



"Siete argumentos sobre las imágenes de IA y los medios generativos". Capítulo 5 de *Artificial Aesthetics* por Lev Manovich y Emanuele Arielli.

Publicado el 20 de Abril 20, 2023.

Traducción a cargo de María Ortiz para la cátedra de Comunicación Transmedia (FSOC - UBA)

Palabras clave: IA, IA generativa, teoría de medios, ciencia de datos, herramientas computacionales, arte digital, medios generativos, cultura visual, estudios de software, futuro de los medios, historia de la IA

Referencia bibliográfica: Manovich, L. y Arielli, E. (2023) "Siete argumentos sobre las imágenes de IA y los medios generativos". Capítulo 5 en *Artificial Aesthetics*, November 2021 – August 2024. <http://manovich.net/index.php/projects/artificial-aesthetics-book>

Parecería que nos encontramos en los comienzos de una verdadera revolución de la creación en los medios: el auge de los medios generativos. Vengo usando herramientas computacionales para arte y diseño desde 1984, y ya llevo vistas varias revoluciones importantes en los medios, incluyendo la introducción de las computadoras Mac y aplicaciones de escritorio para la creación y edición de medios, el desarrollo de gráficos y animación 3D fotorealista, el ascenso de la web después de 1993, y el auge de las redes sociales después de 2006. La nueva revolución de la IA generativa parece ser tan significativa como cualquiera de estas. De hecho, es posible que sea tan significativa como la invención de la fotografía en el siglo diecinueve o la adopción de la perspectiva lineal en el arte occidental en el siglo dieciséis.

(Si eres nuevo en este tema, he aquí una breve historia: La revolución de los medios generativos se viene desarrollando hace más de 20 años. Los primeros papers sobre IA que proponían el uso del vasto y no estructurado universo web de textos, imágenes y otros artefactos culturales para entrenar computadoras en la realización de tareas varias ya habían aparecido en 1999-2001. En 2015 los métodos "deep dream" y "style transfer" de Google llamaron la atención: de repente las computadoras podían crear nuevas imágenes artísticas imitando los estilos de muchos artistas famosos. El lanzamiento de DALL-E en enero 2021 marcó otro hito: ahora las computadoras podían sintetizar imágenes a partir de una descripción textual. Midjourney, Stable Diffusion, y DALL-E 2 contribuyeron a la aceleración de esta evolución en 2022. Ahora las imágenes sintéticas podrían tener muchas estéticas, desde el foto realismo hasta la de cualquier clase de medio físico o digital, incluyendo mosaicos, pinturas al óleo, fotografía de calle o el render de gráficos por computadora 3D. El código para producir estas imágenes, referido como "modelo" en el campo de la inteligencia artificial se hizo público en agosto de 2022, desencadenando un aluvión de experimentos y acelerando su desarrollo.)

En este y el próximo capítulos voy a describir una serie de características de los *medios visuales generativos* en sus formas actuales que creo son particularmente importantes o novedosas. Algunos de mis argumentos aplican también a los medios generativos en general, pero la mayoría hacen foco en los *medios visuales* - reflejando mi propia experiencia de uso, casi todos los días, de algunas herramientas populares para generar imágenes con IA como Midjourney y Stable Diffusion (a veces también Runway ML) entre mediados de 2022 a principios de 2023. Pero primero definamos los términos principales.

Los Términos

En este texto, "artista" o "creador" se refiere a cualquier persona cualificada que crea objetos culturales en cualquier medio o sus combinaciones. Los términos "medios generativos," "medios generados por la IA", "IA generativa," y "medios sintéticos" son intercambiables. Se refieren al proceso de creación de nuevos objetos mediáticos, como imágenes, animación, video, texto, música, modelos y escenas 3D, y otros tipos de medios, recurriendo a redes neuronales profundas. Las redes neuronales también se usan para generar elementos específicos y tipos de contenido, como rostros humanos fotorrealistas, poses y movimientos, además de los objetos propiamente dichos. Y se pueden usar en la edición de medios, para reemplazar una porción de una imagen o de un video con otro contenido que encaje espacialmente.

Estas redes están entrenadas en base a grandes colecciones de objetos mediáticos ya existentes. Los tipos de redes neuronales artificiales más populares para la generación de objetos de medios incluyen modelos de difusión, modelos texto-a-imagen, redes adversariales generativas (GAN), y transformadoras. Para la generación de imágenes fijas y dinámicas usando redes neuronales. Los términos generación de imágenes, imágenes sintéticas, imágenes IA, y visuales IA pueden usarse de modo intercambiable.

Observemos que la palabra "generativa" también puede usarse de distintos modos para referirse a la creación de artefactos culturales mediante cualquier proceso algorítmico (y no sólo redes neuronales), incluso un proceso basado en reglas que no utiliza computadoras. Este es el uso típico que tienen hoy en día las frases "arte generativo" y "diseño generativo" en el discurso cultural y de medios populares. En este capítulo estoy usando "generativo" de un modo más restrictivo para designar métodos de redes profundas y apps para la generación de medios que recurren a estos métodos.

1. La 'IA' como percepción cultural

No existe una tecnología específica o un único proyecto de investigación llamado "IA". Sin embargo, podemos rastrear cómo nuestra percepción cultural de este concepto evolucionó en el tiempo y a qué se refería en cada período. En los últimos 50 años, siempre que una habilidad o destreza supuestamente exclusiva de los humanos es automatizada por medio de tecnología computacional, nos referimos a ella como 'IA'. Sin embargo, ni bien esta automatización se perfecciona y se vuelve exitosa, tendemos a dejar de referirnos a ella como un "caso de IA". En otras palabras, "IA" se refiere a tecnologías y metodologías que automatizan habilidades cognitivas humanas y están empezando a funcionar pero todavía no del todo. La "IA" ya estaba presente en las primeras herramientas computacionales para medios. El primer sistema de dibujo y diseño interactivo, el Sketchpad de Ivan Sutherland (1961-1962), tenía una función que cerraba automáticamente cualquier círculo o rectángulo

que el usuario hubiera comenzado a dibujar. O sea, Sketchpad sabía lo que el/la usuario estaba tratando de hacer. En la muy amplia interpretación del término explicitada arriba, esto ya era, sin duda, un ejemplo de “IA”.

Mi primera experiencia con un programa para pintar de escritorio en una Apple II fue en 1984, y era verdaderamente alucinante mover el ratón y ver las pinceladas de pintura simuladas

aparecer en pantalla. Sin embargo hoy ya no consideramos más que esto sea “IA”.

Otro ejemplo sería la función de seleccionar automáticamente el contorno de un objeto del Photoshop. Esta función fue agregada hace muchos años - y es también “IA” en un sentido amplio, aunque nadie se referiría a ella de este modo hoy en día. La historia de los sistemas y herramientas de medios digitales está llena de estos ‘momentos IA’ – al principio causan asombro, luego se dan por sentado y después de un rato son olvidados como “IA”. (En los libros de historia de la IA, este fenómeno es conocido como “el efecto IA”.) En este momento, el término ‘IA creativa’ hace referencia, exclusivamente, a métodos desarrollados recientemente mediante los cuales las computadoras transforman algunos inputs en nuevas salidas o outputs para medios (por ejemplo en los modelos texto-a-imagen) y técnicas específicas (por ejemplo, determinados tipos de redes neuronales profundas). Tenemos que recordar, sin embargo, que estos métodos no son ni los primeros ni los últimos en la larga historia y en el futuro de simular las habilidades de los humanos para las artes o de asistir a los humanos en la creación de medios.

2. “Make it New”: IA y el modernismo

Después de ser entrenadas en base a billones de páginas de texto o miles de millones de imágenes de arte y fotografía tomadas de la web, las redes neuronales logran generar textos y visuales nuevos del nivel de escritores, artistas, fotógrafos, o ilustradores profesionales muy competentes. Estas capacidades de las redes de los sistemas de IA están distribuidas a lo largo de billones de conexiones entre miles de millones de neuronas artificiales en vez de estar determinadas por algoritmos estándar. En otras palabras, hemos desarrollado una tecnología que en términos de complejidad es extremadamente similar al cerebro humano. No terminamos de entender cómo funciona nuestra tecnología IA, al igual que no logramos aprehender del todo el intelecto y la creatividad humanas.

La generación actual de sistemas de IA generativa, como el GPT y Stable Diffusion, son entrenados en base a enormes y muy diversos conjuntos de datos, que consisten en miles de millones, a veces millones de millones de textos individuales, o pares de imágenes y textos. Es igualmente interesante limitar el conjunto de datos de entrenamiento a un área específica dentro del universo más amplio de la historia cultural humana, o a un conjunto específico de artistas dentro de un período histórico específico. *Unsupervised* por Refik Anadol Studio (2022) es un proyecto de arte e IA que ejemplifica estas posibilidades. El proyecto usa redes neuronales entrenadas con el conjunto de datos de las decenas de miles de piezas de arte de la colección del MoMA¹. Esta colección, en mi opinión, es una de las mejores representaciones del período más creativo y experimental en la historia visual humana - cien años de arte moderno (1870 - 1970) - así como muchos ejemplos importantes de exploraciones artísticas en las décadas subsiguientes. La colección captura

¹ MoMA: Museum of modern Art, o Museo de Arte Moderno de Nueva York

los febres e implacables experimentos de los artistas modernistas para crear nuevos lenguajes visuales y de comunicación y "hacerlo nuevo."



Unsupervised, Refik Anadol Studio (2022). Una selección de cuadros de la animación.

En la superficie, la lógica del modernismo parecería oponerse diametralmente al proceso de entrenamiento de los sistemas de la IA generativa. Los artistas modernos querían distanciarse del arte clásico y de sus características distintivas como la simetría visual, las composiciones jerárquicas y el contenido narrativo. Su arte estaba basado en el rechazo fundamental de todo lo que lo había antecedido (al menos en teoría, como lo expresaban en sus manifiestos). Las redes neuronales, en cambio, son entrenadas del modo opuesto, ya que aprenden de la cultura y del arte histórico creados hasta hoy. Una red neuronal es análoga a un artista muy conservador que estudia en el "meta" "museo sin muros" que alberga el arte histórico.

Pero todos sabemos que la teoría y la práctica del arte no son la misma cosa. Los artistas modernos no rechazaban completamente el pasado y todo lo que les precedió. En cambio, el arte moderno se desarrolló reinterpretando y copiando imágenes y formas de viejas tradiciones artísticas, como los grabados japoneses (van Gogh), la escultura africana (Picasso), y los íconos rusos (Malevich). Por lo tanto, los artistas sólo rechazaban los paradigmas dominantes del arte elevado de su tiempo, el arte realista y el de salón, pero no el resto de la historia del arte humano. De hecho el modernismo era profundamente historicista: en vez de inventar todo de cero, innovó adaptando determinadas estéticas más antiguas a los contextos del arte contemporáneo. (En el caso del arte geométrico abstracto creado en la década de 1910, estos artistas usaron imágenes que ya venían siendo usadas en el campo de la psicología experimental para estudiar la sensación y percepción visual humana. Para un análisis detallado de estas relaciones entre el arte moderno y la psicología experimental, ver Paul Vitz y Arnold Glimcher, *Modern art and Modern Science: The Parallel Analysis of Vision*, 1983.)

Cuando se trata de la IA artística, no debería enceguecernos el modo en que estos sistemas han sido entrenados. Sí, las redes neuronales artificiales son entrenadas en base a obras de arte humano y artefactos culturales creados con anterioridad. Sin embargo sus producciones recientes no son réplicas o simulaciones mecánicas de lo que ya fue creado. En mi opinión, se trata a menudo de artefactos culturales genuinamente nuevos con contenidos, estéticas, o estilos no vistos anteriormente.

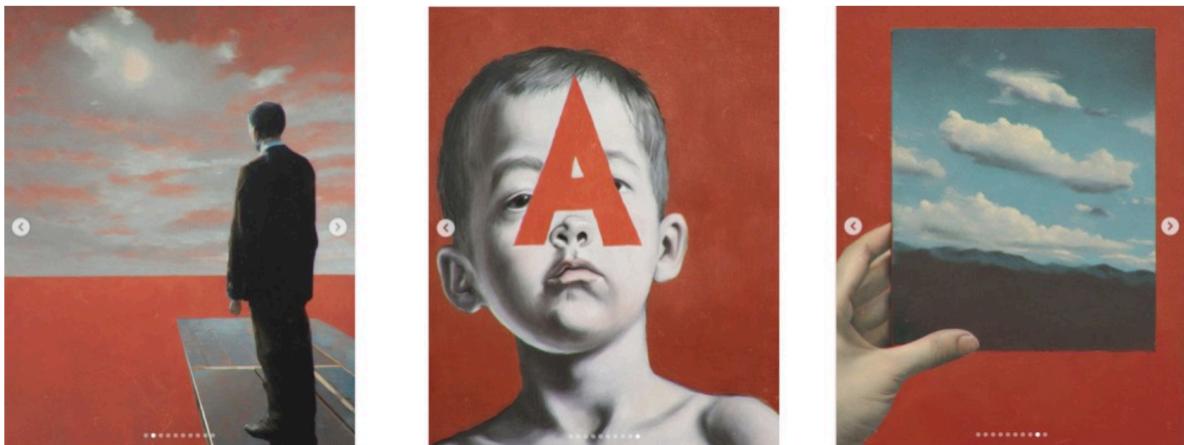
Por supuesto, el simple hecho de ser novedoso no alcanza para que un artefacto automáticamente se vuelva cultural o socialmente interesante o significativo. De hecho, muchas definiciones de "creatividad" concuerdan en este punto, que se trata de la creación de algo que es a la vez original y que vale la pena o es útil.

Sin embargo, en este momento estimar cuál porcentaje de todos los artefactos novedosos producidos por la IA generativa son a su vez útiles o significativos para la cultura más general no es un proyecto factible. Por un lado, no estoy al tanto de ningún esfuerzo sistemático de usar estos sistemas para "completar" una matriz gigante de todo el contenido y posibilidades estéticas con sus correspondientes millones de prompts, elaborados específicamente. En cambio, es probable que, como sucede en las otras formas de la cultura popular, sólo un número reducido de posibilidades sean las que se repiten una y otra vez por millones de usuarios, dejando una larga cola de otras posibilidades sin realizar. Entonces, si lo que se está realizando en la práctica es sólo una diminuta fracción del enorme universo de artefactos potenciales de la IA, no podemos hacer grandes afirmaciones sobre la originalidad o utilidad del resto del universo.

3. Medios Generativos y Arte de base de datos

Algunos artistas de IA como Anna Ridler, Sarah Meyohas y Refik Anadol utilizan en sus obras redes entrenadas con bases de datos específicas. Muchos otros artistas, diseñadores, arquitectos y tecnólogos usan redes reunidas por compañías o institutos de investigación que ya fueron entrenadas con conjuntos de datos muy vastos (e.g, Stable Diffusion), para luego hacerles ajustes y ponerlas a punto con sus propios datos.

Por ejemplo, el artista Lev Pereulkov ajustó el modelo 2.1 de Stable Diffusion usando 40 cuadros de los famosos artistas "no-conformistas" que trabajaban en la Unión Soviética a partir de la década de 1960 (Erik Bulatov, Ilya Kabakov, etc). La serie de imágenes que Pereulkov *Artificial Experiments 1–10* (2023) creó con su red a medida es una pieza de arte original que captura las características del arte de estos artistas, así como también su excepcional semántica surrealista y ridícula sin repetir de cerca ninguna de las obras existentes. Sus "ADN" capturados por la red, en cambio, permiten nuevos significados y conceptos visuales.



Lev Pereulkov, *Artificial Experiments 1–10*, 2022. Tres imágenes de la serie de 10 compartidas en Instagram.

La mayoría de las millones de personas comunes y de profesionales creativos que usan herramientas de medios generativos las usan de modo estándar, como vienen, sin ulteriores ajustes. Esto podría cambiar en el futuro cuando las redes que usan nuestros propios datos se vuelvan más fáciles de usar. Pero independientemente de todas estas cuestiones específicas, todos los artefactos culturales de reciente creación producidos por redes entrenadas tienen una lógica en común.

A diferencia de los dibujos, esculturas y pinturas tradicionales los artefactos de medios generativos no son creados desde cero. Tampoco son el resultado de la captura de algún fenómeno de tipo sensorial, como fotos, videos, o grabaciones de sonidos. En cambio, están construidas en base a un enorme archivo de otros artefactos de medios. Este mecanismo generativo vincula a los medios generativos con géneros y procesos artísticos anteriores.

Podemos compararlo con la edición de cine, que apareció alrededor de 1898, o incluso aún antes, con la fotografía compuesta, que era popular en el siglo 19. También podemos considerar obras específicas que son especialmente relevantes, como el film collage experimental *A Movie* (Bruce Conner, 1958) o muchas de las instalaciones de Nam June Park en las que aparecen fragmentos editados de secuencias de TV.

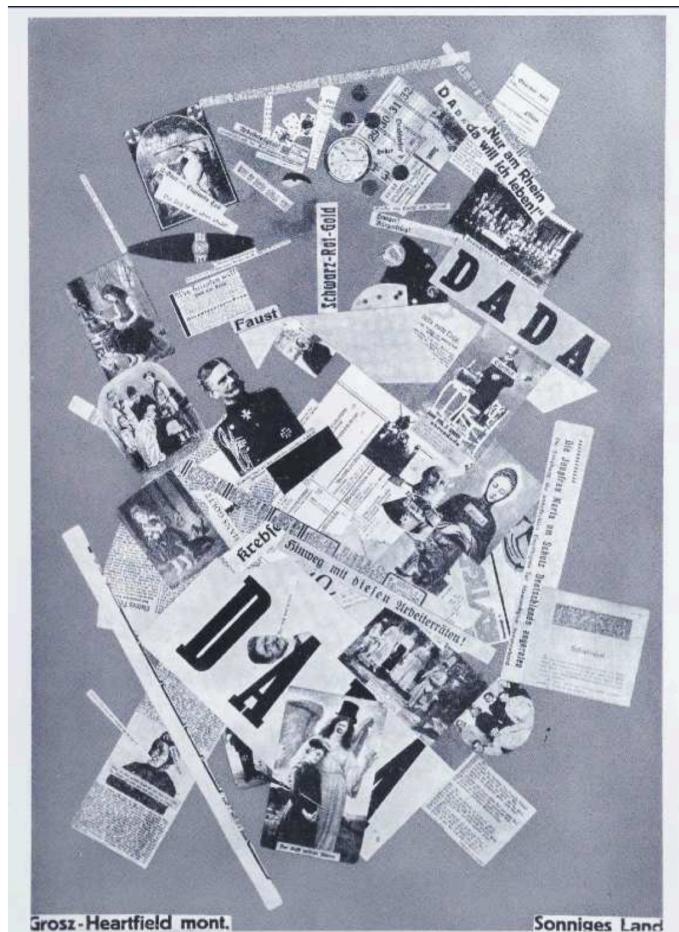
Viendo proyectos como *Unsupervised* o *Artificial Experiments 1-10* en el contexto de este método de creación de medios y sus variaciones históricas nos ayudará a entender estas y muchas otras piezas de arte de IA como objetos artísticos que dialogan con el arte del pasado, más que como puras novedades tecnológicas o piezas de entretenimiento.

Veo muchos momentos y períodos relevantes en los que puedo escudriñar la historia del arte, la cultura visual y de medios en busca de otros usos prominentes de este procedimiento. Son relevantes para los medios generativos actuales no sólo porque los artistas en estos períodos usaban el procedimiento, sino también porque el motivo del uso en todos los casos era consistente. *Una nueva acumulación y accesibilidad de masas de artefactos culturales llevó a los artistas a crear nuevas formas de arte impulsadas por esta acumulación.* Voy a describir algunos de estos ejemplos.

Artistas de la red y digitales crearon hacia finales de los 90s y principios del 2000 numerosas obras en respuesta al universo de la web, nuevo y en rápida expansión. La pieza *_readme* (1998) de Health Bunting, por ejemplo, es una página web que contiene el texto de un artículo sobre el artista, con cada palabra enlazada a una dominio web existente que se corresponde con dicha palabra. *Shredder 1.0* de Mark Napier (también de 1998) presenta un montaje dinámico de elementos que comprende numerosos sitios web - imágenes, textos, código HTML, y enlaces.

Volviendo más atrás en el tiempo, encontramos un amplio paradigma cultural que también era una reacción a la acumulación de artefactos de arte y de cultura históricos en colecciones de medios de fácil accesibilidad. Este paradigma es conocido como "post-modernismo". Los artistas y diseñadores post-modernos a menudo usaban bricolaje y creaban obras que consistían de citas y referencias al arte del pasado, rechazando así el foco del modernismo en la novedad y rompiendo con el pasado.

Si bien existen muchas explicaciones posibles para la emergencia del paradigma postmoderno en los 1960s y 1980s, una es particularmente relevante para nuestra discusión. La acumulación de artefactos de arte y medios anteriores en colecciones estructuradas y accesibles como bibliotecas de diapositivas, archivos de películas, libros de texto de historia del arte con muchas fotos de las obras y otros formatos - en los que se reunían diferentes períodos históricos, movimientos, y creadores - inspiró a los artistas a empezar a crear bricolajes a partir de estas referencias y a citarlos profusamente.



Fotomontaje por John Heartfield, 1919.

¿Y el "modernismo" de las décadas de 1910 y 1920? Mientras que su énfasis generalmente estaba en la originalidad y novedad, uno de los procedimientos que desarrolló en busca de la novedad fue el uso de citas directas tomadas del vasto universo de los medios visuales contemporáneos, que en ese momento se encontraban en rápida expansión. Los grandes titulares, por ejemplo, y la inclusión de fotos y mapas lograban que los periódicos tuvieran mayor impacto visual; aparecían en ese entonces revistas con una nueva orientación visual, como *Vogue* y *Time*, lanzadas en 1913 y 1923, respectivamente; y por supuesto, el nuevo medio del cine seguía su desarrollo.

En respuesta a la intensificación visual de la cultura de masas, a principios de la década de 1910 Georges Braque y Pablo Picasso empezaron a incorporar fragmentos de periódicos, posters, papel de empapelar y telas en sus cuadros. Unos años después, John Heartfield, George Grosz, Hannah Hoch, Aleksandr Rodchenko, y otro puñado de artistas comenzaron a desarrollar la técnica del foto collage, que devino otro método para crear nuevos artefactos mediáticos a partir de imágenes existentes en los medios.

Obras de arte contemporáneas que emplean redes neuronales entrenadas en bases de datos culturales, como *Unsupervised or Artificial Experiments 1-10*, continúan la larga tradición de crear una nueva forma de arte en base a la acumulación de imágenes y otros medios. De este modo, estas obras de arte le siguen abriendo nuevas posibilidades al arte y a sus técnicas, particularmente aquellas a las que me referí anteriormente como "database art" (ver mi artículo "Database as a Symbolic Form", 1998). La introducción de nuevos métodos en la lectura de bases de datos culturales y la creación de nuevas narrativas a partir de éstas es parte de esta expansión.

Por lo tanto, *Unsupervised* no es una pieza que crea collages en base a imágenes existentes, como hacían los artistas modernistas de los 20s, ni los cita extensamente, como hacían los artistas postmodernos de los 1980s. En cambio, el grupo entrena una red neuronal para extraer patrones de las decenas de miles de obras del MoMA. La red entrenada luego genera nuevas imágenes que comparten los mismos patrones pero no se parecen a ninguno de los cuadros específicos. A lo largo de la animación, viajamos por el espacio de estos patrones, explorando varias regiones del universo del arte contemporáneo. (Para más detalles sobre los métodos de entrenamiento de la red GAN usados por Refik Anadol Studio, ver "Creating Art with Generative Adversarial Network: Refik Anadol's Walt Disney Concert Hall Dreams," 2022).

Los *Artificial Experiments 1–10* de Pereulkov usan una técnica distinta para generar nuevas imágenes usando la base de datos existente de imágenes. El artista eligió solo cuarenta cuadros de artistas que tienen en común algunas características clave. Desarrollaron su arte de oposición durante los últimos años de la sociedad comunista (Unión Soviética, 1960s-1980s). También vivían en la misma cultura visual. En mi recuerdo, esta sociedad estaba dominada por dos colores: el gris (representando la monotonía de la vida urbana) y el rojo de la propaganda.

Además Pereulkov eligió cuadros que compartían otro rasgo: "Por norma elegí pinturas que conceptualmente se vinculan de algún modo con el lienzo o con el espacio delimitado por éste. Obtuve el cuadro *New Accordion* de Kabakov, que presenta aplicaciones de papel sobre el lienzo" (comunicación personal con Pereulkov, 04/16/2023). En la puesta a punto

del modelo de Stable Diffusion, Pereulkov usó también descripciones textuales de cada cuadro, elaboradas a medida. Para enseñarle al modelo los lenguajes visuales específicos de los artistas elegidos, agregó a estas descripciones términos como "trazo grueso," "iluminación roja," "fondo azul," y "círculos chatos".

Claramente cada uno de estos pasos representa una decisión conceptual y estética. En otras palabras, la clave del éxito de *Artificial Experiments 1–10* es la creación de una base de datos de este tipo. Esta obra demuestra cómo la puesta a punto de una red neuronal existente entrenada con miles de millones de pares de imagen y texto (como lo es Stable Diffusion) puede lograr que esta red siga las ideas de un artista; los sesgos y el ruido de una red masiva como esta pueden superarse y minimizarse, y no necesitan dominar nuestra propia imaginación.

4. De la representación a la predicción

Históricamente los seres humanos crearon imágenes de escenas existentes o imaginarias mediante una serie de métodos, desde el dibujo manual a la gráfica por computadora 3D (ver abajo la explicación de los métodos). Junto a la IA generativa de medios, emerge un método fundamentalmente nuevo. Las computadoras usan grandes conjuntos de datos de representaciones existentes en varios medios para predecir nuevas imágenes (fijas y animadas).

Uno puede proponer diferentes caminos históricos conducentes a los medios generativos de hoy en día, o dividir la línea de tiempo histórica en diferentes fases. Una trayectoria posible sería:

1. Creación manual de representaciones (por ejemplo, dibujo con una variedad de instrumentos, grabado, etc). Procesos y partes más mecánicas a veces eran llevadas a cabo por asistentes humanos que típicamente se formaban en los estudios de sus maestros - ya existía cierta delegación de funciones.
2. Creación manual de representaciones con asistencia de dispositivos (por ejemplo, máquinas de perspectiva, cámara lúcida). De la creación manual a la creación manual + dispositivo. En esta fase algunas funciones son delegadas a dispositivos mecánicos y ópticos.
3. Fotografía, rayos X, vídeo, captura volumétrica, sensores remotos, fotogrametría. Del uso manual al registro de información mediante el uso de máquinas. De asistentes humanos a asistentes maquinados.
4. Gráfica por computadora 3D (3D CG). Se define un modelo 3D en una computadora y se usan algoritmos que simulan los efectos de fuentes de luz, sombras, niebla, transparencia, translucidez, texturas naturales, profundidad de campo, efecto esfumados por movimiento, etc. Del registro automático a la simulación.
5. IA Generativa. Mediante el uso de conjuntos de datos de medios para predecir imágenes fijas y en movimiento. De la simulación a la predicción.

“Predicción” es el término que generalmente usan los investigadores de la IA en sus publicaciones para describir los métodos generativos mediáticos visuales. Por lo tanto, mientras que este término puede usarse figurativamente y es evocador, es también lo que

de hecho sucede, científicamente, al usar herramientas generativas de imágenes. Cuando se trabaja con un modelo de IA de texto-a-imagen, la red neuronal intenta predecir las imágenes que mejor se corresponden con el texto ingresado. Desde ya no pretendo sugerir que el uso de términos ya aceptados, como ‘generative media’, sea inapropiado. Pero si queremos entender la diferencia entre los métodos de síntesis de medios visuales propios de la IA y los otros métodos de representación desarrollados en la historia de la humanidad, el uso del concepto de ‘predicción’ y por lo tanto la referencia a los sistemas de IA como ‘medios predictivos’ refleja bien esta diferencia.

5. Traducciones de Medios

Existen varios métodos para crear “medios generativos”. Uno de ellos transforma una entrada de medios humanos mientras que retiene el mismo tipo de medio. Un texto ingresado por un usuarix, por ejemplo, puede ser resumido, reescrito, ampliado, etc. La salida, al igual que la entrada, es un texto. Alternativamente, en el método de generación imagen-a-imagen, una o más imágenes de entrada son usadas para generar imágenes nuevas.

Sin embargo existe otro camino que es igualmente intrigante desde una perspectiva histórica y teórica. Los “medios generativos” pueden crearse mediante la “traducción” automática de contenidos entre distintos tipos de medios. Esto es lo que sucede, por ejemplo, al usar Midjourney, Stable Diffusion u otro servicio de generación de imágenes de IA, ingresar un mensaje o un prompt, y la IA genera una o más imágenes en respuesta. El texto es ‘traducido’ a imagen.

Como no se trata de una traducción literal uno-a-uno, he usado la palabra “traducción” entre comillas. En cambio, la entrada de un medio instruye a una red neuronal a predecir una salida apropiada en otro medio. Se puede decir, también, que dicha entrada está siendo ‘mapeada’ en otra salida en otros medios. El texto es mapeado en nuevos estilos de texto, imágenes, animación, vídeo, modelos 3D y música. El vídeo se convierte en modelos 3D o animación. Las imágenes son “traducidas” a texto, y así sucesivamente. El método de traducción de texto-a-imagen es el que actualmente se encuentra más avanzado, pero varias otras formas finalmente lo alcanzarán. La traducción (o mapeo) entre un medio y otro no es un concepto nuevo. Dichas traducciones se han realizado a lo largo de la historia manualmente, a menudo con intención artística. Novelas que han sido adaptadas como obras de teatro y películas, cómics que fueron adaptados para series de televisión, textos de ficción o no-ficción ilustrados con imágenes, etc. Cada una de estas traducciones consistió en un acto cultural deliberado demandando habilidades profesionales y conocimiento de los medios apropiados. Gracias a las redes neuronales artificiales, algunas de estas traducciones hoy pueden realizarse en una escala masiva de modo automático, convirtiéndose en un nuevo medio de comunicación y de creación cultural. Por supuesto, la adaptación de una novela para una película por parte de un equipo humano no es lo mismo que la generación automática de visuales a partir del texto de una novela por parte de una red, pero para muchos casos simples, la traducción automática de medios puede funcionar bien. Lo que antaño era un acto artístico calificado devino ahora una capacidad tecnológica disponible para todxs. Nos puede entristecer todo lo que se pierde como resultado de la automatización – y democratización – de esta operación cultural crítica: habilidades, algo que uno podría llamar ‘originalidad artística profunda’ o ‘creatividad profunda’, y así

sucesivamente. Pero estas pérdidas podrían ser solo temporarias, si las habilidades de la “IA cultural” siguen mejorando para generar contenido más original y para mejor entender el contexto.

Como la mayoría de la gente de la sociedad actual puede leer y escribir en al menos un idioma, los métodos de traducción de texto-a-otro medio son hoy en día los más populares. Estos incluyen los modelos texto-a-imagen, texto-a-animación, texto-a-3D, y texto a música. Estas herramientas de IA pueden ser usadas por cualquiera que pueda escribir o mediante el uso de software de traducción accesible para crear un mensaje en cualquiera de los lenguajes que estas herramientas mejor comprenden en un momento determinado. Otros mapeos de medios, sin embargo, pueden ser igualmente interesantes para los creadores profesionales de medios. A lo largo de la historia cultural humana, varias traducciones entre tipos de medios han recibido atención. Estas incluyen las traducciones entre video y música hecha por VJs en clubes, largas narrativas literarias convertidas en películas y series de televisión, textos ilustrados con imágenes en medios varios, como grabados, números vueltos imágenes (arte digital), textos describiendo pinturas (tradición ekphrasis, que data de la Grecia antigua), mapeos entre sonidos y colores (especialmente popular en el arte moderno), etc.

El desarrollo continuo de modelos de IA para mapeos entre medios de todo tipo, sin privilegiar el texto, tiene el potencial de ser extremadamente fructífero, y espero que proliferen las herramientas que puedan lograr esto. Dichas herramientas serán muy útiles, tanto para artistas profesionales como para otros creadores. Siendo yo mismo un artista, no estoy sosteniendo que la “IA cultural” del futuro será capaz de igualar, por ejemplo, las interpretaciones innovadoras de Hamlet de directores de vanguardia como Peter Brook o las asombrosas películas abstractas de Oscar Fishinger que exploraban las correspondencias entre música e imagen. Alcanza con que las nuevas herramientas de mapeo de medios de la IA estimulen nuestra imaginación, nos provean de nuevas ideas, y nos permitan explorar variaciones numerosas de diseños específicos.

6. Lo estereotípico y lo único

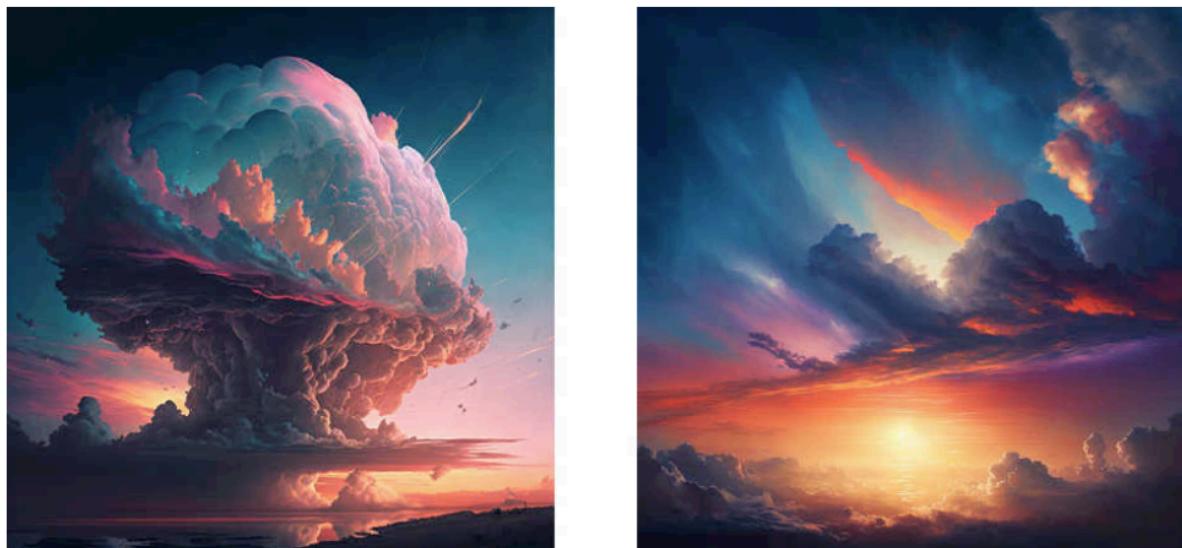
El proceso de creación humana moderno y el proceso de medios generativo de la IA predictiva parecerían funcionar de modo similar. Las redes neuronales son entrenadas usando colecciones no estructuradas de contenidos culturales, como miles de millones de imágenes y sus descripciones o billones de páginas web y de libros. Una red neuronal aprende las asociaciones entre las partes constitutivas de estos artefactos (como cuáles son las palabras que suelen aparecer juntas) así como los patrones y estructuras que tienen en común. Una red entrenada luego usa estas estructuras, patrones y ‘átomos culturales’ para crear nuevos artefactos cuando se lo pedimos. Dependiendo de lo que le pidamos, estos artefactos creados por la IA podrán asimilarse a lo que ya existe, o no.

De modo similar, nuestra vida es un proceso de entrenamiento cultural continuo, supervisado y no supervisado. Tomamos clases de arte y de historia del arte, visitamos sitios web, videos, revistas, y catálogos de muestras, visitamos museos, y viajamos para absorber nueva información cultural. Cuando nos ‘comandamos’² a realizar nuevos

² en el inglés original el verbo es “prompt” que es el mismo término que se sustantiva y se usa para referirse a la indicación mediante un mensaje de texto que se le da, en línea de comando, a la IA para que genere sus imágenes / productos

artefactos culturales, nuestras propias redes neuronales biológicas (hasta la fecha, infinitamente más complejas que cualquier red de IA) generan dichos artefactos en base a lo que hasta ese punto llevamos aprendido: patrones generales que observamos, plantillas para realizar cosas particulares, y a menudo partes concretas de artefactos ya existentes. En otras palabras, nuestras creaciones podrían contener tanto réplicas exactas de artefactos previamente observados y cosas nuevas que representamos usando plantillas que hemos aprendido, como la proporción áurea o el uso de colores complementarios.

Las redes neuronales de la IA que se suelen usar en la generación de imágenes a menudo tienen, por defecto, un estilo “de la casa”. (Este es el término que efectivamente usan los desarrolladores de MidJourney). Cuando uno no especifica explícitamente ningún estilo, la IA generará la imagen usando esta estética ‘por defecto’.



Ejemplos generados por Midjourney version 4 usando el prompt de texto “cielo matutino.”

Para evitar los productos por defecto, es necesario agregar términos a los prompts o indicaciones brindadas por línea de comando a la IA, que especifiquen una descripción del medio, el tipo de iluminación, los colores y el sombreado, o una frase como “en el estilo de” seguido por el nombre de un/una artista, ilustrador/a, fotógrafo/a, diseñador/a de moda o arquitecto bien conocido/a. A continuación, dos ejemplos de estos *prompts* que probé, y las imágenes generadas por Midjourney en respuesta a estos prompts. Los términos usados para definir las características de un estilo particular están resaltadas en verde.



(Imagen generada con Midjourney v3)

Prompt 1:

"aeropuerto gigante moderno de 1965 del futuro en Siberia fabricado de agua y hielo, pintado sobre un panel grande de madera por Hieronymus Bosch, colores pastel brillante con resaltados blancos, lente 23f, mucho detalle --ar 4:3 --s 1250 —test"



(Imagen generada con Midjourney v4)

Prompt 2:

“Foto de dos alumnxs rusxs de secundaria, piel clara iluminación de estudio muy suave, lente de 50mm, monocromo, tonos plateados, alta calidad, ultra realista --v 4 --q 2”

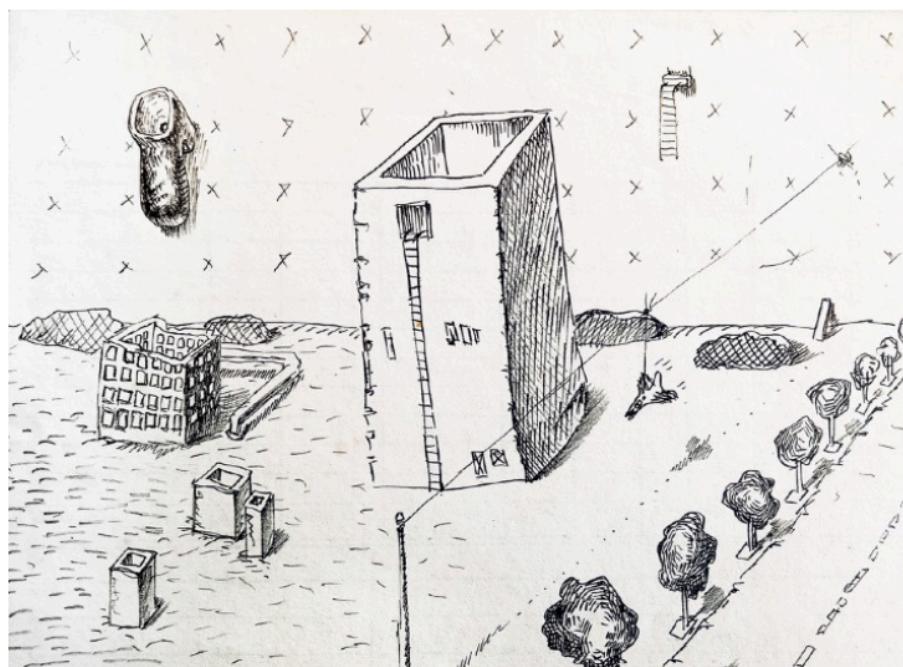
Esta imagen también ilustra el punto que sostengo más adelante en el capítulo: “Los nuevos artefactos mediáticos generados por la IA son frecuentemente más estereotípicos o idealizados de lo que queríamos.” Al poder *simular muchas miles de estéticas y estilos ya existentes e interpolar entre ellos para crear nuevos híbridos*, la IA es más capaz, en este sentido, que cualquier creador humano individual. Sin embargo, actualmente también los y las creadores humanos con talento y mucha experiencia tienen una ventaja significativa. Si bien ambos humanos e inteligencia artificial son igualmente capaces de imaginar y representar objetos y escenas inexistentes y existentes, a diferencia de los generadores de imágenes de la IA, las imágenes creadas por los humanos pueden incluir contenido muy particular, detalles minúsculos singulares, y estéticas distintivas de un modo que supera las capacidades de la IA. En otras palabras, hoy en día un amplio grupo de ilustradores, fotógrafos, y diseñadores puede representar todo lo que puede hacer una red neuronal entrenada (aunque les insuma mucho, mucho más tiempo) pero además pueden también

crear objetos, composiciones o estéticas que la red neuronal en este momento no puede crear. De modo igualmente importante, pueden representar objetos, rostros, composiciones, etc únicas – en oposición a las versiones generalmente más estandarizadas o a menudo más comunes o idealizadas generadas por la IA.

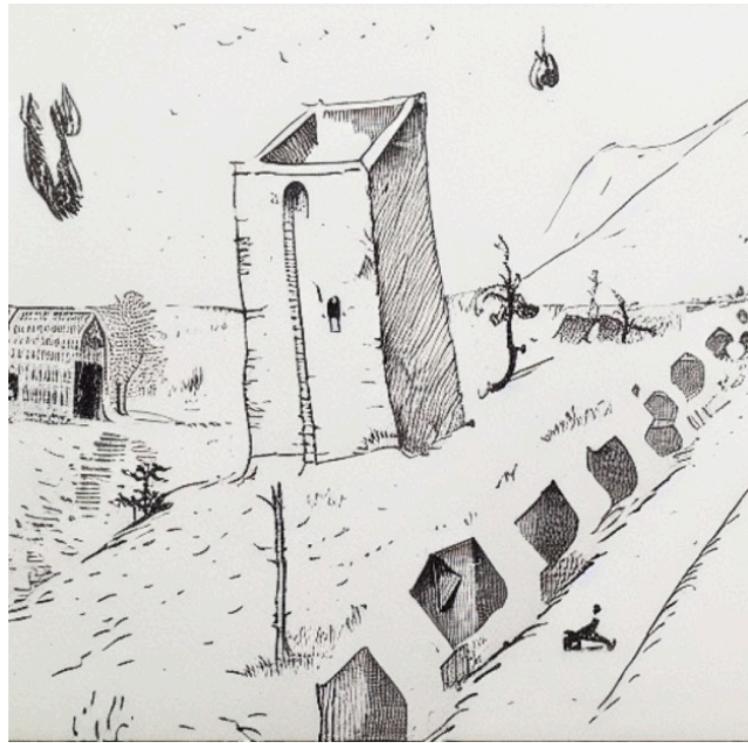
¿Cuál es la causa de esta brecha en los contenidos y en la estética entre creadores humanos y artificiales? Durante el proceso de entrenamiento de una red neuronal artificial, los ‘átomos culturales’, estructuras, y patrones que ocurren con mayor frecuencia en el conjunto de datos de entrenamiento son aprendidos con mucho éxito. Por otro lado, los ‘átomos’ y estructuras que son raros en los datos de entrenamiento, o que sólo aparecen una vez, apenas son aprendidos o directamente, no son ni siquiera analizados. Estos no entran dentro del universo cultural artificial aprendido por la IA. Por lo tanto, cuando le pedimos a la IA que los sintetice, es incapaz de hacerlo.

A raíz de esto, las IA de texto-a-imagen como Midjourney, Stable Diffusion o RunwayML no pueden generar dibujos en mi estilo, ampliar mis dibujos agregando partes generadas a nuevo o reemplazar porciones específicas de mi dibujo con contenido nuevo dibujado en mi estilo (eso es, no pueden realizar “outpainting” or “inpainting” útil en las fotos digitales de mis dibujos). Estas herramientas de IA, en cambio, generan objetos más genéricos de lo que suelo dibujar habitualmente o producen algo que es solamente ambiguo y sin embargo poco interesante.

De ningún modo pretendo afirmar que el estilo y el mundo mostrados en mis dibujos sean completamente únicos. Ellos son también el resultado de los encuentros culturales específicos que tuve, de cosas que observé, de cosas que me fui dando cuenta. Pero como son inusuales (y por lo tanto impredecibles), la IA tiene dificultades para simularlas, al menos sin un entrenamiento adicional usando mis dibujos.



Lev Manovich, sin título dibujo, lapicera sobre papel, 1981-1982



Uno de mis intentos de generar una nueva versión de esta imagen con Stable Diffusion, Otoño 2022.

Aquí nos topamos con lo que veo como *el principal obstáculo que enfrentan los creadores cuando usan la IA generativa de medios*:

Los nuevos artefactos mediáticos generados por la IA son frecuentemente más estereotípicos o idealizados de lo que querríamos.

Esto puede afectar a cualquiera de las dimensiones de la imagen – los elementos del contenido, la iluminación, el sombreado, la atmósfera, la estructura espacial, y detalles de las formas 3D, entre otras. A veces esto es aparente inmediatamente, en cuyo caso uno puede intentar corregirlo, o ignorar los resultados. Muchas veces, sin embargo, estas ‘sustituciones’ son tan sutiles que no llegamos a detectarlas sin observación minuciosa, o en algunos casos, el uso de una computadora para analizar cuantitativamente numerosas imágenes. En otras palabras, los nuevos modelos de la IA generativa de medios, de modo similar a la disciplina estadística desde sus comienzos en el siglo 18 y al campo de la ciencia de datos desde finales de los 2010s, son buenos para procesar ítems y patrones que ocurren con frecuencia en los conjuntos de datos pero no saben qué hacer con los poco frecuentes y poco comunes. Cabe esperar que los investigadores de la IA puedan resolver el problema en un futuro, pero al tratarse de una cuestión tan fundamental, no deberíamos anticipar una solución en lo inmediato.

7. Tema y estilo

En el arte, la relación entre forma y contenido ha sido largamente discutida y teorizada. Esta breve sección no pretende ahondar en todos estos debates, ni plantear discusiones con las teorías más relevantes. Me gustaría, en cambio, considerar cómo estos conceptos se interpretan en la “cultura generativa” de la IA. Pero en vez de usar ‘forma’ y ‘contenido’

usaré otro par de términos que son más comunes en las publicaciones sobre IA y en las conversaciones en línea: *tema* y *estilo*.

A primera vista, las herramientas de IA parecen capaces de distinguir entre el tema y el estilo de cualquier representación dada. En los modelos de texto-a-imagen, por ejemplo, podés generar innumerables imágenes referidas al mismo tema. Agregar los nombres de artistas, medios, materiales, y períodos de la historia del arte específicos es todo lo que hace falta para que el mismo tema sea representado de modo distinto y en correspondencia con la indicación dada. Los filtros de Photoshop empezaron a separar el tema del estilo ya en los 1990s, pero las herramientas de la IA generativa para medios son más potentes. Por ejemplo, al especificar “pintura al óleo” en un prompt, las pinceladas simuladas variarán en tamaño y dirección en la imagen generada según los objetos representados. Las herramientas de IA para medios parecen “entender” la semántica de la representación, en oposición a los anteriores filtros, que simplemente ejecutaban la misma transformación en cada región de la imagen, independientemente del contenido. Por ejemplo, cuando introduce “un cuadro por Malevich” y “un cuadro por Bosch” en la misma línea de comando, Midjourney generó una imagen del espacio que contenía formas abstractas a la Malevich, así como muchas figuras pequeñas humanas y animales, en escala para lograr el efecto perspectiva, como en los famosos cuadros de Bosch.



Imagen generada en Midjourney usando el prompt “pintura por Malevich y Bosch,” Otoño 2022.

Las herramientas de IA agregan automáticamente a la imagen contenido no especificado en el prompt de texto, además de representar lo solicitado. Esto ocurre con frecuencia cuando el prompt incluye una frase como “en el estilo de” o “por”, seguida del nombre de algún artista visual o fotógrafo. En un experimento, usé el mismo prompt con la herramienta de IA Midjourney 148 veces, sumando cada vez el nombre de un fotógrafo distinto. El tema en el prompt se mantuvo básicamente igual – un paisaje vacío con algunos edificios, un camino y postes eléctricos llegando hasta el horizonte. En algunos casos, el agregado del nombre del fotógrafo no tuvo ningún efecto sobre los elementos – como el contraste, la perspectiva, y la atmósfera – de la imagen generada, que se adecuaba a nuestro concepto intuitivo de estilo. Pero cada tanto, Midjourney también modificaba el contenido de la imagen. Por ejemplo, cuando obras conocidas de un fotógrafo particular presentaban figuras humanas en poses específicas, la herramienta agregaba, ocasionalmente, dichas figuras a mis fotos. (Al igual que con Malevich y Bosch, las fotos eran transformadas para encajar en la composición espacial del paisaje, en vez de duplicarse mecánicamente.) Otras veces Midjourney también cambió el contenido de mi imagen para corresponderse con el período histórico en el que un fotógrafo conocido sacó sus fotos más conocidas.

Según mis observaciones, cuando le pedimos a Midjourney o a una herramienta similar que cree una imagen en el estilo de un artista específico, y el tema que describimos en el prompt coincide con los temas típicos del artista, los resultados pueden ser muy exitosos. Sin embargo, cuando el tema del prompt dista mucho del de las imágenes típicas del artista, la representación del tema en este estilo a menudo falla.

El uso del prompt “por Caspar David Friedrich --v 5” en Midjourney genera imágenes que capturan el estilo del artista suficientemente bien. Fuente: <https://www.midlibrary.io/styles/caspar-david-friedrich>.



El uso del prompt “peonias en descomposición por Caspar David Friedrich” en Midjourney genera imágenes que simulan características importantes del estilo del artista, como las combinaciones de colores fríos con la atmósfera dramática. Pero en otros sentidos, las imágenes generadas se alejan significativamente del estilo del artista. El tipo de líneas, la representación de los detalles, y las composiciones simétricas de estas imágenes IA no aparecerían jamás en un auténtico cuadro de Friedrich. A menudo, la IA puede insertar también algunos objetos de aspecto genérico, como las formaciones rocosas en la esquina superior derecha de la primera imagen.



A modo de resumen, para simular con éxito un estilo visual determinado usando las actuales herramientas de IA, puede llegar a ser necesario cambiar el contenido que se quiera representar. *No todos los temas pueden representarse en cualquier estilo con éxito y de modo satisfactorio.* Además, la IA a menudo logra aprender algunos rasgos del estilo de un artista, pero otros no.

Estas observaciones, creo, complejizan la oposición binaria entre los conceptos de 'contenido' y 'estilo'. Para algunos artistas, la IA puede extraer al menos algunos aspectos de su estilo a partir de ejemplos de su obra para luego aplicarlos a diferentes tipos de contenido. Pero para otros artistas, aparentemente, su estilo no puede separarse de su contenido. (Esto es apenas una observación inicial a ser posteriormente desarrollada, tras la realización de más experimentos con Midjourney.)

Para mí, este tipo de observaciones y reflexiones se encuentran entre las razones más importantes para usar las tecnologías de nuevos medios como la IA generativa mediática y para aprender cómo funcionan. Por supuesto, como artista practicante y teórico del arte, hace mucho tiempo vengo pensando las relaciones entre tema y estilo (o contenido y forma) pero la posibilidad de llevar a cabo experimentos sistemáticos como el que acabo de describir trae ideas nuevas y nos permite *leer la historia cultural y el arte retrospectivamente y de nuevas maneras.*