PROMPT

"Crea una ilustración de moda dinámica que encarne la esencia del movimiento y el estilo. Utiliza la proporción áurea para componer una pieza visualmente cautivadora que captura el flujo y la energía de las últimas tendencias de alta costura."

"Create a dynamic fashion illustration that embodies the essence of movement and style. Male model, use the golden ratio to compose a visually captivating piece that captures the flow and energy of the latest couture trends

EN MEDIO DE LA RUPTURA: UN ANÁLISIS DE LA INFLUENCIA DE LA INTELIGENCIA ARTIFICIAL EN LA REPRESENTACIÓN EN EL SISTEMA MODA

In the Midst of Disruption: An Analysis of the Influence of Artificial Intelligence on Representation in the Fashion System

Carlos Ramos Velásquez

Escuela de Diseño Gráfico - Universidad Nacional de Colombia
Gestor editorial SENNOVA-CMTC

RESUMEN

DESDE LOS INICIOS DE LA INTELIGENCIA artificial (IA) hasta la actualidad, hemos sido testigos de una fascinante evolución en la complejidad de los modelos y la sofisticación de los algoritmos. Este progreso, impulsado por la constante innovación, encuentra su terreno de experimentación en campos inesperados, siendo la adquisición de DeepMind por Google o la compra del 50% de OpenAI por Microsoft, ejemplos de cómo los viejos jugadores del mundo de la tecnología se interesan por pequeños jugadores que tienen como bandera el desarrollo de la IA. El recorrido a continuación nos lleva a la intersección entre la tecnología y la moda, explorando cómo los modelos de difusión transforman la creación visual, desde la representación hasta la simulación de texturas y poses. Este avance reconfigura no solo el aspecto estilístico, sino que también plantea interrogantes éticos y económicos sobre el impacto en profesionales creativas como los ilustradores y fotógrafos, dando forma a un nuevo y desconocido panorama digital. A medida que las fronteras entre la inteligencia humana y artificial se desdibujan, surge una reflexión sobre el equilibrio entre la innovación tecnológica y el mantenimiento de la integridad artística, así como sobre el futuro del trabajo en un entorno cada vez más digitalizado.

Palabras clave: Stable Diffusion, Modelos de difusión, Modelos generativos, ilustración de moda.

ABSTRACT

From the early days of artificial intelligence (AI) to the present, we have witnessed a fascinating evolution in the complexity of models and sophistication of algorithms. This progress, driven by constant innovation, finds its experimental ground in unexpected fields, with examples such as Google's acquisition of DeepMind or Microsoft's investment in 50% of OpenAI showcasing how major players in the tech world are investing in startups with AI development as their flagship. The journey ahead takes us to the intersection of technology and fashion, exploring how diffusion models transform visual creation, from representation to the simulation of textures and poses. This advancement not only reconfigures stylistic aspects but also raises ethical and economic questions about the impact on creative professionals such as illustrators and photographers, shaping a new and unknown digital landscape. As the boundaries between human and artificial intelligence blur, there emerges a reflection on the balance between technological innovation and the preservation of artistic integrity, as well as the future of work in an increasingly digitized environment.

Keywords: *Stable Diffusion, Diffusion Models, Generative Models, Fashion Illustration.*

INTRODUCCIÓN

Durante los últimos 12 meses, desde el 30 de noviembre de 2022 —cuando la empresa OpenAI lanzó al público su modelo de lenguaje en un chat conocido como el ya mítico ChatGPT—, se ha puesto a la Inteligencia Artificial en el en centro de los temas de dominio público, está presente en las charlas de café y en los titulares de los noticieros. Hemos sido testigos de la creación de un fenómeno disruptivo comparable en potencia a la aparición de internet o de los *smartphones*, pero con un contenido de mística y amenaza existencial similar a los relatos apocalípticos de ciencia ficción. Sin embargo, más allá de la espectacularidad mediática actual, es crucial entender qué tipo de herramienta estamos adoptando, cuáles son las utilidades reales que se pueden obtener del uso de estos modelos, cuáles son los insumos para su entrenamiento, su accesibilidad y sus aplicaciones prácticas.

El presente artículo tiene como objetivo explicar e introducir al lector en el uso de herramientas de generación de imágenes para su aplicación en la moda. Se abordará la estructura de los modelos de difusión de código abierto basados en "texto-imagen" e "imagen-imagen", centrándonos en el caso de uso específico de Stable Difussion XL. Se examinará el rendimiento de este modelo en comparación con sus competidores más cercanos tanto de código abierto como con licencia. También se explorará la generación de textos de instrucción del modelo (*Prompts*), la creación de restricciones a partir de texto (*Negative prompts*), la generación de imágenes utilizando otras imágenes como referencia, y herramientas complementarias para la generación de poses, así como la simulación de drapeados, texturas, acabados, entre otros aspectos. Además, se abordará el uso de herramientas de generación de imágenes como ControlNet, Magic Poser, LoRAs, y el entrenamiento de modelos personalizados.

Finalmente, exploraremos las implicaciones que estos modelos están teniendo en la actualidad, tanto a nivel laboral, ético como existencial. El objetivo es determinar si estas herramientas son verdaderamente disruptivas o si su impacto se limita a una fiebre mediática en busca de capitalizar beneficios corporativos, es decir, otra burbuja tecnológica más.

CONTEXTO

El concepto de Inteligencia Artificial (IA) fue acuñado en 1956 por el matemático y científico de la computación John McCarthy. Este término se utilizó para describir el tipo de computación que buscaba replicar la inteligencia humana en diversas tareas. Es importante destacar que la noción de IA no es reciente, sino que tiene un largo historial de estudios y aplicaciones. Inicialmente, los algoritmos se empleaban para resolver tareas como partidas de ajedrez, cálculos probabilísticos meteorológicos. En sus inicios, las IA se centraban en procesos lógicos e iterativos, aparentemente limitadas a trabajos repetitivos con un bajo nivel de creatividad (*Narrow AI*, IA estrecha)[1]. Sin embargo, en la actualidad, las IA pueden generar imágenes sorprendentemente realistas, imitar estilos artísticos específicos y desafiar las estructuras laborales de los profesionales creativos. Este cambio drástico plantea la necesidad imperante de comprender el origen, el uso y el impacto de esta tecnología.

La digitalización y la producción de imágenes asistidas por computadora se remontan a la década de 1970, pero las primeras imágenes

1. La IA se clasifica en tres categorías principales: *Narrow AI, Broad AI* y *General AI*. Narrow AI, también conocida como IA especializada, se centra en realizar tareas específicas de manera eficiente, como el reconocimiento de voz o la traducción automática. Por otro lado, General AI aborda una gama más amplia de actividades y puede aprender múltiples tareas, siendo capaz de aplicarse en diversos contextos (multimodales). Finalmente, Super AI, o Super Inteligencia Artificial, representa la capacidad de una máquina para realizar cualquier tarea intelectual humana, siendo adaptable y poseyendo un entendimiento similar al humano en una amplia variedad de dominios. Estas clasificaciones reflejan el nivel de especialización y versatilidad de las aplicaciones de IA, se establecen y se pueden revisar a profundidad en el artículo *What are the 3 types of AI? A guide to narrow, general, and super artificial intelligence*, publicado en codebot.com (Escott, 2017)

Figura 1. Imágenes generadas por el modelo DeepDream de Google en 2016. Fuente: Google DeepDream

generadas por una IA sin la guía simultánea humana comenzaron a surgir en 2015. Modelos de empresas como IBM o Google, como Deep Dream, demostraron su capacidad para generar imágenes que, a nivel semiótico, podían entenderse como algo legible. En sus primeras manifestaciones, las imágenes generadas por estos modelos parecían formar un collage turbulento que transitaba por el 'valle inquietante"[2]. En aquel entonces, muchos espectadores anticipaban que estos modelos alcanzarían el nivel de producción de un autor humano, posiblemente a principios de la década de 2030. Sin embargo, la aceleración de modelos estadísticos, el perfeccionamiento de algoritmos en visión por computadora, el reconocimiento facial y de objetos, el código abierto y la explosión de modelos de aprendizaje profundo durante la pandemia de COVID-19 propiciaron el rápido desarrollo de imágenes generadas por IA.

En la actualidad contamos con modelos como MidJourney, DALL-E, Stable Difussion, FireFly, entre otro centenar que a diario reportan avances y mejoras en sus modelos logrando unos resultados que maravillan y preocupan de igual manera. Vamos a abordar en este artículo el modelo Stable Difussion que se ha liberado a la comunidad y es el que cuenta con más recursos *open source* y *papers* publicados.

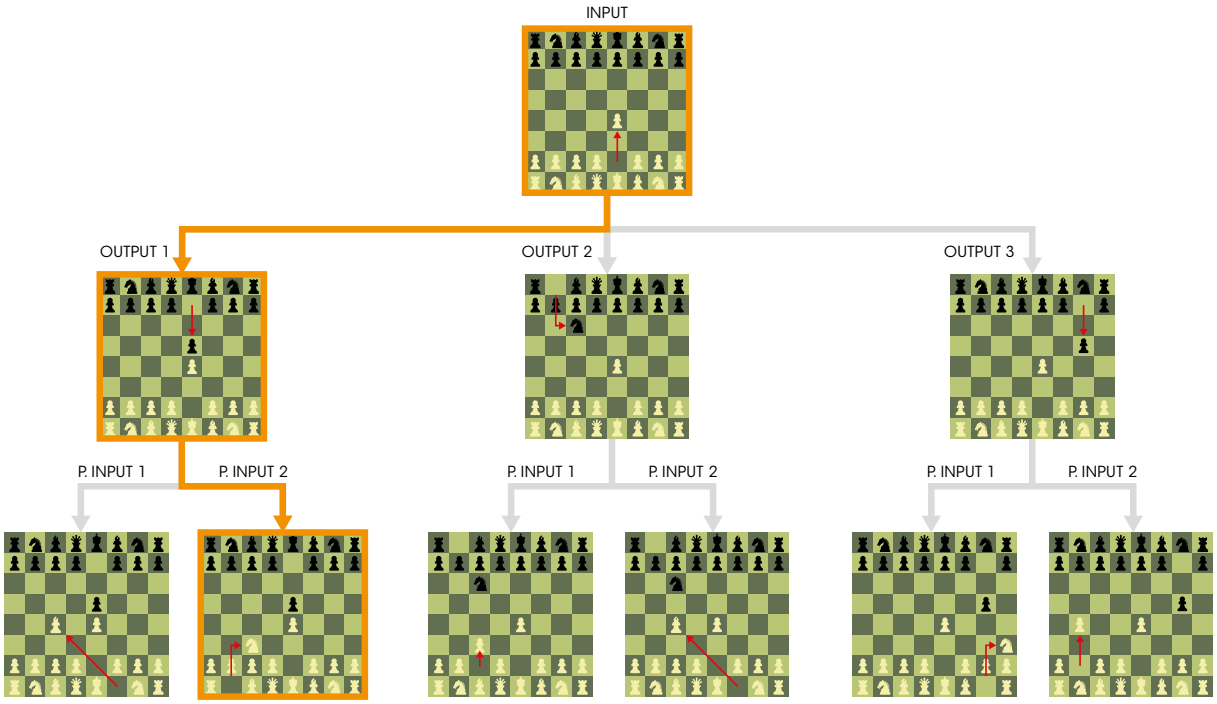
DEFINIENDO LA INTELIGENCIA ARTIFICIAL

Lo que comúnmente denominamos como IA se refiere, en realidad, a modelos de estructuras de datos inherentes al aprendizaje automatizado. Estos modelos se implementaron inicialmente en el desarrollo de juegos de computadora como el ajedrez, el *go* y el *sudoku*[3], luego pasó a desempeñar un papel importante en las primeras cadenas de montaje automatizadas, como las de la industria automotriz a partir de la década de 1960 —*Narrow AI*—. Su programación consistía en estructuras de datos discretas con entradas y salidas clara-

2. El "Valle Inquietante" se refiere a la sensación de incomodidad o extrañeza que experimentamos cuando una representación artificial, como un robot o una imagen generada por computadora, se acerca mucho a parecerse a un ser humano, pero no lo suficiente como para ser completamente convincente. La teoría sugiere que a medida que la semejanza aumenta, la afinidad del observador también lo hace, pero solo hasta cierto punto. Cuando la similitud es demasiado cercana pero aún hay diferencias notables, se produce una disminución repentina de la afinidad, creando una sensación de inquietud. Este fenómeno es conocido como el "Valle Inquietante". El concepto es originalmente producto de los estudios del Psicoanálisis y es acuñado por Freud, quien lo intepretó como "Lo siniestro", pero fue Masahiro Mori quien lo trasladara a la robótica (Mori, 1970).

3. Es notable destacar que los significativos avances en el campo del *machine learning* han ocurrido principalmente en el desarrollo de juegos. Los juegos ofrecen escenarios complejos que involucran diversas variables y probabilidades, generando volúmenes considerables de datos de entrenamiento. Este contexto proporciona la base para la creación de modelos de IA más precisos. Un ejemplo paradigmático de esta conexión es la adquisición de DeepMind en 2022 por parte de Google, una empresa especializada en el desarrollo de IA poderosas que ha demostrado su destreza al vencer a expertos en juegos como ajedrez y el desafiante juego *Go*. Además, es imposible olvidar el histórico enfrentamiento en 1996 entre Gary Kasparov, campeón de ajedrez, y la supercomputadora de IBM, DeepBlue, que marcó la primera derrota de un maestro del ajedrez frente a una IA.

Figura 2. Esquema ilustrativo del funcionamiento de una estructura de datos donde el sistema sale con las figuras blancas generando un estímulo (*input*), al que se le brindan una gama de respuestas (*outputs*). En el entrenamiento, al modelo se le enseñan las posibles mejores jugadas para que tome la ruta a la victoria con mayor probabilidad de éxito (Autoría: Elaboración propia).



mente identificadas y definidas. Por ejemplo, para un movimiento de una pieza en un juego, se almacenaban las variables de salida que dependían de la posición de inicio, las piezas en juego y los posibles movimientos por ficha. De esta manera, el programa exploraba opciones dentro de un conjunto de datos según la entrada del usuario (ver Figura 2). A medida que los modelos se volvieron más complejos, con conjuntos de datos más robustos, surgieron estructuras más dinámicas que permitieron calcular y considerar con mayor precisión las probabilidades de un resultado; con el tiempo, esto se denominó entrenamiento reforzado (*Reinforcement learning*).

La rama de la IA que explora no solo la generación sino el entendimiento y percepción artificial de imágenes se conoce como Visión por Computadora, que abarca la sensórica, la óptica y las arquitecturas de reconocimiento de imágenes y objetos[4]. Las estructuras de datos más recurrentes en los modelos de generación y reconocimiento de imágenes son las redes neuronales convolucionales (Convolutional Networks) y los modelos de difusión (Latent diffusion models LDM). Las redes convolucionales están basadas en el funcionamiento fisiológico de la visión, es decir, cómo el ojo percibe las siluteas, los bordes, límites, iluminación y estímulos en general. El modelo se compone de una serie de capas y "neuronas" que operan diferentes datos entre sí para determinar qué de acuerdo a su dataset[5] está

4. La experimentación con diferentes tejidos de órganos de visión pudo determinar qué tipo y función tenía cada célula y cómo se experimentaba el fenómeno de la visión a nivel fisiológico y cerebral. En el estudio de (Wiesel, 1962), se explora a partir del análisis de tejidos de los órganos de visión de los gatos, midiendo estímulos y respuestas en diferentes niveles, concluyendo que en la visión lo que define el límite es la posibilidad de definir contrastes y de ese modo establecer contornos discernibles.

5. En el ámbito de la minería de datos, el término '*dataset*' se refiere a un conjunto de datos organizado y clasificado que contiene valores estructurados. En otras palabras, se trata de colecciones de datos organizadas según criterios específicos.

percibiendo (Figura 3). Por su parte, los modelos de difusión, se basan en el ruido y la probabilidad para generar imágenes, a partir de extensos datasets. Los modelos de difusión tienen como base el procesamiento natural del lenguaje NPL aplicado a la generación de imágenes en diferentes grados de realismo, y por el análisis de datos entender patrones y generarlos.

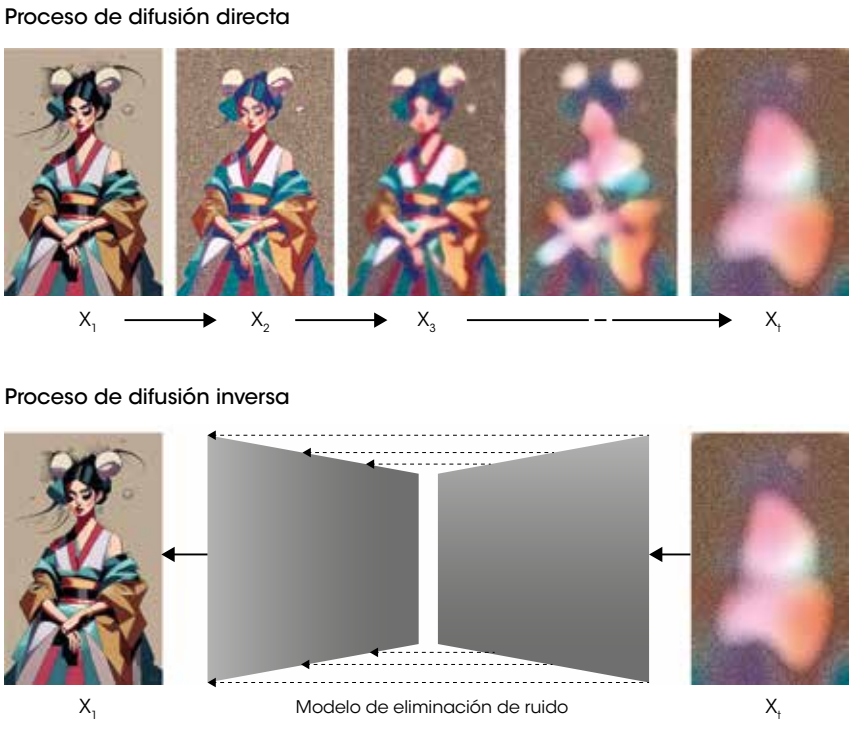
EXPLORANDO STABLE DIFFUSION

Stable Diffusion (en adelante SD) es el producto estrella de la *startup* inglesa Stabiliy.ai, y en pocas palabras es un modelo de difusión que pertenece a la categoría de modelos generativos. Este enfoque específico utiliza técnicas avanzadas para la generación de imágenes, siendo especialmente relevante en aplicaciones relacionadas con las representaciones visuales (dibujo, pintura, fotografía, render, video, etc). Los modelos de difusión, como SD, tienen la capacidad de transformar la representación visual, simulando texturas, poses y otros aspectos estilísticos. Su uso ha redefinido la creación visual en el contexto del diseño y la publicidad, abriendo nuevas posibilidades creativas y planteando cuestionamientos éticos y económicos en la industria.

SD es un ejemplo de modelo LDM, que durante su entrenamiento, se diseñan para aprender a eliminar capas sucesivas de ruido gaussiano[6] aplicado a las imágenes del dataset fuente.

6. En el contexto de procesamiento de imágenes, añadir ruido gaussiano implica introducir pequeñas variaciones aleatorias a los valores de los píxeles de una imagen. Estas variaciones siguen la distribución gaussiana, lo que significa que la mayoría de las variaciones son pequeñas y el impacto general es similar al "ruido" que se encuentra en muchos fenómenos naturales.

Figura 3. Los difusores son modelos probabilísticos en donde a partir de procesos secuenciales, va generando distorsiones en los datos. En su inicio, se añade distorsión de datos en forma de difusión directa, luego se puede a partir de datos distorsionados se entrenan en forma inversa para construir imágenes, o difusión inversa. Modelo: Ideogram v.1 (Autoría: Elaboración propia).



Es un modelo que tiene dos componentes que son los que elaboran el proceso de generación, el primero se conoce como *Text Encoder* o "codificador de texto" que forma parte de un modelo más grande denominado CLIP. Este codificador transforma el texto de entrada en una representación numérica, asignando un vector a cada palabra o token en el texto. La salida es una lista de estos vectores, donde cada vector representa la información semántica de la palabra o token correspondiente en el texto original. El otro componente del modelo se conoce como Image creator, que a su vez se compone de dos módulos: el creador de información de la imagen o *Image information creator*, componente que trabaja en lo que se llama el "espacio de información de imágenes" o "espacio latente" que es iterativo en distintos "pasos" que propician la difusión[8]. Este espacio es una representación abstracta donde la información se organiza de manera que sea más manejable para el modelo. Finalmente el *Image Decoder* condensa la información procesada para representar los diferentes tokens en una imagen[9]. La última versión de SD (v.9, Stable Diffusion XL) agrega además un módulo de refinamiento del modelo que permite generar imágenes como mayor precisión.

El modelo Stable Diffusion fue entrenado utilizando diferentes conjuntos de datos, como LAION2B-es y laion-high-resolution[10]. En las rondas finales, se utilizó un subconjunto llamado LAION-Aesthetics v.2 5+, compuesto por 600 millones de imágenes con subtítulos. Este subconjunto se seleccionó basándose en la predicción de LAION-Aesthetics Predictor v.2, que estimaba que los humanos darían una puntuación de al menos 5 sobre 10 a estas imágenes. Para mejorar la calidad, se excluyeron imágenes de baja resolución y aquellas identificadas por LAION-5B-WatermarkDetection con una alta probabilidad de llevar una marca de agua. También se eliminó un 10% del condicionamiento de texto en las rondas finales para mejorar la orientación de difusión sin clasificador. El entrenamiento se realizó con 256 GPU Nvidia A100 en Amazon Web Services durante 150,000 horas, con un costo de 600,000 USD[11].

Finalmente, SD se ha formulado en un principio como una iniciativa *open source*[12], que permite que los usuarios entrenen sus propios modelos (LoRAs), que se agreguen diferentes modelos de refinamiento y que se creen diferentes formas para generar imágenes. Stability.ai ha proporcionado una interfaz de control de SD llamada Dream Studio, que otorga 25 créditos gratuitos (ocasiones de crear imágenes), con las que se pueden crear cerca de 100 imágenes, luego, por 10 dólares se pueden obtener 1000 créditos más que equivalen a cerca de 5000 imágenes. SD también tiene en su sitio web una interfaz sencilla por la cual se pueden generar imágenes y ofrece suscripciones para planes en los que se ofrecen mejoras a las funciones. Si bien se reco-

7. En términos técnicos, este componente utiliza una red neuronal llamada UNet y un algoritmo de programación. La característica destacada es que opera de manera más eficiente que los modelos de difusión anteriores que trabajaban directamente en el espacio de píxeles de una imagen.

8. La "difusión" en este contexto se refiere al procesamiento paso a paso de la información en este componente. Es un proceso gradual que transforma la información en este espacio de representación abstracta, lo cual es crucial para generar una imagen de alta calidad. La salida de este componente luego se utiliza en el siguiente paso del modelo, el decodificador de imágenes, para finalmente producir la imagen final.

9. Para ver de forma más detallada el proceso técnico de generación de imágenes, se recomienda ver la entrada de blog de GitHub *The Illustrated Stable Diffusion* (Alammar, 2022). Recuperado de: https://jalammar.github.io/illustrated-stable-diffusion/)

10. LAION es el conjunto de datos utilizado en el entrenamiento de Stable Difussion, es producto de una organización alemana sin ánimo de lucro del mismo nombre que recibe fondos de Stability.ai. Los modelos LAION están entrenados con imágenes de licencia libre de diferentes repositorios de internet como Wikipedia Commons, Flickr, WordPress, entre otros que reunió un *dataset* de 12 millones de imágenes ((Baio, 2022). Recuperado de: https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/).

11. (Wiggers, 2022). "This startup is setting a DALL-E 2-like AI free, consequences be damned". *TechCrunch*, 2022. Recuperado de: https://techcrunch.com/2022/08/12/a-startup-wants-to-democratize-the-tech-behind-dall-e-2-consequences-be-damned/

12. La filosofía *open source* se refiere a la práctica de compartir el código fuente de un software de manera abierta y gratuita. En el contexto de Stable Diffusion, el término *open source* sugiere que el código y los pesos del modelo están disponibles públicamente para que otras personas lo utilicen, estudien o modifiquen. Esto fomenta la transparencia, la colaboración y la mejora continua, ya que la comunidad puede contribuir al desarrollo del modelo, identificar posibles mejoras o adaptarlo a diferentes aplicaciones.

Figura 4. Notebook de Google Colab, al activar cada casilla, se instala virtualmente en un Drive el modelo ControlNet para operar Stable Difussion.

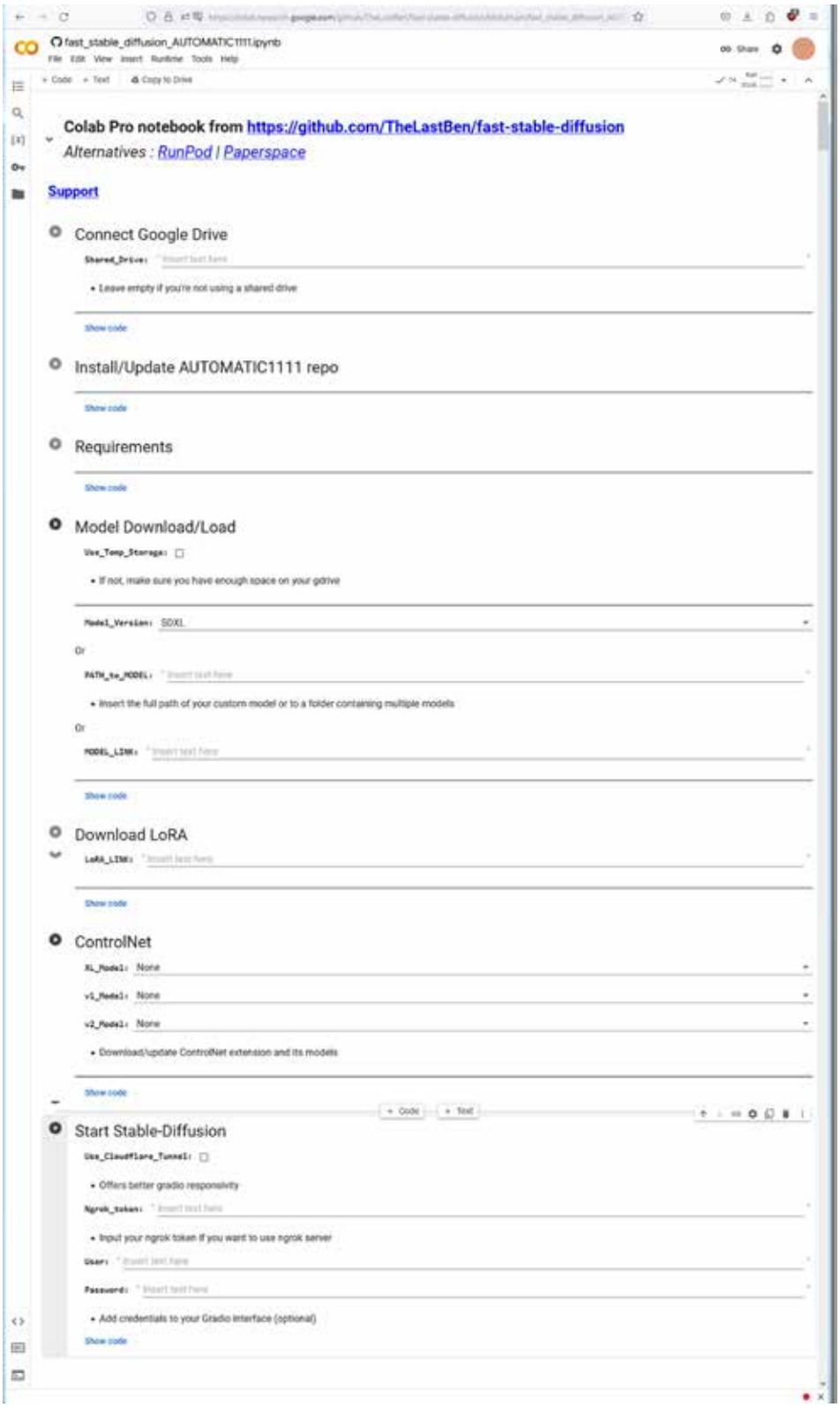
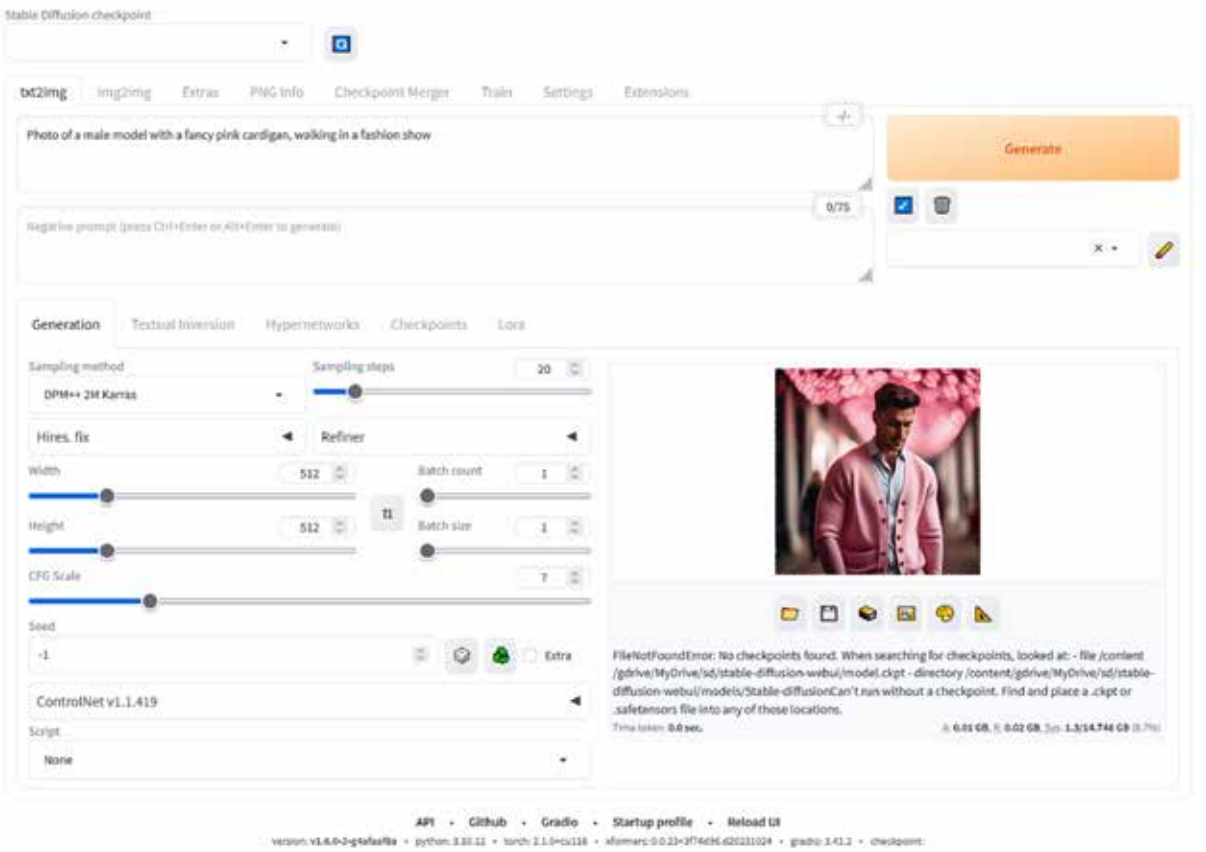


Figura 5. Interfaz de Stable Difussion y ControlNet.



mienda para familiarizarse con la generación de imágenes, otra opción, ésta si de carácter gratuito y que vamos a explorar es ControlNet, interfaz que permite utilizar SD permitiendo controlar aspectos específicos de la imagen como el prompt (instrucción), tamaño de la imagen, resolución, modelo de entrenamiento, entre otros.

UTILIZANDO CONTROLNET

En palabras sencillas, ControlNet es como un tutor especializado para modelos de difusión. Por ejemplo, suponiendo que se tiene un modelo de IA como SD que puede crear imágenes de gran calidad, pero se quiere que también se tomen en cuenta ciertas condiciones adicionales, como el tipo de textura o el estilo artístico deseado; aquí es donde entra ControlNet. Esta "arquitectura de red neuronal" crea una copia segura del modelo original y otra copia que puede aprender esas condiciones especiales. La "convolución cero" es un truco para asegurarse de que el modelo no pierda el control deseado antes de aprender nuevas cosas. Entonces, básicamente, ControlNet asegura que el modelo original, que ya está optimizado en la creación de imágenes, no se vea afectado negativamente cuando se le están enseñando nuevas cosas con conjuntos de datos más pequeños[13]. Esto es sobresaliente porque significa que se pueden entrenar y ajustar los modelos incluso en dispositivos más pequeños, haciéndolos más accesibles y seguros.

La versión de ControlNet más accesible al público y que no necesita de una capacidad de hardware excesiva es la demo pública

13. Para ver una explicación a mayor detalle técnico de ControlNet, se puede visitar el repositorio *Adding Conditional Control to Text-to-Image Diffusion Models* (Lvmin Zhang, 2023). Recuperado de: https://doi.org/10.48550/arXiv.2302.05543

AUTOMAT 1111 (Figura 4) que se encuentra en un notebook de Google Colab[14], que es un repositorio de GitHub, que al instalarse, corre utilizando la nube de Google Drive. Posteriormente y seguidos los pasos que indica el notebook, se generará un link de acceso a la interfaz de creación de imagen (ver Figura 5). En la interfaz se visualizan tanto la opción de creación de imágenes a texto (txt2img), de imagen a imagen (img2img), funciones de inpainting que permiten poner o quitar elementos de una imagen, la salida y el tipo del archivo, un módulo de entrenamiento. En el módulo inferior se ven el tipo de modelo en uso, el número de pasos en los que se genera la imagen, la semilla, controles para el tamaño de la imagen (alto y ancho), el número de imágenes deseado y por último un módulo de ControlNet, donde también se pueden operar las opciones txt2img e img2img.

FUNCIÓN DE GENERACIÓN DE TEXTO A IMAGEN (TXT2IMG)

En un modelo generativo de difusión, el modo "txt2img" se refiere a la transformación de texto (txt) a imágenes (img). En este contexto, "generativo" significa que el modelo puede crear nuevas instancias de datos, y "difusión" hace referencia a la difusión de la información a través del modelo.

En el modo "txt2img", el modelo toma texto como entrada y genera imágenes como salida (*Propmting*). Este tipo de modelo es comúnmente utilizado en tareas de generación de imágenes condicionadas por texto. Por ejemplo, podrías proporcionar una descripción textual, como "un paisaje montañoso al atardecer", y el modelo intentará generar una imagen que coincida con esa descripción.

Los modelos generativos de difusión a menudo utilizan arquitecturas complejas, como GPT (Transformador Generativo Preentrenado)[15], para realizar esta tarea. Estos modelos son entrenados en grandes conjuntos de datos que contienen pares de texto e imágenes para aprender las correspondencias entre descripciones y representaciones visuales.

PROMPT

El término "prompt" se refiere a la entrada dada al modelo para solicitar una respuesta específica. Es esencialmente la instrucción o la entrada de inicio que se presenta al modelo para guiar su generación. En el caso del modo "txt2img", el prompt sería el texto que describe lo que se espera que el modelo genere como imagen.

Por ejemplo, si estás utilizando un modelo generativo de difusión y quieres que genere una imagen de un gato durmiendo en un sofá, tu prompt podría ser la descripción textual: "Un gato descansando plácidamente en un sofá". El modelo utilizaría este prompt para generar una imagen que se ajuste a esa descripción[16].

La creación de instrucciones para los modelos de IA se conoce como *Prompt Enginnering* y básicamente se trata de la organización de texto para que un modelo genere un resultado deseado. Para generar un prompt eficiente se debe tener cuenta una estructura jerarquizada de palabras. En el ejemplo anterior la palabra "Núcleo" o core es la palabra "Gato" (Figura 6). Por lo general la palabra núcleo va en la primera frase del prompt hará parte de la composición central y será lo primero que vaya generando el modelo.

14. https://colab.research.google.com/github/TheLastBen/fast-stable-diffusion/blob/main/fast_stable_diffusion_AUTOMATIC1111.ipynb?authuser=2#scrollTo=PjzwxTkPSPHf

15. En términos generales la sigla GPT corresponde a los Transformadores generativos preentrenados o *Generative Pre-Trained Transformer* o "Generativo significa que puede crear nuevos datos, en este caso texto, a semejanza de sus datos de entrenamiento. Preentrenado significa que el modelo ya ha sido optimizado en función de estos datos, lo que significa que no necesita compararlos con sus datos de entrenamiento originales cada vez que se le solicita. Y Transformador es un tipo poderoso de algoritmo de red neuronal que es especialmente bueno para aprender relaciones entre largas cadenas de datos, por ejemplo oraciones y párrafos" ((Perrigo, 2023). "*The A to Z of Artificial Intelligence*", en TIME. Recuperado de: https://time.com/6271657/a-to-z-of-artificial-intelligence/

16. El texto se apoya en la guía básica para prompt enginering de stability.ai (StabilityAI, 2022), recuperado de: https://beta.dreamstudio.ai/prompt-guide

Figura 6. Versiones de un mismo prompt cambiando de núcleos y añadiendo diferentes complementos en cada iteración (Autoría: Elaboración propia, usando Controlnet y SDXL).

PROMPT 1

Un gato descansando plácidamente en un sofá

NÚCLEO

A cat resting peacefully on a sofa

CORE

PROMPT 2

Una acuarela de un gato descansando plácidamente en un sofá

NÚCLEO

A watercolor of a cat resting peacefully on a sofa

CORE

PROMPT 3

Una acuarela de un gato descansando plácidamente en un sofá, estilo Art Noveau, etiqueta de vino

COMPLEMENTO: Estilo, formato

A watercolor of a cat resting peacefully on a sofa, art nouveau style, wine label

COMPLEMENT: Style, format

PROMPT 4

Una acuarela de un gato descansando plácidamente en un sofá, estilo Art Noveau, etiqueta de vino, letrero "*Le chat noir*"

COMPLEMENTO: Estilo, formato

A watercolor of a cat resting peacefully on a sofa, art nouveau style, wine label, lettering "Le chat noir"

COMPLEMENT: Style, format

Figura 7. Método de precisión por pesos asignados (Autoría: Elaboración propia, usando Controlnet y SDXL).

PROMPT

Una acuarela: 1 de un gato: -1 descansando plácidamente en un sofá, estilo Art Noveau, etiqueta de vino, letrero "*Le chat noir*": 1

*A watercolor: 1 of a cat: -1 resting peacefully on a sofa, art nouveau style, wine label, lettering "*Le chat noir: 1*"*

La palabra core puede variar por ejemplo si se busca hacer énfasis en una técnica, por ejemplo "una pintura en acuarela de un gato descansando plácidamente en un sofá". En este caso será "acuarela" en lo que se centre el modelo y luego en el sujeto "gato". El core del prompt puede complementarse con especificaciones de estilos, técnicas o artistas[17], por ejemplo: "una acuarela de un gato descansando plácidamente en un sofá, estilo art noveau, etiqueta de vino.", en donde "art noveau" y "etiqueta de vino" funcionarían como complementos. Hay modelos como DALLE 3 o Ideogram que han mejorado sus funciones de texto, por lo que se podrán probar combinaciones como "una acuarela de un gato descansando plácidamente en un sofá, estilo art noveau, etiqueta de vino, con el letrero le chat noir".

PROMPTING AVANZADO

SD ha implementado un sistema de identificación de ajustes que valora las prioridades de precisión en cada elemento del *prompt,* dándole un peso específico (*Weigth*) es decir, pondera un resultado deseado con uno (1) y uno no deseado con (-1), teniendo variaciones intermedias entre cero punto cinco (0.5) y (-0.5). Por ejemplo, si usamos el *prompt* de la etiqueta del gato y queremos que se preste más atención en la técnica y el *lettering*, y menos atención en detalles del gato pondremos "una acuarela: 1 de un gato: -1 descansando plácidamente en un sofá, estilo art noveau, etiqueta de vino con el letrero le chat noir: 1" (Figura 7). También existen formas de ajustar la precisión de los elementos que no queremos ver en un resultado con los *negative prompts*.

NEGATIVE PROMPTS

El término *negative prompt* se refiere a una técnica en la que se proporciona al modelo una instrucción que indica lo que no se quiere en la imagen generada. Es una forma de guiar el proceso creativo del modelo al señalar aspectos específicos que deben evitarse. Por ejemplo, si se está utilizando un modelo generativo de imágenes y deseas que genere una imagen de un photoshoot sin un objeto particular, podrías usar un negative prompt como: "Evitar incluir cualquier tipo de persona en fondo de la imagen". El modelo utilizaría esta instrucción para generar una imagen de paisaje sin personas (Figura 8).

El uso de *negative prompts* puede ayudar a refinar y controlar las salidas del modelo al proporcionar indicaciones tanto de lo que se desea como de lo que se debe evitar en la generación de la imagen. Es una técnica que puede ser útil para personalizar y ajustar los resultados del mo-

17. Veremos las implicaciones de uso de diferentes artistas y modelos cuando lleguemos a los asuntos éticos de la IA.

Figura 8. Elaboración de una imagen con el uso de *negative prompts*, en donde se eliminan a partir de instrucciones elementos con el fin de generar una imagen más editorial de una misma instrucción
(Autoría: Elaboración propia, usando Controlnet y SDXL).

PROMPT

Foto de un modelo masculino luciendo un cardigan rosa

Photo of a male model wearing a pink cardigan.

PN: Pose casual, fondo casual, modelo masculino musculado

NP: Casual pose, casual background, muscled male model

PN: Pose casual, fondo casual, modelo masculino musculado, gafas, saturación de color, artefactos jpeg

NP: Casual pose, casual background, muscled male model, glasses, colour saturation, jpeg artifacts

delo según las preferencias específicas del usuario. Existen listados completos de *negative promts*, producto de las diferentes experimentaciones que hacen los usuarios con los modelos.

FUNCIÓN DE GENERACIÓN DE IMAGEN A IMAGEN (IMG2IMG)

La opción "img2img" se refiere a la capacidad del modelo para generar imágenes directamente a partir de otras imágenes, en lugar de depender de texto descriptivo como en el modo "txt2img". En esencia, "img2img" significa "imagen a imagen". En este modo, en lugar de proporcionar una descripción textual como entrada, alimentas al modelo una imagen y esperas que genere una imagen relacionada de alguna manera. Esto puede incluir tareas como la creación de variantes de una imagen dada, la mejora de la calidad de la imagen original o cualquier otra tarea relacionada con la generación de imágenes a partir de imágenes (Figura 9).

El modelo, al aprender de grandes conjuntos de datos de imágenes, intenta entender patrones visuales y características en las imágenes de entrada para generar imágenes coherentes y relevantes como salida. El modo "img2img" permite que el modelo genere imágenes directamente basándose en la información visual proporcionada por otras imágenes, sin depender de descripciones de texto. Vale aclarar que si bien las descripciones de texto pueden no ser necesarias, los modelos suelen incluir un módulo de *prompts* para reforzar el resultado del modelo img2img.

El modelo de SD, tiene la capacidad de generar imágenes a partir de distintas imágenes fuente, mapas de profundidad, bordes de objetos o contornos de un boceto, mapas de segmentación, esqueletos o modelos de pose, etc. Veremos cómo funciona con cada ejemplo (Figura 10).

El resultado del img2img también se puede hacer más preciso, cuando se intervienen los siguientes parámetros:

Figura 9. En la primera versión se pierden detalles de rostro y manos, en la segunda versión mejorada desde el modo img2img de ControlNet. (Autoría: Elaboración propia, usando Ideogram, ControlNet y SDXL).

PROMPT

Crea una ilustración de moda dinámica. Utiliza la composición de proporción áurea, últimas tendencias de alta costura

Create a dynamic fashion illustration. Use the golden ratio composition, latest couture trends

- ***Canny:*** Se refiere al operador de detección de bordes *Canny*. Este operador identifica los bordes en una imagen aplicando un algoritmo de detección de bordes. En el contexto de modelos generativos, puede referirse a la detección o manipulación de bordes en las imágenes generadas.
- ***Seed:*** se refiere a la semilla utilizada en la inicialización de números aleatorios. En modelos generativos, la semilla puede influir en la generación de contenido aleatorio, permitiendo cierto grado de reproducibilidad. Cambiar la semilla puede llevar a diferentes resultados, pero al usar la misma semilla, se puede replicar un resultado específico.
- ***Steps:*** Hace referencia a pasos o iteraciones en el proceso de difusión del modelo. En modelos generativos de difusión, el proceso de difusión implica una serie de pasos donde la información se propaga gradualmente a través de la red generativa. A medida que avanza en estos pasos, la generación de la imagen se desarrolla, captu-

Figura 10. Diferentes orígenes de imagen y sus correspondientes resultados usando img2img de ControlNet. (Autoría: Elaboración propia, usando Ideogram, ControlNet y SDXL).

PROMPT

Foto de una modelo femenina asiática luciendo un vestido de lana

Photo of a asian female model wearing a wool dress


Origen:
Otra imagen o ilustración


Origen:
Mapa de profundidad


Origen:
Mapa de segmentación


Origen:
Contornos - Boceto

rando detalles y complejidades. Si bien a mayor número de pasos el resultado es más detallado, también toma más tiempo en procesamiento.

LoRAs

El control del resultado no sólo puede hacerse más preciso con los procedimientos ya comentados, también se puede hacer desde el entrenamiento. Para modelos como SD existen complementos como modelos más pequeños especializados en el entrenamiento de ciertas características específicas como una técnica, un estilo, un modelo, un acabado en particular, etc.

LoRA (*Low-Rank Adaptation)* es una técnica de entrenamiento para ajustar modelos de SD. Suelen ser de 10 a 100 veces más pequeños que los modelos de puntos de control convencionales, lo que los hace muy atractivos para personas que tienen una extensa colección de modelos[18]. Son los sucesores de Dreambooth[19], red neural que es potente pero usa *datasets* extensos y una gran capacidad de cómputo (2-7 GB). Por otro lado, los LoRas utilizan capacidades más pequeñas (alrededor de 100 KB), pero sus capacidades son limitadas. Son ideales para uso no local, ya que se optimizan para uso en línea. Los LoRAs es una excelente manera de personalizar modelos de arte de IA sin ocupar demasiado espacio de almacenamiento local.

Existen LoRAs específicos para la representación de moda, entrenados con imágenes de drapeados, materiales, poses, etc. El repositorio más popular en cuando uso y variedad es Civitai[20], donse se albergan modelos open source, para uso en SD.

18. Para mejor comprensión y uso de LoRAs para ControlNet y Stable Difusion (Andrew(Username), 2023), puede revisarse: https://stable-diffusion-art.com/lora/

19. Las diferencias técnicas que existen entre el entrenamiento de Dreambooth y LoRAs, puede revisarse en: (Gözükara, 2023), SDXL LoRA vs SDXL DreamBooth Training Comparison Results, en *Medium*, https://medium.com/@furkangozukara/sdxl-lora-vs-sdxl-dreambooth-training-results-comparison-794257a91c60

20. https://civitai.com/

APLICACIONES AL SISTEMA MODA

Los modelos de difusión tienen una utilidad significativa en el mundo de la moda, especialmente en la representación y creación de diseños innovadores. Aquí hay algunos argumentos que respaldan su utilidad:

- **Generación Creativa:** Los modelos de difusión permiten la generación creativa de patrones y diseños únicos. Al aplicar pequeños cambios a modelos de referencia, estos modelos pueden crear variaciones sorprendentes y originales que son fundamentales en la industria de la moda, donde la innovación y la singularidad son altamente valoradas.
- **Exploración de Estilos:** Facilitan la exploración de una amplia gama de estilos y tendencias. Los diseñadores pueden ajustar y adaptar estilos existentes para crear nuevas interpretaciones y fusiones, lo que abre oportunidades para la experimentación y la diversificación en la moda.
- **Optimización del Proceso Creativo:** Al ofrecer un equilibrio entre el tamaño del archivo y la potencia de entrenamiento, modelos como LoRA permiten a los diseñadores optimizar su proceso creativo. Pueden personalizar modelos de manera eficiente sin abrumar su almacenamiento local, lo que resulta beneficioso para la productividad.
- **Inspiración y Tendencias:** Los modelos de difusión pueden servir como herramientas inspiradoras al capturar y reflejar las tendencias emergentes. Los diseñadores pueden utilizar estos modelos para identificar elementos de moda populares y adaptarlos a sus propias creaciones.
- **Manejo Eficiente de Recursos:** Dada la capacidad de los modelos de difusión para generar resultados impactantes

con tamaños de archivo más manejables, son una opción eficiente en términos de recursos. Esto es crucial para diseñadores independientes o empresas emergentes que pueden enfrentar limitaciones de almacenamiento y recursos computacionales.

- **Personalización y Diversidad:** Permiten la personalización y diversificación de la moda al adaptar estilos a diferentes preferencias y audiencias. Esto contribuye a la creación de colecciones más inclusivas y representativas de una amplia variedad de gustos y culturas.

Los modelos de difusión en la moda ofrecen a los diseñadores una herramienta poderosa para la creatividad, la exploración de estilos y la optimización de procesos, contribuyendo así a la evolución continua e innovación dentro de la industria.

CONCLUSIONES

Quizá más que con otro avance tecnológico, seguir la pista de la actualidad en la IA es casi que imposible, por lo menos en etapas recientes. Los avances suceden casi que a diario y la tecnología que hoy se está lanzando, al cabo de 2 o 3 meses ya ha sido superada por otro modelo o por otra versión mejorada de sí misma. La cantidad de empresas que invierten en modelos generativos y aprendizaje profundo va desde jugadores pequeños, que son los que más fácil tienen la experimentación, hasta empresas grandes en las que si bien la innovación es más calculada y lenta, si tienen las instalaciones con la capacidad de computo suficiente para el entrenamiento de modelos con inmensos *datasets*. De los modelos que generaban imágenes confusas e indistinguibles, hemos pasado en menos de 5 años a crear imágenes que compiten con profesionales formados en sus respectivos campos.

Esta circunstancia plantea retos en el mercado laboral del diseño. Quienes primero lo están resintiendo son los diseñadores gráficos e ilustradores, quienes están inconformes con ciertas circunstancias que han traído el uso de modelos. Por ejemplo, desde 2022, diferentes grupos de ilustradores y artistas formaron manifestaciones y movimientos en contra del uso de la IA[21], alegando que el uso de estas herramientas incentivaba la destrucción del trabajo, la apropiación indebida intelectual por el uso de imágenes para entrenar modelos de artistas y movimientos específicos. Pero la manifestación no se queda ahí, puesto que muchos artistas están entablando demandas contra las empresas desarrolladoras de IA, por utilizar su obra para entrenar modelos generativos.

Los artistas alegan justamente en cuanto al uso de su obra de forma indebida. Lo que no se ha calculado es el nivel de oferta-demanda que tiene la ilustración. La obra de un artista tiene un tiempo de formación y consolidación, crear un estilo propio es una labor que puede durar incluso años y el artista está en todo su derecho a ponerle precio monetario e intelectual a su obra. Ahora bien, no todas las personas tienen la comprensión de esta circunstancia, pero tampoco tienen la capacidad monetaria para pagar una ilustración, la combinación de ambas cosas hace que el uso indiscriminado de las IA produzca resultados sin un grado alto de algo que llamaremos en adelante "Alfabetización visual"[22].

Para crear una imagen contundente con IA en esta etapa de su desarrollo, que contar con una gama de conocimientos variados, en técnicas pictóricas, historia del arte y movi-

21. (Noveck & O'Brien, 2023), Artistas visuales luchan contra empresas de IA para proteger su obra, en *Los Ángeles Times.* Recuperado de: https://www.latimes.com/espanol/entretenimiento/articulo/2023-08-31/artistas-visuales-luchan-contra-empresas-de-ia-para-proteger-su-obra

22. El término *Visual Literacy* o Alfabetización Visual, se refiere al grupo de habilidades y técnicas para encontrar, evaluar y formar criterios respecto al uso y creación de imágenes. Estas habilidades son estudiadas por diferentes instituciones y universidades, que se ocupan de la creación y contenido de las imágenes abordándolas desde sus aspectos técnicos y semióticos. Statton. D (2019) Teaching students to critically read digital images: a visual literacy approach using the DIG method, *Journal of Visual Literacy*, 38:1-2, 110-119, DOI: 10.1080/1051144X.2018.1564604

mientos artísticos, técnicas de fotografía, renderizado, entre otras. Este conocimiento específico garantiza que las imágenes generadas tengan la calidad que se aproxime a las generadas por la habilidad humana.

Por otro lado, los ilustradores y visualizadores pueden a partir del entrenamiento de modelos, mejoras sus procesos de creación, aminorar los tiempos muertos o las actividades repetitivas propias de técnicas manuales o digitales[23]. No es la primera vez que un avance tecnológico cuestiona las bases de una profesión. La llegada de la fotografía en su momento hizo que la gente creyera que la pintura iba a desaparecer, o que la llegada de la radiodifusión iba a acabar con las orquestas, cosas que no pasaron. Lo que si sucedió fue que tanto la pintura como la música encontraron otros caminos de exploración y lenguajes que propiciaron la llegada de las vanguardias de principios del siglo XX.

No solo los ilustradores sienten el impacto disruptivo de la aparición de los modelos generativos. La profesión del modelaje también ha sentido el cambio producido por el entrenamiento de modelos con imágenes reales para nutrir *datasets* y generar imágenes de un modelo sin que este se encuentre presente en una sesión de fotos. Si a eso se le suman los avances de escaneo 3D y la digitalización de volúmenes, podemos hablar que en un futuro no solo se podrán crear imágenes estáticas, sino piezas audiovisuales, desfiles, automatización de la metrología y las formas de patronaje, etc.

A pesar de todas estas objeciones éticas y profesionales, los modelos siguen avanzando ya sea en sus formas *open source*, o en sus formas monetizadas, como las implementaciones que hacen en los programas de diseño y CAD que están haciendo actualmente empresas como Autodesk o Adobe. Lo que plantea una irreversibilidad e improbabilidad de cálculo del impacto en las profesiones. Son tecnologías que han llegado para quedarse y nuestra respuesta debe ir más allá de una actitud ludita[24], que si bien en cierto grado es positiva —puesto que hace que nos cuestionemos y tengamos suspicacias al respecto—, serán las personas que adapten estas herramientas quienes tengan ventaja sobre quienes se opongan o no tengan la capacidad de reinventarse. Por eso se propone desde este artículo que en la formación se incluya no solamente el uso de las herramientas, sino la formación en implicaciones éticas de su uso, además de la ya mencionada "Alfabetización visual" en la que se formen personas con habilidades técnicas suficientes para el manejo profesional de los modelos y unas bases éticas y de habilidades blandas que permitan no solo resultados estéticos, sino empáticos y más humanos.

A partir, de la escritura de este artículo ya se están evidenciando avances en la generación de imágenes en simultáneo, es decir, en una misma interfaz se puede ir dibujando y el modelo va reinterpretando a partir de un *prompt* el resultado de la imagen. También se estima que los avances irán más por el campo de la generación de video, ya que la optimización en la velocidad de los procesos reducirá los tiempos de renderizado. Se recomienda siempre estar al tanto de los avances e implementarlos en los flujos de trabajo, ya sea a nivel formativo, profesional o comercial.

23. En el artículo ¿Qué lugar ocupa la creatividad en la industria ante el avance de la IA? (Rubbiani, 2023), plantea que a pesar de los avances de la IA, las industrias creativas aún tienen la posibilidad de generar proyectos en los que la empatía, hacen la comunicación efectiva de ideas. Tomado de: https://www.latinspots.com/sp/noticia/qu-lugar-ocupa-la-creatividad-en-la-industria-ante-el-avance-de-la-ia/68381

24. Se denominaron "luditas" a los movimientos de artesanos que en la revolución industrial de mediados de siglo XVIII se opusieron a la mecanización y aceleración del trabajo y de la producción. Los luditas se opusieron a la industrialización no solo a partir de un pensamiento o postura, sino desde la acción directa, es decir, quemaban máquinas y fábricas, y boicoteaban cualquier avance mecanizado. El término se ha mantenido hasta el día de hoy y se le asigna a las personas que se oponen o sospechan de los avances tecnológicos. No sólo los artesanos fueron *luditas*, hubo personas con formación que hicieron parte y defendieron las posturas del ludismo, como se puede ver en el ensayo de "Los rompedores de máquinas" del historiador Eric Hobsbawm, donde defendía la labor de los artesanos y trabajadores manuales ingleses ((Hobsbawm, 1952). *The machine breakers*. Recuperado de: https://libcom.org/article/machine-breakers-eric-hobsbawm)

REFERENCIAS BIBLIOGRÁFICAS

Alammar, J. (Noviembre de 2022). *The Illustrated Stable Diffusion*. Obtenido de https://jalammar.github.io/illustrated-stable-diffusion/

Andrew(Username). (22 de Noviembre de 2023). *What are LoRA models and how to use them in AUTOMATIC1111*. Obtenido de https://stable-diffusion-art.com/lora/

Baio, A. (30 de Octubre de 2022). *Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator.* Obtenido de https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/

Escott, E. (24 de Octubre de 2017). *What are the 3 types of AI? A guide to narrow, general, and super artificial intelligence*. Obtenido de https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible

Gözükara, F. (20 de Septiembre de 2023). *SDXL LoRA vs SDXL DreamBooth Training Results Comparison*. Obtenido de https://stable-diffusion-art.com/lora/

Hobsbawm, E. (1952). *The machine breakers.* Obtenido de https://libcom.org/article/machine-breakers-eric-hobsbawm

Lvmin Zhang, A. R. (10 de Noviembre de 2023). *Adding Conditional Control to Text-to-Image Diffusion Models.* Obtenido de https://arxiv.org/abs/2302.05543

Mori, M. (1970). The uncanny valley. *Energy*, 33-35.

Noveck, J., & O'Brien, M. (31 de Agosto de 2023). *Artistas visuales luchan contra empresas de IA para proteger su obra*. Obtenido de https://www.latimes.com/espanol/entretenimiento/articulo/2023-08-31/artistas-visuales-luchan-contra-empresas-de-ia-para-proteger-su-obra

Perrigo, B. (13 de Abril de 2023). *The A to Z of Artificial Intelligence.* Obtenido de https://time.com/6271657/a-to-z-of-artificial-intelligence/

Rubbiani, D. (8 de Diciembre de 2023). *Qué lugar ocupa la creatividad en la industria ante el avance de la IA*. Obtenido de https://www.latinspots.com/sp/noticia/qu-lugar-ocupa-la-creatividad-en-la-industria-ante-el-avance-de-la-ia/68381

StabilityAI. (7 de Diciembre de 2022). *Basics of Prompt Engineering.* Obtenido de Basics of Prompt Engineering

Wiesel, D. H. (Enero de 1962). *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex* . Obtenido de doi: 10.1113/jphysiol.1962.sp006837.

Wiggers, K. (12 de Agosto de 2022). *This startup is setting a DALL-E 2-like AI free, consequences be damned*. Obtenido de https://techcrunch.com/2022/08/12/a-startup-wants-to-democratize-the-tech-behind-dall-e-2-consequences-be-damned/?guccounter=1