas de las personas que aparecen, posiciones corporales inverosímiles, piezas de ropa incompletas, facciones poco definidas o con errores.

Para crear una imagen basada en las fotografías de Ansel Adams (Fig. 7 y 8) se utiliza el siguiente *prompt*: «a photograph in black and white of a View from above of plateau with some pine trees, snow covered mountain in background in long peak, rocky mountain national park with a 35 mm objective».

Figura 16. Imágenes generadas en Midjourney describiendo la figura. 7



Fuente: Midjourney, 2024.

Como podemos observar en este caso se han conseguido cuatro imágenes realistas, que cumplen perfectamente con lo pedido a Midjourney.

El siguiente paso es incluir en estos *prompts* que sean imágenes al estilo de los autores. Para ello, se elige una de las imágenes generadas para a partir de ella crear las nuevas incluyendo en el texto (Fig. 17 y 18).

Figura 17. Imagen generada en Midjourney



Fuente: Midjourney, 2024.

Figura 18. Imagen generada en Midjourney



Fuente: Midjourney, 2024.

El efecto de incluir en la descripción que las imágenes sean del estilo de Martin Parr es que se producen cuatro fotografías (Fig. 15) con los mismos rasgos positivos en la representación en la que se ha basado la IA para su generación, es decir los elementos inanimados y el agua están bien imitados, pero también cuenta con las mismas carencias que anteriormente, es decir, cuerpos humanos no del todo bien definidos, con algunas partes cortadas o deformes.

Figura 19. Imágenes generadas en Midjourney describiendo la figura 6



Fuente: Midjourney, 2024.

En cuanto a las imágenes concebidas al estilo de Ansel Adams (Fig. 20), se puede decir que siguen con la misma literalidad la descripción hecha para su predecesora y consiguen simular sin ningún tipo de problema los distintos elementos básicos de una fotografía del autor.

Figura 20. Imágenes generadas en Midjourney describiendo la figura 8



Fuente: Midjourney, 2024.

La tabla propuesta a continuación se integra en nuestro estudio con el fin de proporcionar un análisis detallado y cuantitativo sobre diversos parámetros cruciales en la evaluación de imágenes generadas mediante inteligencia artificial. Cada categoría examinada, es evaluada en un espectro entre 1 a 5, proporcionando así una métrica comparativa objetiva. Esta tabla facilita la comprensión y establece un marco comparativo para futuras investigaciones.

Tabla 1. Análisis de Realismo y Coherencia de las Imágenes Generadas

	Adecuación/aproximación a una fotografía	Coherencia de elementos
Paisajes 1 ^{er} plano	12345	12345
Paisajes 2 ^o plano	12345	12345
Rostros 1 ^{er} plano	12345	12345
Rostros 2 ^o plano	12345	12345
Extremidades 1 ^{er} plano	12345	12345
Extremidades 2 ^o plano	12345	12345
Objetos 1 ^{er} plano	12345	12345
Objetos 2 ^o plano	12345	12345
Texturas	12345	12345
Simulación del estilo	12345	12345
Iluminación	12345	12345
Profundidad de campo	12345	12345
Selección de lente	12345	-----
Eye Level	12345	-----

Fuente: Elaboración propia, 2024.

6. Discusión y Conclusiones

6.1. Discusión

La habilidad de escribir indicaciones efectivas para interactuar con las herramientas de creación de contenido se está convirtiendo en un área de conocimiento crucial. Según Oppenlaender (2023), la efectividad en la redacción de estas indicaciones no solo se basa en el conocimiento técnico de las redes neuronales, sino también en la experiencia y familiaridad de la persona con diferentes modificadores de

solicitud (los *prompts*). Hemos constatado que tener una comprensión detallada de cómo funcionan estas redes y cómo se pueden manipular sus entradas para obtener los resultados deseados es fundamental para aprovechar al máximo su potencial. Coincidimos con la idea de que la ingeniería rápida de mensajes es un aprendizaje y una habilidad, porque no es inmediatamente evidente cómo escribir de manera efectiva los *prompts* y qué palabras clave resultan buenos modificadores (Rogers et al., 2020). Creemos que la interpretación y el correcto uso de los *prompts* es algo que exige experiencia y nos planteamos si dentro de estas habilidades se encuentran las nomenclaturas más técnicas y profesionales, o los *prompts* han de limitarse a descripciones más generales.

Entre los desafíos que identificamos a través de nuestro análisis, encontramos que una introducción restringida no intencionada de datos es un elemento que limita la explotación de la IA. Estas restricciones de información suponen una brecha técnica importante para llegar a los resultados deseados. Además de esto hay que tener en cuenta que el diseño y programación de las IA están diseñados por las empresas desarrolladoras y llevado a cabo por humanos, y por lo tanto existen los sesgos. Estamos de acuerdo con que el sesgo es un problema humano, no de máquina, por lo que el enfoque para abordarlo no puede basarse exclusivamente en soluciones técnicas (Schetinger et al., 2023). Pero también creemos que se ha demostrado con el experimento que las limitaciones técnicas todavía son notables y no se solucionan todas las alucinaciones que genera la IA mediante nuevos *prompts* que traten de modificar estos resultados.

Cuando se diseña un contenido audiovisual a través de la IA, con un criterio dirigido a un objetivo concreto, ya sea personal o por encargo, pero no aleatorio, es importante tener los conceptos del resultado al que pretendemos llegar bien definidos, con características específicas de estructura, estilo y mensaje. Estas características son clave para lograr objetivos funcionales del contenido, como transmitir un mensaje específico, contar una historia o generar una respuesta emocional en la audiencia (Brisco et al., 2023). Sin embargo, actualmente nos enfrentamos a ciertas limitaciones. Estos sistemas no siempre comprenden completamente los atributos estructurales y de comportamiento detallados en el texto. Hemos observado que, sin especificar una petición de estilo concreto, las escenas generadas pueden tener un aspecto surrealista y onírico, y que no siguen las convenciones típicas del mundo real o del contenido audiovisual tradicional. En la práctica, esto significa que si un director de cine o un creador de contenido intenta usar una IA de texto a imagen puede llegar en primera instancia a resultados que parecen poco realistas o artísticamente abstractos. Aunque esto puede ser beneficioso para proyectos creativos que buscan un estilo visual novedoso, único o surrealista, no es adecuado para proyectos que requieran representaciones precisas y realistas basadas en conceptos detallados.

Con el desarrollo de la IA, nos enfrentamos a una situación donde el dominio del amateurismo muy común sobre todo en redes sociales favorece la posibilidad de que personas sin ninguna formación artístico/técnica puedan desarrollar sus ideas más fácilmente. Los sistemas basados en IA son una extensión de nuestro trabajo cognitivo humano, y tienen el potencial de aumentar y apoyar nuestra creatividad. Pero la capacidad de generar ideas creativas es una condición insuficiente para la innovación (López et al., 2019). Coincidimos con López en que para crear contenidos de interés a través de IA los creadores de contenido y tomadores de decisiones ya sean usuarios de redes o profesionales de los medios necesitan no solo generar y concluir en ideas creativas, sino también obtener resultados con calidad para que se produzca la innovación y el impacto deseado.

La tecnología de transformar texto en imagen está generando una previsible y creciente interacción con estos sistemas. A medida que esta relación continúe, evolucionarán nuestras habilidades comunicativas con la IA. Esta nueva relación con la informática puede tener un efecto en nuestro comportamiento, lenguaje, conocimientos y habilidades. De manera similar a cómo la World Wide Web cambia la sociedad (tanto como la sociedad cambia la Web), la interacción con la IA en lenguaje natural dará forma a la futura sociedad digital, la forma en que interactuamos con las computadoras y la forma en que trabajamos (Oppenlaender, 2022).

Sin embargo, consideramos que la inclusión de nuevas tecnologías en una estructura de equipo profesional puede traer aspectos críticos que necesitan un mínimo de conciencia por parte de los creadores de contenido. Debido a la amplia gama de conocimientos específicos que los diseñadores necesitan para mantenerse al día en esta sociedad tan tecnológicamente dinámica, las nuevas herramientas y competencias suponen un desafío para crear productos que estén bien alineados con las necesidades del usuario (Figoli et al., 2022).

6.2. Conclusiones

En esta investigación se ha demostrado que la IA ya se ha introducido como un asistente creativo y de productividad de manera coherente con la realidad actual. Sus beneficios permitirán que tenga cada vez una mayor presencia, facilitará la personalización de contenidos y en un futuro se volverá esencial para optimizar la eficiencia operativa durante la producción.

El arte generativo permite una fusión entre la creatividad humana y la de la máquina, ofreciendo una nueva perspectiva sobre la expresión artística. El dominio y la habilidad con que se utilizan los *prompts* refuerza la creatividad. Su accesibilidad ha permitido que prospere una nueva generación de creativos autodidactas y sin preparación especializada.

Actualmente la IA generativa ya permite, dependiendo de los recursos y el tiempo dedicado, la creación de contenido más o menos profesional. En cuanto a la creación de representaciones visuales fotográficas fieles a la realidad a partir del uso de un texto descriptivo, las imágenes generadas a través del modelo Midjourney mostraron paralelismos muy evidentes con respecto de la obra de los autores propuestos. Igualmente, demostró una gran precisión a la hora de interpretar el texto propuesto.

Aun así, centrando el análisis en la interpretación de ciertos conceptos más técnicos como la profundidad de campo o una óptica determinada, los resultados no son del todo satisfactorios, al igual que con la representación de elementos humanos complejos en grandes cantidades y sin ser parte del plano principal de la fotografía.

Por el momento los modelos de IA tienen limitaciones en cuanto a la flexibilidad para realizar ajustes finos o modificaciones específicas que un profesional puede necesitar, lo que limita su utilidad en entornos de producción especializados. La inconsistencia en los resultados generados por IA y la falta de previsibilidad pueden ser problemáticas en un entorno profesional donde se requiere consistencia y fiabilidad.

Estamos en los inicios de una nueva generación de herramientas y es pronto para evaluar cuáles finalmente se harán imprescindibles o una vez pasado el efecto novedad tendrán una nueva evolución a otra tecnología y pasarán de moda. El tiempo resolverá si finalmente estos proyectos colaborativos resultan en obras que fusionan la capacidad creativa humana con la eficiencia analítica de la IA, o si predominará un futuro audiovisual efectista y de bajo presupuesto donde las IA coparán la mayor parte de los procesos.

Referencias

- Adams, A. (1942). National Archives. <https://www.archives.gov/espanol/ansel-adams>
- Arana Arrieta, E., Mimenza Castillo, L. y Narbaiza Amillategi, B. (2020). Pandemia, consumo audiovisual y tendencias de futuro en comunicación. *Revista de Comunicación y Salud*, 10(2), 149–183. [https://doi.org/10.35669/rcys.2020.10\(2\).149-183](https://doi.org/10.35669/rcys.2020.10(2).149-183)
- Boden, M. A. y Edmonds, E. A. (2009). What is generative art? *Digital Creativity*, 20(1–2), 21–46. <https://doi.org/10.1080/14626260902867915>
- Brisco, R., Hay, L. y Dhimi, S. (2023). Exploring the role of text-to-image AI in concept generation. *Proceedings of the Design Society*, 3, 1835–1844. <https://doi.org/10.1017/pds.2023.184>
- Chen, L., Wang, P., Dong, H., Shi, F., Han, J., Guo, Y., Childs, P. R. N., Xiao, J. y Wu, C. (2019). An artificial intelligence based data-driven approach for design ideation. *Journal of Visual Communication and Image Representation*, 61, 10–22. <https://doi.org/10.1016/j.jvcir.2019.02.009>
- Cobb, P. J. (2023). Large Language Models and Generative AI, Oh My! *Advances in Archaeological Practice*, 11, 363–369). *Cambridge University Press*. <https://doi.org/10.1017/aap.2023.20>
- Crowson, K., Biderman, S., Kornis, D. y Stander, D. (2023). VQGAN-CLIP: Open Domain Image Generation and Editing with Natural Language Guidance. *Arxiv. Cornell University*. <https://doi.org/10.48550/arXiv.2204.08583>
- Elharrouss, O., Almaadeed, N., Al-Maadeed, S. y Akbari, Y. (2020). Image Inpainting: A Review. *Neural Processing Letters*, 51, 2007–2028. Springer. <https://doi.org/10.1007/s11063-019-10163-0>
- Evans, Z., Carr, C., Taylor, J., Hawley, S. H. y Pons, J. (7 febrero 2024). Fast Timing-Conditioned Latent Audio Diffusion. *Arxiv. Cornell University*. <https://doi.org/10.48550/arXiv.2402.04825>
- Figoli, F. A., Mattioli, F. y Rampino, L. (2022). AI in the design process: training the human-AI collaboration. *Proceedings of the 24th International Conference on Engineering and Product Design Education 2022. The design society*. <https://doi.org/10.35199/EPDE.2022.61>
- Forrester Consulting (5 octubre 2017). *The Machine on your Team: New study shows how marketers are adapting in the Age of AI*. <https://www.prnewswire.com/news-releases/the-machine-on-your-team-new-study-shows-how-marketers-are-adapting-in-the-age-of-ai-300531385.html>
- Fu, T.-J., Hu, W., Du, X., Wang, W. Y., Yang, Y. y Gan, Z. (2023). Guiding Instruction-based Image Editing via Multimodal Large Language Models. *Arxiv. Cornell University*. <https://doi.org/10.48550/arXiv.2309.17102>
- Gatys, L. A., Ecker, A. S. y Bethge, M. (2016). *Image Style Transfer Using Convolutional Neural Networks*. Computer Vision Foundation. https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Gatys_Image_Style_Transfer_CVPR_2016_paper.pdf
- Jayanthiladevi, A., Raj, A. G., Narmadha, R., Chandran, S., Shaju, S. y Krishna Prasad, K. (2020). AI in Video Analysis, Production and Streaming Delivery. *Journal of Physics: Conference Series*, 1712(1). <https://doi.org/10.1088/1742-6596/1712/1/012014>
- Lee, S. (2023). Transforming Text into Video: A Proposed Methodology for Video Production Using the VQGAN-CLIP Image Generative AI Model. *International Journal of Advanced Culture Technology*, 11(3), 225–230. <https://doi.org/10.17703/IJACT.2023.11.3.225>
- Liu, V. y Chilton, L. B. (2022, April 29). Design Guidelines for Prompt Engineering Text-to-Image Generative Models. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3491102.3501825>
- López, C. E., Miller, S. R. y Tucker, C. S. (2019). Exploring biases between human and machine generated designs. *Journal of Mechanical Design, Transactions of the ASME*, 141(2). <https://doi.org/10.1115/1.4041857>
- Mirowski, P. W., Mathewson, K. W., Pittman, J. y Evans, R. (2023). Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. *CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3544548.3581225>
- Molina-Siles, P. y Giménez Ribera, M. (2023). Inteligencia artificial y creatividad para la generación de imágenes arquitectónicas a partir de descripciones textuales en Midjourney. Emulando a Louis I. Kahn, *EGA Expresión Gráfica Arquitectónica*, 28(49), 238–251. <https://doi.org/10.4995/ega.2023.19294>

- Momot, I. (2022). *Artificial Intelligence in Filmmaking Process Future Scenarios*. [Bachelor's thesis]. <https://urn.fi/URN:NBN:fi:amk-2022052712497>
- Nightingale, S. J. y Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences of the United States of America*, 119(8). <https://doi.org/10.1073/pnas.2120481119>
- Oppenlaender, J. (2022). The Creativity of Text-to-Image Generation. *ACM International Conference Proceeding Series*, 192–202. <https://doi.org/10.1145/3569219.3569352>
- Oppenlaender, J. (2023). A taxonomy of prompt modifiers for text-to-image generation. *Behaviour and Information Technology*. <https://doi.org/10.1080/0144929X.2023.2286532>
- Parr, M. (1997). Martin Parr's official website. <https://www.martinparr.com/>
- Rogers, A., Kovaleva, O. y Rumshisky, A. (2020). *A Primer in BERTology: What we know about how BERT works*. Arxiv. Cornell University. <https://doi.org/10.48550/arXiv.2002.12327>
- Schetinger, V., Di Bartolomeo, S., El-Assady, M., McNutt, A., Miller, M., Passos, J. P. A. y Adams, J. L. (2023). Doom or Deliciousness: Challenges and Opportunities for Visualization in the Age of Generative Models. *Computer Graphics Forum*, 42(3), 423–435. <https://doi.org/10.1111/CGF.14841>
- Son, J.-W., Han, M.-H. y Kim, S.-J. (2019). Artificial Intelligence-Based Video Content Generation. *Electronics and Telecommunications Trends*. <https://doi.org/10.22648/ETRI.2019.J.340304>
- Sosa, R. y Gero, J. S. (2016). Multi-dimensional creativity: A computational perspective. *International Journal of Design Creativity and Innovation*, 4(1), 26–50. <https://doi.org/10.1080/21650349.2015.1026941>
- Steinfeld, K. (2023). Clever little tricks: A socio-technical history of text-to-image generative models. *International Journal of Architectural Computing*, 21(2), 211–241. <https://doi.org/10.1177/14780771231168230>
- Wang, X., Li, Y., Zhang, H. y Shan, Y. (2021). Towards Real-World Blind Face Restoration with Generative Facial Prior. Arxiv. Cornell University. <https://doi.org/10.48550/arXiv.2101.04061>
- Zhang, C. y Peng, Y. (2018). Stacking VAE and GAN for Context-aware Text-to-Image Generation. *2018 IEEE 4th International Conference on Multimedia Big Data, BigMM 2018*. <https://doi.org/10.1109/BIGMM.2018.8499439>
- Zhang, L., Chen, Q., Hu, B. y Jiang, S. (2020). Text-Guided Neural Image Inpainting. *MM 2020 - Proceedings of the 28th ACM International Conference on Multimedia*, 1302–1310. <https://doi.org/10.1145/3394171.3414017>