
UNIVERSIDAD DE GRANADA

Máster Universitario en
Gestión y Tecnologías de Procesos de Negocio
Curso académico 2024-2025

*Trabajo Fin de Máster*

# ESTUDIO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL PARA DISEÑO DIGITAL Y PUBLICIDAD

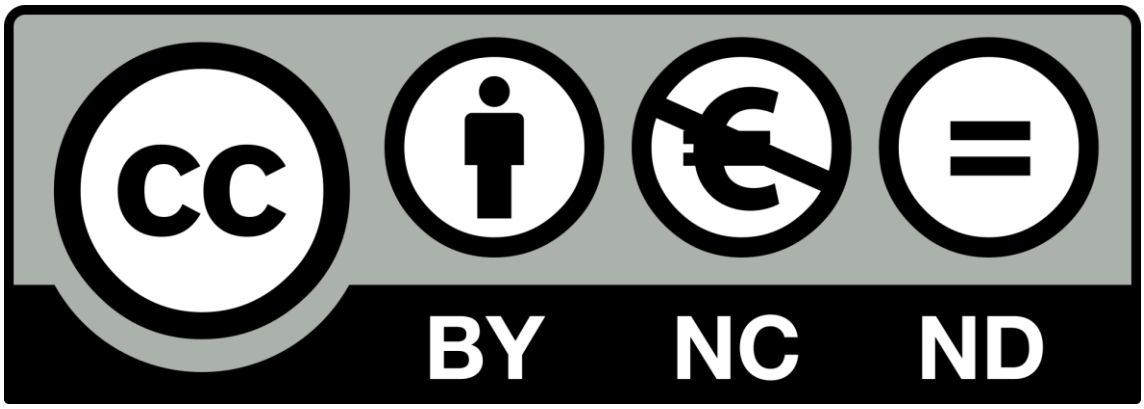
Nicole Nicho Príncipe

Tutor:

Germán Arroyo Moreno

**DETECCIÓN DEL PLAGIO**

La Universidad utiliza el programa ***Turnitin Feedback Studio*** para comparar la originalidad del trabajo entregado por cada estudiante con millones de recursos electrónicos y detecta aquellas partes del texto copiadas y pegadas. **Copiar o plagiar en el TFM se considera falta grave, y puede conllevar la expulsión definitiva de la Universidad**.



# Estudio de herramientas de inteligencia artificial para Diseño Digital y Publicidad

Nicole Nicho Príncipe

**Palabras clave:** inteligencia artificial, diseño digital, ComfyUI, Stable Diffusion, generación de imágenes, IA generativa, automatización de flujos de trabajo

## Resumen

Este trabajo de fin de máster analiza el impacto de las herramientas de inteligencia artificial en los procesos creativos del diseño digital y la publicidad, con especial atención a ComfyUI, una plataforma de generación de imágenes basada en nodos que permite trabajar con modelos como Stable Diffusion, LoRA y ControlNet. El objetivo principal es ofrecer una guía práctica que facilite la comprensión y aplicación de ComfyUI, especialmente para usuarios sin experiencia técnica.

La metodología combina una revisión teórica sobre los fundamentos de la inteligencia artificial en el ámbito creativo con el desarrollo de una serie de casos prácticos que exploran distintos flujos de trabajo desde la generación de imágenes a partir de texto, hasta la utilización de modelos personalizados y técnicas de control avanzado como poses, profundidad o estilos gráficos. Para ello, se emplean modelos oficiales como SD1.5, SDXL, SDXL Turbo, Epic Realism y se documentan los procedimientos de instalación, configuración y uso de la herramienta.

Como resultado, se presenta un conjunto de casos de uso accesibles y reproducibles, que demuestran el potencial de ComfyUI para la creación visual mediante IA. Este trabajo pretende servir de puente entre las capacidades técnicas de estas herramientas y su aplicación práctica en procesos creativos, proporcionando un recurso útil tanto para profesionales interesados en la adopción de la IA en el sector creativo.

# Study of Artificial Intelligence Tools for Digital Design and Advertising

Nicole Nicho Príncipe

**Keywords:** artificial intelligence, digital design, ComfyUI, Stable Diffusion, image generation, generative AI, workflow automation

## Abstract

This master's thesis analyzes the impact of artificial intelligence tools on creative processes in digital design and advertising, with a particular focus on ComfyUI a node-based image generation platform that enables the use of models such as Stable Diffusion, LoRA, and ControlNet. The main objective is to provide a practical guide that facilitates the understanding and application of ComfyUI, especially for users without a technical background.

The methodology combines a theoretical review of the fundamentals of artificial intelligence in creative fields with the development of practical case studies that explore various workflows, from text-to-image generation to the use of custom models and advanced control techniques such as pose, depth, or graphic styles. Official models such as SD1.5, SDXL, SDXL Turbo, and Epic Realism are used, and detailed procedures for the installation, configuration, and use of the tool are documented.

As a result, a set of accessible and reproducible use cases is presented, demonstrating the potential of ComfyUI for AI-driven visual creation. This work aims to bridge the gap between the technical capabilities of these tools and their practical application in creative processes, providing a valuable resource for professionals interested in adopting AI in the creative sector.

## AGRADECIMIENTOS

Quiero expresar mi más profundo agradecimiento a todas las personas que me han acompañado y apoyado a lo largo de este proceso académico y personal.

Hace casi un año inicié una nueva etapa fuera de mi país, enfrentando el reto de estar lejos de mi familia, amigos y mi zona de confort. Este camino ha estado lleno de aprendizajes, desafíos y crecimiento, tanto profesional como personal.

A mi familia, por su amor incondicional y su apoyo constante, incluso desde miles de kilómetros de distancia. Gracias por permitirme extender mis alas, incluso cuando eso significó alejarme de casa. Este logro no es solo mío; es también el reflejo de sus sueños depositados en mí.

A mis padres, que me han acompañado con cada paso, no para que viva con miedo, sino para que me atreva a soñar, enfrentar dificultades y construir una vida de la que pueda sentirme orgullosa.

A mi novio, por siempre estar en cada estación, por sostenerme en los días grises y celebrar conmigo mis logros por más pequeños que sean.

A mis seres queridos que ya no están físicamente, pero cuya presencia ha sido constante en mi corazón. A mi abuela, especialmente, por creer en mí más de lo que yo misma lo hacía.

A mi tutor, por su orientación, dedicación y apoyo constante a lo largo del desarrollo de este trabajo.

# ÍNDICE GENERAL

# ÍNDICE DE FIGURAS

# ÍNDICE DE TABLAS

# 1. INTRODUCCIÓN

En la era digital, la inteligencia artificial (IA) ha adquirido un papel fundamental en diversos sectores, transformando la manera en que se crean, gestionan y optimizan procesos. Desde la automatización industrial hasta la personalización del contenido en plataformas digitales, la IA ha demostrado ser una tecnología clave para mejorar la eficiencia y la creatividad en múltiples disciplinas. En particular, el ámbito del diseño digital y la publicidad ha experimentado una evolución significativa con la incorporación de herramientas impulsadas por IA, que permiten generar imágenes, optimizar composiciones visuales y mejorar la personalización del contenido de manera automatizada.

El desarrollo de herramientas impulsadas por IA ha transformado la manera en que los profesionales del diseño y la publicidad crean contenido. Antes, la generación de material visual dependía principalmente del talento, la experiencia y el uso de software de edición manual. Ahora, gracias a modelos avanzados de generación de imágenes y análisis de datos, estos procesos pueden agilizarse sin sacrificar la calidad ni la creatividad. Para sacar el máximo provecho de estas tecnologías, los profesionales especialmente en los sectores de innovación y diseño deben comprender su funcionamiento, y aprender a utilizarlas de manera efectiva para potenciar su creatividad y mejorar sus procesos de trabajo.

Dentro de este contexto, el presente proyecto se enfoca en analizar herramientas específicas que aplican inteligencia artificial en el diseño digital, con un énfasis particular en ComfyUI. Esta plataforma permite crear, modificar y mejorar imágenes mediante redes neuronales, y destaca por su enfoque intuitivo basado en nodos, lo que facilita su uso incluso para personas con poca experiencia técnica.

A través de este trabajo, se busca ofrecer una visión teórica y práctica sobre el impacto de estas herramientas en el ámbito del diseño y la publicidad, proporcionando tanto una guía aplicada para su uso como una comprensión más profunda de su funcionamiento y potencial creativo.

## 1.1. Motivación

En la actualidad, la inteligencia artificial (IA) ha demostrado ser una herramienta clave en diversos sectores, transformando la manera en que se gestionan y ejecutan procesos creativos y productivos. En este contexto, las herramientas de IA aplicadas al diseño digital y la publicidad están revolucionando el panorama de la creatividad, permitiendo a los profesionales generar contenido de manera más eficiente, innovadora y personalizada.

A medida que estas herramientas siguen evolucionando, se hace imprescindible entender cómo funcionan y cómo se pueden implementar en los flujos de trabajo actuales. Esto no solo contribuye a optimizar los procesos, sino que también abre nuevas posibilidades en términos de personalización, lo que lleva a una mejora continua en los resultados.

Este trabajo tiene como objetivo principal estudiar el funcionamiento de ComfyUI dentro del ámbito del diseño digital y la publicidad, desde su instalación, configuración, funcionalidades y características, además de ejecutar una demostración práctica de la herramienta. Además, se realizará una comparación con otras herramientas similares disponibles en el mercado, con el fin de evaluar las ventajas y limitaciones de cada una de ellas. También se explorarán los antecedentes técnicos relacionados con los conceptos que fundamentan estas herramientas de IA, tales como las redes neuronales, el aprendizaje automático y otros modelos avanzados que hacen posible la creación de contenido mediante IA.

La motivación detrás de este trabajo surge de mi interés por comprender cómo la inteligencia artificial está transformando el sector del diseño y la publicidad. Como estudiante de un máster en Gestión y Tecnologías de Procesos de Negocio, considero fundamental explorar no solo las capacidades técnicas de estas herramientas, sino también su impacto en la creatividad, la toma de decisiones y la forma en que los profesionales deben adaptarse para mantenerse competitivos. En un entorno donde la innovación avanza rápidamente, resulta crucial no solo adoptar nuevas tecnologías, sino también entenderlas a fondo para utilizarlas de manera efectiva.

Con base en los conocimientos adquiridos durante el máster y la creciente tendencia hacia la adopción de IA en la industria, este trabajo tiene como objetivo no solo comprender el potencial de estas herramientas, sino también proporcionar un recurso práctico que sea útil tanto para estudiantes como para profesionales que deseen aplicarlas en el diseño y la publicidad. A través de este proyecto, espero contribuir con un enfoque que ayude a los profesionales a integrar estas herramientas de forma consciente y efectiva en sus procesos creativos.

## 1.2. Objetivos

El objetivo principal de este trabajo de fin de máster es:

Analizar y evaluar las capacidades de las herramientas de inteligencia artificial en el diseño digital y la publicidad, con énfasis en ComfyUI, mediante una guía práctica que facilite su comprensión, explorando su funcionamiento, configuración y aplicación en distintos flujos de trabajo creativos.

Para alcanzar este objetivo general, se establecen los siguientes objetivos específicos:

- Describir el contexto actual y la evolución de la inteligencia artificial aplicada al diseño digital y la publicidad, identificando los principales avances tecnológicos y su relevancia en el sector.
- Explorar y comparar las principales herramientas de IA disponibles en el mercado para la generación y edición de imágenes en diseño digital, centrándose en sus funcionalidades clave.
- Documentar en profundidad el uso de ComfyUI, incluyendo su instalación, configuración, estructura basada en nodos, integración con diferentes modelos, comparación de parámetros de control y ejemplos de flujos de trabajo aplicados a distintos contextos creativos.

## 1.3. Estructura de la memoria

El presente proyecto consta de siete capítulos, tal y como se describen a continuación:

**Capítulo 1: Introducción, Motivación y Objetivos**

En el primer capítulo se establece el contexto general del trabajo y la relevancia de la inteligencia artificial en el diseño digital y la publicidad. Se presenta la motivación detrás del trabajo, destacando la importancia de comprender herramientas como ComfyUI y su impacto en la actualidad. Se expone la motivación que llevó a desarrollar este estudio, así como los objetivos generales y específicos que se buscan alcanzar.

**Capítulo 2: Estado del arte**

En el segundo capítulo se analizan antecedentes y trabajos relacionados sobre la aplicación de inteligencia artificial en contextos creativos, estableciendo la base investigadora y académica del proyecto.

**Capítulo 3: Fundamentos de la inteligencia artificial y aprendizaje profundo**

En el tercer capítulo se presenta los fundamentos teóricos necesarios para entender cómo funcionan las herramientas de generación de contenido mediante inteligencia artificial. Se explican conceptos esenciales como la inteligencia artificial (IA), el aprendizaje automático (machine learning), las redes neuronales artificiales y el aprendizaje profundo (deep learning). También se introduce el papel de las redes neuronales convolucionales (CNN) en el procesamiento de imágenes, así como el funcionamiento de los modelos de lenguaje.

**Capítulo 4: Modelos de difusión y ecosistema tecnológico**

En el cuarto capítulo se detallan los modelos de difusión, con especial énfasis en Stable Diffusion y sus variantes más recientes, como SDXL y SDXL Turbo. Se explica su funcionamiento, arquitectura modular y principales parámetros de control. Además, se analizan tecnologías complementarias como LoRA (Low-Rank Adaptation), ControlNet y también se profundiza en el diseño de interfaces basadas en nodos, lo cual resulta clave para el uso práctico de herramientas como ComfyUI.

**Capítulo 5: Análisis de herramientas del mercado**

En el quinto capítulo se realiza una exploración y comparación de diferentes herramientas de IA utilizadas en el diseño digital y la publicidad. Se presentan sus características principales y se incluyen tanto herramientas open-source como propietarias, para evaluar su funcionalidad, facilidad de uso y personalización.

**Capítulo 6: Análisis de ComfyUI**

En el sexto capítulo se desarrolla una exploración práctica de ComfyUI, centrada en su instalación, configuración inicial y uso funcional. A través de casos prácticos se prueban flujos de trabajo que permiten evaluar diferentes funcionalidades como generación de imágenes, transformación de estilos, uso de ControlNet y LoRA, destacando su potencial para proyectos creativos que requieren un alto grado de control técnico. Además, se realiza un análisis detallado de parámetros clave como Sampler, Steps, CFG, Seed y Denoise, evaluando su impacto en la generación de imágenes para distintos productos, lo que aporta una guía práctica para optimizar resultados según objetivos creativos específicos.

**Capítulo 5: Conclusiones y líneas futuras de trabajo**

En el capítulo final se presentan las conclusiones del trabajo con base en los objetivos planteados. Se analizan los hallazgos más relevantes, así como las principales ventajas y limitaciones observadas en el uso de ComfyUI. Por último, se proponen posibles mejoras y líneas de investigación futuras que podrían complementar o extender el trabajo realizado.

# 2. REVISIÓN DEL ESTADO DEL ARTE

## 2.1. Antecedentes/ Trabajos relacionados

En este capítulo se presentará como antecedentes las diversas investigaciones de proyectos similares que han sido desarrollados en temas relacionados a la presente investigación.

**Herencia Campaña (2024),** titulado en Multimedia por la Universitat Politècnica de Catalunya, realizó un trabajo titulado La IA en la profesión del creativo, con el objetivo de desarrollar una guía práctica para que los diseñadores gráficos y web utilicen herramientas de inteligencia artificial (IA) como DALL-E, Stable Diffusion y Midjourney de manera efectiva en la generación de imágenes. El estudio buscó evaluar las ventajas y limitaciones de estas herramientas, así como identificar las mejores prácticas para la creación de prompts que optimicen los resultados.

Como metodología, se empleó un enfoque ágil e iterativo, realizando pruebas comparativas con las herramientas mencionadas para analizar su capacidad de interpretar descripciones textuales y generar imágenes de calidad. Se evaluaron aspectos como la precisión, el realismo, la creatividad y la coherencia con los prompts.

Los resultados demostraron que Midjourney destaca por su alta calidad visual y detalle artístico, mientras que Stable Diffusion sobresale en realismo y DALL-E en versatilidad creativa. Sin embargo, todas las herramientas presentaron limitaciones en la generación de textos dentro de las imágenes. La guía desarrollada proporciona instrucciones detalladas para la creación de prompts efectivos, incluyendo ejemplos y estrategias para cada herramienta.

Finalmente, se concluyó que la IA puede transformar el proceso creativo, liberando a los diseñadores de tareas repetitivas y permitiéndoles enfocarse en aspectos más estratégicos. Se recomendó actualizar periódicamente la guía debido a la rápida evolución de estas tecnologías y explorar su integración en otras áreas del diseño.

**Ji Wang (2023),** graduada en Publicidad y Relaciones Públicas por la Facultad de Ciencias Sociales, Jurídicas y de la Comunicación, presentó un Trabajo de Fin de Grado titulado “La inteligencia artificial en la industria publicitaria”, con el objetivo de investigar el impacto de la inteligencia artificial (IA) en el sector publicitario y explorar las formas prácticas de aplicar sus herramientas en el contexto actual. El estudio también analizó la evolución histórica de la IA y su influencia en la vida cotidiana, abordando aspectos como la ética, la educación, la representación audiovisual en los medios, la aparición de influencers digitales, el metaverso, la creatividad artificial y el futuro de esta tecnología.

Como metodología, se empleó un estudio documental cualitativo basado en la revisión de fuentes secundarias, como artículos académicos, libros y sitios web especializados, contrastando la hipótesis inicial: *"La aplicación de la inteligencia artificial en la industria publicitaria tiene el potencial de optimizar y mejorar las actividades y campañas, aunque su implementación plantea una serie de dificultades que deben ser tenidas en cuenta para gestionarlas de forma responsable"*.

Entre los resultados destacados se encontró que la IA ofrece beneficios significativos, como la personalización de campañas, el análisis de grandes volúmenes de datos para predecir comportamientos del consumidor y la automatización de tareas repetitivas. Sin embargo, también se identificaron desafíos éticos, como la privacidad de los datos, la discriminación algorítmica y el uso malintencionado de herramientas como los deepfakes. Además, se analizaron casos prácticos de herramientas como GPT-4, DALL-E 2 y influencers virtuales, que demuestran cómo la IA está transformando la creatividad y la interacción con las audiencias.

Finalmente, se confirmó la hipótesis inicial, evidenciando que la IA optimiza procesos publicitarios pero requiere un enfoque responsable para mitigar riesgos. Como conclusión, se destacó la necesidad de equilibrar la innovación tecnológica con la transparencia, la educación en competencias digitales y la adaptación continua de los profesionales del sector. Además, se proyectó un futuro donde la IA seguirá evolucionando, ofreciendo oportunidades para la industria publicitaria, siempre que se gestionen sus implicaciones éticas y sociales de manera adecuada.

**Muñoz Murillo (2023),** titulada en Ingeniería de Sonido e Imagen, realizó una tesis titulada "Inteligencia Artificial aplicada al Análisis de Imágenes", con el objetivo de evaluar el rendimiento de herramientas de IA generativa (Playground AI, Clipdrop y Lexica) en la creación de imágenes mediante técnicas de prompting, así como desarrollar modelos de clasificación de imágenes médicas mediante redes neuronales convolucionales.

En el ámbito de la generación de imágenes, el estudio analizó cómo diferentes estructuras de prompts afectaban la calidad y coherencia de los resultados. Se comprobó que los prompts organizados por categorías específicas (estilo visual, género, iluminación) mejoraban significativamente la precisión de las imágenes generadas. Sin embargo, también se identificaron limitaciones importantes, particularmente en escenarios complejos que requerían composiciones detalladas o perspectivas no convencionales.

Los resultados demostraron que, mientras las herramientas basadas en prompting tradicional son útiles para generación básica, presentan carencias críticas para aplicaciones profesionales como la falta de control granular sobre los parámetros de generación, incapacidad para integrar múltiples modelos especializados, y rigidez para

automatizar flujos de trabajo complejos. Estas limitaciones subrayan la necesidad de interfaces más avanzadas y personalizables.

Finalmente, el estudio estableció una metodología cuantitativa para evaluar herramientas generativas, sentando las bases para futuras investigaciones sobre sistemas más sofisticados, además la autora propuso como línea futura la exploración de interfaces avanzadas basadas en nodos para superar las limitaciones identificadas en herramientas de prompting tradicionales.

**Lara Artolazábal (2023)**, titulado en Ingeniería Informática en Tecnologías de la Información por la Universidad Miguel Hernández de Elche, realizó un trabajo fin de grado titulado: "Generación de imágenes con Inteligencia Artificial: revisión de la situación actual y aplicaciones prácticas", con el objetivo de analizar y comparar los modelos más relevantes de generación de imágenes mediante IA, evaluando sus ventajas, limitaciones y aplicaciones prácticas.

Como metodología, se llevó a cabo un estudio exhaustivo de los modelos DALL-E, Midjourney y Stable Diffusion, utilizando pruebas comparativas con los mismos prompts para evaluar la calidad, coherencia y detalle de las imágenes generadas. Además, se exploraron herramientas complementarias de Stable Diffusion, como ControlNet y Dreambooth, para entrenar modelos personalizados y mejorar el control sobre los resultados.

Los resultados demostraron que Midjourney destaca por generar imágenes más artísticas y detalladas, mientras que DALL-E ofrece una interfaz más intuitiva y funciones avanzadas de edición. Stable Diffusion, aunque menos preciso en prompts complejos, sobresale por su naturaleza de código abierto y la flexibilidad que brinda a los usuarios para personalizar y ampliar sus capacidades. La implementación de Dreambooth permitió entrenar el modelo con imágenes propias, logrando resultados personalizados, y ControlNet facilitó un mayor control en la generación de imágenes mediante guías estructurales.

Finalmente, se concluyó que la generación de imágenes con IA es un campo en rápido avance, con herramientas cada vez más accesibles y potentes. Se sugirió explorar futuras aplicaciones, como la generación de vídeos y modelos 3D, así como mejorar la usabilidad para acercar estas tecnologías al público general. Además, se destacó la importancia de considerar los aspectos éticos y legales derivados del uso de estas herramientas en el ámbito creativo.

**Gil Viñarás (2023),** titulado en Filosofía por la Universidad de Zaragoza, realizó un trabajo de fin de grado titulado "Acercamiento filosófico a las imágenes generadas por inteligencia artificial. Análisis y experimentación técnica de imágenes con el programa

Midjourney", cuyo objetivo principal fue analizar desde una perspectiva filosófica y experimental las imágenes generadas por inteligencia artificial (IA), específicamente utilizando el programa Midjourney. El trabajo se propuso demostrar que las imágenes creadas por IA presentan sesgos, reproducen una realidad estadística cargada de elementos comunes y clichés, y pueden apropiarse de imágenes culturales preexistentes para generar otras "nuevas". Además, el autor planteó la necesidad de desarrollar un Test de Turing adaptado a las imágenes generadas por IA, para distinguir sus características frente a la creación humana.

Como metodología, Gil Viñarás combinó una revisión bibliográfica interdisciplinar (centrada en aspectos técnicos de la IA, discursos sobre la imagen y el cruce entre arte y tecnología) con experimentos prácticos generando imágenes mediante prompts en Midjourney. A través del análisis y comparación de los resultados, se buscaron tendencias y se formularon hipótesis sobre el tipo de expresión y los límites de la creatividad de la IA.

Los resultados obtenidos confirmaron en términos generales las hipótesis iniciales: las imágenes generadas por IA tienden a reproducir patrones estadísticos, reflejando sesgos y lugares comunes que las alejan de la capacidad humana de generar lo verdaderamente novedoso y concreto. El trabajo destaca la diferencia sustancial entre la mirada estadística de la IA y la mirada humana, capaz de introducir diferencia y movimiento en la historia de las imágenes. Esta diferencia es considerada esencialmente humana y de suma importancia para el desarrollo cultural.

Finalmente, el autor concluye que es necesario plantear un Test de Turing específico para imágenes generadas por IA, que permita identificar sus limitaciones y carencias, especialmente en lo relativo a la expresión de características humanas. Se sugiere que futuras investigaciones amplíen la recogida de datos empíricos y exploren la aplicación de estos análisis en otros contextos culturales y tecnológicos.

**Amate Bonachera (2024),** doctor en el Grupo de Investigación Genius de la Universidad de Diseño, Tecnología e Innovación (UDIT), realizó un trabajo titulado "La IA generativa en el diseño gráfico y la publicidad: Herramientas y estrategias emergentes", cuyo objetivo principal fue examinar el impacto transformador de la inteligencia artificial generativa en los sectores del diseño gráfico y la publicidad, así como identificar los beneficios, desafíos éticos y ambientales, y proponer estrategias para su integración sostenible.

Como metodología, el autor combinó una revisión exhaustiva de la literatura académica sobre modelos avanzados de IA (como GANs, modelos de difusión y transformadores) con el análisis de casos prácticos de campañas publicitarias recientes, como la línea Teen de Mango y el anuncio navideño de Coca-Cola, ambos realizados con IA generativa. Además, se integró un enfoque interdisciplinar que abarca el diseño, la ética

tecnológica y el análisis de datos, permitiendo un marco teórico-práctico para evaluar las oportunidades y riesgos de estas tecnologías.

Los resultados del estudio muestran que herramientas como MidJourney, DALL-E y Adobe Firefly están democratizando el acceso a procesos creativos, permitiendo a profesionales y creativos independientes generar contenido visual hiperrealista y personalizado a una velocidad sin precedentes. Esto ha redefinido conceptos como la autoría y la originalidad, ya que los diseñadores y publicistas pasan a ser curadores y colaboradores, más que creadores exclusivos. Sin embargo, el trabajo también identifica riesgos importantes, como la sobresaturación de contenido genérico, la dependencia excesiva de algoritmos, el refuerzo de sesgos culturales y los desafíos en torno a la propiedad intelectual y el impacto ambiental.

Finalmente, el autor concluye que la IA generativa puede enriquecer la creatividad y transformar las industrias creativas si se utiliza de manera responsable y ética. Se recomienda la implementación de normativas claras, estrategias sostenibles para mitigar el impacto ambiental y una colaboración interdisciplinar entre diseñadores, tecnólogos y expertos en ética, asegurando que la innovación se alinee con valores fundamentales como la autenticidad, la equidad y la sostenibilidad.

# 3. FUNDAMENTOS DE LA INTELIGENCIA ARTIFICIAL Y APRENDIZAJE PROFUNDO

Dado que este proyecto se centra en analizar el impacto de herramientas como ComfyUI en el diseño digital y la publicidad, es fundamental comprender primero los conceptos clave que las sustentan. En este capítulo se explican las bases de la inteligencia artificial, el aprendizaje profundo y otros términos relacionados, que nos ayudarán a entender cómo funcionan estas tecnologías y por qué están revolucionando el proceso creativo.

## 3.1. Conceptos básicos de IA y aprendizaje automático

### 3.1.1. Inteligencia Artificial (IA)

La inteligencia artificial (AI) es un subcampo de la informática dedicada a la creación de algoritmos capaces de simular el comportamiento humano cognitivo, como el aprendizaje, el razonamiento lógico y la toma de decisiones (Russell & Norvig, 2016). Desde su inicio, la IA ha tratado de proporcionar a las máquinas la capacidad de realizar tareas que tradicionalmente necesitaban inteligencia humana, como el reconocimiento de patrones, el reconocimiento del lenguaje natural y la solución de problemas difíciles (Cárdenas, 2023). El término inteligencia artificial fue creado en 1956, y desde entonces, su alcance se ha ampliado para incluir desde la simulación de comportamiento inteligente hasta el desarrollo de agentes racionales con la capacidad de alcanzar el resultado más alto posible (DataSciencetest, s. f.).

El aprendizaje automático y el aprendizaje profundo se destacan entre los subcampos actuales de IA. El aprendizaje automático permite a las máquinas aprender de los datos y mejorar sus resultados sin una programación explícita para cada tarea, mientras que el aprendizaje profundo utiliza redes neuronales artificiales con múltiples capas para procesar grandes volúmenes de datos y extraer información compleja de datos de entrada (Cabanelas Omil, 2019). Estas tecnologías han permitido aplicaciones avanzadas como el procesamiento del lenguaje natural, la visión por computadora y la generación de contenido original, que han dañado los campos como la medicina, la industria y la educación (Mayol, 2024).

A pesar de los avances, la IA plantea cuestiones éticas, socioeconómicas y de seguridad ligados con la transparencia del algoritmo, el sesgo en los datos y la toma de decisiones humanas como lo han señalado diversos autores (Cárdenas, 2023; Cabanelas Omil, 2019). Según Russell y Norvig (2016) la investigación actual no solo se centra en mejorar las capacidades técnicas de la IA, sino también en crear nuevos marcos normativos y éticos para que su integración en la sociedad se haga responsablemente.

### 3.1.2. Aprendizaje Automático (Machine Learning)

El aprendizaje automático (ML) es un subcampo de la inteligencia artificial que se enfoca en el desarrollo de algoritmos y modelos estadísticos que permiten a las computadoras aprender y mejorar su desempeño en tareas específicas a partir de datos, sin ser programadas explícitamente para cada función (Russell & Norvig, 2016; Cárdenas, 2023). Esta capacidad de aprendizaje autónomo se basa en la identificación y generalización de patrones en grandes volúmenes de datos, lo que permite la toma de decisiones y el pronóstico con un alto grado de certeza (Uc Castillo et al., 2025).

El aprendizaje automático ha experimentado un crecimiento exponencial desde que el término se introdujo en 1959, debido a los avances en la capacidad computacional y la disponibilidad de datos de masa, así como la mejora de los algoritmos y técnicas (Uc Castillo et al., 2025; Mayol, 2024). Los principales enfoques incluyen el aprendizaje supervisado, no supervisado y por refuerzo, cada uno con aplicaciones específicas según la naturaleza de los datos y objetivos del problema (DataScientest, 2022; Russell & Norvig, 2016).

El aprendizaje supervisado utiliza conjuntos de datos etiquetados para entrenar modelos que puedan clasificar o predecir resultados sobre datos nuevos, mientras que el aprendizaje no supervisado busca descubrir estructuras ocultas o agrupamientos en datos no etiquetados. Por su parte, el aprendizaje por refuerzo se basa en la interacción con un entorno para maximizar una recompensa acumulada mediante prueba y error (Uc Castillo et al., 2025; DataScientest, 2022).

Entre las técnicas más utilizadas en ML se encuentran los árboles de decisión, máquinas de vectores de soporte (SVM), redes neuronales artificiales y métodos de agrupamiento (clustering), que se aplican en varios campos como la visión artificial, el procesamiento del lenguaje natural y la medicina (Angra & Ahuja, 2017). En particular, el aprendizaje profundo (deep learning), que emplea redes neuronales con múltiples capas, ha revolucionado la capacidad de los sistemas para procesar datos complejos y no estructurados, como imágenes y texto (Russell & Norvig, 2016; Uc Castillo et al., 2025).

Estas técnicas permiten automatizar procesos, mejorar la precisión en diagnósticos médicos, optimizar sistemas financieros y personalizar experiencias en plataformas digitales, entre otras aplicaciones (Mayol, 2024). Podemos decir entonces que el aprendizaje automático se ha consolidado como una herramienta fundamental para impulsar la innovación tecnológica y científica del siglo, transformando sectores clave y abriendo nuevas posibilidades en el análisis de datos, la automatización y el desarrollo de soluciones inteligentes

### 3.1.3. Redes Neuronales Artificiales

Las redes neuronales artificiales (RNA) son redes informáticas inspiradas en la estructur a y operación del cerebro humano, y están diseñadas para procesar datos utilizando una

red de unidades interconectadas llamadas neuronas artificiales. Estas neuronas reciben entradas ponderadas, las transforman a través de funciones de activación y transmiten señales a otras neuronas, permitiendo la identificación y generalización de patrones complejos en los datos (Arnal, 2018; Tirado Picado, 2024). Las RNA se caracterizan por su capacidad de aprendizaje a partir de datos, adaptándose y mejorando su desempeño en tareas específicas sin necesidad de programación explícita para cada caso (Aggarwal, 2018).

El desarrollo histórico de las RNA se inicia con el modelo matemático propuesto por McCulloch y Pitts en 1943, que representaba neuronas mediante funciones lógicas binarias. Posteriormente, en 1958, Rosenblatt introdujo el perceptrón, la primera red neuronal capaz de aprender mediante ajuste de pesos sinápticos. La Figura 1 muestra la estructura de un perceptrón simple, donde se observan las entradas, los pesos sinápticos, el sesgo, la función de activación y la salida resultante. Con el tiempo, la incorporación de arquitecturas multicapa y algoritmos como la retropropagación permitió a las RNA resolver problemas no lineales y manejar datos de alta dimensionalidad, lo que impulsó su aplicación en diversos campos (Campos Wright & Trujillo Casañola, 2021; Tirado Picado, 2024).

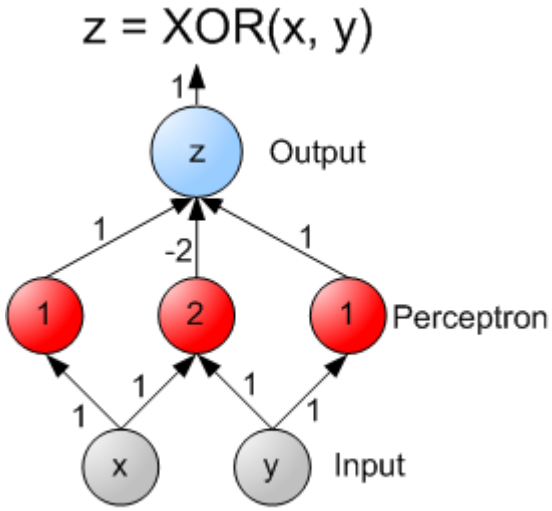



**Figura 1.** Esquema de un perceptrón simple. Fuente: Wikimedia Commons. (2004)

Las RNA se clasifican en varios tipos según su arquitectura y función. Las redes **feedforward** son las más básicas, con flujo unidireccional de información desde la capa de entrada hasta la de salida, utilizadas comúnmente en tareas de clasificación y regresión. Las redes **recurrentes** incorporan conexiones cíclicas que les permiten procesar datos secuenciales o temporales, como series de tiempo o lenguaje natural. Además, las redes **convolucionales** están especializadas en el procesamiento de datos con estructura espacial, como imágenes, siendo fundamentales en el aprendizaje

profundo (Aggarwal, 2018; Tirado Picado, 2024).

Actualmente, las RNA forman la base del aprendizaje profundo (deep learning), que utiliza redes con múltiples capas ocultas para extraer representaciones jerárquicas de datos complejos. Esta capacidad ha revolucionado la inteligencia artificial, permitiendo avances significativos en automatización, personalización y toma de decisiones inteligentes en sectores tan diversos como la salud, las finanzas y la tecnología (Aggarwal, 2018; Tirado Picado, 2024).

### 3.1.4. Deep Learning

El aprendizaje profundo, o deep learning, es una rama del aprendizaje automático que utiliza redes neuronales artificiales con varias capas para abordar problemas complejos, especialmente cuando se trata de datos no estructurados como imágenes, texto o audio (Goodfellow et al., 2016; Chollet, 2018).

A diferencia de los métodos tradicionales, que suelen depender mucho de que un experto diseñe manualmente las características importantes, el deep learning permite que los propios modelos aprendan representaciones jerárquicas y abstractas a medida que procesan la información a través de sus múltiples capas.

El crecimiento y popularidad del deep learning se deben a varios factores: la disponibilidad de grandes conjuntos de datos, algoritmos más eficientes para entrenar redes como la retropropagación y el uso de hardware especializado, como las GPU, que facilitan el manejo de modelos con millones de parámetros (Goodfellow et al., 2016).

Gracias a estas condiciones, han surgido arquitecturas muy avanzadas que pueden resolver tareas que antes se consideraban demasiado difíciles para la inteligencia artificial. Entre las arquitecturas más destacadas están:

- **Redes neuronales profundas (DNN)**: Modelos genéricos formados por muchas capas.
- **Redes convolucionales (CNN):** Diseñadas para trabajar con datos que tienen estructura espacial, como imágenes y videos.
- **Redes recurrentes (RNN):** incluyendo variantes como LSTM y GRU: Pensadas para secuencias y datos temporales, como el lenguaje o series de tiempo.
- **Redes generativas adversariales (GAN):** Que se usan para crear datos sintéticos muy realistas.
- **Transformers:** La arquitectura más avanzada en generación de texto y comprensión del lenguaje, que ha demostrado un rendimiento superior frente a

modelos previos en tareas como la traducción y el análisis semántico (Goodfellow et al., 2016; Chollet, 2018).

Una gran ventaja del deep learning es su habilidad para descubrir patrones complejos y relaciones no lineales en grandes volúmenes de datos, logrando en muchos casos resultados igual o mejores que los humanos en áreas como reconocimiento de imágenes, traducción automática y diagnóstico médico basado en imágenes (Chollet, 2018). Por eso, hoy en día es la base de sistemas inteligentes en campos como la visión por computadora, el procesamiento del lenguaje, la robótica y las recomendaciones personalizadas.

El deep learning ha revolucionado el campo de la inteligencia artificial, permitiendo soluciones cada vez más sofisticadas para problemas complejos y el manejo eficiente de grandes cantidades de datos. Su capacidad para generalizar y adaptarse, junto con la evolución constante de sus arquitecturas, aseguran que seguirá siendo clave en la innovación tecnológica y científica en los próximos años (Goodfellow et al., 2016; Chollet, 2018).

### 3.1.5. Redes Neuronales Convolucionales (CNN)

Las redes neuronales convolucionales (CNN) son una de las arquitecturas más importantes dentro del aprendizaje profundo, especialmente diseñadas para trabajar con datos que tienen una estructura espacial, como imágenes o señales en varias dimensiones. Lo que las hace tan poderosas es su capacidad para aprender automáticamente diferentes niveles de representación de los datos mediante filtros convolucionales.

Estos filtros identifican características tanto locales como globales a lo largo de varias capas, lo que permite a las CNN reconocer patrones complejos sin necesidad de que un experto seleccione manualmente qué características analizar (Krichen, 2023). En cuanto a su estructura, una CNN típica está compuesta principalmente por tres tipos de capas:

- **Capas convolucionales:** Aplican filtros que detectan características importantes como bordes y texturas en la imagen, capturando patrones locales esenciales.
- **Capas de agrupamiento (pooling):** Reducen la cantidad de datos al resumir regiones cercanas, lo que facilita el procesamiento y mejora la capacidad del modelo para generalizar.
- **Capas totalmente conectadas:** Combinan toda la información extraída para realizar tareas específicas, como clasificar imágenes o detectar objetos, transformando las características en resultados concretos.

Estas capas trabajan de forma conjunta para transformar la información inicial en representaciones cada vez más abstractas, facilitando que el modelo realice la tarea

deseada. Como se muestra en la Figura 2, se presenta una ilustración simplificada de una red neuronal convolucional (CNN) utilizada para la clasificación de imágenes.

Se muestra el flujo de datos desde la imagen de entrada, que es procesada inicialmente por la capa de entrada, pasando por varias capas ocultas donde se aplican filtros que extraen características específicas como bordes y texturas. Estas capas sucesivas permiten detectar características de mayor nivel hasta llegar a la capa de salida, que genera una distribución de probabilidad sobre las posibles clases por ejemplo, coche, autobús o avión.

Además el aprendizaje de una CNN se basa en la retropropagación, un algoritmo que calcula los gradientes del error a través de la red y ajusta los pesos de los filtros utilizando técnicas de optimización como el descenso de gradiente, con el objetivo de minimizar la diferencia entre la predicción y el valor real (Krichen, 2023).

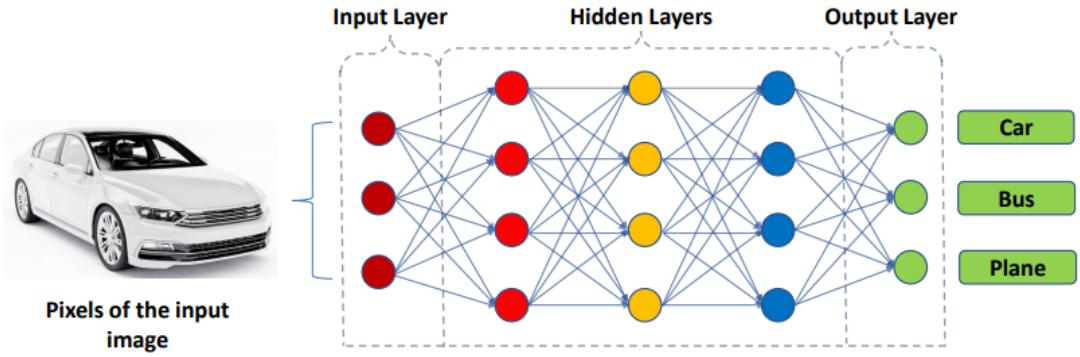



**Figura 2.** Ilustración simplificada de una red neuronal convolucional (CNN) para clasificación de imágenes. Fuente: Krichen (2023)

Hoy en día, las CNN son fundamentales en muchos campos, especialmente en visión por computadora, donde se utilizan para tareas como reconocimiento y clasificación de imágenes, detección y segmentación de objetos, análisis de video e incluso en diagnósticos médicos asistidos por imágenes. Su éxito radica en que pueden aprender directamente de los datos y adaptarse a problemas complejos, llegando en muchos casos a igualar o superar el rendimiento de expertos humanos (Krichen, 2023).

## 3.2. Modelos de Lenguaje y Generación de Contenido

### 3.2.1. Modelos de Lenguaje de Gran Tamaño (LLM)

En los últimos años, los Modelos de Lenguaje de Gran Tamaño (LLM, por sus siglas en inglés) se han posicionado como una de las innovaciones más influyentes dentro del procesamiento del lenguaje natural (PLN) y de la inteligencia artificial. Gracias a su entrenamiento con enormes volúmenes de texto, estos modelos pueden comprender, generar y analizar lenguaje humano de manera contextualizada y coherente. Esta

capacidad los convierte en herramientas sumamente versátiles para tareas que van desde la generación de texto automático y la traducción, hasta la respuesta a preguntas, la síntesis de información y la programación asistida (Luo et al., 2025; Atkinson-Abutridy, 2024).

La base tecnológica que permite el funcionamiento de estos modelos es la arquitectura transformer, presentada por Vaswani et al. en 2017. Este enfoque marcó un antes y un después al dejar atrás las redes neuronales recurrentes y convolucionales, apostando por un mecanismo de autoatención que facilita la comprensión de relaciones complejas y de largo alcance dentro de un texto. Gracias a ello, los modelos pueden procesar grandes cantidades de información de manera simultánea y eficiente (Devlin et al., 2019; Atkinson-Abutridy, 2024). A partir de esta arquitectura, han surgido modelos ampliamente conocidos como BERT, GPT, LLaMA y PaLM, que no solo mejoraron el rendimiento en tareas de PLN, sino que también permitieron el desarrollo de aplicaciones como asistentes conversacionales y generadores de código (Devlin et al., 2019; Luo et al., 2025).

El entrenamiento de un LLM suele dividirse en dos fases: el preentrenamiento y el ajuste fino (fine-tuning). En la primera etapa, el modelo aprende de manera general a partir de grandes corpus de texto sin etiquetar, con tareas como predecir la siguiente palabra o completar oraciones incompletas. Luego, en la segunda fase, se afina el modelo con datos más específicos, lo que le permite adaptarse mejor a contextos particulares o tareas concretas (Atkinson-Abutridy, 2024; Luo et al., 2025).

Una de las principales características de los modelos de lenguaje de gran tamaño (LLM) es su magnitud, ya que cuentan con miles de millones e incluso billones de parámetros. Esta gran capacidad les otorga la habilidad de realizar tareas complejas, como el razonamiento y la transferencia de conocimientos entre diferentes dominios (Luo et al., 2025; Bommasani et al., 2021). No obstante, dicha escala también genera retos significativos, tanto en el ámbito técnico como en el ético. Entre ellos se encuentran el alto consumo energético, la presencia de sesgos en los datos de entrenamiento, y la dificultad para comprender cómo toman decisiones los modelos (Bommasani et al., 2021; Atkinson-Abutridy, 2024).

### 3.2.2. Generación de Imágenes con IA

La generación de imágenes mediante inteligencia artificial ha revolucionado la creación de contenido digital, permitiendo la síntesis autónoma de imágenes, música y texto, y transformando tanto el ámbito creativo como el científico (Goodfellow et al., 2014; Kingma & Welling, 2014; Gao et al., 2024). Este progreso se debe a la evolución de modelos generativos como las Redes Generativas Antagónicas (GANs), los Autocodificadores Variacionales (VAEs) y, más recientemente, los modelos de difusión. Cada uno de estos enfoques ha aportado soluciones particulares a los desafíos de la síntesis de imágenes, como la mejora de la calidad visual, la diversidad y la estabilidad del entrenamiento.

Las GANs, introducidas por Goodfellow et al. (2014), emplean una estructura de competencia entre un generador y un discriminador para producir imágenes realistas, mientras que los VAEs, propuestos por Kingma y Welling (2014), adoptan un enfoque probabilístico que facilita la manipulación del espacio latente y la interpolación de atributos visuales. Sin embargo, los modelos de difusión han emergido como una de las técnicas más vanguardistas en la generación de imágenes, permitiendo la creación de imágenes realistas y variadas a través de procesos estocásticos iterativos (Ho et al., 2020; Gao et al., 2024). Estos modelos funcionan refinando progresivamente ruido aleatorio hasta formar imágenes coherentes, lo que les otorga una notable capacidad para abordar problemas complejos de manipulación y edición de imágenes.

**Figura 3.** Ejemplos de imágenes generadas con IA

Gao et al. (2024) destacan que los modelos de difusión no solo han alcanzado el estado del arte en calidad visual, sino que también han ampliado el espectro de aplicaciones posibles, desde la generación creativa hasta la resolución de problemas reales en manipulación de imágenes. Además, se subraya la importancia de las arquitecturas específicas y las técnicas desarrolladas para optimizar estos modelos, lo que ha permitido su rápida adopción en la industria y la investigación. Así, la IA generativa, especialmente a través de los modelos de difusión, se consolida como una herramienta esencial para la innovación en sectores como el arte, la medicina, la educación y la ciencia.

### 3.2.3. IA generativa

La inteligencia artificial generativa (IAG) es una rama emergente dentro del campo de la inteligencia artificial que se enfoca en la creación automática de contenido original como textos, imágenes, audio o video a partir de datos previos o instrucciones específicas. A diferencia de otros enfoques más tradicionales, centrados en clasificar información o hacer predicciones, la IAG se basa en aprender patrones complejos dentro de los datos y generar resultados nuevos que se alineen con esos patrones. Gracias a esta capacidad, ha transformado la forma en que se produce y consume contenido, automatizando procesos creativos y adaptando materiales a las necesidades de distintos sectores (Cortés Hernández et al., 2024).

En sus inicios, los sistemas generativos se apoyaban en reglas fijas y plantillas, pero el desarrollo de nuevas arquitecturas basadas en redes neuronales profundas permitió un salto cualitativo. Modelos como las redes generativas antagónicas (GANs), los autocodificadores variacionales (VAEs) y los modelos de difusión han logrado generar contenido con altos niveles de realismo y variedad. Con la llegada de modelos basados en transformadores, como GPT o DALL-E, se amplió aún más el alcance de la IAG, haciendo posible la generación de contenido que combina diferentes tipos de datos (texto, imagen, audio), y llevándola a una escala mucho más amplia (Gao et al., 2024).

Según el tipo de datos que procesan y su arquitectura, existen distintas clases de modelos generativos. Las GANs, por ejemplo, son especialmente eficaces para crear imágenes o sonidos que parecen reales. Los VAEs permiten explorar y modificar características internas de los datos, lo que abre posibilidades creativas interesantes. Por su parte, los modelos de difusión han destacado recientemente por su capacidad de generar imágenes de altísima calidad, superando en algunos casos la estabilidad y el control de las GANs (Gao et al., 2024).

Actualmente, la inteligencia artificial generativa es considerada una tecnología clave en la transformación digital, con potencial para democratizar el acceso a herramientas creativas, optimizar procesos industriales y potenciar la innovación en múltiples disciplinas. Sin embargo, su adopción también plantea desafíos éticos y sociales, como la veracidad de los contenidos generados, el sesgo algorítmico y la protección de los derechos de autor, lo que requiere marcos regulatorios y reflexión ética sobre los límites y oportunidades de la creación automatizada (Cortés Hernández et al., 2024; Gao et al., 2024).

# 4. MODELOS DE DIFUSIÓN Y ECOSISTEMA TECNOLÓGICO

Los modelos de difusión han transformado la manera en que se generan imágenes con inteligencia artificial, haciendo posible convertir simples descripciones en resultados visuales sorprendentes. En este capítulo se explora cómo funcionan estos modelos, con especial atención a Stable Diffusion y sus distintas versiones. También se presentan herramientas y tecnologías complementarias, como LoRA y ControlNet, que permiten ampliar las posibilidades creativas y técnicas. Todo ello forma parte de un ecosistema cada vez más accesible para artistas, diseñadores y desarrolladores.

## 4.1. Modelos de difusión: teoría y variantes

### 4.1.1. Modelos de Difusión y Stable Diffusion

Los modelos de difusión han surgido como una de las aproximaciones más innovadoras y efectivas en la generación de imágenes mediante inteligencia artificial. Estos modelos, introducidos formalmente por Ho, Jain y Abbeel (2020), se inspiran en procesos termodinámicos, donde una imagen se degrada progresivamente añadiendo ruido gaussiano en múltiples pasos, y luego una red neuronal aprende a invertir este proceso, eliminando el ruido paso a paso hasta reconstruir una imagen coherente y de alta calidad. Esta metodología, denominada Denoising Diffusion Probabilistic Models (DDPM), ha demostrado una capacidad sobresaliente para generar imágenes realistas, superando a arquitecturas previas como las redes generativas antagónicas (GANs) en términos de estabilidad y diversidad de muestras generadas.

La arquitectura de los modelos de difusión se basa en dos fases principales: la fase de difusión, en la que se añade ruido a los datos originales hasta obtener una distribución completamente aleatoria, y la fase de denoising, en la que el modelo, entrenado mediante aprendizaje profundo, aprende a revertir este proceso de manera progresiva. Ho et al. (2020) demostraron que, al optimizar una función de costo basada en una cota variacional y emplear técnicas de score matching, los modelos de difusión pueden alcanzar resultados de síntesis de imágenes comparables o superiores a los mejores modelos GAN de su época, tanto en calidad visual como en métricas objetivas como FID e Inception Score.

Sobre esta base teórica, Rombach et al. (2022) propusieron una extensión denominada Latent Diffusion Models (LDM), que optimiza la eficiencia computacional realizando el proceso de difusión no en el espacio de las imágenes originales, sino en un espacio latente comprimido aprendido por un autoencoder. Esta innovación permitió el desarrollo de **Stable Diffusion**, un modelo capaz de generar imágenes de alta resolución a partir de descripciones textuales de forma rápida y con menores requerimientos de hardware. Stable Diffusion ha democratizado el acceso a la IA generativa avanzada, permitiendo su integración en aplicaciones creativas, científicas e industriales, y facilitando la personalización y el control sobre el contenido generado.

En la actualidad, los modelos de difusión y sus variantes latentes, como Stable Diffusion, han transformado el panorama de la inteligencia artificial generativa. Su capacidad para producir imágenes realistas, diversas y controlables ha impulsado su adopción en campos tan variados como el arte digital, la medicina, la investigación científica y la industria del entretenimiento. Además, la naturaleza modular y abierta de modelos como Stable Diffusion ha fomentado una rápida evolución y adaptación a nuevas tareas, consolidándolos como una herramienta clave en la automatización creativa y la innovación tecnológica.

### 4.1.2. Arquitectura Modular en Modelos de Difusión

Los modelos de difusión modernos, como Stable Diffusion, están diseñados con una arquitectura modular que permite flexibilidad y personalización en la generación de imágenes. Los componentes principales incluyen:

- **Tokenizer:** Convierte el texto del prompt en tokens numéricos, que son la entrada para el modelo de lenguaje (Hugging Face, s.f.).
- **Text Encoder:** Transforma los tokens en vectores de embeddings que capturan el significado semántico del texto, utilizando modelos como CLIP (Hugging Face, s.f.; OpenVINO™ Documentation, 2023a).
- **UNet:** Red neuronal convolucional que realiza el proceso de denoising, eliminando el ruido de las representaciones latentes para generar imágenes coherentes (Hugging Face, s.f.).
- **VAE (Variational Autoencoder):** Transforma las imágenes a un formato comprimido conocido como espacio latente, y luego las reconstruye a partir de esa representación. Esto hace que el proceso de generación sea más eficiente y que se puedan controlar mejor ciertos aspectos visuales (Hugging Face, s.f.; OpenVINO™ Documentation, 2023a).

Esta arquitectura modular permite ajustar y reemplazar componentes individuales para modificar el comportamiento del modelo sin necesidad de reentrenar todo el sistema (Hugging Face, s.f.).

### 4.1.3. Parámetros de Control

Los modelos de difusión permiten modificar varios parámetros que afectan directamente el resultado de la imagen, ofreciendo al usuario más control sobre el proceso creativo.

- **Steps (número de pasos):** Determina cuántas iteraciones de eliminación de ruido o denoising se realizan. Más pasos suelen mejorar la calidad de la imagen pero aumentan el tiempo de cómputo (Hugging Face, s.f.; OpenVINO™ Documentation, 2023b).

- **Sampler (scheduler):** Son algoritmos numéricos que determinan cómo se elimina el ruido en cada paso del proceso de generación de imágenes. Estos algoritmos implementan un método numérico distinto para resolver el proceso de eliminación de ruido, lo que afecta la velocidad y la calidad de la imagen generada.

  Entre los más utilizados se encuentran DDIM (Denoising Diffusion Implicit Models), Euler, LMS (Linear Multistep), PLMS (Pseudo Linear Multistep), DPM++, PNDM y UniPC. La elección del sampler puede influir en el resultado final y depende de las preferencias y necesidades del usuario (Hugging Face, s.f.; Aiarty, 2024).

- **CFG Scale (Classifier-Free Guidance Scale):** Controla la adherencia de la imagen generada al prompt textual. Si se configuran valores más altos se refuerza la fidelidad al texto, mientras que valores más bajos permiten mayor creatividad (Hugging Face, s.f.).

- **Seed (semilla):** Es un valor numérico que define el punto de partida del ruido aleatorio en la generación de imágenes. Al reutilizar la misma semilla junto con los mismos parámetros y el mismo prompt, es posible obtener exactamente la misma imagen cada vez. Esto permite repetir resultados de forma precisa y controlar las variaciones que pueden surgir en el proceso (Hugging Face, s.f.; OpenVINO™ Documentation, 2023b).

### 4.1.4. Stable Diffusion 1.5

Stable Diffusion 1.5 (SD 1.5) es un modelo de difusión latente de código abierto diseñado para generar imágenes realistas a partir de descripciones textuales. Desarrollado por Stability AI y la comunidad académica (Rombach et al., 2022), representa una evolución significativa en la generación de imágenes mediante inteligencia artificial, equilibrando accesibilidad computacional y calidad visual.

SD 1.5 democratizó el acceso a la generación de imágenes mediante IA, sirviendo como base para investigaciones en mejora de resolución, control composicional y mitigación de sesgos. Su arquitectura modular y eficiente ha inspirado desarrollos posteriores, como Stable Diffusion XL (SDXL), que amplían sus capacidades (Stability AI, 2022).

A continuación, se detallan sus aspectos clave:

**Arquitectura y Funcionamiento**

SD 1.5 se basa en tres componentes principales:

**Autoencoder (VAE):**

- El VAE se encarga de transformar imágenes de alta resolución como las de 512×512 píxeles en una representación latente mucho más compacta

generalmente de 64×64×4, lo que permite disminuir significativamente la cantidad de datos a manejar y la carga computacional (Rombach et al., 2022; Stability AI, 2022).

- En la fase de generación, el VAE decodifica la representación latente final y la transforma de nuevo en una imagen en el espacio de píxeles, asegurando que los detalles visuales sean preservados en la reconstrucción (Rombach et al., 2022; Stability AI, 2022).

**U-Net:**

- U-Net es la red neuronal principal encargada de llevar a cabo el proceso iterativo de eliminación de ruido, también llamado denoising, dentro del espacio latente.
- Opera mediante una arquitectura de codificador-decodificador con conexiones de salto, lo que le permite captar tanto información global como detalles locales en la imagen latente.
- Utiliza mecanismos de atención cruzada (cross-attention) para integrar la información textual proveniente del codificador de texto, guiando así el proceso de generación para que la imagen resultante sea coherente con el prompt proporcionado (Rombach et al., 2022; Stability AI, 2022)

**Codificador de texto (CLIP ViT-L/14):**

- El codificador de texto, basado en la arquitectura CLIP, transforma el prompt textual en una representación semántica numérica (embedding).
- Estos embeddings se utilizan para condicionar el proceso de denoising, permitiendo que la U-Net genere imágenes que reflejen fielmente la descripción textual dada (Rombach et al., 2022; Stability AI, 2022).

## Entrenamiento

Según Rombach et al. (2022), el entrenamiento de Stable Diffusion se sustenta en una función de pérdida orientada a la reconstrucción. En este proceso, la red U-Net aprende a estimar el ruido que se ha incorporado a la representación latente de la imagen en cada etapa del proceso de difusión, y el modelo se optimiza minimizando la diferencia entre el ruido real añadido y el ruido que el modelo predijo que se mide con una función llamada error cuadrático medio. Este enfoque permite que el modelo aprenda a revertir el proceso de adición de ruido y a generar imágenes coherentes a partir de ruido aleatorio, guiado por información textual (Rombach et al., 2022).

## Proceso de generación

En el artículo de Rombach et al. (2022), se describe que en cada paso del proceso de generación, el modelo predice el ruido presente en la representación latente y utiliza un *scheduler* (programador de pasos) para actualizar su estado, acercándolo progresivamente a la imagen final. Este proceso se basa en la predicción del ruido

realizada por la red U-Net y en la aplicación de una función de actualización que varía según el *scheduler* elegido, como DDIM o PLMS, entre otros. El modelo opera mediante un proceso iterativo de denoising, en el que parte de ruido aleatorio y lo va refinando paso a paso hasta obtener una imagen coherente con el texto de entrada. De este modo, el modelo aprende a reconstruir imágenes guiado por la información textual proporcionada como condición.

En la Figura 4 se representa la arquitectura de un Modelo de Difusión Latente (LDM), base de Stable Diffusion 1.5. La imagen original se codifica en un espacio latente comprimido, donde se aplica el proceso de difusión para eliminar progresivamente el ruido. Una red U-Net, guiada por mecanismos de atención cruzada, permite condicionar la generación con texto u otras entradas. Finalmente, la imagen se decodifica nuevamente al espacio de píxeles y es así, que esta arquitectura permite generar imágenes de alta calidad de manera más eficiente.

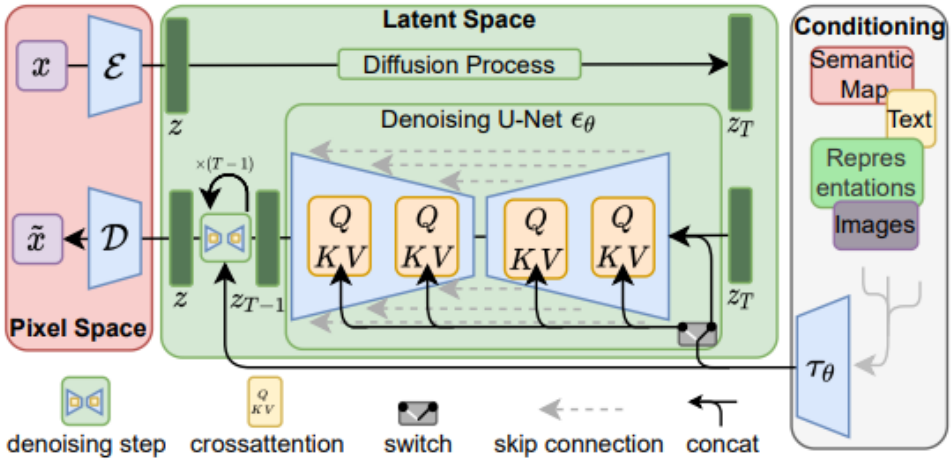



**Figura 4.** Arquitectura general de un Modelo de Difusión Latente (LDM), que opera en un espacio latente comprimido y utiliza atención cruzada para condicionar la generación de imágenes a partir de texto u otras representaciones. Fuente: Rombach et al. (2022)

### Mejoras en Stable Diffusion 1.5

SD 1.5 introduce avances respecto a versiones anteriores:

**Entrenamiento en alta resolución:**

SD 1.5 fue ajustado (fine-tuned) durante 595,000 pasos adicionales a una resolución de 512×512 píxeles, utilizando el dataset LAION-Aesthetics v2 5+. Este conjunto de datos fue filtrado específicamente para priorizar imágenes estéticamente atractivas y con baja probabilidad de contener marcas de agua, lo que contribuye a una mayor calidad visual en las imágenes generadas.

**Dropout de condicionamiento textual:**

Durante el entrenamiento, se omitió intencionadamente el 10% de los textos de entrada (text-conditioning dropout). Esta técnica mejora la capacidad del modelo para generar imágenes incluso cuando el prompt textual no está presente, facilitando el uso de classifier-free guidance y permitiendo un mayor control sobre la generación de imágenes.

**Eficiencia computacional:**

Al operar en el espacio latente, SD 1.5 reduce significativamente el consumo de memoria y los requisitos computacionales, lo que permite su ejecución en hardware convencional, como GPUs con 4 GB de VRAM. Esta eficiencia es una de las razones principales por las que Stable Diffusion se ha popularizado para aplicaciones personales y de investigación.

**Limitaciones**

Stable Diffusion 1.5 presenta diversas limitaciones inherentes a su diseño y datos de entrenamiento. El modelo no logra fotorealismo perfecto, tiene dificultades para generar texto legible en las imágenes y puede fallar en tareas de composición compleja o en la representación precisa de rostros y personas.

Además, su rendimiento es óptimo principalmente con descripciones en inglés, y el proceso de compresión latente implica cierta pérdida de información visual. Finalmente, el modelo refleja sesgos culturales y lingüísticos presentes en su dataset, lo que puede limitar la diversidad de los resultados (Stability AI, 2022).

### 4.1.5. Stable Diffusion XL (SDXL)

SDXL (Stable Diffusion XL) desarrollado por la empresa Stability AI, es una evolución de los modelos de difusión latente, y fue desarrollada con el objetivo de generar imágenes de alta resolución de hasta 1024×1024 píxeles. Esta capacidad se logra gracias a una arquitectura optimizada y al uso de técnicas de condicionamiento innovadoras. Es un modelo de código abierto, accesible para la comunidad a través de plataformas como Hugging Face.

En su artículo publicado en 2023 en donde se presenta SDXL, Podell et al. explican cómo el modelo fue creado y cómo esta nueva versión incorpora cinco innovaciones clave que le permiten crear imágenes más realistas, detalladas y versátiles que las versiones anteriores.

## Mejoras en Stable Diffusion XL

### Arquitectura y Funcionamiento

A diferencia de versiones anteriores de Stable Diffusion, SDXL utiliza una versión de la arquitectura U-Net con una capacidad de procesamiento tres veces mayor, lo que le permite procesar información visual de forma más avanzada y detallada.

Para lograr una comprensión más precisa del texto, SDXL integra dos modelos distintos: OpenCLIP ViT-bigG, que destaca por captar matices textuales más sutiles, y CLIP ViT-L, enfocado en interpretar el significado general del contenido. Esta combinación le permite interpretar las descripciones de forma más profunda y precisa, enriqueciendo así el resultado final.

Otra gran novedad es su enfoque en dos etapas: primero se genera una imagen inicial basada en el texto, y luego un modelo refinador la mejora, corrige imperfecciones y añade más detalle. Gracias a esta estrategia, SDXL puede producir imágenes muy nítidas y coherentes, incluso en resoluciones de hasta 1024×1024 píxeles, superando ampliamente las limitaciones técnicas del pasado (Podell et al., 2023).

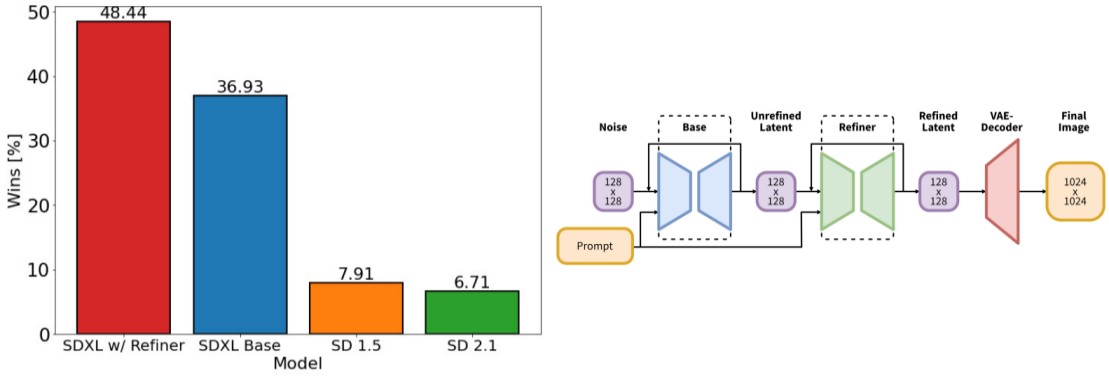



**Figura 5.** Comparación entre las preferencias de usuario para SDXL y versiones anteriores (Stable Diffusion 1.5 y 2.1). A la derecha, visualización del pipeline en dos etapas utilizado por SDXL. Fuente: Podell et al. (2023)

### Micro-Conditioning

Uno de los problemas clásicos en los modelos anteriores era su dificultad para ajustarse a distintos tamaños o composiciones de imagen. También ocurrían recortes indeseados en los objetos generados, producto de los recortes aleatorios durante el entrenamiento.

Para resolver estos problemas, SDXL introduce dos innovaciones clave:

1. **Size-Conditioning (Condicionamiento por tamaño):**

   Consiste en informar al modelo del tamaño original de cada imagen durante el entrenamiento. Cada valor es codificado por separado usando funciones de Fourier. Luego, estas codificaciones se concatenan en un único vector y este vector se suma a la codificación del tiempo de difusión (timestep embedding) en

el modelo. Así, el modelo aprende a generar imágenes en distintas resoluciones sin perder calidad.

Los resultados mostraron que esta técnica mejora tanto la calidad como la variedad de las imágenes generadas. (Podell et al., 2023).

2. **Crop-Conditioning (Condicionamiento por recorte):**

Consiste en indicar al modelo desde dónde se recortó la imagen original. Durante la generación, se puede fijar el recorte en (0,0) para centrar el objeto principal y evitar recortes artificiales, como cabezas cortadas o elementos incompletos. Esto permite generar imágenes más completas y centradas. (Podell et al., 2023).

SDXL también recibe información sobre el punto exacto desde el cual se recortó cada imagen (coordenadas de la esquina superior izquierda). Esto reduce los efectos negativos de los recortes aleatorios utilizados como aumento de datos, permitiendo mayor control sobre la posición de los elementos visuales. Durante la generación, se puede fijar el recorte en (0,0) para centrar el objeto principal y evitar recortes artificiales, como cabezas cortadas o elementos incompletos (Podell et al., 2023).

**Entrenamiento Multi-Aspecto**

A diferencia de versiones anteriores que solo trabajaban con imágenes cuadradas, SDXL fue entrenado con imágenes de distintas formas: cuadradas, verticales, horizontales, etc. Durante el entrenamiento, las imágenes se agrupan por su proporción, lo que le permite al modelo aprender a generar imágenes en todos estos formatos.

Este enfoque no solo amplía su utilidad, desde ilustraciones hasta portadas o banners, sino que también evita esas distorsiones incómodas que solían aparecer cuando los modelos anteriores intentaban forzar imágenes en formatos que no conocían (Podell et al., 2023).

**Autoencoder Mejorado**

El autoencoder es la parte del modelo que comprime y luego reconstruye las imágenes para facilitar el proceso de generación. En SDXL, este componente de tipo VAE: Variational Autoencoder ha sido rediseñado desde cero y reentrenado con técnicas más avanzadas y una mayor cantidad de datos.

El resultado son imágenes con más nitidez, texturas mejor conservadas y una fidelidad visual mucho mayor. Aunque pasen por un proceso de compresión, las imágenes mantienen su calidad y responden mejor a lo que el usuario tenía en mente (Podell et al., 2023).

### Integración de Componentes

SDXL integra los elementos anteriores en un flujo de generación dividido en dos fases.

1. **Generación Inicial:** El modelo base produce una imagen latente a partir del texto, utilizando el micro-conditioning para interpretar con precisión cada componente del prompt. La imagen generada contiene la estructura general, colores base y disposición de elementos.

2. **Refinamiento:** Una de las principales innovaciones de SDXL es la incorporación de un pipeline de refinamiento post-hoc, que utiliza un modelo especializado para mejorar la fidelidad visual de las imágenes generadas. Para ello, se añade una cantidad controlada de ruido gaussiano a la imagen latente generada, lo que permite que el modelo refinador tenga margen para mejorar detalles sin alterar la composición general. El modelo refinador es un U-Net optimizado específicamente para aumentar la calidad visual en regiones de alta frecuencia espacial, enfocándose en bordes, texturas y detalles finos. Este refinador fue entrenado para corregir artefactos y mejorar la coherencia visual local, particularmente en áreas complejas como rostros, manos o fondos (Podell et al., 2023).

En la Figura 6. se observa una comparación entre los resultados obtenidos con el pipeline de una sola etapa de SDXL y aquellos generados con el pipeline de dos etapas que incluye el modelo de refinamiento. Las diferencias se evidencian especialmente en los detalles ampliados, donde el modelo refinado presenta mayor nitidez, coherencia y calidad. Esto reafirma la utilidad del modelo de refinamiento como parte clave de la arquitectura de SDXL.

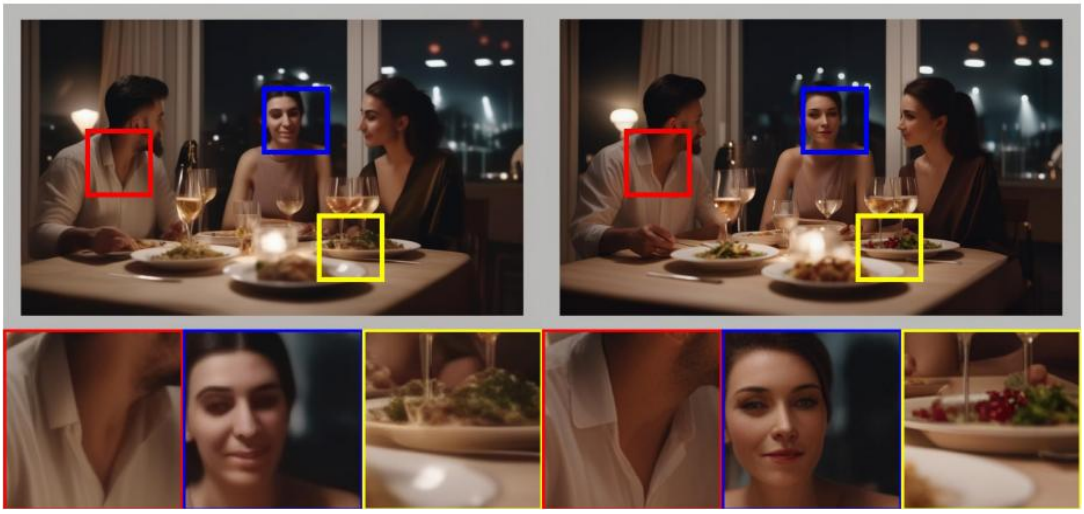


**Figura 6.** Comparación adicional entre el pipeline de una y dos etapas de SDXL. A la izquierda, imágenes generadas solo con SDXL; a la derecha, las mismas imágenes mejoradas mediante el modelo de refinamiento. Fuente: Podell et al. (2023).

Esta arquitectura modular permite que SDXL produzca imágenes con alta fidelidad al prompt, manteniendo coherencia estructural y mejorando significativamente el realismo y la riqueza de detalles respecto a versiones anteriores. A continuación, muestra una comparación directa entre SDXL y versiones anteriores de Stable Diffusion (1.5 y 2.1), generando tres muestras aleatorias por modelo a partir de los mismos prompts y configuraciones de muestreo. Esta comparación permite observar de forma clara las mejoras visuales de SDXL, tanto en calidad de imagen como en coherencia de composición.


**Figura 7** Comparación de los resultados generados por SDXL y versiones anteriores de Stable Diffusion. Fuente: Podell et al. (2023).

### Limitaciones

A pesar de los avances introducidos por SDXL en la generación de imágenes de alta resolución y mayor fidelidad visual, el modelo sigue presentando limitaciones.

Entre ellas destacan las dificultades para interpretar prompts complejos o con múltiples objetos, la incapacidad para generar texto legible dentro de las imágenes y la persistencia de ciertos sesgos culturales y lingüísticos derivados de los datos de entrenamiento.

Además, la generación de detalles anatómicos precisos y la representación de rostros humanos pueden seguir mostrando errores, y el uso de SDXL requiere mayores recursos computacionales respecto a versiones anteriores (Podell et al., 2023).

## 4.1.6. Stable Diffusion XL Turbo (SDXL Turbo)

SDXL Turbo es un modelo generativo de texto a imagen desarrollado por Stability AI y presentado por Sauer et al. (2023). Se basa en la arquitectura SDXL 1.0, pero presenta

la innovación principal un método de entrenamiento llamado Adversarial Diffusion Distillation (ADD). Este método permite sintetizar imágenes de alta calidad en tan solo uno a cuatro pasos de inferencia, una mejora significativa respecto a los modelos de difusión tradicionales, que requieren decenas o cientos de pasos para lograr resultados comparables.

**Arquitectura y Funcionamiento**

La base técnica de SDXL Turbo se sustenta en el método conocido como Adversarial Diffusion Distillation (ADD), el cual fusiona dos enfoques clave: la destilación de modelos de difusión y el aprendizaje adversarial. Esta combinación permite un proceso de entrenamiento optimizado que prioriza la generación de imágenes con alta calidad visual (Sauer et al., 2023).

Durante el entrenamiento, intervienen tres componentes principales:

- El modelo maestro (teacher): un modelo de difusión avanzado con pesos congelados, cuya función es proporcionar predicciones precisas como referencia.
- El modelo estudiante (student): basado en una red U-Net preentrenada, aprende a generar imágenes rápidamente imitando al maestro.
- El discriminador: una red entrenable que evalúa si las imágenes generadas son realistas, incentivando al estudiante a mejorar sus resultados.

En conjunto, el estudiante aprende a generar imágenes partiendo de ruido, guiado tanto por las predicciones del maestro como por las evaluaciones del discriminador. Gracias a este funcionamiento, SDXL Turbo permite generar imágenes coherentes y visualmente detalladas en solo 1 a 4 pasos de inferencia, superando ampliamente a anteriores modelos en velocidad y sin sacrificar calidad (Sauer et al., 2023; Stability AI, 2023).

**Entrenamiento**

El entrenamiento de SDXL Turbo se basa en dos tipos de funciones de pérdida complementarias:

- **Pérdida de destilación de puntuaciones (Score Distillation Loss):**
  Permite transferir el conocimiento del modelo maestro al estudiante. Las predicciones del maestro, creadas durante el proceso de denoising tambien conocido como eliminación de ruido, son una referencia directa para el modelo estudiante. Esta técnica permite que el estudiante adquiera la capacidad de generar imágenes con estructura y coherencia semántica similares al modelo maestro, pero en muchos menos pasos (Sauer et al., 2023).
- **Pérdida adversarial (Adversarial Loss):**
  El discriminador, que evalúa si una imagen parece real o generada, está condicionado tanto por el texto como por las características visuales. Esto

incentiva al modelo estudiante a producir imágenes que no solo reflejen adecuadamente el texto, sino que además sean indistinguibles de una imagen real (Sauer et al., 2023).

En la Figura 8. observamos que el proceso de entrenamiento de SDXL Turbo combina un modelo estudiante, un modelo maestro congelado y un discriminador. Se observa que el estudiante aprende a reconstruir imágenes a partir de versiones ruidosas, guiado por la comparación con las salidas del maestro y la evaluación del discriminador, optimizando simultáneamente una pérdida de destilación y una pérdida adversarial.

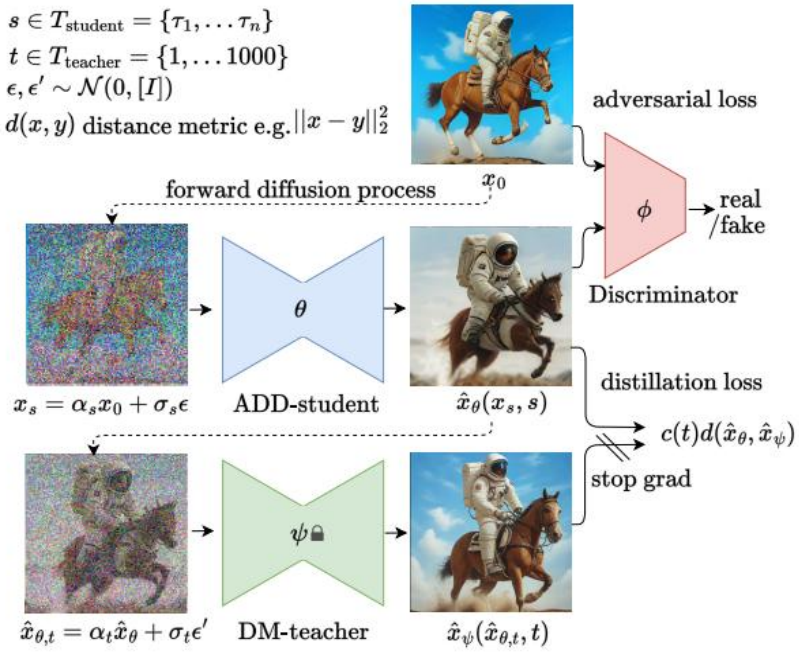



**Figura 8.** Arquitectura del entrenamiento en SDXL Turbo mediante Adversarial Diffusion Distillation (ADD). Fuente: Sauer et al. (2023).

## Proceso de Generación

Una vez entrenado, SDXL Turbo se destaca por su eficiencia en el momento de generación o inferencia. Gracias al uso del método ADD, el modelo es capaz de generar imágenes en tan solo 1 a 4 pasos, lo que representa una mejora radical frente a los modelos de difusión anteriores, que requerían decenas de pasos para obtener resultados similares (Sauer et al., 2023).

Durante la inferencia:

- Se parte de una representación latente, generada a partir del texto y ruido inicial.
- El modelo utiliza un proceso de denoising muy reducido, ajustado con precisión durante el entrenamiento.

- En un solo paso (o unos pocos), se obtiene una imagen final que mantiene una alta calidad visual, detalle y coherencia.

Este enfoque permite aplicaciones de generación en tiempo real, como interfaces interactivas, herramientas creativas y animación, que antes eran inviables con modelos de difusión estándar (Sauer et al., 2023; Stability AI, 2023).

**Mejoras en Stable Diffusion XL Turbo**

La fusión de la destilación de puntuaciones (score distillation) y la pérdida adversarial es la clave del avance de SDXL Turbo. Este enfoque permite que el modelo sintetice imágenes fotorrealistas a partir de texto en solo un paso, manteniendo la calidad y la fidelidad al prompt.

Además, la arquitectura permite condicionar tanto en texto como en imagen, facilitando tareas de texto a imagen y de imagen a imagen. Estudios de preferencia con usuarios muestran que SDXL Turbo supera a otros modelos de generación rápida en calidad visual y alineación con el texto, especialmente cuando se utilizan varios pasos de muestreo (Sauer et al., 2023).

En la Figura 9. podemos observar que aunque ADD-XL realiza la generación en tan solo 4 pasos, en contraste con los 50 pasos de SDXL Base, mantiene una alta calidad visual, incluso superando en ciertos aspectos el realismo de su modelo maestro.

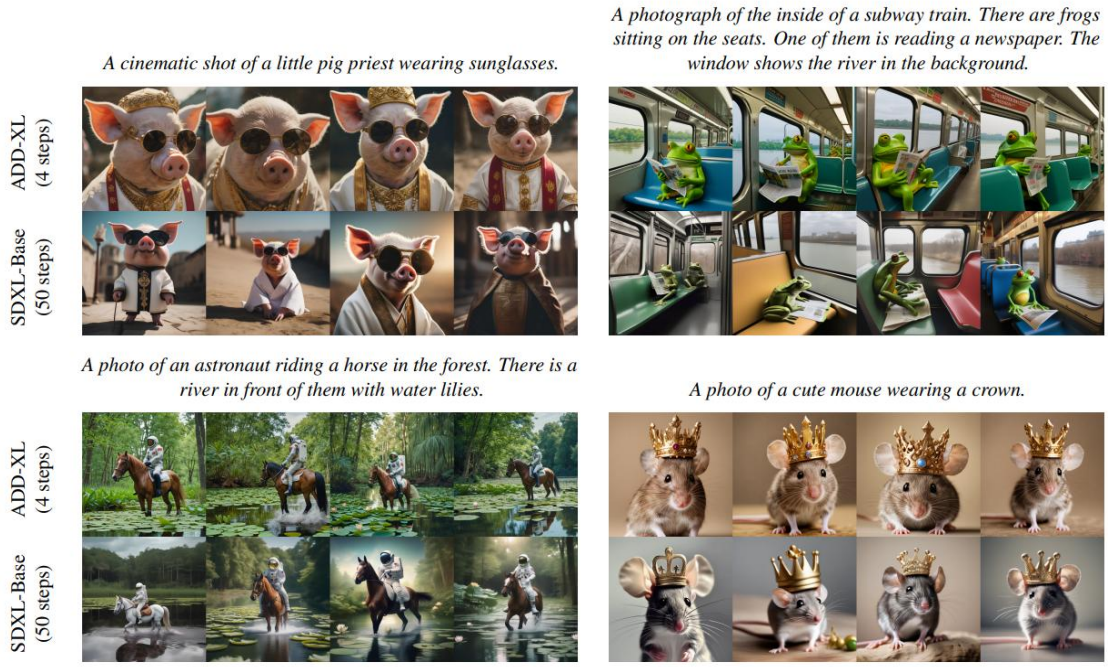



**Figura 9.** Comparación cualitativa entre SDXL Base (modelo maestro) y ADD-XL (modelo estudiante). Fuente: Sauer et al. (2023).

**Limitaciones**

A pesar de su eficiencia, SDXL Turbo presenta limitaciones. El modelo está optimizado para una resolución fija de 512×512 píxeles y puede tener dificultades para generar texto legible o detalles anatómicos precisos en las imágenes.

Además, no utiliza parámetros de control como guidance scale o negative prompts durante la generación, y la compresión latente puede provocar pérdida de detalles visuales. Por último, como otros modelos generativos, puede reflejar sesgos presentes en los datos de entrenamiento y no está diseñado para tareas que requieran precisión factual o representación fiel de personas reales (Stability AI, 2023).

## 4.2. Herramientas, tecnologías y frameworks complementarios

### 4.2.1. Flux y Flux.1 Kontext

FLUX es una familia de modelos generativos desarrollada por Black Forest Labs, diseñada para unificar la generación y edición de imágenes en un solo sistema. A diferencia de modelos anteriores que requerían herramientas separadas para diferentes tareas visuales, FLUX permite tanto la creación de imágenes desde texto como la edición iterativa de imágenes existentes mediante instrucciones en lenguaje natural (Black Forest Labs, 2024a). Su variante más avanzada que se anunció en mayo 2025, FLUX.1 Kontext, representa un salto cualitativo en la consistencia visual y la velocidad de generación, superando limitaciones de modelos como DALL·E 3 y Stable Diffusion 3 (Black Forest Labs, 2025a).

**Arquitectura y Funcionamiento**

La arquitectura de FLUX se fundamenta en tres componentes técnicos principales:

- **Autoencoder convolucional:**

  El autoencoder de FLUX comprime las imágenes en un espacio latente de 16 canales logrando reconstrucciones de imagen con mayor fidelidad visual que otros modelos populares y mejorando significativamente la reconstrucción visual ya que las imágenes generadas se ven más nítidas y detalladas, acercándose mejor a su versión original. (Black Forest Labs, 2024b).

- **Transformer de difusión con flow matching:**

  A diferencia de los modelos de difusión tradicionales que predicen ruido, FLUX utiliza flow matching, una técnica que aprende directamente la trayectoria óptima desde el ruido hasta la imagen final. Esto permite una convergencia más rápida y estable durante el entrenamiento (Black Forest Labs, 2024a).

- **Bloques de atención híbridos:**

FLUX implementa una combinación de bloques de atención "doble stream los cuales procesan simultáneamente tokens de imagen y texto y bloques "single stream" que refinan exclusivamente la representación visual. Esta estructura se optimiza mediante embeddings posicionales rotatorios tridimensionales, que codifican tanto la posición espacial como la temporal de cada token (Black Forest Labs, 2024b).

Cada bloque de atención incluye capas fusionadas que mejoran la eficiencia computacional, permitiendo escalar el modelo hasta 12 mil millones de parámetros sin comprometer el rendimiento en hardware convencional (Black Forest Labs, 2024b).

**Entrenamiento**

El proceso de entrenamiento de FLUX.1 Kontext combina varias estrategias innovadoras:

- **Concatenación de secuencias:**

  En FLUX.1 Kontext la imagen de referencia y la imagen que se quiere generar se convierten en tokens latentes, piezas de información compacta que representan el contenido visual de cada imagen, mediante un codificador automático.

  Lo innovador del modelo consiste en unir en una misma secuencia los tokens de la imagen de entrada, los de la imagen objetivo y la instrucción textual correspondiente. Esta estructura permite al modelo aprender a realizar dos tipos de tareas con un solo sistema porque por un lado puede editar una imagen existente siguiendo una instrucción y por otro también puede generar una imagen completamente nueva cuando solo se proporciona texto. Además, la forma en que se organizan estos tokens permite trabajar con imágenes de diferentes tamaños y proporciones sin necesidad de cambiar la arquitectura del modelo (Black Forest Labs, 2025a).

- **Objetivo de flow matching rectificado:**

  El entrenamiento de FLUX.1 Kontext se basa en una técnica llamada flow matching, que puede entenderse como enseñar al modelo a “navegar” desde una imagen llena de ruido hasta una imagen clara y coherente, siguiendo la mejor ruta posible.

  Para lograr esto, el modelo aprende a predecir en cada paso la velocidad y dirección en la que debe transformar el ruido para acercarse a la imagen objetivo y esto se expresa como una función matemática de pérdida que mide qué tan bien el modelo predice ese camino óptimo entre el ruido y la imagen final (Black Forest Labs, 2025a).

- **Destilación adversarial (LADD):**

FLUX.1 Kontext resuelve el problema de lentitud de modelos de difusion tradicionales que requieren muchos pasos para generar imágenes de gran calidad, al usar una técnica llamada Latent Adversarial Diffusion Distillation (LADD).

En este proceso, el modelo se entrena adicionalmente con un crítico o un discriminador que evalúa si las imágenes generadas son suficientemente realistas, este modelo aprende a engañar a este crítico y así produce imágenes de alta calidad en muchos menos pasos y en solo 3 a 5 segundos, lo que permite usarse en varias herramientas y aplicaciones que requieren una mayor velocidad de generación (Black Forest Labs, 2025a).

En la Figura 10. se muestra cómo se procesan la imagen de entrada, la imagen de contexto y el texto mediante codificadores automáticos y flujos paralelos de texto e imagen, que luego se combinan para generar la imagen final editada o creada (Black Forest Labs, 2025a).

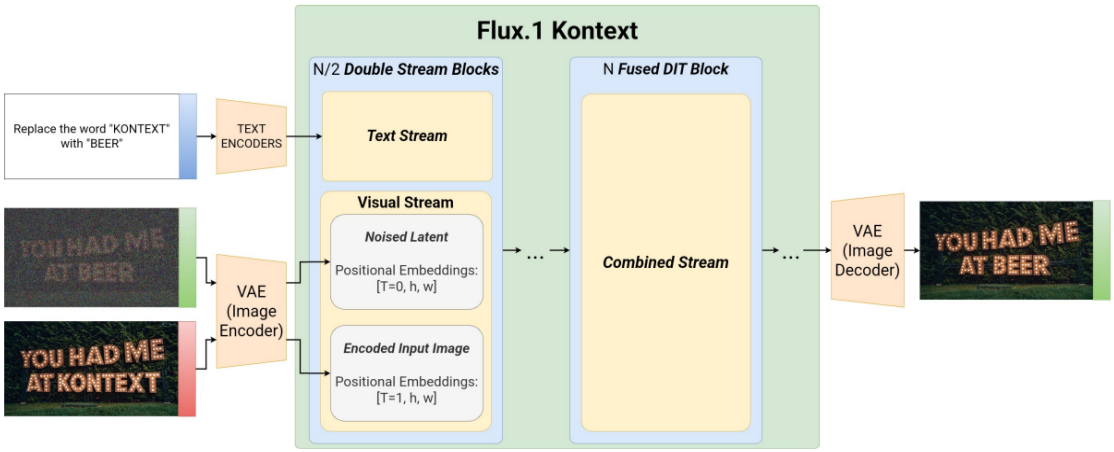



**Figura 10.** Vista general de alto nivel de FLUX.1 Kontext. Fuente: Black Forest Labs, (2025a)

## Mejoras de Flux.1 Kontext

FLUX.1 Kontext expande las capacidades del modelo base mediante un enfoque de generación en contexto, que permite modificar imágenes existentes o crear nuevas manteniendo la coherencia visual a lo largo de múltiples iteraciones. Entre sus principales innovaciones se encuentran:

- **Transferencia desde texto a imagen:**

  Durante su fase de entrenamiento, FLUX.1 Kontext se construye sobre un modelo previamente capacitado para la generación de imágenes a partir de texto, lo que le permite conservar la facultad de producir composiciones visuales completamente nuevas basadas únicamente en descripciones escritas, además de editar imágenes existentes (Black Forest Labs, 2025a).

- **Consistencia de personajes y objetos:**

  Una de las contribuciones más significativas de FLUX.1 Kontext es su capacidad para mantener la identidad visual de personajes y objetos a través de múltiples ediciones. Como se muestra en la Figura 1 de la documentación, el modelo puede transformar un personaje (por ejemplo, un pájaro) a través de diferentes escenarios (un bar, un cine, un supermercado) manteniendo sus características distintivas, lo que facilita aplicaciones como la creación de storyboards y narrativas visuales (Black Forest Labs, 2025a).

- **Edición local y global:**

  El modelo unifica dos capacidades tradicionalmente separadas: la edición local (modificar elementos específicos manteniendo el contexto) y la generación completa en contexto (crear nuevas escenas basadas en una referencia). Esto permite flujos de trabajo donde, por ejemplo, se puede eliminar un objeto, cambiar la ubicación de la escena y modificar las condiciones climáticas en pasos sucesivos, como se ilustra en la Figura 2 de la documentación (Black Forest Labs, 2025a).

Como se observa en la Figura 11. del artículo técnico, FLUX.1 Kontext permite editar imágenes de forma iterativa mediante instrucciones en lenguaje natural, manteniendo la coherencia de personajes, pose y estilo a lo largo de múltiples modificaciones (Black Forest Labs, 2025a)


**Figura 11.** Edición iterativa guiada por instrucciones en FLUX.1 Kontext. Fuente: Black Forest Labs, (2025a)

**Limitaciones**

A pesar de sus avances en generación de imágenes de calidad y con mayor velocidad, FLUX.1 Kontext presenta algunas limitaciones importantes.

Entre ellas la calidad visual puede degradarse tras múltiples ediciones sucesivas, generando artefactos que requieren reiniciar el proceso. Además, su comprensión contextual es limitada en tareas que demandan conocimientos especializados, como escenas históricas o detalles anatómicos precisos. La precisión y coherencia dependen en gran medida de la claridad del prompt lo que dificulta su uso para usuarios sin experiencia en ingeniería de prompts (Black Forest Labs, 2024b).

Por último, al igual que otros modelos generativos Flux puede amplificar sesgos presentes en los datos de entrenamiento y no está diseñado para ofrecer información factual (Black Forest Labs, 2025b).

### 4.2.2. OpenAI

Es una entidad de investigación en inteligencia artificial que lidera el desarrollo de modelos avanzados, con énfasis en la seguridad, alineación ética y utilidad social de sus sistemas. Su investigación abarca desde el procesamiento de texto hasta la generación de imágenes, audio y video, con un enfoque especial en modelos multimodales capaces de integrar diferentes tipos de información (OpenAI, 2025a; OpenAI, 2025b).

Cabe señalar que la información utilizada para este apartado se obtuvo de fuentes oficiales de OpenAI, como su blog de investigación y documentos técnicos publicados públicamente (OpenAI, 2024). Estas fuentes no ofrecen detalles completos sobre la arquitectura interna o los procesos específicos de entrenamiento, por lo que el análisis se limita a un enfoque descriptivo de las capacidades, aplicaciones y limitaciones que han sido comunicadas abiertamente.

**Arquitectura y Funcionamiento**

OpenAI ha desarrollado modelos como DALL·E, CLIP y la serie GPT incluyendo GPT-4o y o3, que pueden comprender y generar contenido en múltiples modalidades, como texto, imagen y audio. Estos modelos utilizan técnicas de aprendizaje profundo y grandes volúmenes de datos para alcanzar capacidades avanzadas de razonamiento, generación creativa y análisis de patrones (OpenAI, 2025b).

En el ámbito visual, OpenAI ha presentado modelos capaces de generar imágenes precisas y fotorrealistas a partir de descripciones textuales, así como transformar imágenes de entrada siguiendo instrucciones naturales (OpenAI, 2025b).

**Entrenamiento**

El entrenamiento de sus modelos se basa en el uso de grandes conjuntos de datos y técnicas de aprendizaje profundo, así como en la incorporación de retroalimentación

humana para mejorar la alineación de los modelos con las expectativas y valores humanos. OpenAI reconoce la importancia de la alineación y la seguridad como desafíos centrales en el desarrollo de IA general (OpenAI, 2025b).

**Mejoras de OpenAI**

Los modelos multimodales de OpenAI destacan por su capacidad para generar imágenes de alta calidad, mantener coherencia entre texto e imagen y ofrecer respuestas relevantes y seguras en tareas de comprensión y generación creativa. Estas capacidades han permitido aplicaciones en áreas como la creación de contenido visual, la asistencia conversacional y la investigación científica (OpenAI, 2025a; OpenAI, 2025b).

**Limitaciones**

OpenAI reconoce que si bien sus modelos han logrado avances significativos aún enfrentan limitaciones en la comprensión contextual profunda y pueden reproducir sesgos derivados de los datos con los que fueron entrenados. Por ello, la alineación con valores humanos y la seguridad en el uso de la IA continúan siendo áreas prioritarias de investigación, con el objetivo de reducir riesgos y brindar un impacto positivo en la sociedad (OpenAI, 2025b).

### 4.2.3. LoRA (Low-Rank Adaptation)

Low-Rank Adaptation (LoRA) es una técnica innovadora que permite adaptar modelos de lenguaje grandes a tareas concretas sin tener que cambiar todos sus parámetros. Fue propuesta por Hu et al. (2021) y su valor está en insertar pequeñas matrices entrenables en ciertas capas del modelo dejando el resto sin ninguna modificación lo que permite ahorrar recursos computacionales y memoria haciendo el proceso mucho más eficiente. La Figura 12 muestra cómo LoRA introduce matrices de bajo rango (A y B) en las capas del modelo Transformer, manteniendo los pesos preentrenados congelados y adaptando el modelo a nuevas tareas de manera eficiente.

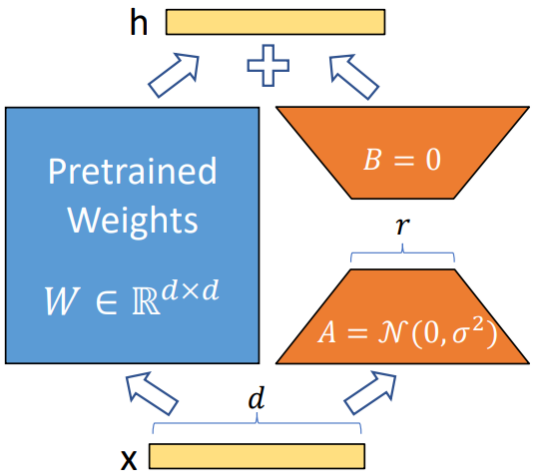



**Figura 12.** Esquema de LoRA: adaptación de rango bajo en modelos de lenguaje. Fuente: Hu et al. (2021)

La idea detrás de LoRA es que las adaptaciones necesarias suelen concentrarse en un espacio de baja dimensión, por lo que basta con actualizar solo una pequeña parte del modelo. En lugar de modificar directamente grandes matrices, se entrena la suma de dos matrices pequeñas, lo que reduce drásticamente la cantidad de parámetros necesarios hasta 10,000 veces menos y disminuye el uso de GPU sin sacrificar rendimiento (Hu et al., 2021).

Otras ventajas es que un mismo modelo preentrenado puede reutilizarse para múltiples tareas simplemente intercambiando los módulos LoRA correspondientes, lo que reduce el almacenamiento y facilita el cambio entre tareas. También disminuye la barrera técnica de entrada ya que al enfocarse solo en un pequeño conjunto de parámetros no es necesario mantener estados para todo el modelo (Hu et al., 2021).

Además gracias a su diseño, los pesos ajustados con LoRA pueden fusionarse con los pesos originales al momento del despliegue, sin añadir latencia en la inferencia. Esto lo hace comparable en velocidad a un modelo completamente ajustado, pero con un uso mucho más eficiente de los recursos (Hu et al., 2021).

### 4.2.4. ControlNet y su Funcionalidad

ControlNet es una arquitectura que mejora los modelos de difusión al permitir un mayor control sobre las imágenes generadas. Funciona como una extensión de modelos generativos como Stable Diffusion, permitiendo que además de descripciones textuales, también se utilicen señales visuales como bordes, posturas, segmentaciones o mapas de profundidad para dirigir de forma más precisa y detallada el resultado final. Esto permite que las imágenes se ajusten mejor a lo que el usuario busca ganando coherencia y mayor detalle (Zhang et al., 2023).

Una característica técnica fundamental de ControlNet es el uso de capas convolucionales inicializadas a cero también conocidas como “zero convolutions”, que permiten incorporar nuevas condiciones sin alterar el modelo base preentrenado. Esta estrategia facilita el entrenamiento eficiente del modelo incluso con conjuntos de datos relativamente pequeños y sin necesidad de hardware excesivamente potente (Zhang et al., 2023).

Con el tiempo han surgido varias variantes de ControlNet especializadas en distintos tipos de condiciones visuales. Por ejemplo, puede usarse OpenPose para capturar la postura de una persona y generar una imagen que la respete, o convertir un simple boceto en una ilustración detallada (Zhang et al., 2023).

Además, es posible combinar varias condiciones en una misma generación, lo que permite obtener resultados más personalizados y precisos. Gracias a esta versatilidad, ControlNet se ha vuelto útil en campos como los videojuegos, la educación, la arquitectura y otros sectores creativos y técnicos (Zhang et al., 2023).

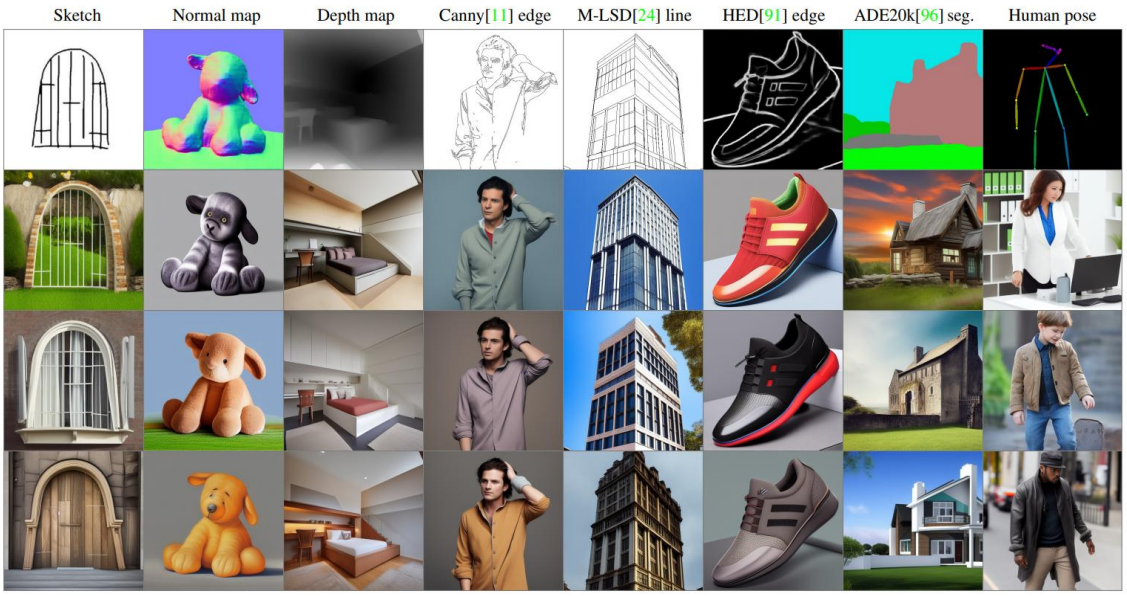



**Figura 13.** Control de Stable Diffusion con diversas condiciones visuales sin usar prompts. La fila superior muestra las condiciones de entrada, las inferiores las salidas generadas usando sólo señales visuales (Zhang et al., 2023).

En definitiva, ControlNet representa un gran avance en la IA generativa al ofrecer un control más estructurado del proceso creativo, haciendo que la generación automática de imágenes sea más personalizable y provechoso.

### 4.2.5. Interfaz basada en nodos (Node-based UI)

Las interfaces basadas en nodos, conocidas en inglés como Node-based UI, son entornos visuales que permiten a los usuarios diseñar y manipular procesos complejos conectando gráficamente diferentes módulos o "nodos". Cada nodo representa una función o acción específica, como cargar un modelo, ajustar parámetros, aplicar transformaciones o generar un resultado final. En lugar de escribir código, los usuarios construyen un flujo de trabajo uniendo estos nodos en un espacio gráfico, lo que facilita ver claramente la secuencia y las relaciones entre las diferentes operaciones (Erfani, 2024).

En lugar de requerir programación textual, los usuarios pueden expresar sus intenciones artísticas y funcionales mediante la manipulación directa de nodos y conexiones, lo que reduce la barrera de entrada y acelera la iteración creativa. Además, la integración de modelos generativos y técnicas de procesamiento de lenguaje natural permite que sistemas node-based como los descritos por Erfani (2024) respondan a instrucciones en lenguaje natural, generando automáticamente grafos de nodos que cumplen con los requisitos del usuario.

Las interfaces basadas en nodos aplicadas a la inteligencia artificial generativa marcan un avance significativo en la manera en que los usuarios se relacionan con la tecnología, al combinar la claridad y adaptabilidad de los sistemas modulares visuales con la

capacidad de los modelos generativos y el procesamiento del lenguaje natural, estas interfaces fomentan la innovación y ayudan a que la generación de contenido digital sea más sencilla de usar, adaptada a las necesidades individuales y optimizada en recursos.

En la Figura 14. se muestra un ejemplo de interfaz basada en nodos aplicada a la generación de imágenes con inteligencia artificial, aquí cada nodo representa una función específica como cargar una imagen, ajustar el tamaño, aplicar un modelo de difusión, o guardar el resultado final, y el usuario puede construir flujos de trabajo complejos conectando gráficamente estos módulos.

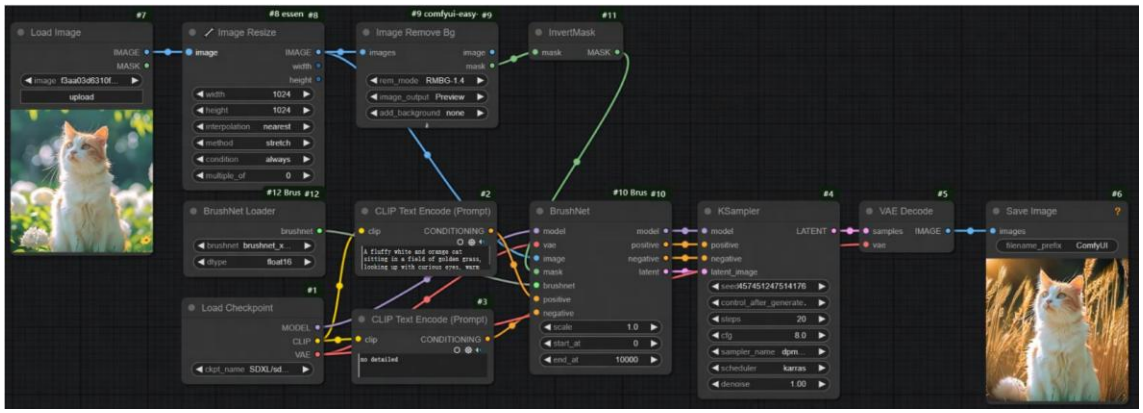



**Figura 14.** Ejemplo de interfaz basada en nodos (Node-based UI) utilizada para la generación de imágenes con inteligencia artificial. Fuente: Huang et al. (2025)

### 4.2.6. Automatización de flujos de trabajo

La automatización de flujos de trabajo o workflow automation en el contexto de la inteligencia artificial generativa permite estructurar y ejecutar procesos complejos de manera más eficiente, reproducible y accesible. En plataformas como ComfyUI, la automatización se implementa mediante flujos visuales construidos con grafos de nodos, donde cada nodo representa una función específica, como cargar modelos, procesar datos o generar salidas visuales.

(Huang et al., 2025) presentan ComfyGPT, un sistema basado en múltiples agentes que automatiza y optimiza estos flujos dentro de ComfyUI, integrando la lógica de nodos con mecanismos de razonamiento autónomo para adaptar dinámicamente el proceso según los objetivos definidos.

En la Figura 15 se presenta el pipeline de ComfyGPT para la generación automática de flujos de trabajo en ComfyUI. El proceso inicia con una instrucción del usuario que pasa por cuatro agentes especializados: ReformatAgent, que evalúa y adapta las consultas; FlowAgent, que crea y corrige el flujo de trabajo utilizando aprendizaje supervisado y optimización; RefineAgent, que mejora la calidad y consistencia del workflow; y ExecuteAgent, que convierte el resultado en un formato compatible con ComfyUI para su ejecución. Este enfoque permite automatizar y optimizar la creación

de flujos visuales complejos, facilitando su uso incluso para usuarios sin experiencia técnica (Huang et al., 2025).

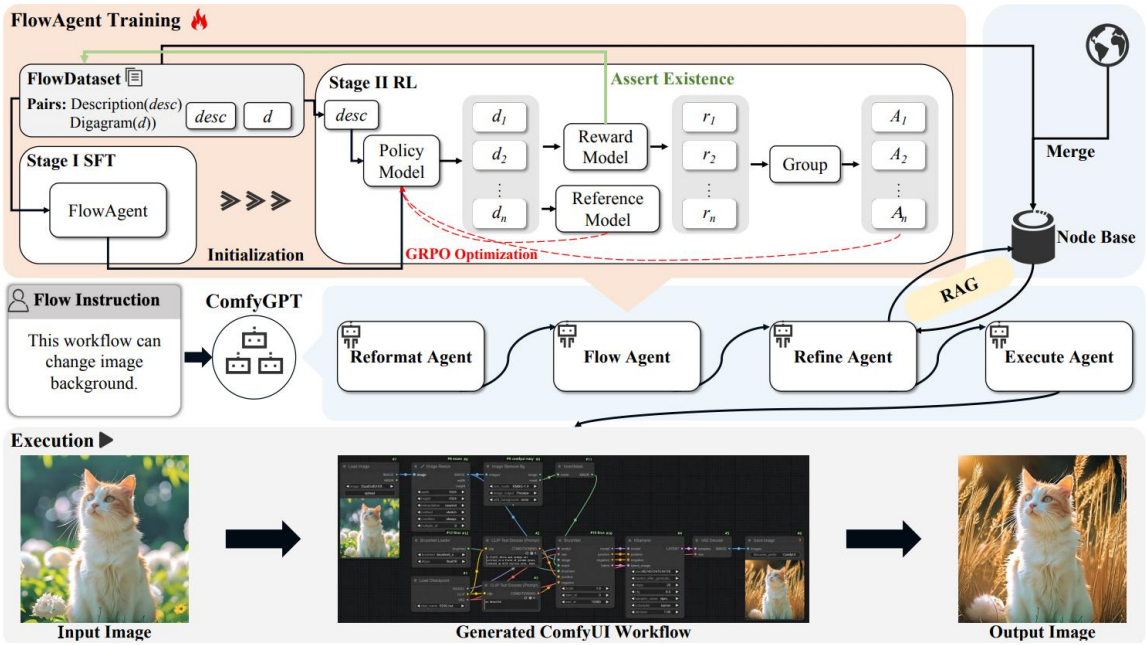



**Figura 15.** Visión general del pipeline de ComfyGPT para la generación automatizada de flujos de trabajo en ComfyUI. Fuente: Huang et al. (2025)

# 5. ANÁLISIS DE HERRAMIENTAS DEL MERCADO

La inteligencia artificial generativa ha transformado profundamente la manera en que se produce contenido visual y se desarrollan ideas creativas en el diseño y la publicidad. Dentro de este contexto, existen diversas plataformas y herramientas que permiten a los profesionales aprovechar estas tecnologías, desde opciones populares y accesibles hasta soluciones más avanzadas y personalizables. Este capítulo se dedica a analizar las plataformas más destacadas, evaluándolas desde una perspectiva técnica y operativa para entender qué tan adecuadas son en ambientes profesionales. La comparación incluye desde herramientas de uso generalizado hasta opciones más personalizables, como ComfyUI, poniendo énfasis en sus fortalezas y limitaciones en contextos reales.

## 5.1. Criterios de comparación

Para evaluar las herramientas de generación de imágenes mediante inteligencia artificial, se han definido los siguientes criterios, fundamentados en estudios recientes que analizan comparativamente el funcionamiento, las posibilidades y el impacto de estas herramientas (Lara Artolazábal, 2023).

- **Control técnico:** Nivel de acceso y ajuste que el usuario puede ejercer sobre los modelos, incluyendo la posibilidad de modificar parámetros avanzados, entrenar modelos propios o construir flujos de trabajo personalizados. Este aspecto es esencial para valorar el grado de intervención y personalización que permiten las distintas herramientas.

- **Usabilidad:** Nivel de facilidad con el que se puede utilizar la herramienta, teniendo en cuenta si su interfaz resulta intuitiva para el público en general o si está orientada a usuarios con experiencia técnica avanzada.

- **Calidad visual:** Coherencia, resolución, nivel de detalle y adecuación de las imágenes generadas para su uso profesional en diseño, arte o publicidad, aspectos señalados como fundamentales en la literatura sobre IA generativa.

- **Personalización:** Capacidad para adaptar la herramienta a necesidades específicas mediante el entrenamiento de modelos, integración de recursos propios o personalización de flujos de trabajo, lo que permite ajustar los resultados a contextos concretos.

- **Rendimiento:** Incluye aspectos como las necesidades de hardware, la velocidad de ejecución y la eficiencia del sistema, además del consumo de recursos tanto locales como en la nube. Estos factores influyen directamente en la factibilidad de implementación en diferentes contextos.

- **Coste y accesibilidad:** Modelo de negocio, dependencia de conexión a Internet y facilidad de acceso para distintos perfiles de usuario, variables que determinan la adopción y democratización de estas tecnologías.

## 5.2. MidJourney

MidJourney es una herramienta de inteligencia artificial generativa desarrollada por un laboratorio independiente, pensada para crear imágenes a partir de descripciones escritas (MidJourney, 2025a). Desde su aparición, se ha convertido en una de las plataformas más utilizadas en los ámbitos creativo y publicitario, gracias a su capacidad para interpretar prompts complejos, adaptarse a distintos estilos artísticos y ofrecer resultados visuales de alta calidad (MidJourney, 2025a). La mayoría de los usuarios acceden a través de Discord, donde interactúan con el bot de MidJourney usando comandos, aunque también existe una versión con interfaz web (MidJourney, 2025a).

La plataforma es utilizada por diseñadores, publicistas, creadores de contenido y artistas digitales para conceptualizar ideas visuales, generar moodboards, producir imágenes para campañas y desarrollar recursos gráficos originales. MidJourney opera bajo un modelo de suscripción, con diferentes planes según el volumen de uso y la velocidad de procesamiento (MidJourney, 2025b).

**Características y funcionalidades clave**

- **Generación de imágenes desde texto:** Convierte descripciones escritas en imágenes detalladas y originales, interpretando tanto conceptos sencillos como ideas abstractas y complejas (MidJourney, 2025a).

- **Variedad de estilos artísticos:** Permite seleccionar entre una amplia gama de estilos, desde fotorrealismo hasta ilustración abstracta, arte digital o anime, adaptándose a distintas necesidades creativas (MidJourney, 2025a).

- **Parámetros personalizables:** El usuario puede ajustar el formato, el nivel de detalle, la creatividad o el grado de fidelidad al prompt mediante comandos, y elegir la resolución de salida (hasta 1792 x 1024 píxeles) (MidJourney, 2025a).

- **Iteración, variación y refinamiento**: Iteración y refinamiento: Por cada prompt, se generan cuatro versiones iniciales. Luego, el usuario puede ampliar, modificar o variar cualquiera de ellas para perfeccionar los resultados (MidJourney, 2025a; Gelato, 2025).

- **Imágenes de referencia:** Es posible subir imágenes propias como parte del prompt, lo cual es útil en trabajos de branding o cuando se busca mantener una línea visual (MidJourney, 2025a).

- **Gestión de proyectos y perfiles:** MidJourney facilita la organización de imágenes generadas mediante galerías y perfiles de usuario, permitiendo mantener la coherencia en proyectos de largo plazo (MidJourney, 2025a).
- **Procesamiento en la nube:** Todo el procesamiento se realiza en servidores remotos, eliminando la necesidad de hardware local potente y garantizando accesibilidad y rapidez (MidJourney, 2025a).
- **Actualizaciones y mejoras continuas:** La plataforma recibe actualizaciones frecuentes que incorporan nuevas funciones, mejoras en la calidad de imagen y optimización del rendimiento (MidJourney, 2025a).

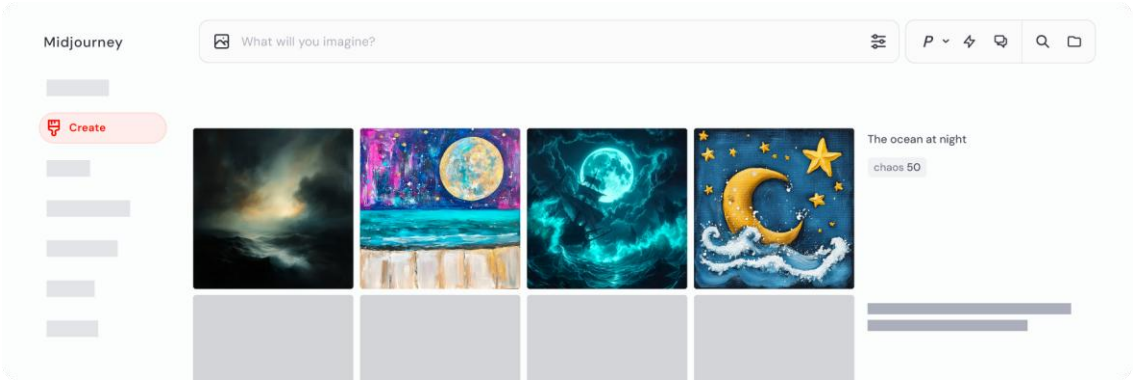



**Figura 16.** Interfaz gráfica de Midjourney mostrando la función de creación de imágenes.
Fuente: MidJourney (2025a)

## Análisis según criterios de comparación

- **Control técnico**: El usuario puede ajustar el resultado mediante prompts y comandos, pero no tiene acceso al modelo subyacente ni puede personalizarlo en profundidad, lo que limita la flexibilidad técnica (MidJourney, 2025a).
- **Usabilidad:** La plataforma destaca por su accesibilidad y facilidad de uso, permitiendo a cualquier usuario generar imágenes de calidad sin conocimientos técnicos. La curva de aprendizaje es baja y la comunidad facilita la integración de nuevos usuarios (MidJourney, 2025a).
- **Calidad visual:** MidJourney genera imágenes visualmente llamativas, diversas y con un alto nivel de detalle, aunque en algunos casos puede presentar dificultades al interpretar indicaciones poco claras o elementos muy específicos (MidJourney, 2025a).

- **Personalización:** La personalización se basa en la creatividad del prompt y algunos parámetros, pero no permite entrenar modelos propios ni integrar datasets personalizados (MidJourney, 2025a).
- **Rendimiento:** El procesamiento en la nube garantiza rapidez y resultados consistentes, sin depender del hardware local. Los planes de suscripción ofrecen diferentes velocidades y prioridades (MidJourney, 2025b).
- **Accesibilidad y costos:** No hay versión gratuita permanente, y el acceso depende de una suscripción mensual. Aunque no se requiere hardware especializado, sí es necesario tener acceso a Discord y conexión a Internet (MidJourney, 2025b).

## 5.3. DALL·E 3

DALL·E 3 es una herramienta desarrollada por OpenAI que permite generar imágenes a partir de descripciones escritas, una técnica conocida como *text-to-image* (OpenAI, 2024a). Desde su lanzamiento en 2023, esta versión ha supuesto un salto considerable respecto a sus antecesores, especialmente en lo que respecta a la comprensión del lenguaje natural y la capacidad para reflejar detalles complejos y relaciones espaciales en las imágenes generadas (OpenAI, 2024a).

Una de sus ventajas es que está disponible de forma gratuita a través de plataformas como ChatGPT y Microsoft Copilot, lo que ha facilitado su uso tanto para profesionales del diseño como para usuarios con menos experiencia técnica (Microsoft, 2025). Su aplicación es muy variada: se utiliza en campos como el diseño gráfico, la educación, la publicidad o la creación de contenido digital, permitiendo transformar ideas abstractas en imágenes útiles para prototipos, ilustraciones o materiales formativos (OpenAI, 2024a).

**Características principales**

- **Generación de imágenes a partir de texto:** A través de prompts escritos, DALL·E 3 es capaz de crear imágenes visualmente detalladas y de alta calidad, gracias a su entrenamiento con grandes volúmenes de datos combinando texto e imagen (OpenAI, 2024a).
- **Entendimiento profundo del lenguaje natural:** El modelo destaca por su capacidad para interpretar matices, relaciones espaciales y características específicas dentro de un prompt, lo que le permite generar resultados que se alinean con lo que el usuario quiere expresar (OpenAI, 2024a).
- **Variedad de estilos y flexibilidad creativa:** Se pueden combinar estilos artísticos y elementos conceptuales en una sola imagen, lo que ofrece libertad

para crear desde ilustraciones realistas hasta arte abstracto o de fantasía (OpenAI, 2024a).

- **Calidad profesional:** Las imágenes pueden alcanzar resoluciones de hasta 1792 x 1024 píxeles, con mejoras notables en texturas, iluminación y fondos, haciéndolas aptas incluso para usos profesionales (OpenAI, 2024a).
- **Mejoras en representación humana y texto:** A diferencia de versiones anteriores, DALL·E 3 maneja de forma más precisa la anatomía humana y la integración de texto dentro de las imágenes (OpenAI, 2024a).
- **Generación de variaciones:** Cada vez que se introduce un mismo prompt, el sistema genera imágenes diferentes, lo cual fomenta la exploración creativa (OpenAI, 2024a).
- **Integración mediante API:** Está disponible una API que permite incluir la herramienta en aplicaciones o plataformas propias, lo que facilita su incorporación en flujos de trabajo personalizados (OpenAI, 2024b).
- **Accesibilidad y disponibilidad multiplataforma:** Puede utilizarse sin coste a través de ChatGPT y Microsoft Copilot, lo que abre su uso a una audiencia global (Microsoft, 2025).

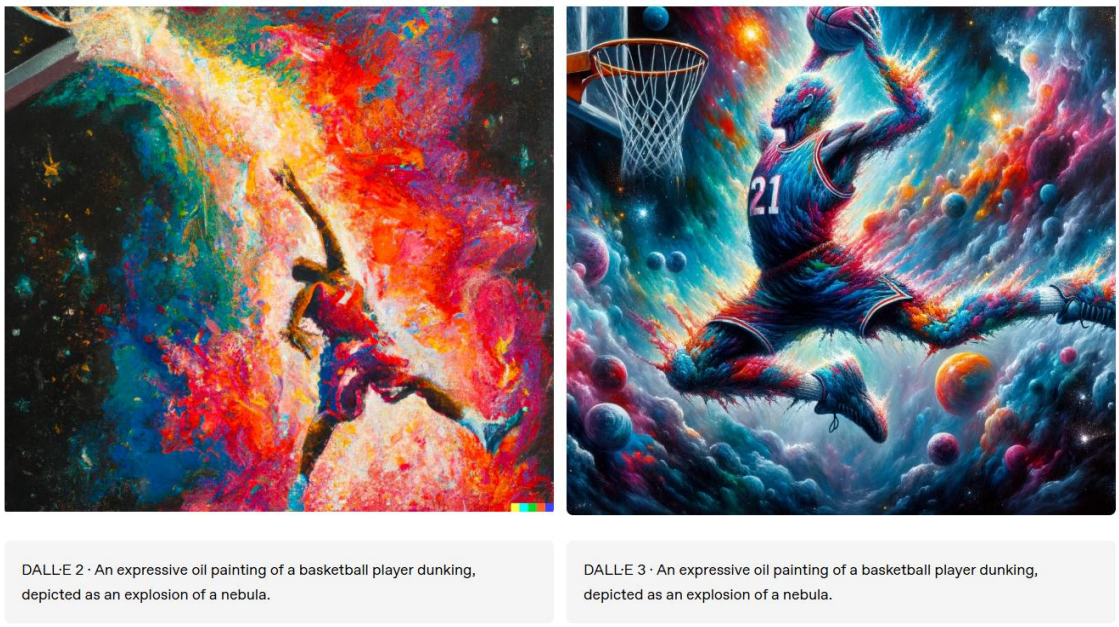



**Figura 17.** Ejemplos de imágenes generadas con DALL E 3 AI a través de un prompt..
Fuente: OpenAI (2024a).

### Análisis según criterios de comparación

- **Control técnico:** Aunque el usuario puede influir en el resultado ajustando los prompts o ciertos parámetros, no tiene acceso al modelo en sí ni puede entrenar

versiones personalizadas, lo que limita su flexibilidad en contextos más técnicos (OpenAI, 2024a).

- **Facilidad de uso:** DALL·E 3 destaca por su accesibilidad y facilidad de uso, con una interfaz amigable e integración directa en ChatGPT o Microsoft Copilot, por lo que no se requiere experiencia previa para comenzar a generar imágenes (OpenAI, 2024a).

- **Calidad visual:** Las imágenes producidas alcanzan altos estándares de calidad, con mejoras en el realismo, el detalle anatómico y la coherencia visual, lo que las hace útiles incluso para publicaciones profesionales pero dependen de la calidad y precision del prompt. (OpenAI, 2024a).

- **Opciones de personalización:** Si bien ofrece variedad en estilos y composiciones, la personalización avanzada está limitada al texto del prompt y no permite alterar el modelo ni usar complementos (OpenAI, 2024a).

- **Rendimiento:** El procesamiento en la nube permite resultados rápidos y consistentes, aunque la velocidad puede verse afectada por la demanda del servicio o por el tipo de plan de acceso (OpenAI, 2024a).

- **Coste y disponibilidad:** DALL·E 3 ahora ofrece acceso gratuito limitado por ejemplo, hasta dos imágenes diarias a través de ChatGPT y Bing. Sin embargo, para un uso más intensivo creando muchas imágenes al día o en entornos profesionales sigue siendo necesario obtener ChatGPT Plus o usar la API de OpenAI (OpenAI, 2024a).

## 5.4. Leonardo AI

Leonardo.AI es una plataforma de IA generativa enfocada en la creación de imágenes y recursos visuales para publicidad, videojuegos y branding. Destaca por su facilidad de uso y la posibilidad de entrenar modelos personalizados a partir de datasets propios, lo que permite adaptar los resultados a estilos y necesidades específicas (Leonardo.AI, 2025).

**Características y funcionalidades clave**

- **Generación de imágenes** basándose en texto, ajustadas a diversos estilos y aplicaciones (Leonardo.AI, 2025). La plataforma procesa descripciones escritas y las convierte en imágenes minuciosas, facilitando la elección entre estilos como el realismo, el arte digital, la ilustración o el pixel art.

- **Entrenamiento de modelos personalizados** para proyectos específicos (Leonardo.AI, 2025). Los usuarios pueden cargar sus propios datasets para crear

modelos que aprendan y reproduzcan un estilo visual particular, ideal para proyectos con identidad gráfica definida.

- **Modificación y adaptación de estilos**: alteración de colores, estructura y características visuales (Leonardo.AI, 2025). Leonardo.AI facilita la mejora de los resultados producidos modificando factores como la iluminación, el enfoque, la gama de colores y la distribución visual para alcanzar una coherencia estética más eficaz.

- **Interfaz web intuitiva** y administración de proyectos y recursos visuales (Leonardo.AI, 2025). Su diseño centrado en el usuario facilita la organización de trabajos por carpetas, la vista previa de resultados y la colaboración dentro de equipos creativos.

- **Procesamiento en la nube** y modelo freemium con opciones premium (Leonardo.AI, 2025). No se requiere instalación local ni hardware potente, y los planes de suscripción permiten acceder a más recursos, velocidad de procesamiento y modelos exclusivos según las necesidades del usuario.

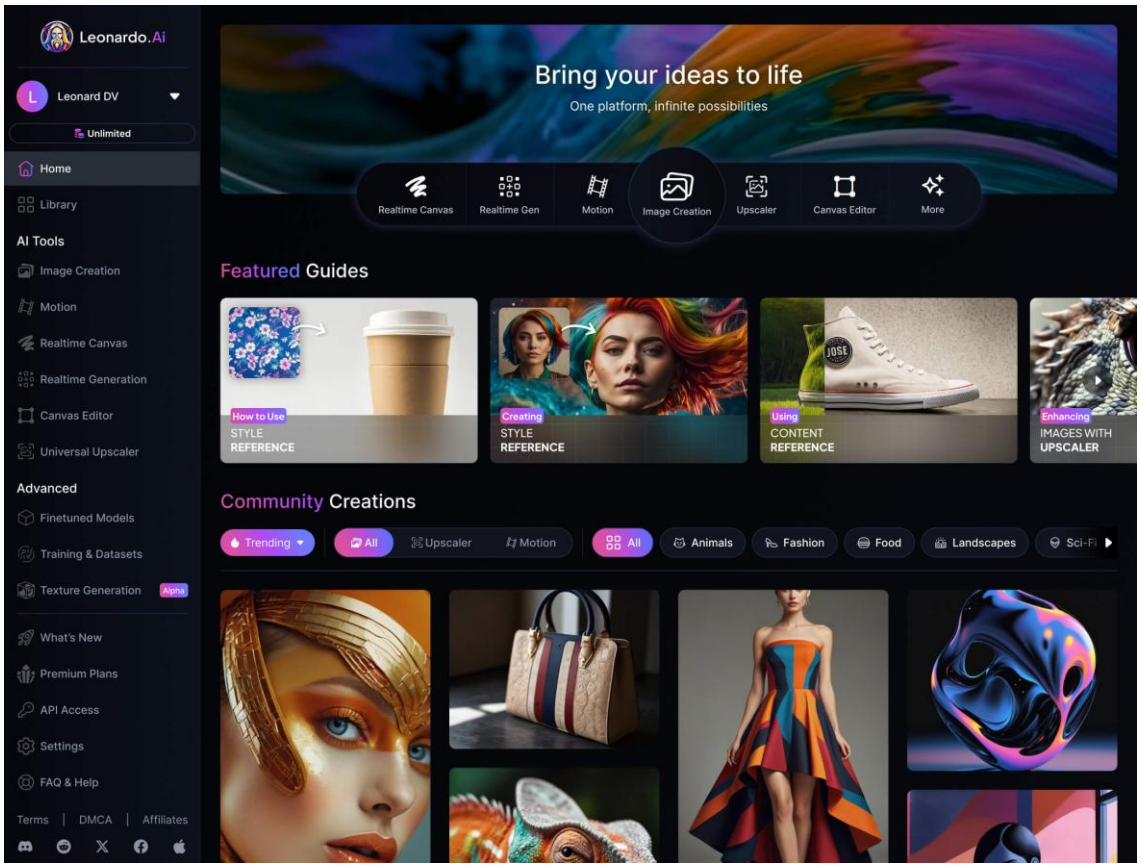



**Figura 18.** Interfaz gráfica de Leonardo AI con la página principal incluyendo funciones y menú.
Fuente: Leonardo.AI. (2025).

## Análisis según criterios de comparación

- **Control técnico:** Permite cierto grado de control técnico, especialmente a través del entrenamiento de modelos personalizados y el ajuste de parámetros visuales. Sin embargo, la personalización avanzada está limitada por las opciones que ofrece la plataforma y no permite modificar el modelo base (Leonardo.AI, 2025).
- **Usabilidad:** La interfaz es amigable y está orientada a creativos y equipos de marketing, facilitando la adopción sin requerir conocimientos avanzados en IA. La gestión de proyectos y assets visuales es sencilla y eficiente, lo que agiliza los flujos de trabajo (Leonardo.AI, 2025).
- **Calidad visual:** Las imágenes generadas son de alta calidad y pueden adaptarse a diferentes estilos, lo que resulta especialmente útil en branding y videojuegos. La calidad final está condicionada tanto por el modelo seleccionado como por los parámetros establecidos (Leonardo.AI, 2025).
- **Personalización:** Ofrece una personalización media-alta gracias al entrenamiento de modelos propios, pero no permite modificar la arquitectura interna ni integrar plugins externos. Esto lo distingue a las soluciones de fuente abierta más técnicas (Leonardo.AI, 2025).
- **Rendimiento:** El procesamiento se realiza en la nube, lo que garantiza rapidez y elimina la necesidad de hardware local potente. Sin embargo, el rendimiento puede depender del plan de suscripción y la demanda de la plataforma (Leonardo.AI, 2025).
- **Coste y accesibilidad:** Opera bajo un modelo freemium, permitiendo acceso básico gratuito y funciones avanzadas mediante suscripción. La accesibilidad es alta, ya que se puede utilizar desde cualquier navegador (Leonardo.AI, 2025).

## 5.5. AUTOMATIC1111 (Stable Diffusion WebUI)

AUTOMATIC1111 es una interfaz gráfica de usuario de código abierto diseñada para ejecutar Stable Diffusion de manera local, permitiendo a los usuarios gestionar y personalizar la generación de imágenes por IA en sus propios equipos (AUTOMATIC1111, 2023a).

Esta herramienta se ha consolidado como una de las soluciones más completas y flexibles para usuarios avanzados que requieren control total sobre los modelos, parámetros y flujos de trabajo, así como privacidad en el procesamiento de datos (AUTOMATIC1111, 2023a).

La interfaz es compatible con sistemas operativos Windows, Linux y Mac, y soporta diferentes arquitecturas de GPU, lo que la hace accesible para una amplia comunidad de desarrolladores y artistas digitales (AUTOMATIC1111, 2022).

**Características y funcionalidades clave**

- **Interfaz web local:** AUTOMATIC1111 permite ejecutar Stable Diffusion directamente en el equipo del usuario a través de una interfaz web, lo que facilita la gestión y ejecución de modelos sin depender de servicios en la nube y garantiza la privacidad y el control de los datos generados (AUTOMATIC1111, 2023a).

- **Soporte para múltiples modelos y versiones:** La herramienta es compatible con una amplia variedad de modelos de Stable Diffusion, incluyendo versiones oficiales, modelos personalizados, así como integraciones avanzadas como LoRA y ControlNet (AUTOMATIC1111, 2023b).

- **Control de parámetros:** AUTOMATIC1111 brinda un control completo sobre los parámetros de generación y esta flexibilidad es fundamental para optimizar resultados y realizar experimentos sistemáticos (AUTOMATIC1111, 2023a).

- **Funciones avanzadas de edición:** La interfaz incorpora herramientas para relleno de áreas específicas de una imagen, extensión de imágenes más allá de sus bordes originales, generación de imagen a imagen, mezcla de estilos y upscaling, lo que amplía considerablemente las posibilidades creativas y de postprocesado (AUTOMATIC1111, 2023a).

- **Extensibilidad y personalización:** El sistema permite la instalación de plugins, scripts y automatizaciones desarrolladas tanto por la comunidad como por el propio usuario lo que facilita la creación de flujos de trabajo personalizados (AUTOMATIC1111, 2023b).

- **Procesamiento por batches y gestión eficiente de recursos:** La herramienta facilita la generación simultánea de múltiples imágenes y optimiza la organización del almacenamiento tanto de los modelos como de los resultados, esto es útil en escenarios que requieren producir un alto volumen de imágenes o realizar evaluaciones comparativas entre diferentes modelos (AUTOMATIC1111, 2023a).

- **Documentación técnica y recursos de soporte:** La herramienta cuenta con una wiki oficial y documentación técnica detallada que abarca desde la instalación y configuración inicial hasta la personalización avanzada y el desarrollo de extensiones. (AUTOMATIC1111, 2023a).

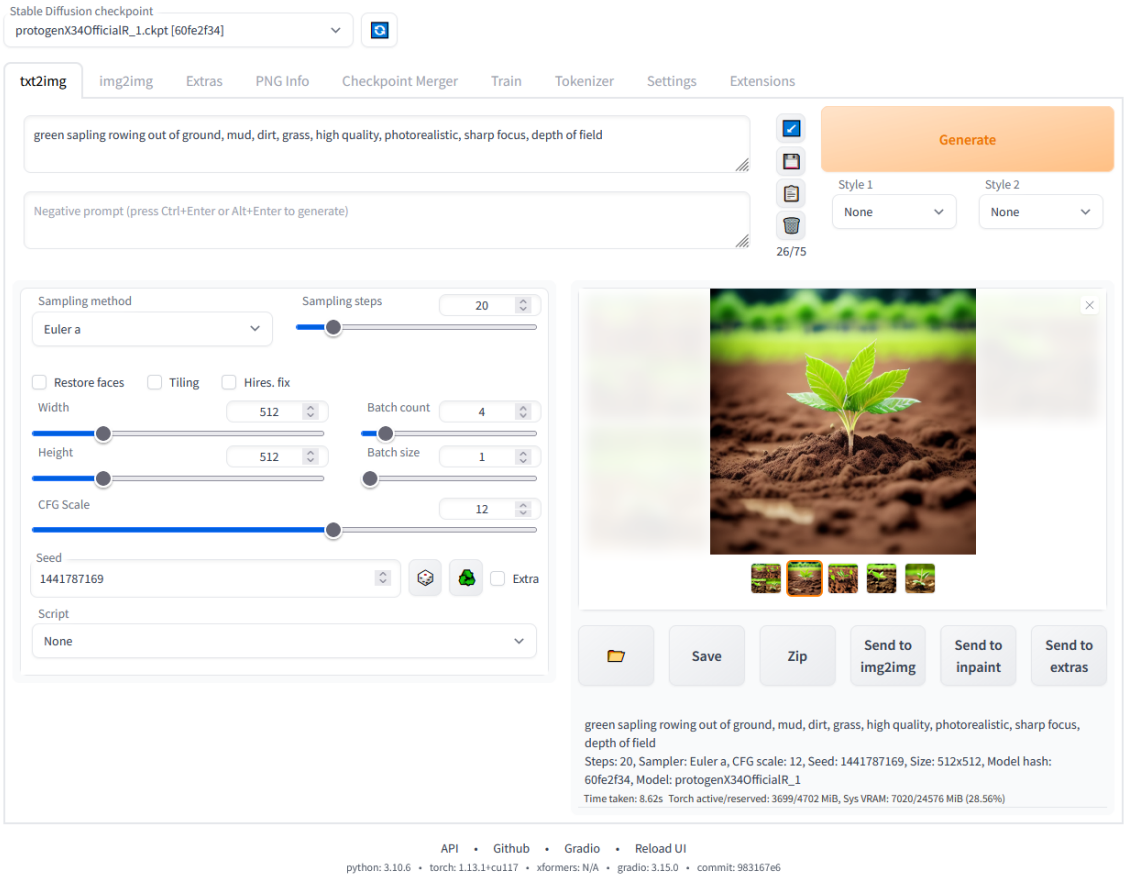



**Figura 19.** Interfaz gráfica de Automatic1111 con implementada usando Gradio library de ejemplo.
Fuente: AUTOMATIC1111. (2023a).

## Análisis según criterios de comparación

- **Control técnico:** AUTOMATIC1111 ofrece un nivel técnico avanzado que permite ajustar a detalle los parámetros de generación, incorporar modelos personalizados e incluso entrenarlos, así como diseñar flujos de trabajo complejos. Por estas características, suele ser la herramienta elegida por quienes desean explorar nuevas metodologías y adaptar el proceso de generación a sus necesidades específicas (AUTOMATIC1111, 2023a).

- **Usabilidad:** La interfaz gráfica facilita la gestión de modelos y parámetros, pero la abundancia de opciones y la necesidad de configuración pueden resultar complejas para usuarios sin experiencia previa en IA o programación. La curva de aprendizaje es pronunciada, aunque la documentación y la comunidad ayudan a mitigar este desafío (AUTOMATIC1111, 2023a).

- **Calidad visual:** Las imágenes generadas alcanzan calidad profesional y pueden adaptarse a estilos y necesidades específicas según el modelo y los parámetros

seleccionados. La flexibilidad de la herramienta permite optimizar los resultados mediante pruebas iterativas y ajustes finos (AUTOMATIC1111, 2023b).

- **Personalización:** AUTOMATIC1111 es altamente personalizable, permitiendo entrenar modelos propios, instalar plugins y scripts, modificar el pipeline y automatizar procesos, lo que la convierte en una plataforma ideal para proyectos experimentales y desarrollos a medida (AUTOMATIC1111, 2023b).
- **Rendimiento:** El desempeño de la herramienta está directamente vinculado a las capacidades del hardware local, en particular a la GPU y la cantidad de memoria disponible (AUTOMATIC1111, 2023a).
- **Coste y accesibilidad:** Es una herramienta gratuita y de código abierto, facilitando su adopción en entornos académicos y profesionales pero el desempeño de la herramienta depende bastante de las capacidades del hardware local (AUTOMATIC1111, 2023a).

## 5.6. ComfyUI

ComfyUI es una interfaz gráfica de usuario de código abierto diseñada específicamente para la generación de imágenes con modelos como Stable Diffusion. Su principal característica es un sistema visual basado en nodos, que permite a los usuarios construir, modificar y visualizar flujos de trabajo personalizados para cada etapa del proceso de generación de imágenes. Esta arquitectura modular facilita tanto la experimentación avanzada como la comprensión de los procesos internos, siendo especialmente útil para usuarios que desean un control granular y transparencia en la manipulación de modelos de IA (ComfyUI, 2024a).

ComfyUI está orientada a usuarios con conocimientos intermedios o avanzados en IA generativa, aunque su diseño visual también favorece el aprendizaje progresivo para principiantes interesados en explorar la personalización y automatización de tareas. Puede ejecutarse localmente en sistemas Windows, Linux y Mac, y es compatible con una amplia variedad de modelos y extensiones, permitiendo la integración de herramientas como LoRA, ControlNet y modelos personalizados (ComfyUI, 2024a).

**Características y funcionalidades clave**

- **Interfaz visual basada en nodos:** ComfyUI utiliza una interfaz gráfica basada en nodos que permite crear flujos de generación de imágenes mediante la conexión visual de módulos. Esto hace que sea más sencillo entender, ajustar y personalizar cada fase del proceso (ComfyUI, 2024a).
- **Compatibilidad con múltiples modelos y técnicas:** Soporta la integración de modelos personalizados de Stable Diffusion, así como técnicas avanzadas como

LoRA, ControlNet y variantes de samplers, permitiendo experimentar con diferentes arquitecturas y estilos (ComfyUI, 2024b).

- **Ajuste detallado de parámetros y nodos:** Ofrece control sobre parámetros clave (pasos de inferencia, semillas, tamaño de imagen, escalado, etc.) y permite crear nodos personalizados para tareas específicas, adaptando el pipeline a las necesidades del usuario (ComfyUI, 2024a).

- **Funciones avanzadas de edición y procesamiento:** Incluye herramientas para inpainting, outpainting, generación de imagen a imagen, mezcla de estilos y upscaling, así como procesamiento por lotes y automatización de tareas repetitivas (ComfyUI, 2024a).

- **Procesamiento local y gestión eficiente de recursos:** Al ejecutarse en el equipo del usuario, garantiza privacidad y control total sobre los datos y modelos, optimizando el uso de GPU y memoria disponible (ComfyUI, 2024a).

- **Extensibilidad y comunidad activa:** Es posible ampliar la funcionalidad mediante scripts, nodos adicionales y plugins desarrollados por la comunidad, lo que favorece la innovación y la adaptación a nuevos retos (ComfyUI, 2024b).

- **Documentación técnica y recursos de soporte:** La plataforma cuenta con documentación oficial, tutoriales y foros de discusión que cubren desde la instalación hasta el desarrollo de nodos personalizados y la resolución de problemas (ComfyUI, 2024a).

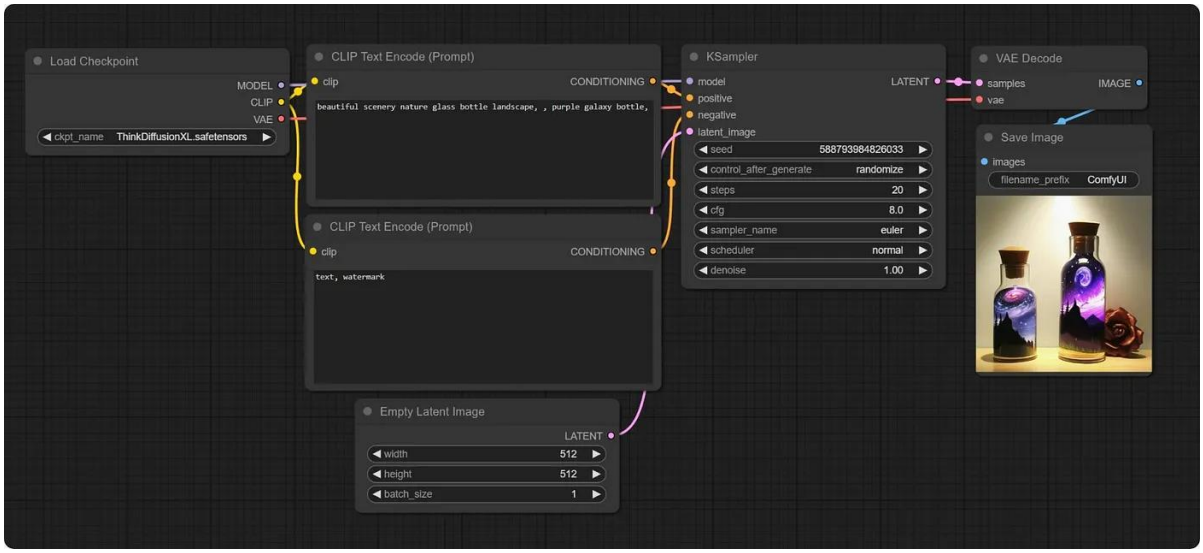



**Figura 20.** Interfaz gráfica de ComfyUI con workflow de ejemplo. Fuente: ComfyUI Documentación Oficial (2025)

- **Control técnico:** ComfyUI destaca por su control técnico avanzado, permitiendo a los usuarios construir y modificar flujos de trabajo personalizados mediante nodos visuales. Este flujo de trabajo puede modificarse, optimizarse o

extenderse, lo que la convierte en una muy buena opción para experimentar, crear modelos personalizados y explorar nuevas metodologías (ComfyUI, 2024a).

- **Usabilidad:** La herramienta ofrece una interfaz visual por nodos lo que ayuda a entender mejor y analizar los procesos. Aunque está orientada a usuarios con conocimientos previos en inteligencia artificial o procesamiento de imágenes, su estructura modular permite una curva de aprendizaje manejable y favorece la comprensión del flujo de trabajo (ComfyUI, 2024a).

- **Calidad visual:** Permite obtener imágenes de alta calidad, adaptándose a necesidades específicas mediante la personalización de cada etapa del flujo. La calidad final depende del modelo usado pero también de la configuración de nodos y especialmente de los parámetros y la combinación de todos estos pueden generar resultados optimizados (ComfyUI, 2024b).

- **Personalización:** Ofrece una personalización muy alta, permitiendo la integración de modelos propios, nodos personalizados y la creación de flujos de trabajo completamente adaptados al proyecto, lo que la convierte en una de las opciones más flexibles para usuarios avanzados (ComfyUI, 2024b).

- **Rendimiento:** El desempeño de ComfyUI está condicionado por las características del hardware local, en especial la GPU. La herramienta está diseñada para sacar el máximo provecho de los recursos disponibles y admite tanto el procesamiento por lotes como la automatización de tareas, aunque su rendimiento puede disminuir en equipos con capacidades limitadas (ComfyUI, 2024a).

- **Coste y accesibilidad:** Es gratuita y de código abierto, pero requiere inversión en hardware y tiempo de aprendizaje. No depende de conexión a Internet para su uso habitual, salvo para descargas de modelos o actualizaciones, lo que la hace accesible para usuarios con necesidades de privacidad y control total sobre sus datos (ComfyUI, 2024a).

## 5.7. Análisis de resultados

Con el objetivo de ofrecer una visión clara y accesible sobre las diferencias entre las principales plataformas de generación de imágenes por inteligencia artificial, se presenta a continuación un cuadro comparativo. Este cuadro resume el desempeño de cada herramienta según los criterios definidos en este capítulo: control técnico, usabilidad, calidad de salida, personalización, rendimiento y coste/accesibilidad.

Para facilitar la interpretación, se ha utilizado un sistema de colores tipo semáforo:

-  Verde indica un desempeño sobresaliente en el criterio evaluado,
-  Amarillo señala un nivel intermedio o aceptable,
-  Rojo identifica áreas con limitaciones o margen de mejora.

Esto permite identificar de más rápidamente los puntos fuertes y débiles de cada plataforma, ayudando a orientar la elección de la herramienta más adecuada según las necesidades específicas de cada usuario o proyecto.

El análisis comparativo en la **Tabla 1**. muestra que cada plataforma responde a perfiles y necesidades diferentes. MidJourney y DALL·E 3 resultan especialmente atractivas para quienes buscan facilidad de uso y resultados de alta calidad sin necesidad de conocimientos técnicos avanzados. Leonardo AI se posiciona como una opción equilibrada, combinando accesibilidad y algunas posibilidades de personalización.

En cambio, AUTOMATIC1111 y ComfyUI están orientadas a quienes prefieren tener un mayor control sobre todo el proceso creativo. Ambas son herramientas más técnicas y flexibles, pensadas para usuarios que quieren experimentar, ajustar parámetros a fondo y trabajar con soluciones de código abierto.

Este panorama general ayuda a entender por qué se eligió ComfyUI para el análisis en profundidad del siguiente capítulo, ComfyUI es una opción potente y versátil para quienes buscan ir más allá de lo que ofrecen las plataformas más cerradas y automáticas.

**Tabla 1.** Cuadro Comparativo entre herramientas de generación de imágenes con IA

| Herramienta / Criterio | Control técnico | Usabilidad | Calidad de salida | Personalización | Rendimiento | Coste y accesibilidad |
|---|---|---|---|---|---|---|
| **MidJourney** |  **Medio** Permite algunos ajustes, pero el control avanzado es limitado. |  **Muy alta** Interfaz sencilla y resultados rápidos sin curva de aprendizaje. |  **Alta** Resultados consistentes, aunque la precisión respecto a prompts muy específicos puede variar. |  **Bajo** Opciones de personalización limitadas al prompt. |  **Alto** Generación rápida en la nube. |  **Suscripción** Requiere pago mensual para uso continuo. |
| **DALL·E 3** |  **Medio** Control técnico limitado, ajustes avanzados no disponibles. |  **Muy alta** Fácil de usar, integración directa en plataformas conocidas. |  **Alta** Imágenes realistas pero la variabilidad depende del prompt y de los procesos internos del modelo |  **Bajo** Personalización limitada, sin ajustes finos. |  **Alto** Procesamiento rápido en la nube. |  **Gratis con límites** Con opción de pago para mayor uso. |
| **Leonardo AI** |  **Medio** Ofrece algunos controles adicionales respecto a otras plataformas. |  **Alta**. Interfaz intuitiva, aunque con más opciones que pueden requerir exploración. |  **Alta** Imágenes de buena calidad, aunque puede variar según el modelo elegido y el prompt |  **Media** Permite cierto nivel de personalización, pero no tan avanzado. |  **Alto** Generación eficiente en la nube. |  **Freemium**. Versión gratuita con límites, opciones de pago para funciones avanzadas. |
| **AUTOMATIC1111** |  **Alto** Gran control sobre parámetros y modelos, ideal para usuarios avanzados. |  **Media** Requiere instalación y conocimientos técnicos básicos. |  **Muy alta** Resultados personalizables y de alta calidad. |  **Muy alta** Permite personalización profunda mediante modelos y configuraciones. |  **Variable** Depende del hardware disponible en el equipo local. |  **Gratis** Software open source sin coste de licencia. |
| **ComfyUI** |  **Muy alto** Control total sobre el flujo de trabajo y parámetros avanzados. |  **Media** Con curva de aprendizaje, requiere instalación y conocimientos técnicos. |  **Muy alta** Imágenes de excelente calidad y gran control sobre el proceso. |  **Muy alta**. Personalización máxima mediante nodos y flujos configurables. |  **Variable** Depende del hardware y configuración local. |  **Gratis** Open source, sin restricciones de uso. |

# 6. ANÁLISIS DE COMFYUI

En este capítulo se presenta un análisis detallado de ComfyUI, una herramienta de generación de imágenes mediante inteligencia artificial basada en modelos de difusión. A diferencia de las demás plataformas analizadas en este trabajo, ComfyUI no funciona como un servicio en la nube ni está disponible para uso inmediato, sino que se trata de una interfaz modular que necesita ser instalada y configurada técnicamente antes de su utilización.

Durante la realización de este Trabajo Fin de Máster, se utilizó un servidor remoto de la Universidad para instalar y ejecutar ComfyUI. Para ello, se empleó la herramienta Podman en modo rootless, simulando un entorno con contenedores similar a Docker, pero más adecuado para entornos académicos y seguros. Esta aproximación permitió montar toda la infraestructura necesaria para ejecutar modelos como Stable Diffusion XL, ControlNet y aplicar técnicas como LoRA o flujos automatizados.

A continuación, se detallan los pasos y consideraciones más relevantes para la instalación, configuración y uso práctico de ComfyUI.

## 6.1. Instalación (servidor)

En esta sección se describen los pasos seguidos para acceder al servidor remoto proporcionado por la Universidad de Granada y preparar el entorno necesario para la posterior instalación de ComfyUI.

El procedimiento se detalla paso a paso y está orientado a usuarios sin conocimientos técnicos, proporcionando una guía clara, explicando cada instrucción utilizada.

**Acceso a un entorno Linux (WSL)**

Se activó WSL (Windows Subsystem for Linux) en el ordenador local con sistema operativo Windows para disponer de una terminal de Linux. Para ello se uso el siguiente comando en el PowerShell:

**wsl --install**

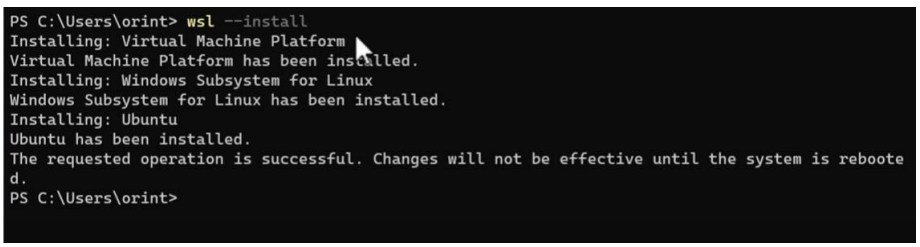



**Figura 21.** Ejemplo de activación de WSL desde el terminal

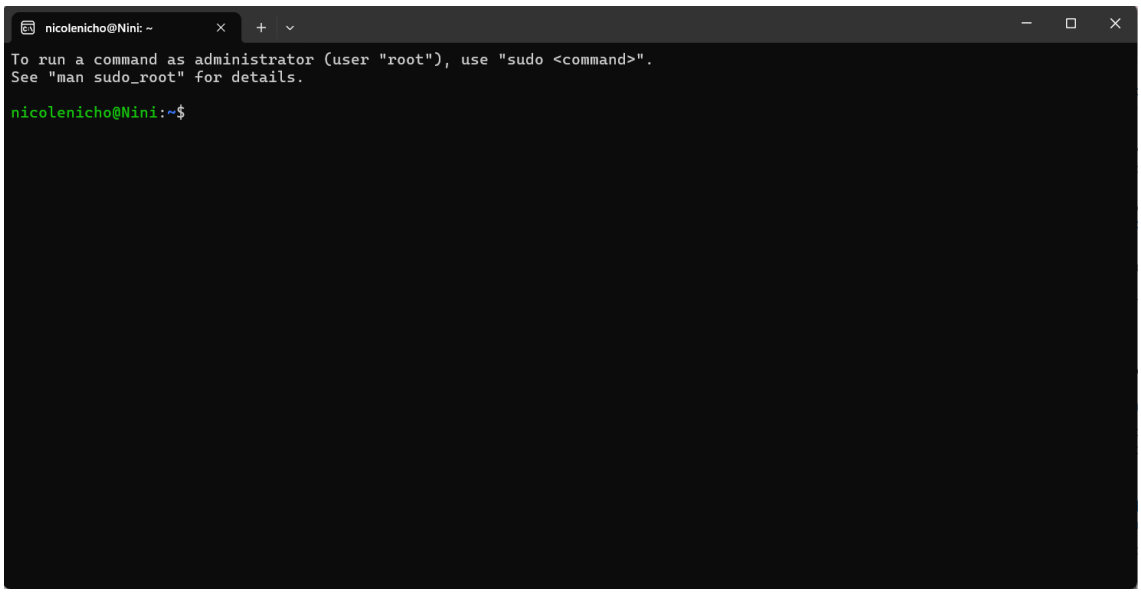



**Figura 22.** Ejemplo de interfaz de WSL desde el terminal

### Conexión inicial al servidor

Desde la terminal de Linux es decir Ubuntu en WSL del paso anterior, se usó el siguiente comando para conectarse al servidor:

```
ssh -p 2222 usuario@servidor
```

- ssh es el comando para conectarse de forma segura a otro equipo a través de Internet.
- -p 2222 indica que se usará el puerto 2222 (en lugar del puerto estándar 22) para conectarse.
- usuario@servidor representa el nombre de usuario y la dirección del servidor al que se desea acceder.

### Cambio de contraseña

Dentro del terminal del servidor y al tratarse del primer acceso, se recomienda que se cambiara la contraseña por protección. El cambio se realizó utilizando este comando:

```
passwd
```

- Este comando permite cambiar la contraseña del usuario actual en el servidor.
- Se debe introducir primero la contraseña actual, y luego la nueva dos veces para confirmarla.

### Crear una clave SSH (en el equipo local)

Para evitar tener que escribir la contraseña cada vez que se accede al servidor, se configura un sistema de acceso por clave pública. Desde la terminal local no desde el servidor, se ejecutó:

**ssh-keygen -t rsa -b 4096**

- ssh-keygen es el comando que genera un par de claves.
- -t rsa indica que se usará el tipo de clave RSA.
- -b 4096 especifica que se creará una clave de 4096 bits (más segura).

Luego simplemente se presionó Enter en las preguntas que aparecieron como ubicación de guardado, contraseña de la clave, etc.

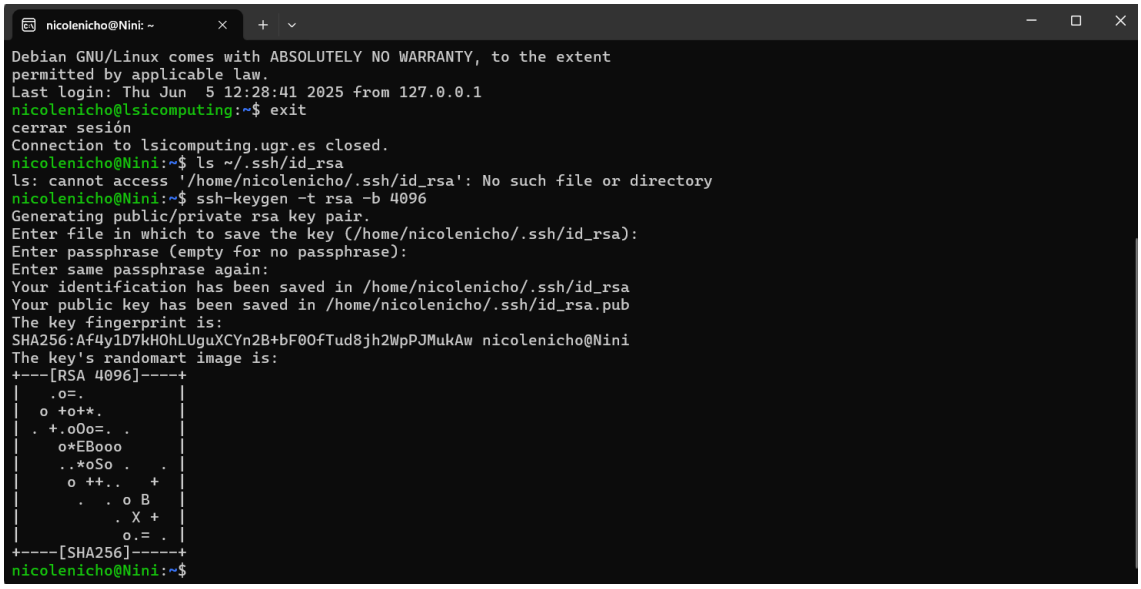



**Figura 23.** Ejemplo de proceso de creación de clave SSH

Esto generó dos archivos:

- id_rsa: la clave privada que no debe compartirse.
- id_rsa.pub: la clave pública que se puede copiar al servidor.

### Copiar la clave pública al servidor

Para autorizar el acceso sin contraseña, se copió la clave pública al servidor usando este comando en la terminal local:

**ssh-copy-id -p 2222 usuario@servidor**

Este comando copia automáticamente la clave pública generada al lugar correcto dentro del servidor para que el acceso sin contraseña funcione.

### Crear configuración SSH personalizada

Para no tener que escribir siempre -p 2222 y el usuario completo, se creó un archivo de configuración SSH. En la terminal local, se ejecutó:

```
nano ~/.ssh/config
```

Y se pegó el siguiente contenido:

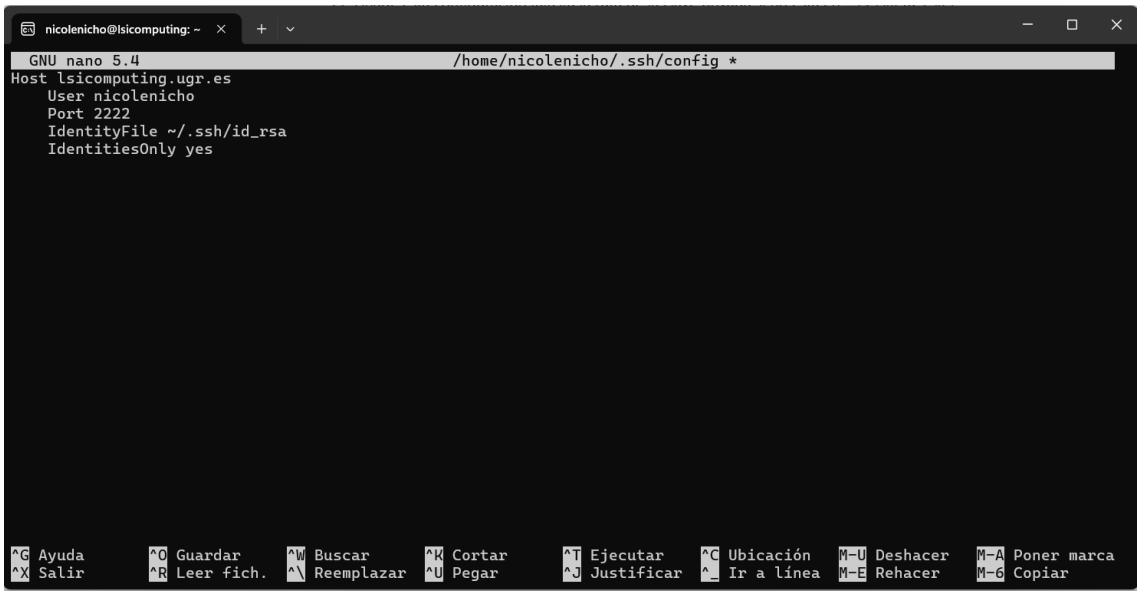



**Figura 24.** Archivo SSH Config de ejemplo

Esto indica que cada vez que se use ssh lsicomputing, el sistema sabrá a qué servidor conectarse, con qué usuario, y por qué puerto.

Guardar y cerrar con Ctrl + O, luego Enter, y luego Ctrl + X.

### Acceder al servidor de forma sencilla

Ahora ya se puede acceder al servidor simplemente escribiendo:

```
ssh [nombre de host]
```

Ya no será necesario ingresar contraseña cada vez.

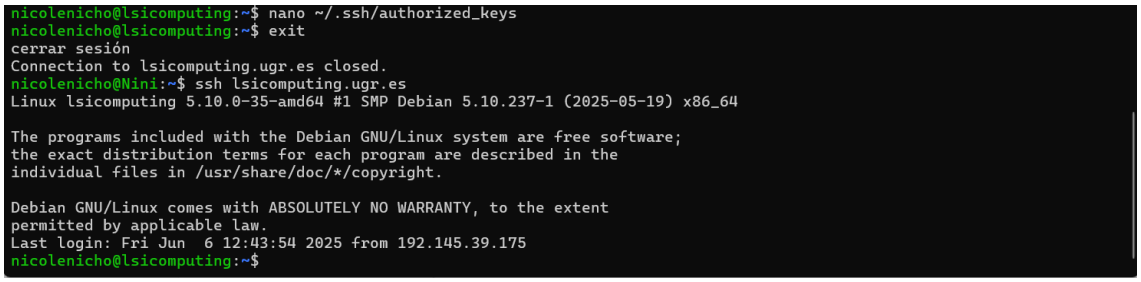



**Figura 25.** Prueba de acceso al servidor sin contraseña

## 6.2. Instalación de ComfyUI

Una vez preparado el entorno base, se procede a la instalación de **ComfyUI**. Aquí se explican las instrucciones necesarias para clonar el repositorio, instalar las dependencias, preparar los volúmenes compartidos y configurar el entorno gráfico para su ejecución en contenedores. También se incluye la estructura de carpetas utilizada para organizar los modelos, entradas y salidas.

### Crear carpetas compartidas para persistencia de archivos

Para que los archivos (modelos, imágenes, configuraciones) no se pierdan al cerrar el contenedor, es necesario crear carpetas en el servidor que serán montadas dentro del contenedor con los siguientes comandos:

```
mkdir -p ~/comfyui/models

mkdir -p ~/comfyui/inputs

mkdir -p ~/comfyui/output

mkdir -p ~/comfyui/models/checkpoints
```

Estas carpetas permiten que los datos persistan fuera del contenedor.

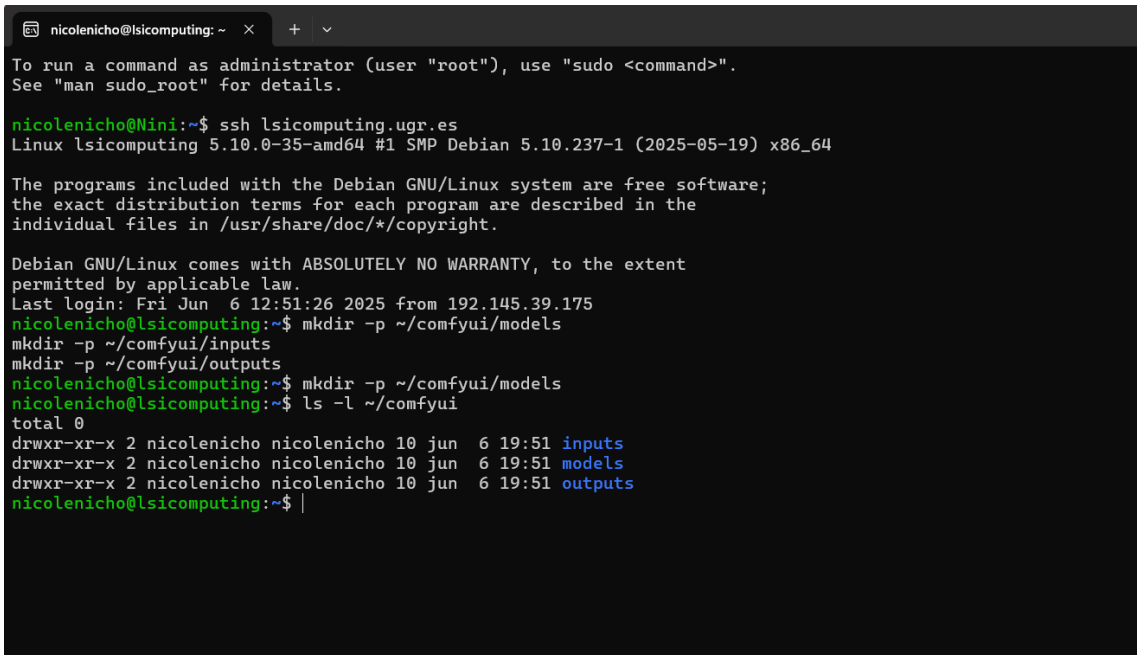



**Figura 26.** Ejemplo de carpetas creadas para ComfyUI

### Subir los archivos necesarios al servidor

Se necesitan tres archivos esenciales para construir y lanzar el contenedor:

- **Dockerfile**: contiene las instrucciones necesarias para construir la imagen del contenedor, incluyendo la instalación de todas las dependencias requeridas por ComfyUI.
- **builder.sh**: es un script que automatiza la construcción de la imagen a partir del Dockerfile utilizando la herramienta Podman.
- **launch.sh**: es un script que permite lanzar el contenedor de forma sencilla, especificando las carpetas locales que se montarán dentro del contenedor y el puerto de acceso a la aplicación.

Por ejemplo, el siguiente comando (que será ejecutado por launch.sh) lanza el contenedor y vincula las carpetas compartidas:

```
podman run --rm -it \ -v ~/comfyui/models:/workspace/models \ -v ~/comfyui/inputs:/workspace/inputs \ -v ~/comfyui/outputs:/workspace/outputs \ -p 8188:8188 \ comfyui
```

Esto vincula las carpetas locales con las del contenedor y expone el puerto 8188 para acceder a ComfyUI.

Una vez preparados los tres archivos en el equipo local, se deben transferir al servidor remoto utilizando el comando scp. A continuación, se muestra el comando utilizado, donde se copia el Dockerfile, builder.sh y launch.sh al directorio ~/comfyui/ del servidor:

**scp -P 2222 Dockerfile builder.sh launch.sh usuario@servidor:~/comfyui/**

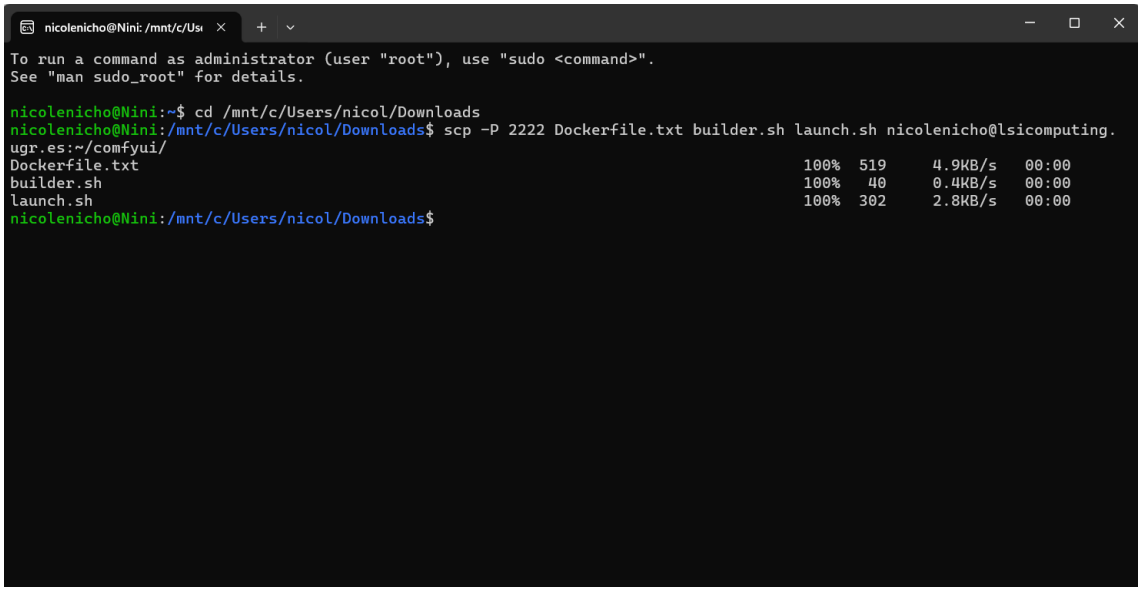



**Figura 27.** Ejemplos de comando para la copia de archivos

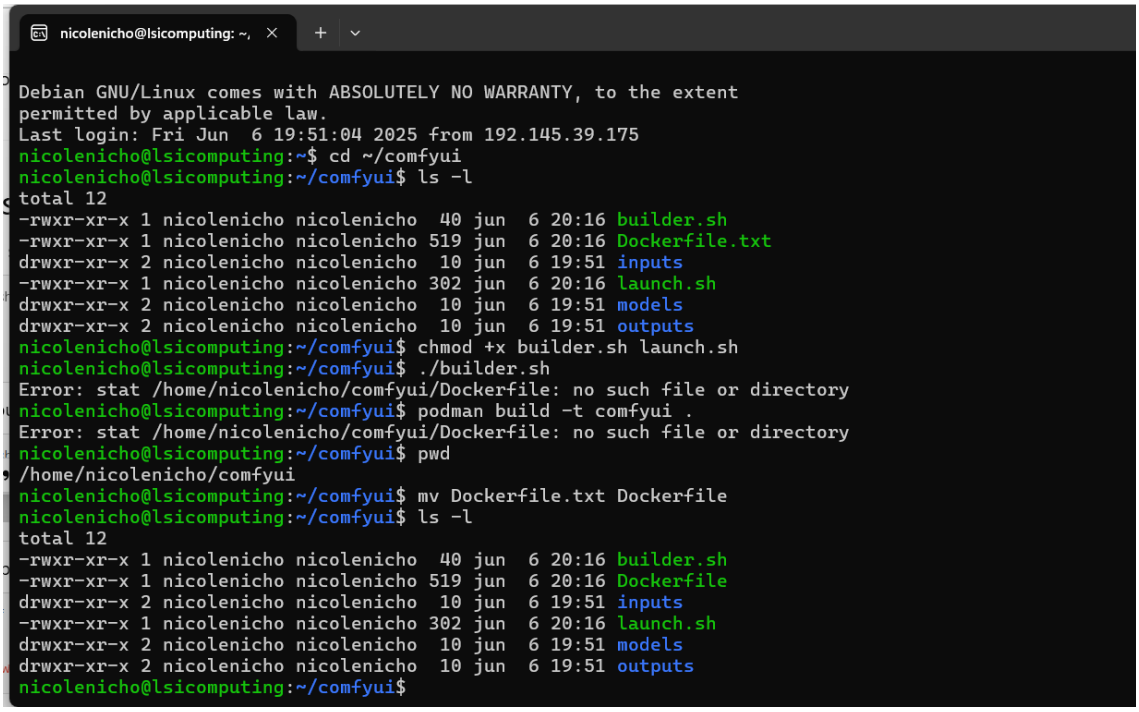



**Figura 28.** Verificación correcta de copia de archivos

## Dar permisos de ejecución a los scripts

En el servidor, entra a la carpeta donde subiste los archivos y asigna permisos:

**cd ~/comfyui chmod +x builder.sh launch.sh**

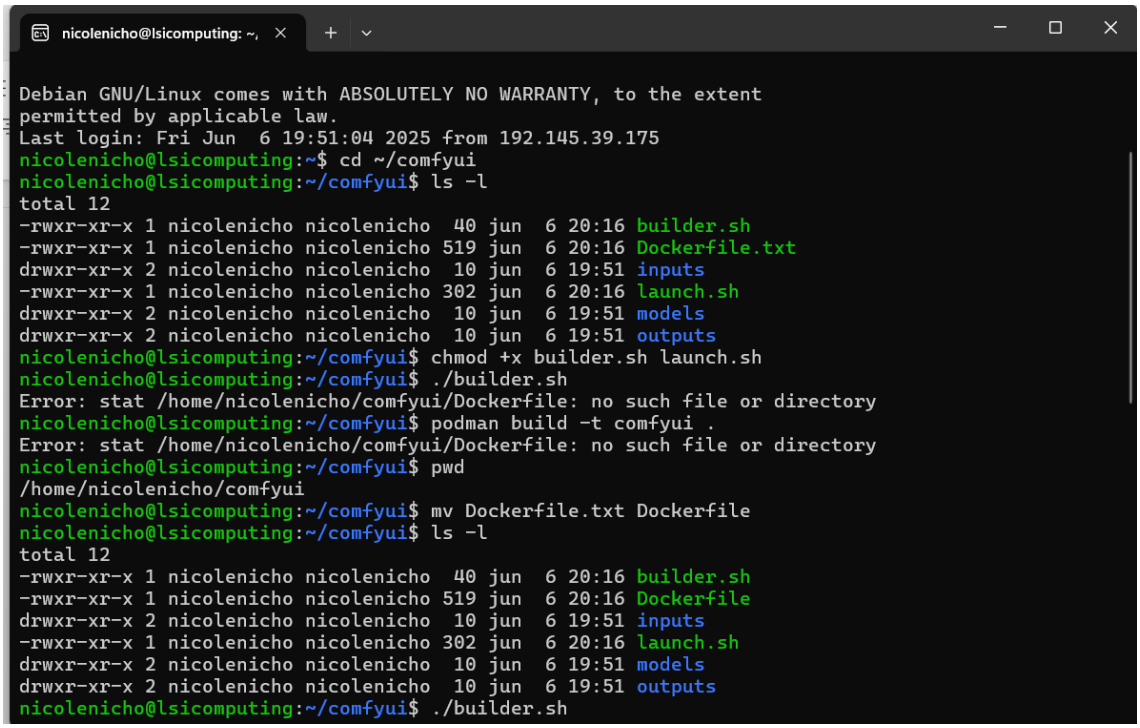



**Figura 29.** Asignación de permisos a archivos builder.sh y launch.sh

## Construir la imagen con Podman

Ejecuta el siguiente comando para construir la imagen de ComfyUI a partir del Dockerfile:

**./builder.sh**

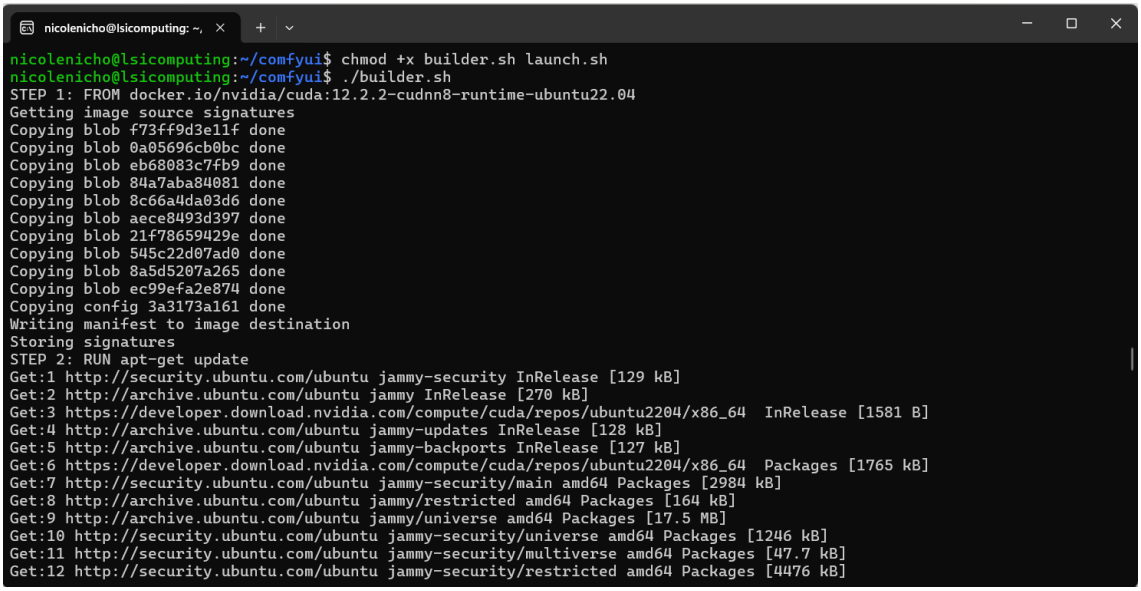



**Figura 30.** Ejemplo de activación de WSL desde el terminal

Este script ejecuta internamente un comando similar a:

```
podman build -t comfyui .
```

**Flujo Alterno con problema detectado durante el build:**

Al ejecutar la construcción (build) de la imagen de ComfyUI, el proceso se detenía en el paso 10 (instalación de Conda), arrojando el siguiente error:

```
write /var/tmp/buildah019225336/layer: no space left on device
```

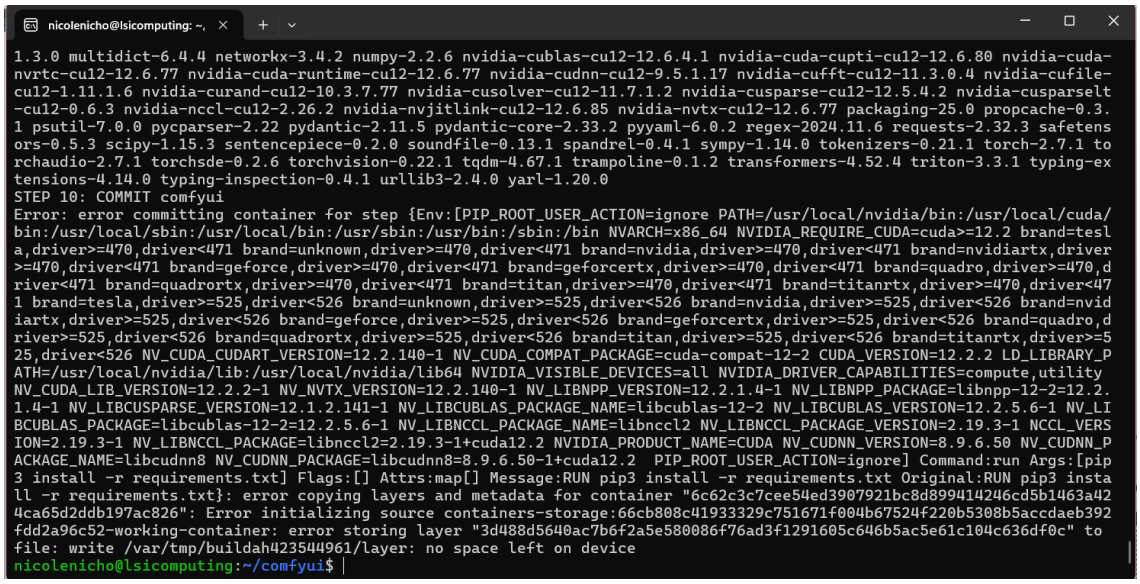



**Figura 31.** Error encontrado durante la etapa de construcción

Este error aparecía a pesar de tener espacio disponible en el directorio personal. La causa era que el sistema usaba el directorio temporal predeterminado /tmp, el cual se encuentra en la partición raíz, la cual estaba casi llena en el servidor de la universidad.

Para resolver el problema, se forzo a Podman y a los procesos internos como Conda a usar un directorio temporal alternativo ubicado dentro del home.

Para crear un directorio temporal dentro del home:

```
mkdir -p ~/tmp
```

Se exporta la variable de entorno TMPDIR antes de ejecutar el build:

```
export TMPDIR=~/tmp
```

Cuando se usa export TMPDIR=~/tmp antes de ejecutar el build con Podman, se le dice al proceso que use esa carpeta como espacio temporal para almacenar archivos durante la construcción de la imagen.

Esto redirigió las operaciones temporales como la descarga y extracción de paquetes al directorio ~/tmp, que sí contaba con suficiente espacio. Tras aplicar esta solución, el proceso de construcción se completó correctamente.

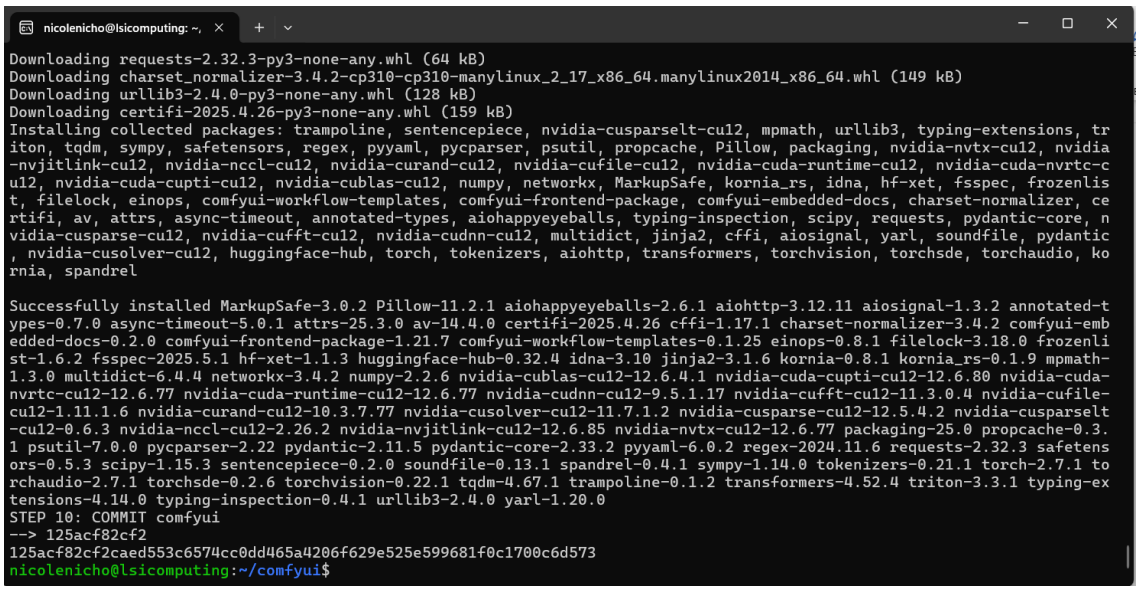



**Figura 32.** Proceso de construcción completado correctamente

## Verificar las imágenes creadas

Para listar las imágenes disponibles en Podman, se usa el siguiente comando:

```
podman image ls
```

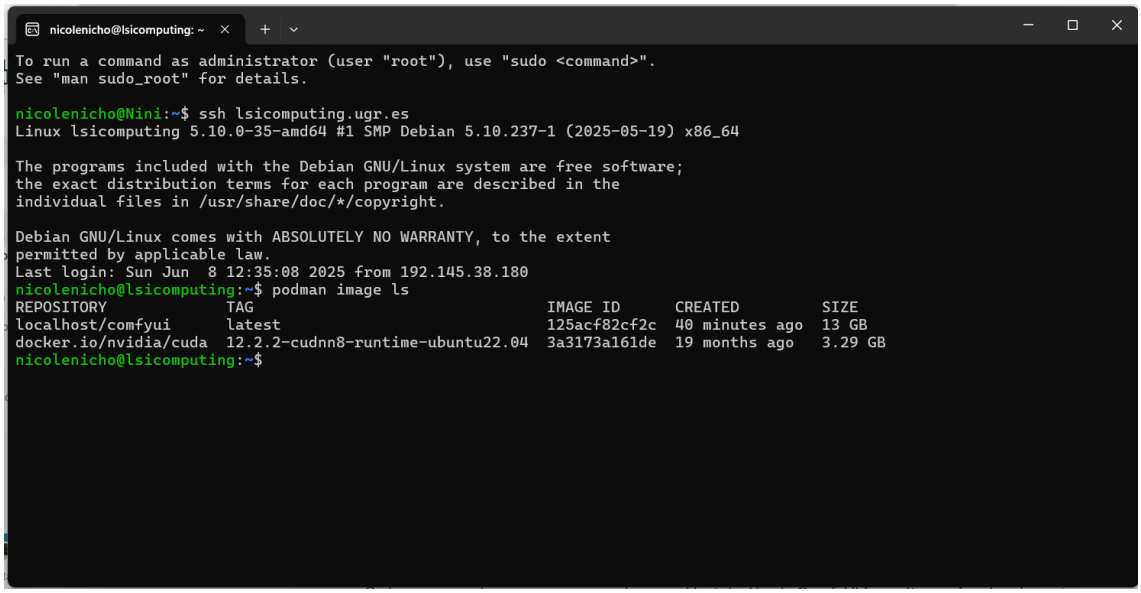



**Figura 33.** Listado de imágenes disponibles en Podman

## Monitorizar la GPU

Para ver qué GPUs están libres y su estado, se usa el siguiente comando:

**nvidia-smi -l**

Se revisa la memoria usada, procesos activos y temperatura para elegir una GPU libre.

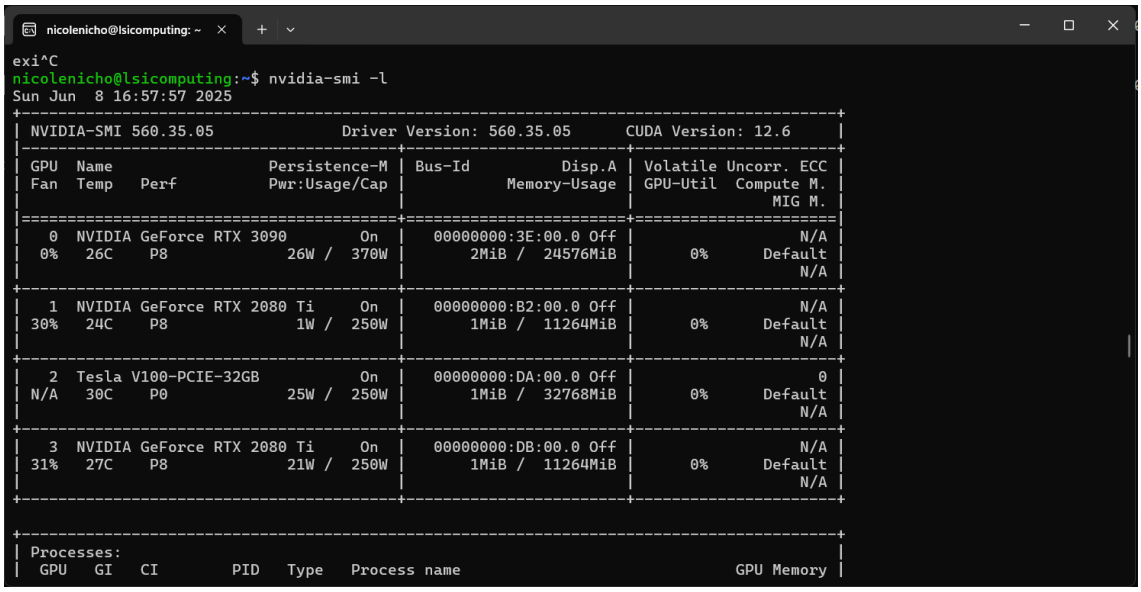



**Figura 34.** Monitoreo de GPUs

## Descarga y colocación del modelo base en ComfyUI

Se descargó el modelo v1-5-pruned-emaonly.safetensors (4.27 GB) desde Hugging Face, que es la versión optimizada para generación de imágenes y en formato seguro .safetensors.

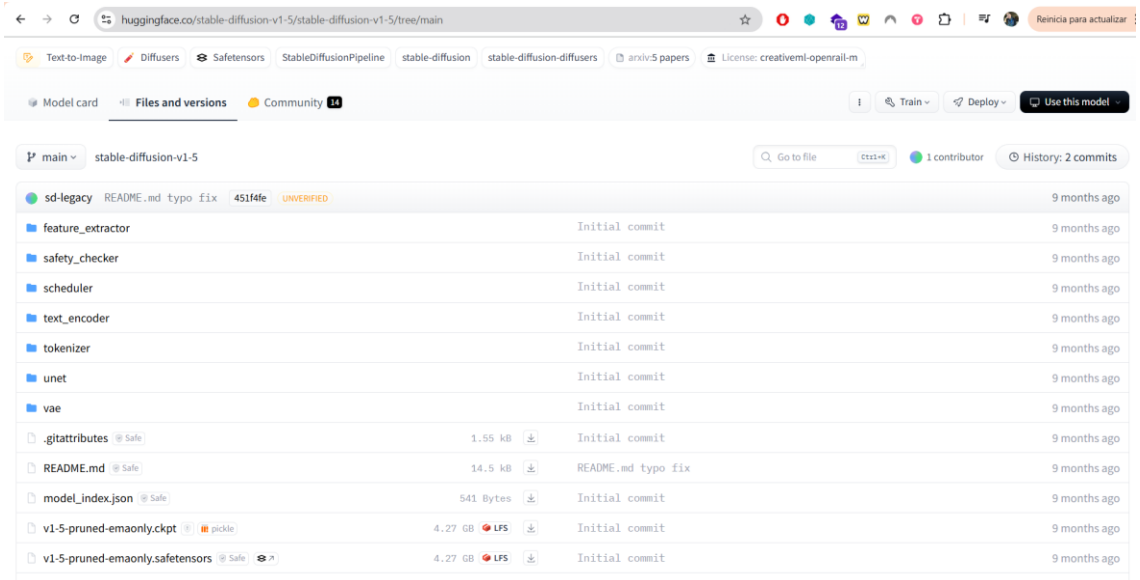


**Figura 35.** Descarga de modelos desde Hugging Face

El archivo se copió desde Windows a la carpeta de modelos en Linux, usando WSL:

```
scp -P 2222 v1-5-pruned-emaonly.safetensors usuario@servidor:
~/comfyui/models/checkpoints
```

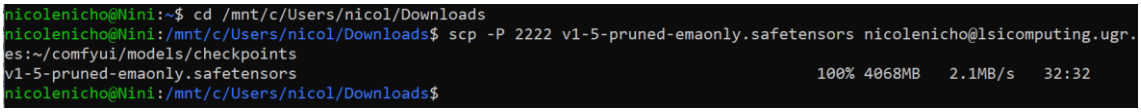



**Figura 36. Copia del modelo descargado a la carpeta de modelos de ComfyUI**

Se verificó que la ruta y el archivo existieran correctamente con el comando:

```
ls ~/comfyui/models/checkpoints/
```

### Montar las carpetas en el contenedor

Si se corre ComfyUI dentro de un contenedor sin montar carpetas compartidas o volúmenes, los archivos que se generen como modelos, imágenes, configuraciones, se almacenan dentro del contenedor y se perderán cuando se detenga o elimine.

Para ello antes de lanzar el contenedor de ComfyUI en el archivo launch.sh se debe montar las carpetas inputs, models y outputs con -v y mapearlas a las rutas que ComfyUI espera dentro del contenedor, se puede acceder al editor a través del comando:

```
nano launch.sh
```

Así los datos que se guarden dentro del contenedor se sincronizarán con esas carpetas en el sistema host y aunque se detenga o se borre el contenedor, los archivos permanecen en el disco y no se pierden.

Por ejemplo, si ComfyUI usa /ComfyUI/models, /ComfyUI/inputs y /ComfyUI/outputs, el comando quedaría:

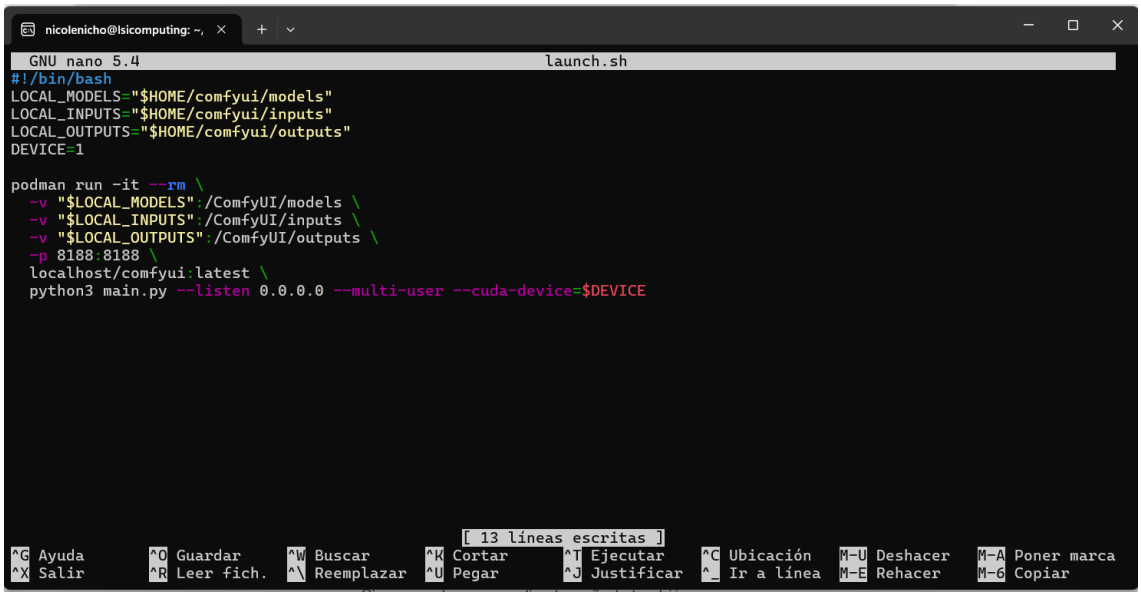



**Figura 37.** Archivo launch.sh con las carpetas compartidas montadas

Se debe tomar en cuenta que según la estructura oficial de ComfyUI:

- **models**: carpeta para almacenar los modelos de IA.
- **inputs**: carpeta donde se colocan imágenes o datos que quieres cargar en ComfyUI .
- **outputs**: carpeta donde ComfyUI guarda las imágenes o resultados generados.

### Lanzar ComfyUI especificando la GPU

El servidor cuenta con varias GPUs. Para seleccionar la GPU a usar, se modificó el script launch.sh para incluir el argumento --cuda-device con el ID de la GPU deseada.

Esto garantiza que ComfyUI use la GPU correcta, respetando cualquier política de reserva de recursos, por ejemplo para el uso del GPU 1:

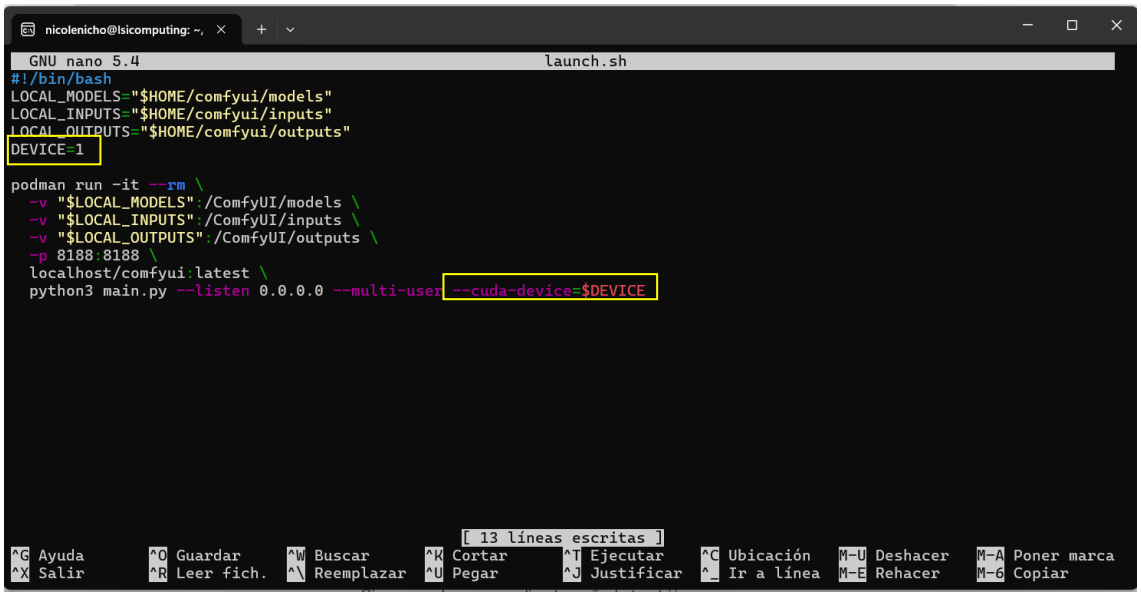



**Figura 38.** Especificación de usa de GPU en el launch.sh

Si ya se tiene el modelo base por ejemplo v1-5-pruned-emaonly.safetensors en la carpeta checkpoints dentro de models y el script launch.sh está correctamente configurado para montar esa carpeta, entonces solo se tiene que ejecutar este comando para iniciar el contenedor con acceso a las carpetas compartidas y el puerto para la interfaz web:

```
./launch.sh
```

Esto arrancará ComfyUI, que detectará automáticamente el modelo y estará listo para usarlo sin necesidad de configuraciones adicionales

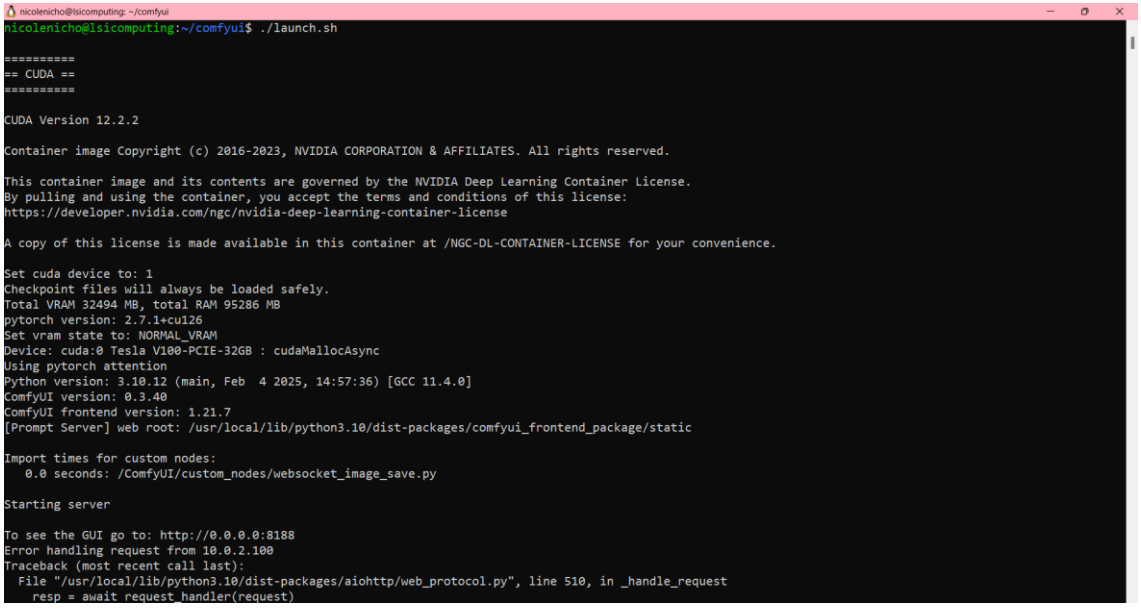



**Figura 39.** Lanzamiento de ComfyUI

**Acceso a la interfaz gráfica de ComfyUI desde local (Tunneling)**

En este caso, el servidor está en una máquina remota que es accesible por SSH en servidor puerto 2222 y ComfyUI corre en un puerto local por defecto 8188 que no está expuesto directamente a Internet. Para acceder a la interfaz gráfica desde el navegador local, se necesita crear un túnel SSH que redirija ese puerto remoto a la máquina local.

El túnel SSH es una conexión segura que permite que el tráfico de un puerto en la máquina local se redirija a un puerto en el servidor remoto o viceversa.

Como ComfyUI corre en el servidor remoto y expone la interfaz en el puerto 8188, se creó un túnel SSH desde la máquina local para redirigir ese puerto y poder abrir la interfaz en el navegador:

**ssh -L 8188:localhost:8188 -p 2222 usuario@servidor**

- -L 8188:localhost:8188: Esto significa que se está creando un túnel SSH local. El puerto 8188 de tu máquina local se redirige al puerto 8188 del localhost en el servidor remoto es decir, el mismo servidor al que te estás conectando, servidor.
- -p 2222: especifica que el puerto 2222 es el que se utiliza para establecer la conexión SSH. Este es el único puerto realmente abierto al exterior del servidor.
- usuario@servidor es el usuario y servidor.

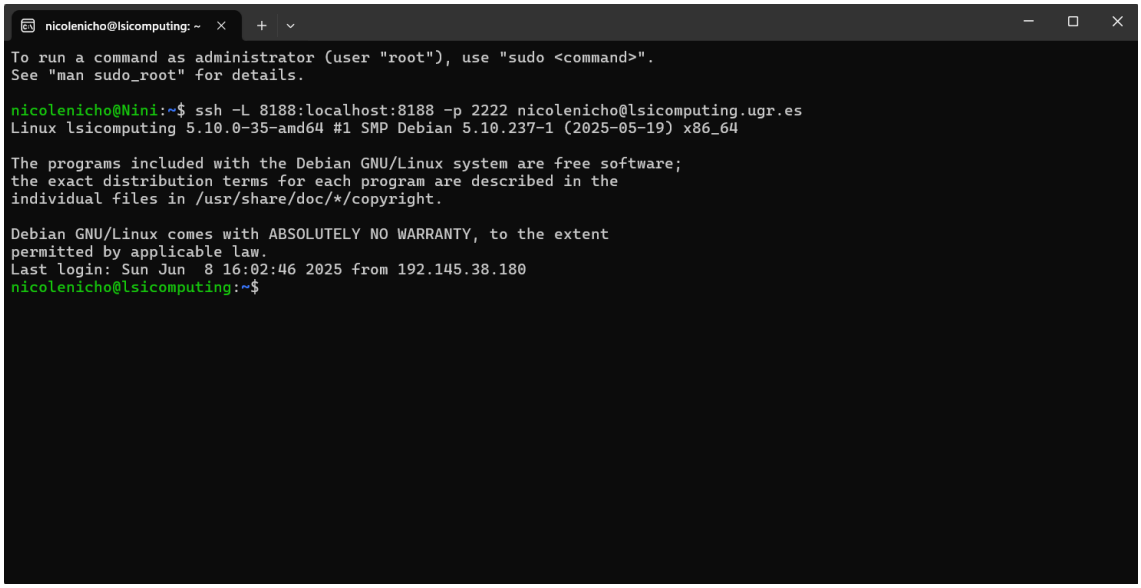



**Figura 40.** Ejemplo de tunneling para acceder a la interfaz gráfica

Esta conexión se mantiene abierta mientras se usa ComfyUI.

Una vez creado el túnel, se debe abrir el navegador en la máquina local:

**http://localhost:8188**

Así se accede a la interfaz gráfica de ComfyUI que corre en el servidor.

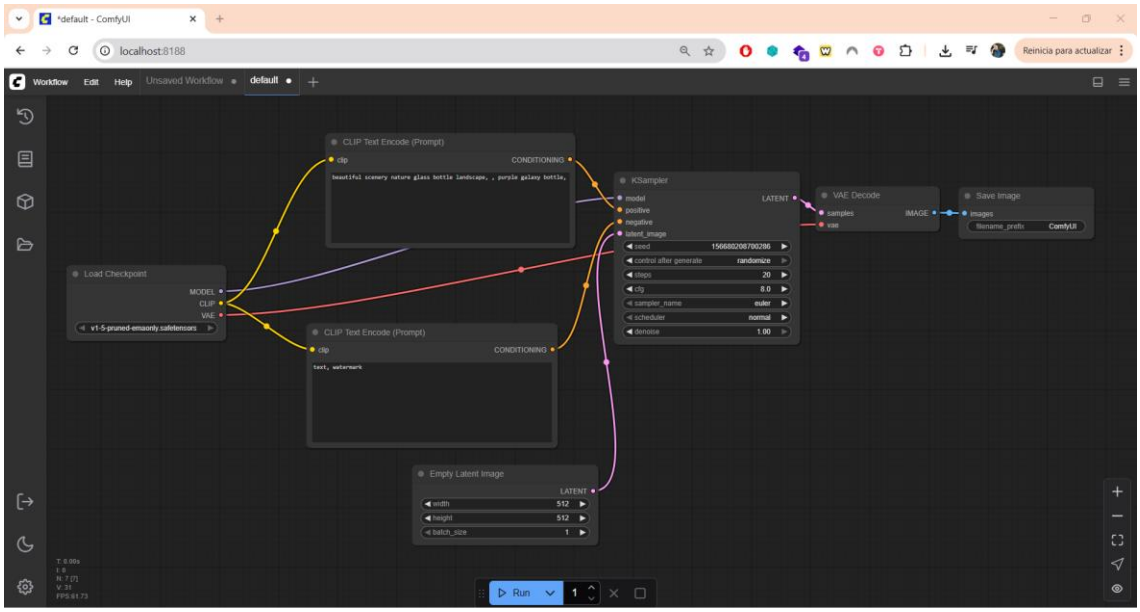



**Figura 41.** Interfaz gráfica de ComfyUI a través del navegador

### Detener el contenedor

Para detener el contenedor, simplemente se presiona Ctrl + C en la terminal donde se está ejecutando.

### Personalización avanzada de la instalación de ComfyUI: gestión de custom nodes y usuarios

Durante la instalación y puesta en marcha de ComfyUI en un entorno de servidor, fue necesario realizar ajustes adicionales para garantizar tanto la correcta gestión de usuarios como la integración de carpetas personalizadas y la compatibilidad con la estructura de almacenamiento del servidor.

Se adaptó el archivo de arranque **launch.sh** para incluir la ruta al directorio **custom_nodes**, asegurando que ComfyUI los cargara correctamente al iniciar. Esto permitió instalar y gestionar paquetes como **ComfyUI-Manager,** para facilitar la gestión de nodos y modelos con su interfaz más amigable. También se verificó el montaje de volúmenes como user_data, necesario para el uso con --multi-user.

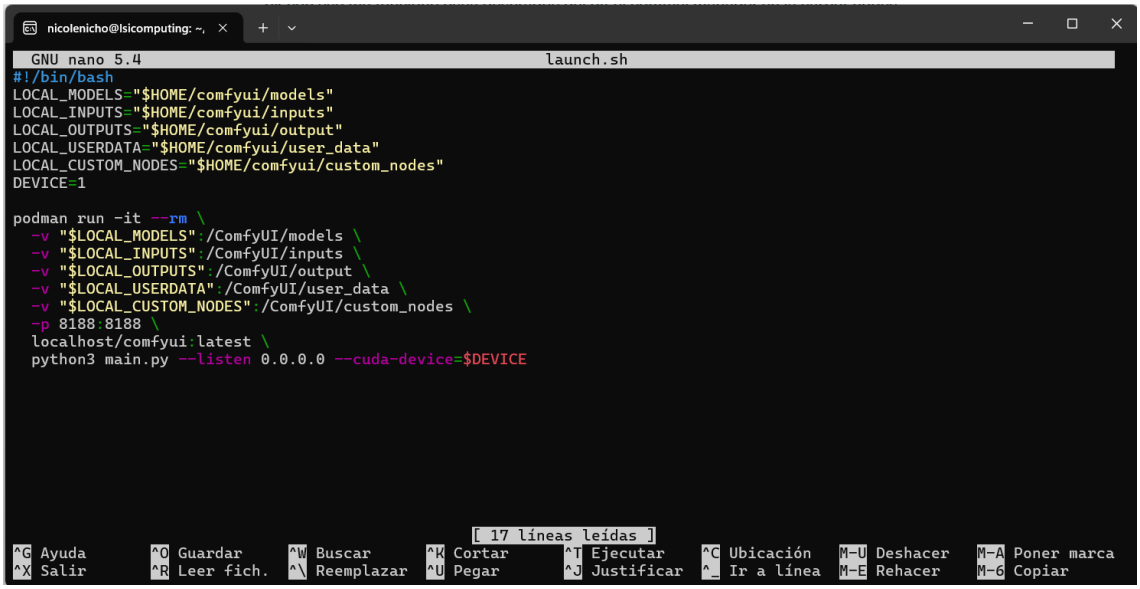



**Figura 42.** Adaptación de archivo launch.sh

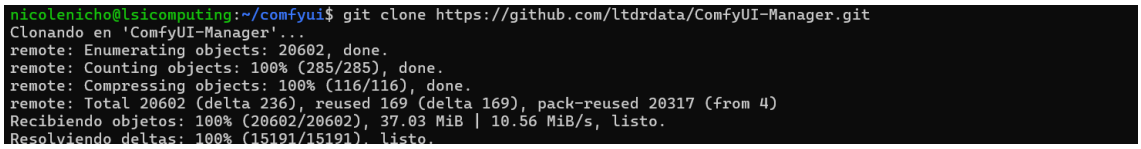



**Figura 43.** Descarga de ComfyUI Manager desde el terminal

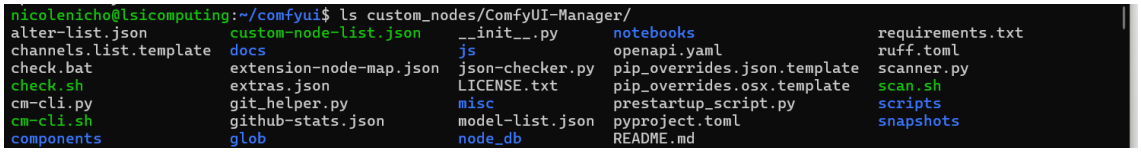



**Figura 44.** Verificación de instalación de ComfyUI

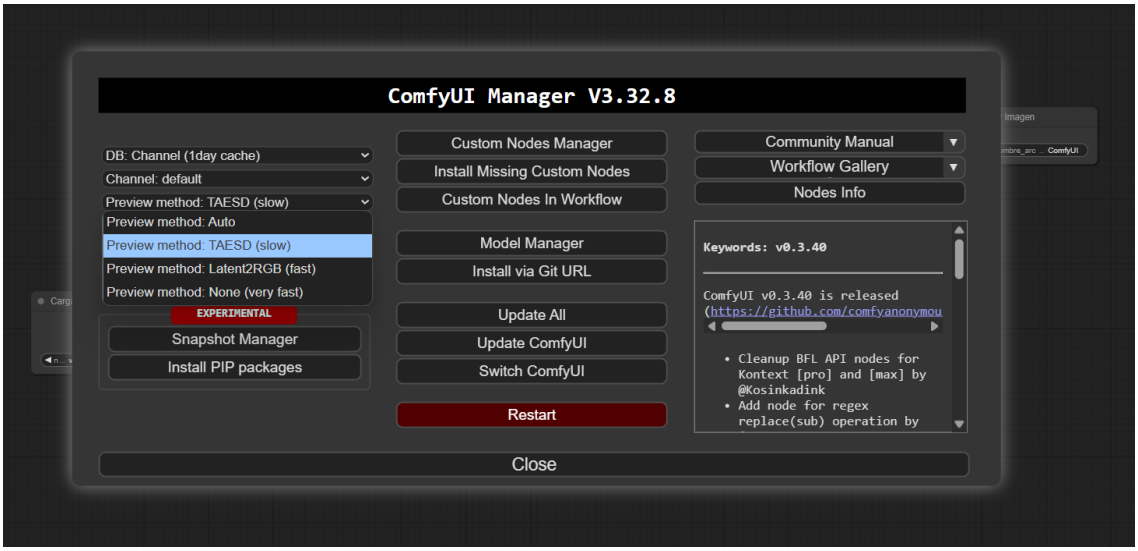



**Figura 45.** Interfaz de ComfyUI Manager desde ComfyUI

Como el sistema user_management no se habilitó automáticamente, se copió manualmente al directorio principal. Se creó la carpeta user_data/default con un config.json válido y se ajustaron permisos para evitar errores como KeyError: **'Unknown user: default'.**

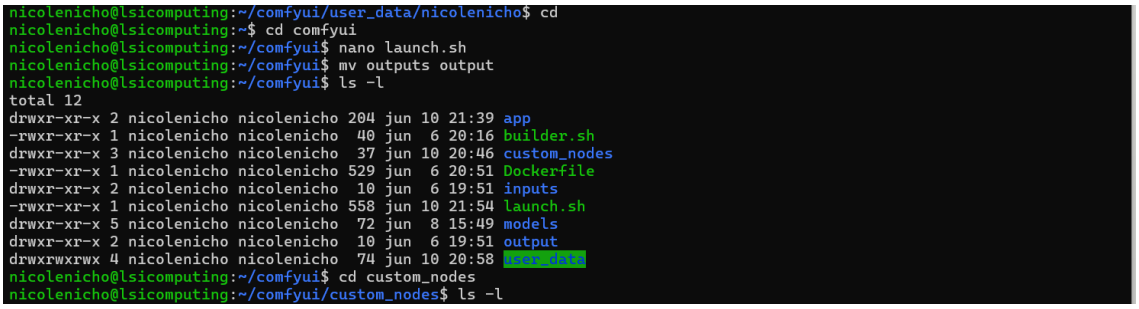



**Figura 46.** Verificación de directorio en ComfyUI

Para comenzar con las pruebas, se descargaron los siguientes modelos oficiales de Stable Diffusion que están descritos en el marco teórico:

- **SD 1.5**: Modelo base estándar, rápido y eficiente
- **SDXL 1.0**: Versión mejorada con mayor calidad de imagen
- **SDXL Turbo**: Versión optimizada para generación rápida

Aunque existen modelos alternativos como Flux.1, que ofrecen características innovadoras en generación de texto y velocidad, este trabajo se centra en los modelos oficiales de Stable Diffusion 1.5, SDXL y SDXL Turbo debido a su amplia adopción, documentación y soporte en la comunidad.

## 6.3. Exploración práctica de ComfyUI

En esta sección se documentan los pruebas prácticas realizadas con ComfyUI, con el objetivo de mostrar sus posibilidades y facilitar su adopción a personas sin experiencia previa en herramientas de inteligencia artificial generativa. Cada pruebas explora una funcionalidad específica, utilizando modelos populares como Stable Diffusion 1.5, SDXL, SDXL Turbo y modelos complementarios como ControlNet, LoRA, entre otros presentando nodos utilizados, flujos de trabajo y análisis de resultados para cada caso.

### 6.3.1. Generación de imágenes desde texto (Text-to-Image)

Este caso práctico consiste en crear imágenes a partir de descripciones escritas también conocidas como prompts. Es la forma más básica y directa de generar imágenes con IA, ideal para comenzar a explorar ComfyUI y entender cómo el texto guía la creación visual. ComfyUI incluye un workflow predefinido para esta función.

Se probó el mismo prompt con diferentes configuraciones:

**Prompt:**

*"A luxurious glass perfume bottle with golden cap, placed on a soft pink pastel background, delicate rose petals scattered around, soft studio lighting, subtle reflections, bokeh highlights, ultra-realistic, high detail."*

**Modelos comparados:**

- SD 1.5
- SDXL 1.0
- SDXL Turbo

**Nodos usados y conexiones:**

- Load Checkpoint: carga el modelo (SD1.5, SDXL o SDXL Turbo).
- CLIP Text Encode (Prompt positivo): convierte el texto descriptivo en vectores.
- CLIP Text Encode (Prompt negativo): convierte el texto para evitar elementos no deseados.
- KSampler: realiza la generación de la imagen latente usando el modelo y los embeddings de texto.
- VAE Decode: convierte la imagen latente en imagen visible.
- Save Image / Viewer: muestra o guarda la imagen generada.

El flujo conecta el modelo cargado a KSampler, que recibe también los vectores de texto positivo y negativo, y envía la imagen latente a VAE Decode para obtener la imagen final.

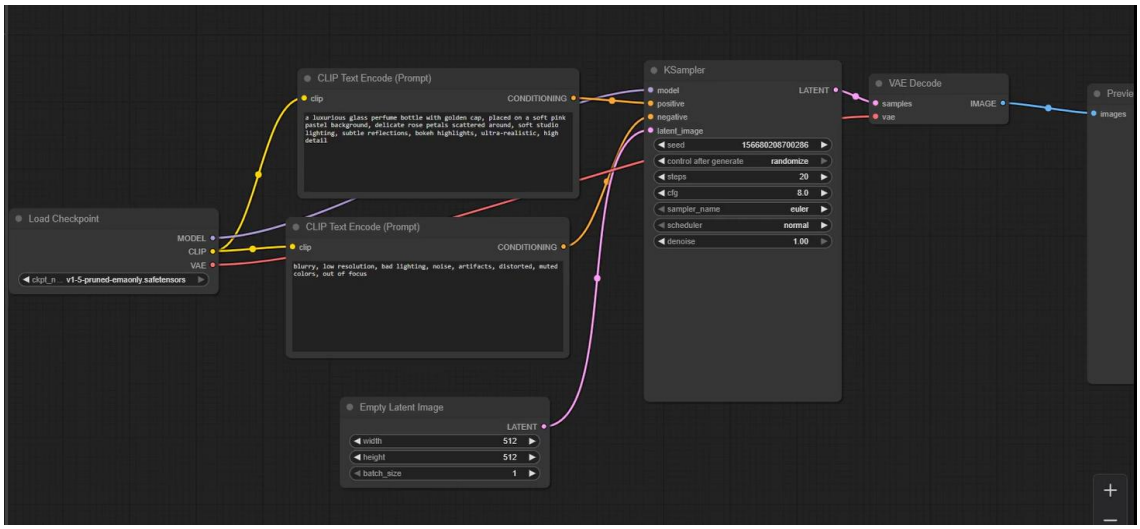



**Figura 47.** Ejemplo de workflow de Text to Image Comfy UI

## Proceso paso a paso:

- En el nodo “Checkpoint Loader”, se seleccionó uno de los modelos descargados: SD1.5, SDXL o SDXL Turbo.
- Se escribe el prompt positivo y negativo en los nodos CLIP Text Encode.
- Se configuran parámetros en KSampler: sampler (Euler o DPM2 Karras), CFG (6-8), número de pasos (20-30).
- Se ejecutó la generación haciendo clic en “Queue Prompt” y se repitió el proceso con cada modelo y diferentes configuraciones.

## Análisis de resultados:

- SD1.5 generó imágenes rápidamente, aunque con menor nivel de detalle y menos realismo en su composición.

**Figura 48.** Resultados de Text to Image para SD1.5 con Euler (izquierda) y Dpm_2 (derecha)

- SDXL produjo imágenes más detalladas y con iluminación más natural, aunque el tiempo de espera fue mayor.

**Figura 49.** Resultados de Text to Image para SDXL con Euler (derecha) y Dpm_2 (izquierda)

- SDXL Turbo permitió obtener resultados en menos tiempo, pero con menor riqueza visual, y sobresaturación en la mayoría de imágenes generadas.

**Figura 50.** Resultados de Text to Image para SD1.5 con Euler (derecha) y Dpm_2 (izquierda)

- En general, SDXL destacó por el equilibrio entre calidad y fidelidad a la descripción, mientras que SDXL Turbo resultó útil para pruebas rápidas pero las imagénes no se generaron con la mejor calidad visual.
- El sampler influye en los detalles y texturas: DPM2 genera resultados con más contraste; Euler a veces da resultados más suavizados.
- Cambiar el CFG scale cambia la fidelidad del resultado al prompt porque valores más altos siguen mejor el texto, pero también pueden introducir artefactos que pueden producir sobresaturación o una imagen no tan realista.

### 6.3.2. Transformación de imagen a imagen (Image-to-Image)

Este caso práctico transforma una imagen base según un prompt, permitiendo modificar o estilizar imágenes existentes manteniendo su composición. ComfyUI incluye un workflow predefinido para esta función llamado “Image to Image”.

Se probó el mismo prompt con diferentes configuraciones:

*“A futuristic cityscape, realistic, skyscrapers with neon lights, advanced technology, ultra-detailed, night, sunset, 8k”*

**Modelos comparados:**

- SD 1.5
- SDXL 1.0
- SDXL Turbo

**Nodos usados y conexiones:**

- Load Image: carga la imagen base.
- Latent Encode: codifica la imagen en espacio latente.
- CLIP Text Encode (positivo y negativo): procesa el prompt.
- Load Checkpoint: carga el modelo.
- KSampler: genera la imagen latente modificada, usando la imagen codificada como punto de partida. Se puede controlar qué tanto de la imagen de entrada se usa para influir en la imagen de salida con el parámetro **denoise**, Si el valor es bajo, la imagen final se parece mucho a la original. Si es alto, la imagen cambia bastante y puede verse muy diferente.
- VAE Decode: convierte la latente a imagen visible.
- Save Image / Viewer: muestra o guarda la imagen final.

El flujo conecta la imagen cargada a Latent Encode, que se conecta a KSampler junto con el modelo y los vectores de texto, para luego decodificar y mostrar la imagen.

**Proceso paso a paso:**

- Se seleccionó el flujo de trabajo “image to image”.
- En el nodo de carga de imagen (“Load Image”), se subió una fotografía de Nueva York.
- Se codifica con Latent Encode y si es necesario antes se conecta a un nodo de Upcale Image para que la imagen se reduzca de tamaño.
- Se escribe el prompt positivo y negativo en CLIP Text Encode.
- Se seleccionó el modelo deseado y se ajustaron los parámetros de muestreo, denoise que se utiliza qué porcentaje de la imagen de entrada se usa para influir en la imagen de salida, CFG y steps.
- Se generó la imagen y se repitió el proceso con los diferentes modelos.

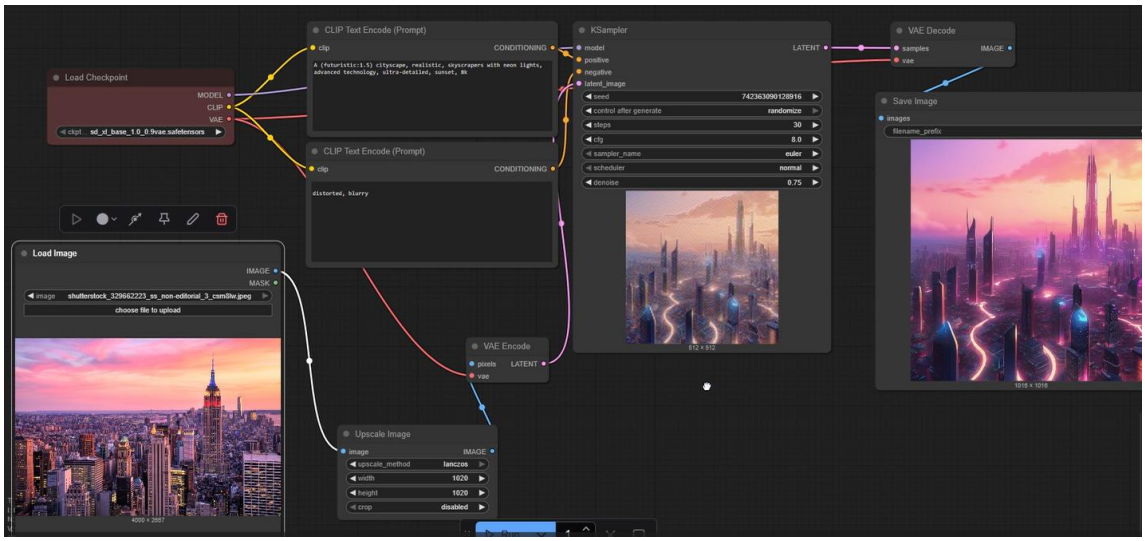



**Figura 51.** Ejemplo de workflow Image to Image

## Análisis de resultados:

- SDXL y SDXL Turbo respetan muy bien la estructura y composición de la imagen original, añadiendo detalles y efectos según el prompt. Sin embargo, SDXL Turbo se apega un poco más al prompt y muestra una ciudad más futuristica con un denoise de 0.5, mientras SDXL no incorpora tantos elementos del prompt y toma más en cuenta la imagen base.

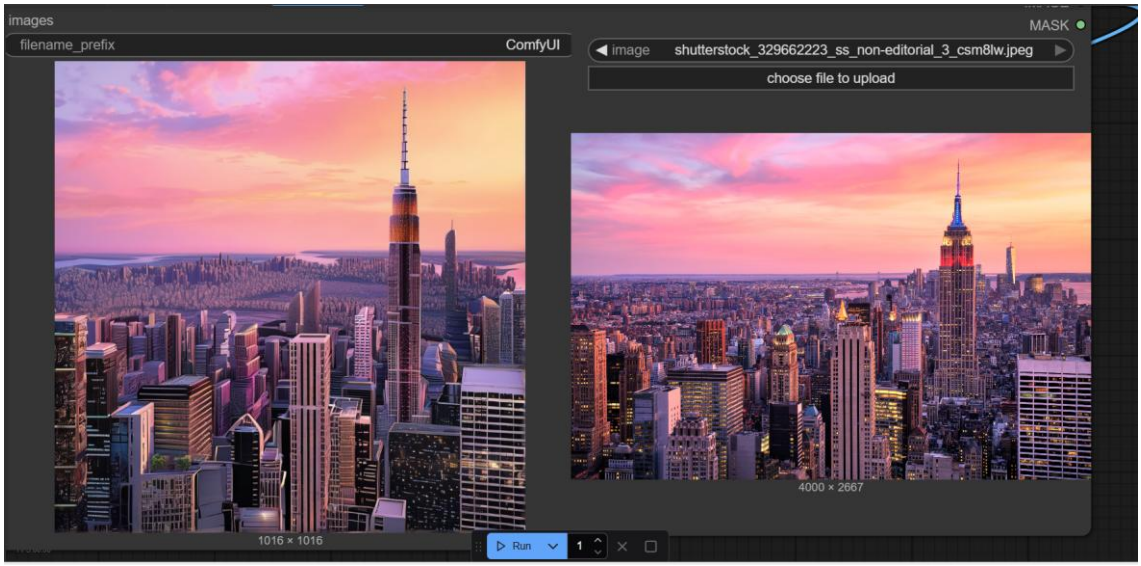



**Figura 52.** Imagen generada con Image to Image para SDXL con Dpm_2 (izquierda) e imagen original subida (derecha)


**Figura 53.** Imagen generada con Image to Image para SDXL Turbo con Dpm_2 (izquierda) e imagen original subida (derecha)

- SD1.5 puede generar resultados más variables, a veces alterando demasiado la imagen base con los mismos parámetros.


**Figura 54.** Imagen generada con Image to Image para SD1.5 con Dpm_2 (izquierda) e imagen original subida (derecha)

- Ajustar el **denoise**, además del CFG y los pasos permite controlar cuánto se modifica la imagen original.
- Este método es útil para crear variaciones o estilizaciones manteniendo la esencia visual.

- **Nota:** Al utilizar un prompt de generación de imágenes para retratos de una mujer usando un vestido traje típico color rojo con el fondo de España, con una imagen de una modelo vistiendo una traje típico, la influencia de la imagen base fue menor en todos los modelos, ya que la IA tendió a reinterpretar el rostro con variaciones y en los 3 modelos las caras se mostraron distorsionadas, sin mucho realismo, y los vestidos no tomaban en cuenta el patron del vestido de la imagen base a pesar de probar diferentes configuraciones como se muestra a continuación:

**Figura 55.** Ejemplo de imágenes generadas (Image to Image) , imagen base a la izquierda, imágenes generadas con modelos SD1.5,SDXL,SDXL Turbo a la derecha

### 6.3.3. Control avanzado con ControlNet

Este caso práctico consiste en probar ControlNet el cual permite condicionar la generación con mapas específicos de bordes, profundidad, poses,etc lo que facilita obtener resultados con estructuras o composiciones específicas.

Se probó el mismo prompt con diferentes modelos de ControlNet:

*" Instagram photo, closeup photo of a 18 year old woman, beautiful face, makeup, city Street bokeh, motion blur"*

**Modelo usado:**

- SD 1.5

**Nodos usados y conexiones:**

- Load Image: carga la imagen guía.

- Preprocesamiento (Canny Edge Detector, Depth Estimation, OpenPose): transforma la imagen en un mapa guía.
- Load ControlNet Model: carga el modelo ControlNet específico (canny, depth, openpose, scribble).
- ControlNet: recibe el mapa guía y modula la generación.
- Load Checkpoint: carga el modelo base.
- CLIP Text Encode: procesa el prompt.
- KSampler: genera la imagen latente condicionada.
- VAE Decode: decodifica la imagen latente.
- Save Image / Viewer: muestra o guarda la imagen.

ControlNet se conecta a KSampler junto con el modelo base y los embeddings de texto, usando el mapa guía para condicionar la generación.

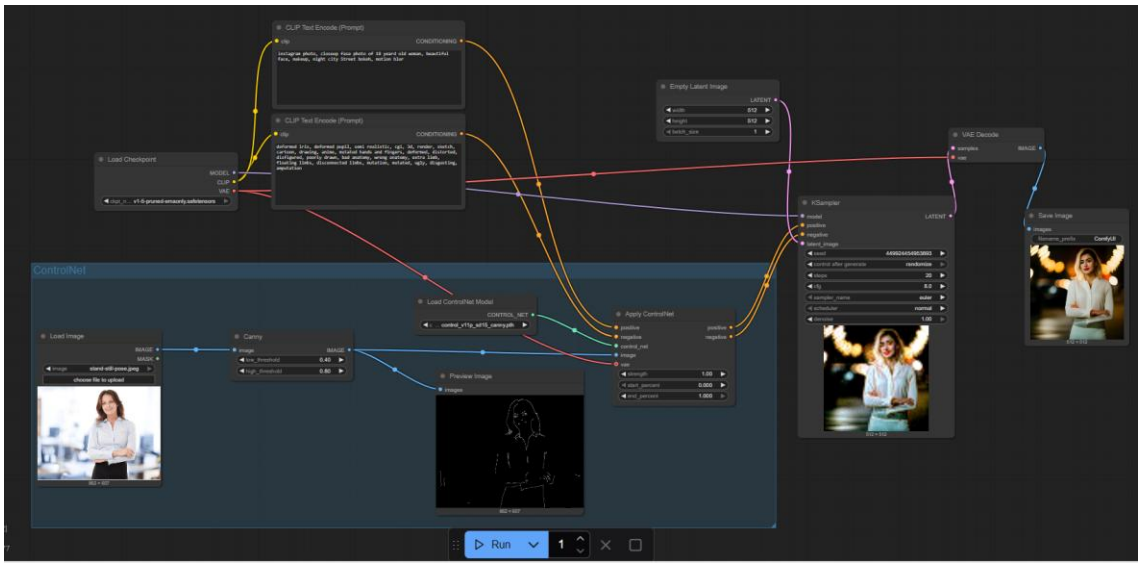



**Figura 56.** Ejemplo de workflow de ControlNet

**Proceso paso a paso:**

- Se descargaron los modelos más populares para SD 1.5:
    - Canny: Detección de bordes
    - Depth: Mapeo de profundidad
    - OpenPose: Detección de poses humanas

- Scribble: Dibujos simples como guía

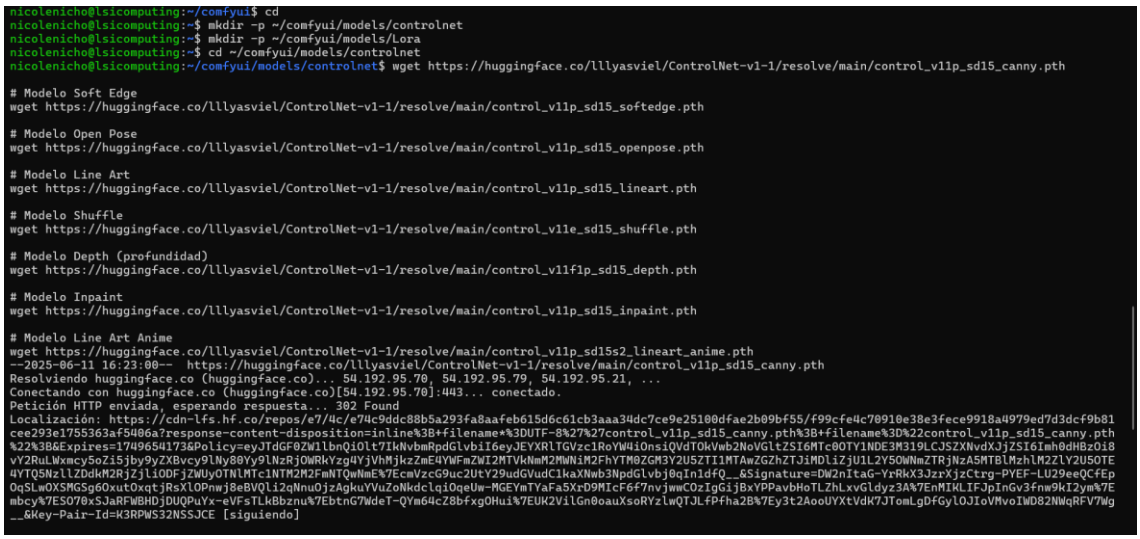



**Figura 57.** Proceso en el terminal para descargar modelos de ControlNet.

- Se instalan los nodos de ControlNet (canny, depth, openpose, scribble) a través de Comfy Manager.

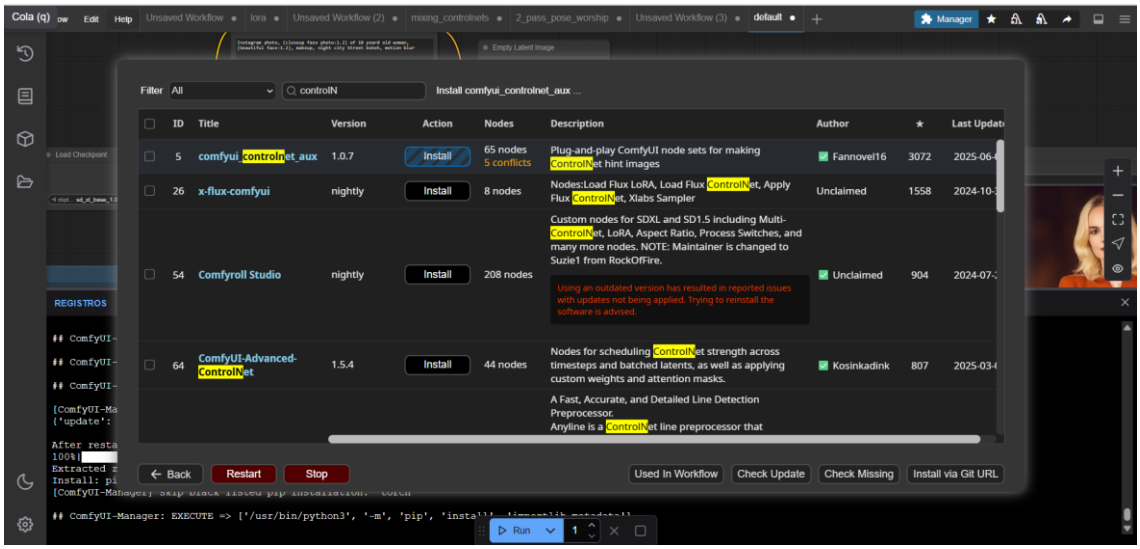



**Figura 58.** Instalacion de nodos de ControlNet a través de ComfyManager

- Se sube una imagen guía (por ejemplo, una mujer de brazos cruzados) y se generaron diferentes mapas dependiendo de lo que se quiere lograr por ejemplo se conecta el nodo canny para el modelo Canny de Control Net.

- Se carga el modelo ControlNet adecuado para el ejemplo anterior por ejemplo se debe seleccionar Canny.

- Se conecta ControlNet a KSampler para condicionar la generación.

- Se escribe el prompt y negativo.
- Se configuran parámetros y se genera la imagen, se probó cada tipo de ControlNet y se observaron los resultados.

**Análisis de resultados:**

- El uso de ControlNet facilitó la generación de imágenes con poses o estructuras definidas.

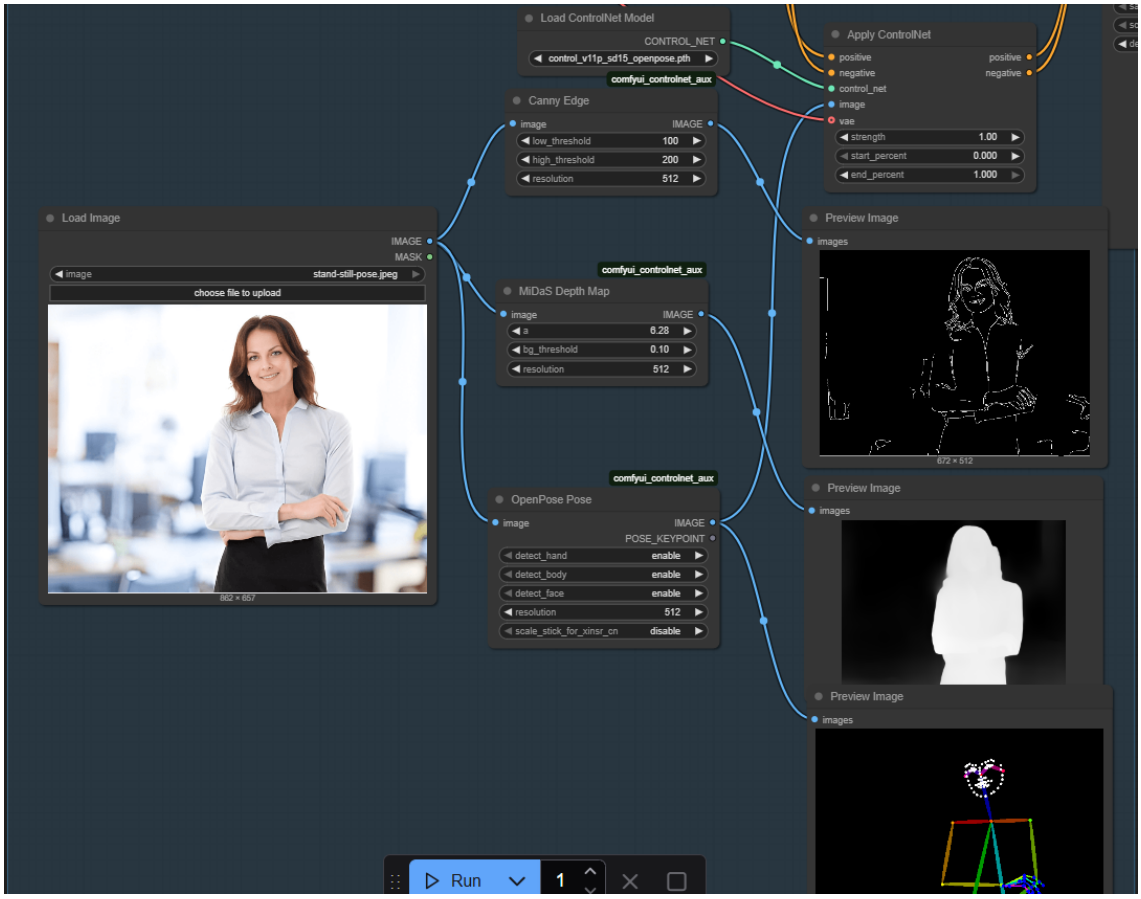



**Figura 59.** Resultados generados del mapa de imagen para Canny, Depth y OpenPose

- El modelo depth fue especialmente efectivo para mantener la pose y añadir realismo tridimensional especialmente en la cara y composicion de la imagen.

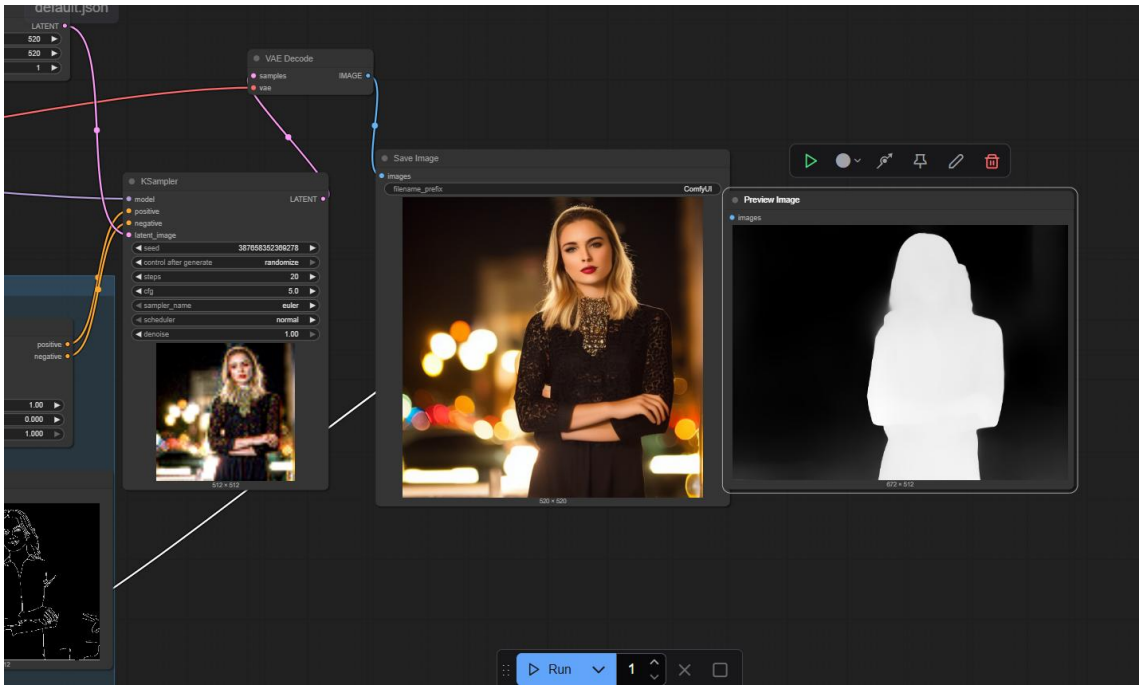



**Figura 60.** Resultado de Control Net: Depth

- Canny permitió controlar los contornos principales, mientras que OpenPose ofreció control sobre la postura del cuerpo.

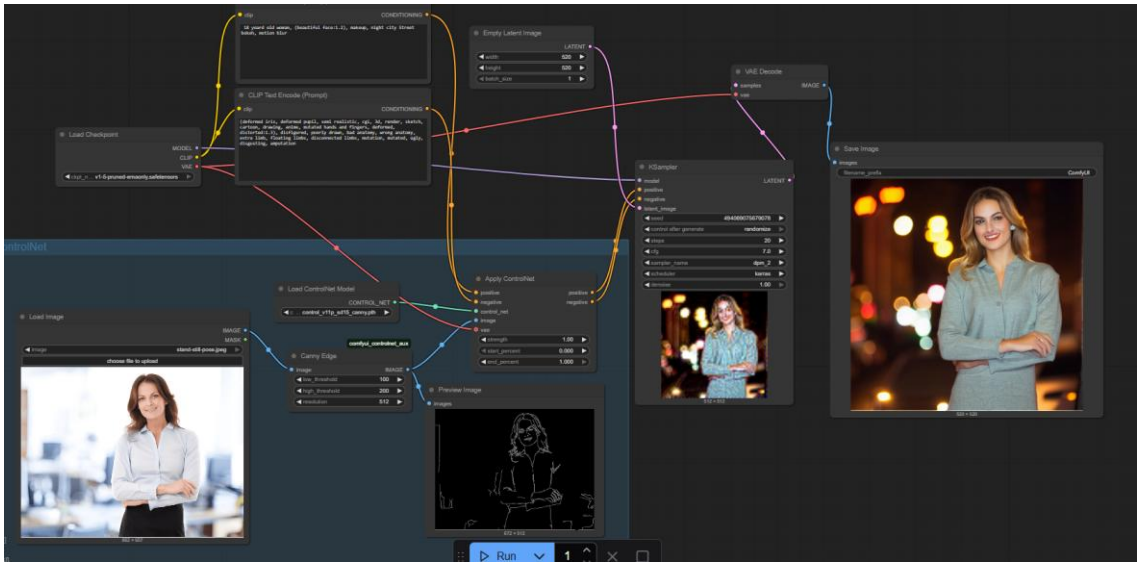



**Figura 61.** Resultado de Control Net: Canny

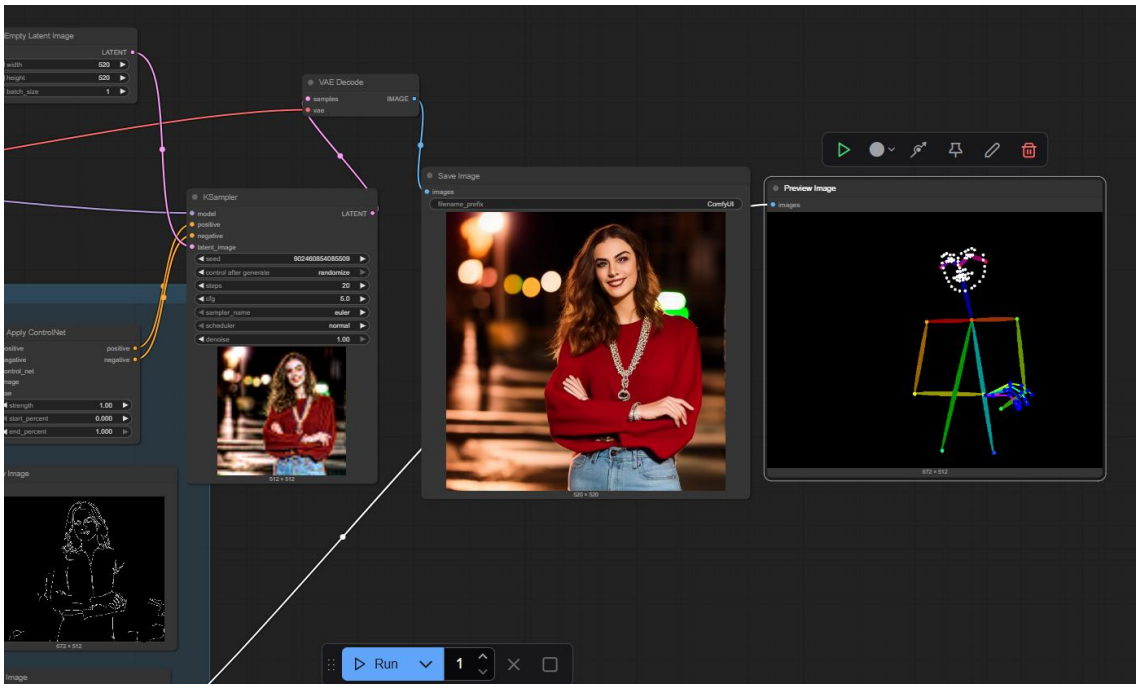



**Figura 62.** Resultado de Control Net: OpenPose

- Cada tipo de mapa aportó diferentes niveles de control y creatividad, permitiendo adaptar la generación a necesidades específicas.
- Al trabajar con la imagen de una persona, se pudo notar que en algunas generaciones, la imagen de la cara se generaba deformada a pesar del prompt negativo.

### 6.3.4. Aplicación de estilos artísticos con LoRA

Los modelos LoRA permiten aplicar estilos artísticos específicos a las imágenes, modificando la apariencia visual según el estilo deseado.

Se probó el mismo prompt con diferentes modelos de Lora:

*"A beagle dog, beautiful landscape, [trigger word of the Lora] style drawing"*

**Modelo usado:**

- **SD 1.5**

**Nodos usados y conexiones:**

- Load Checkpoint: carga el modelo base.
- Load LoRA: carga el modelo LoRA.

- CLIP Text Encode: procesa el prompt.
- KSampler: genera la imagen latente condicionada.
- VAE Decode: decodifica la imagen latente.
- Save Image / Viewer: muestra o guarda la imagen.

Se conecta entre el nodo Load Checkpoint y los nodos CLIP Text Encode para modificar la generación.

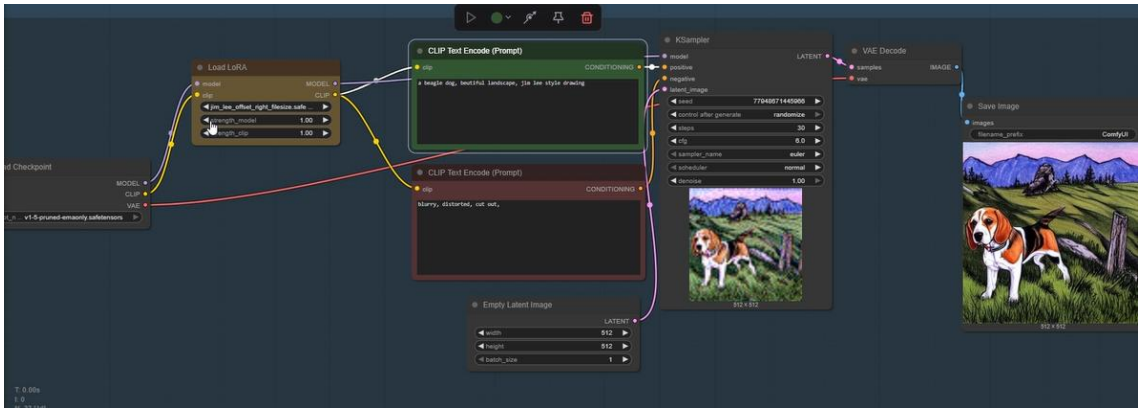



**Figura 63.** Ejemplo de workflow de Lora Comfy UI

## Proceso paso a paso:

- Se descargan e instalan los modelos LoRA deseados.

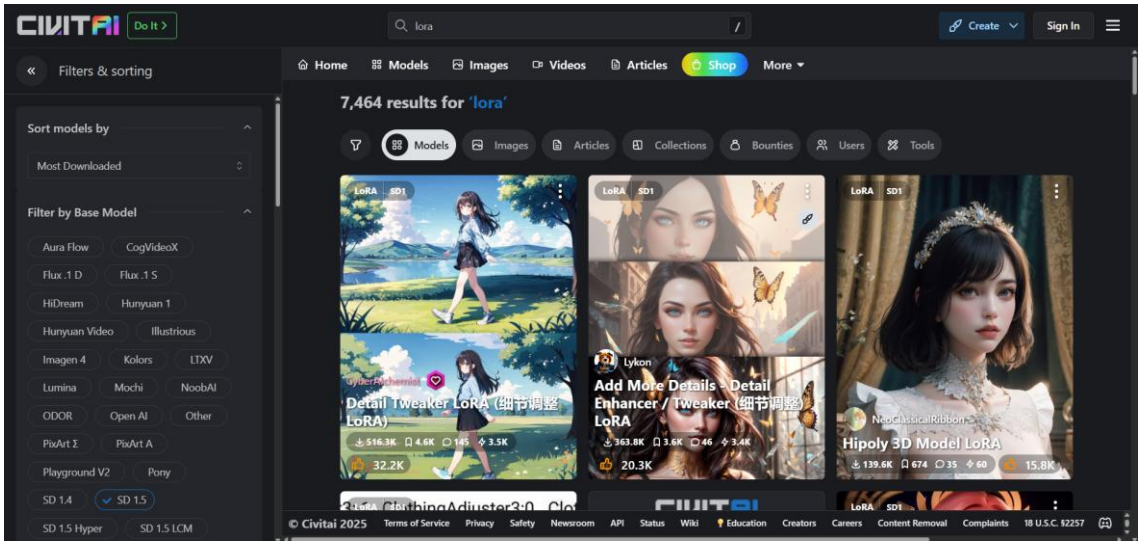



**Figura 64.** Ejemplos de Lora Models en CivitAI

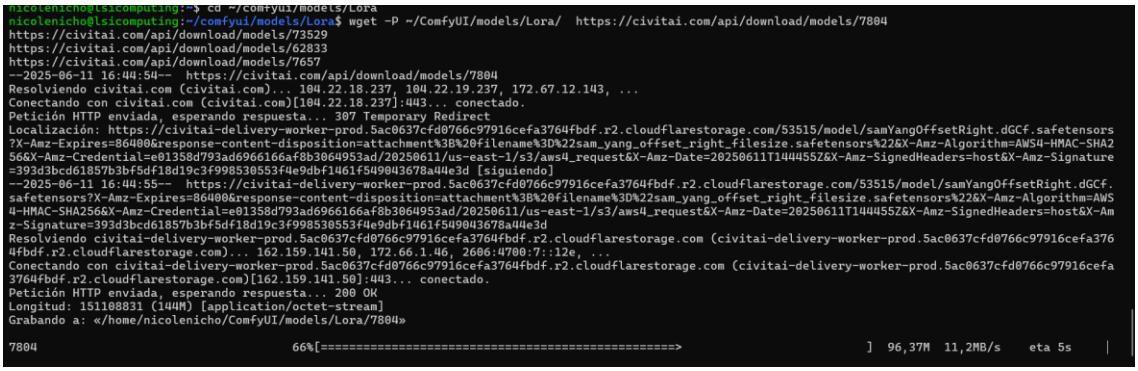



**Figura 65.** Ejemplo de descarga en terminal de Lora Models de CivitAI

- Se añade el nodo Load LoRA en el flujo, conectándolo correctamente.
- Se escribe el prompt con la palabra clave (trigger) del LoRA para activar el estilo.
- Se ejecuta la generación para cada modelo de Lora y se analizan cada uno de los resultados.

**Análisis de resultados:**

- Los LoRA cambiaron radicalmente el estilo visual, adaptando la imagen al universo artístico elegido como Sam Yang, Ghibli o cómic y el éxito dependía de usar correctamente las palabras clave (triggers) asociadas a cada LoRA

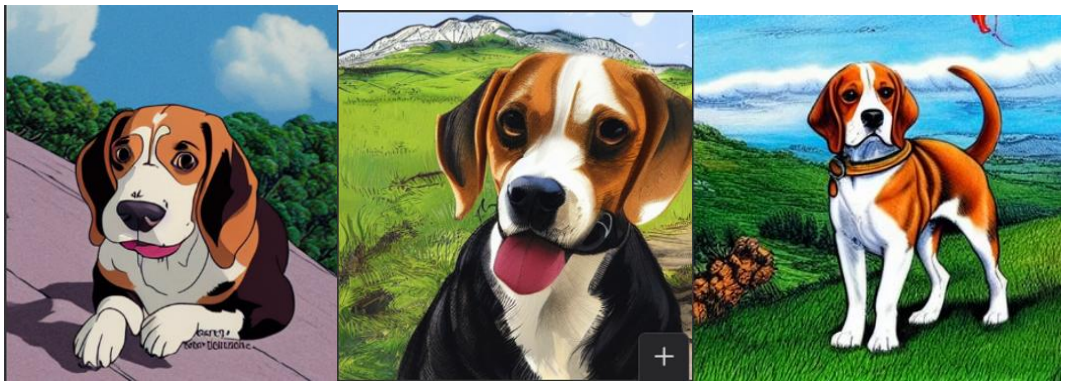


**Figura 66.** Imágenes generadas con modelos de Lora: Ghibli Style (izquierda), Sam Yang style (centro), Jim Lee Comic style (derecha)

- Cuando se usaron junto a ControlNet, se logró controlar tanto la pose como el estilo, generando imágenes muy personalizadas.

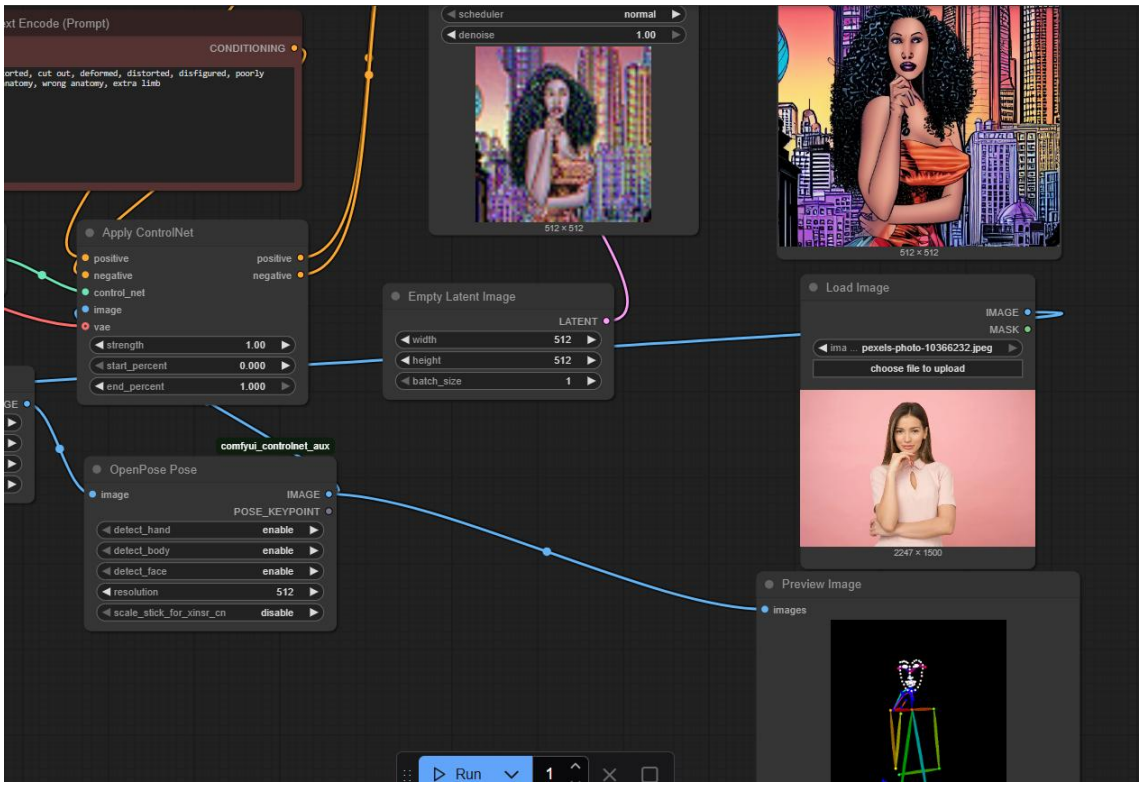



**Figura 67.** Imágenes generadas con modelos de Lora: Jim Lee Comic Style y ControlNet : OpenPose

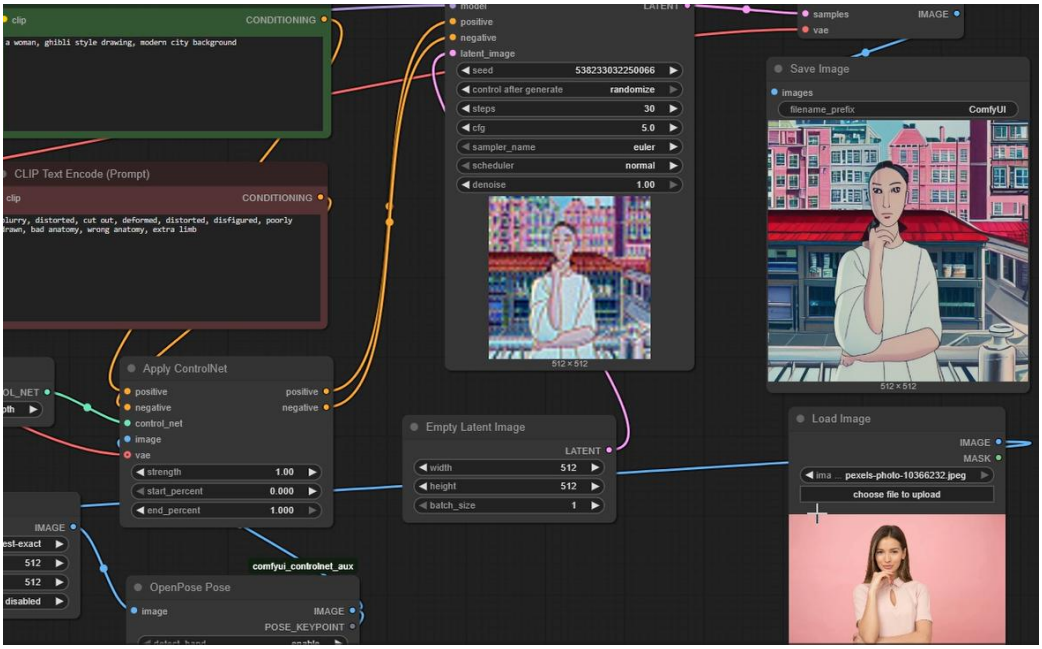



**Figura 68.** Imágenes generadas con modelos de Lora: Ghibli Style y ControlNet : OpenPose

### 6.3.5. Creación de patrones y uso de Node Painter

Este caso práctico consiste en probar Node Painter que permite dibujar patrones o bocetos que se usan como guía para la generación, útil para diseño, arquitectura o ingeniería.

Se probó diferentes prompts con diferentes bocetos:

*"A flower pattern" "A metal circuit, electronics" "a beautiful landscape with mountains, a tree and a lake"*

**Modelo usado:**

- SD 1.5

**Nodos usados y conexiones:**

- Load Checkpoint: carga el modelo base.
- Load LoRA: carga el modelo LoRA.
- CLIP Text Encode: procesa el prompt.
- Node Painter: nodo para dibujar directamente en la interfaz.
- ControlNet (scribble): usa el dibujo como mapa guía.
- KSampler: genera la imagen latente condicionada.
- VAE Decode: decodifica la imagen latente.
- Circular VAE Decode y Repeat Image Batch: nodos para crear imágenes seamless es decir sin cortes, los patrones se pueden repetir horizontalmente,verticalmente o ambos.
- Save Image / Viewer: muestra o guarda la imagen.

**Proceso paso a paso:**

Se instaló Node Painter desde Comfy Manager.

Se utilizó Node Painter para crear bocetos sencillos (flores, circuitos, texturas).

El dibujo se conectó al nodo ControlNet (scribble) para guiar la generación de la imagen con un flujo de texto a imagen pero con ControlNet Scribble como modelo.

Se emplearon nodos como "circular VAE decode" que reemplaza al VAE decode y "repeat image batch" que se configuro entre el VAE y el nodo Save Image para crear patrones continuos (seamless).

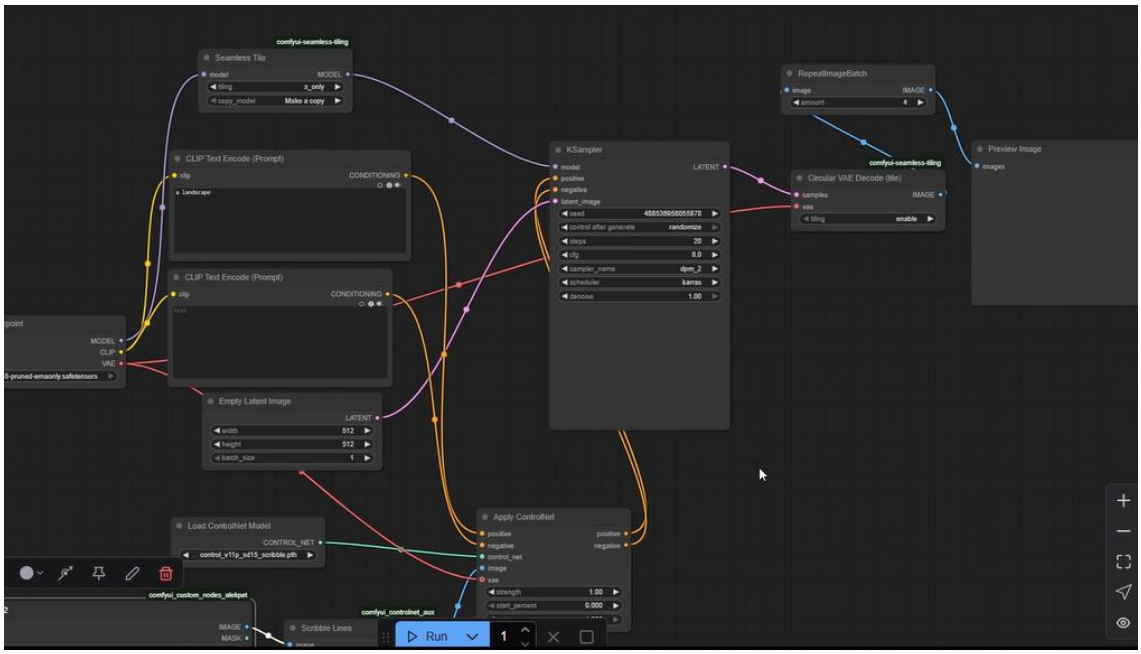



**Figura 69.** Ejemplo de workflow con nodos Seamless Tile Repeat y Image Batch Comfy UI

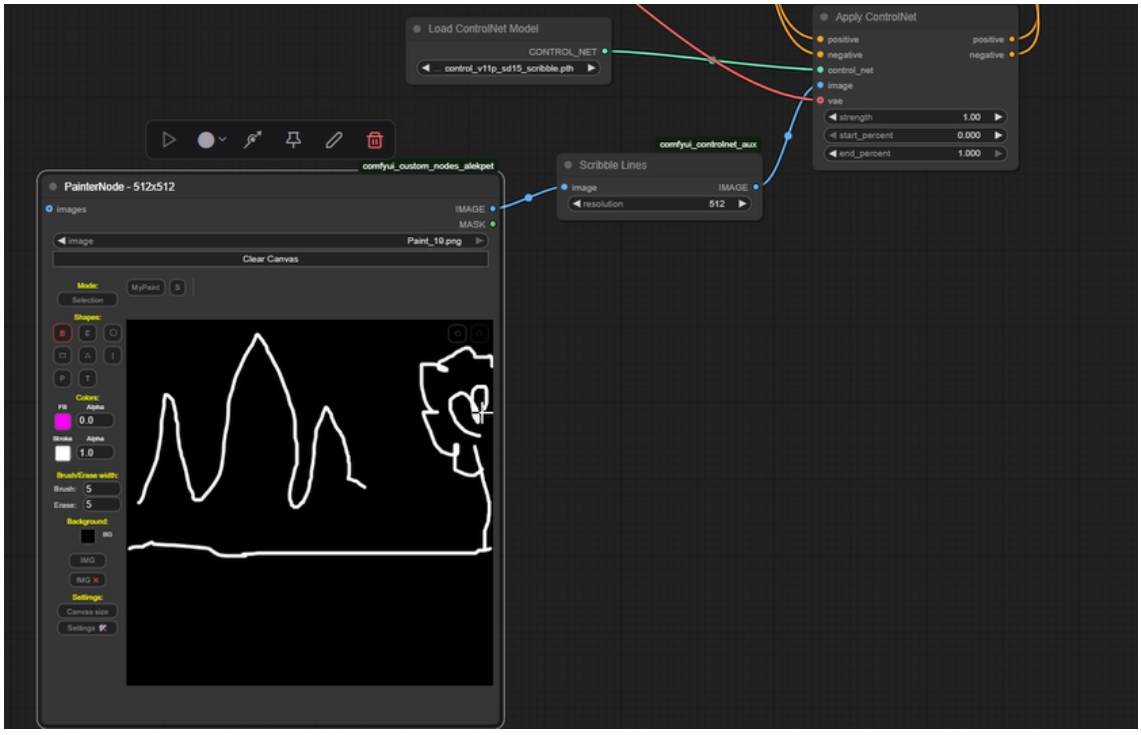



**Figura 70.** Ejemplo de workflow con Control Net Scribble Comfy UI

## Análisis de resultados:

- Incluso con dibujos simples dibujados en Scribble, la IA generó imágenes complejas y coherentes siguiendo el patrón.

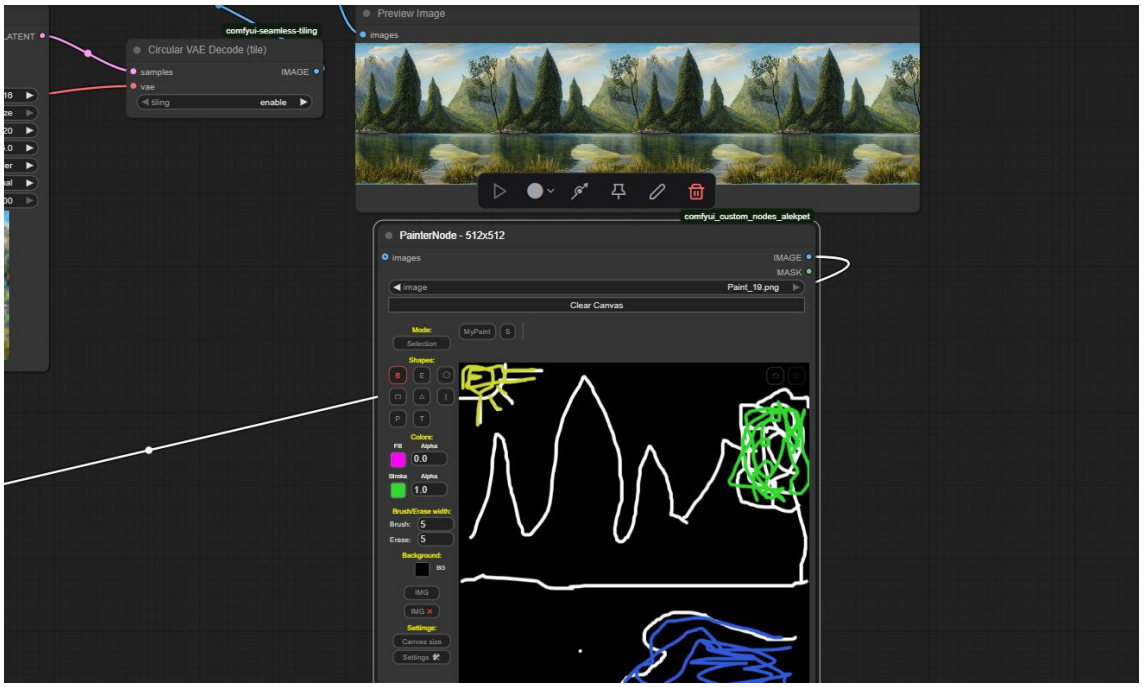



**Figura 71.** Resultados Control Net Scribble Comfy UI con prompt de paisaje

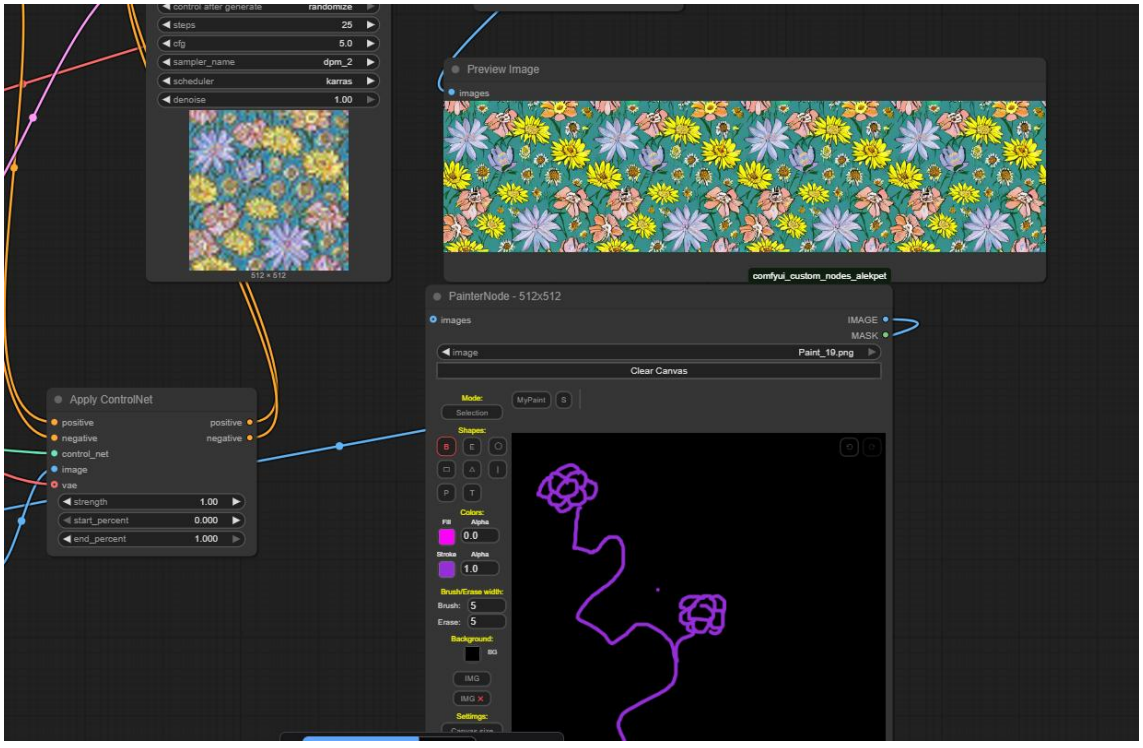



**Figura 72.** Resultados Seamless Control Net Scribble Comfy UI con prompt de patrón floral

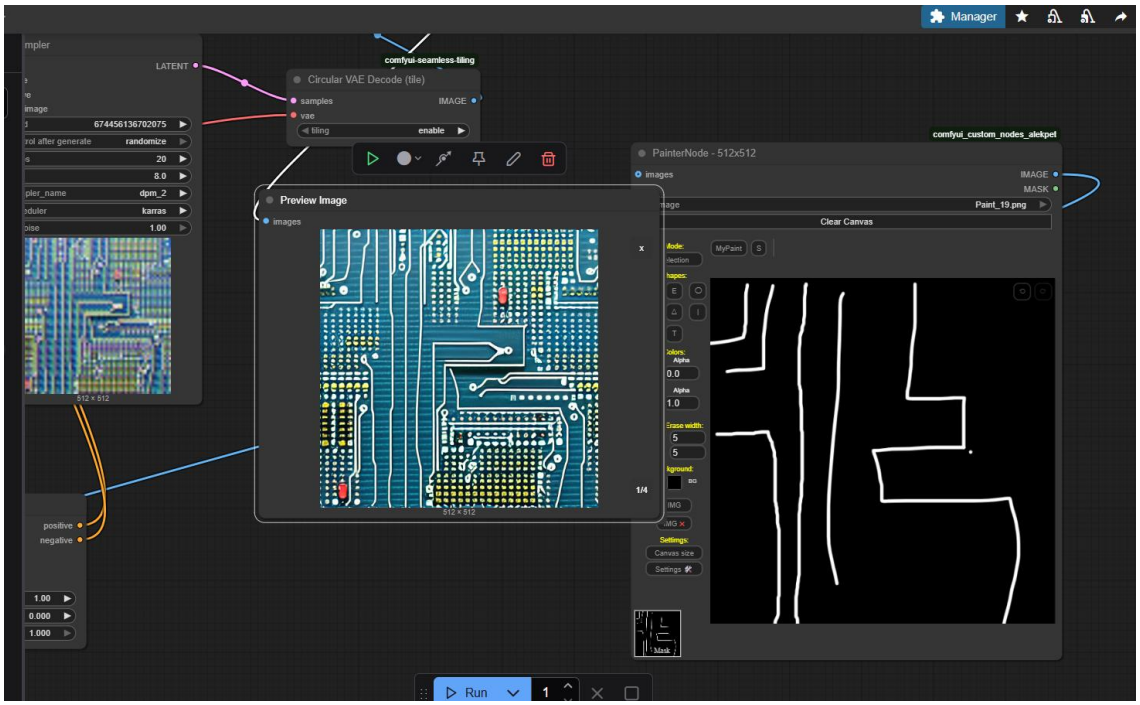



**Figura 73.** Resultados Seamless Control Net Scribble Comfy UI con prompt de circuito de metal

• La función seamless permitió crear imágenes que se pueden repetir sin cortes, ideales para fondos o texturas para uso profesional como arquitectos para programas de diseño 3D y generar texturas, o diseñadores para crear patrones que puedan usar en sus diferentes procesos creativos.

### 6.3.6. Comparativa para retratos realistas con epiCRealism XL

epiCRealism XL es un modelo comunitario ha sido desarrollado y afinado por la comunidad para mejorar la generación de imágenes fotorrealistas, con especial énfasis en la calidad y realismo de los rostros humanos, y el objetivo de esta prueba es compararlo con los modelos oficiales SD1.5, SDXL , SDXL Turbo que como se ha visto en las pruebas anteriores han tenido dificultad al generar rostros realistas a tráves de un workflow simple como “Text to Image”.

Se probó el mismo prompt con diferentes modelos:

*"A photorealistic portrait of a young woman with dark hair, wearing a traditional flamenco dress, standing in front of the Alhambra palace in Granada, warm natural sunlight, intrincate details on the dress, soft background bokeh, ultra detailed skin texture, cinematic lightning, 8k resolution."*

El prompt negativo:

*“deformed iris, deformed pupil, semi realistic, cgi, 3d, render, sketch, cartoon, drawing, anime, mutated hands and fingers, deformed, distorted, disfigured, poorly*

*drawn, bad anatomy, wrong anatomy, extra limb, floating limbs, disconnected limbs, mutation, mutated, ugly, disgusting, amputation."*

**Modelos comparados:**

- SD 1.5
- SDXL 1.0
- SDXL Turbo

**Nodos usados y conexiones:**

- Load Checkpoint: carga el modelo (SD1.5, SDXL o SDXL Turbo).
- CLIP Text Encode (Prompt positivo): convierte el texto descriptivo en vectores.
- CLIP Text Encode (Prompt negativo): convierte el texto para evitar elementos no deseados.
- KSampler: realiza la generación de la imagen latente usando el modelo y los embeddings de texto.
- VAE Decode: convierte la imagen latente en imagen visible.
- Save Image / Viewer: muestra o guarda la imagen generada.

El flujo Text to Image conecta el modelo cargado a KSampler, que recibe también los vectores de texto positivo y negativo, y envía la imagen latente a VAE Decode para obtener la imagen final.

**Proceso paso a paso:**

- Se descarga e instala epiCRealism XL parecido a la instalacion de los anteriores modelos, el modelo se descargó desde CivitAI: https://civitai.com/models/277058?modelVersionId=1522905 .
- Se usa el flujo Text-to-Image con el prompt explicado anteriormente y también se puede abrir el workflow directamente desde ComfyUI
- Se genera la imagen con SD1.5, SDXL, SDXL Turbo y epiCRealism XL.

**Análisis de resultados:**

- epiCRealism XL produce los rostros más realistas, con detalles faciales precisos, con mayor naturalidad, y de acuerdo al prompt.

**Figura 74.** Resultados de imágenes generadas con Epic Realism

- Los modelos oficiales especialmente SDXL Turbo y SD 1.5 presentan a veces distorsiones, inconsistencias o menor detalle. Mientras que SDXL produce imágenes de gran calidad pero menos naturales a comparación de epiCRealism XL en la relación al rostro y la composición de la imagen.

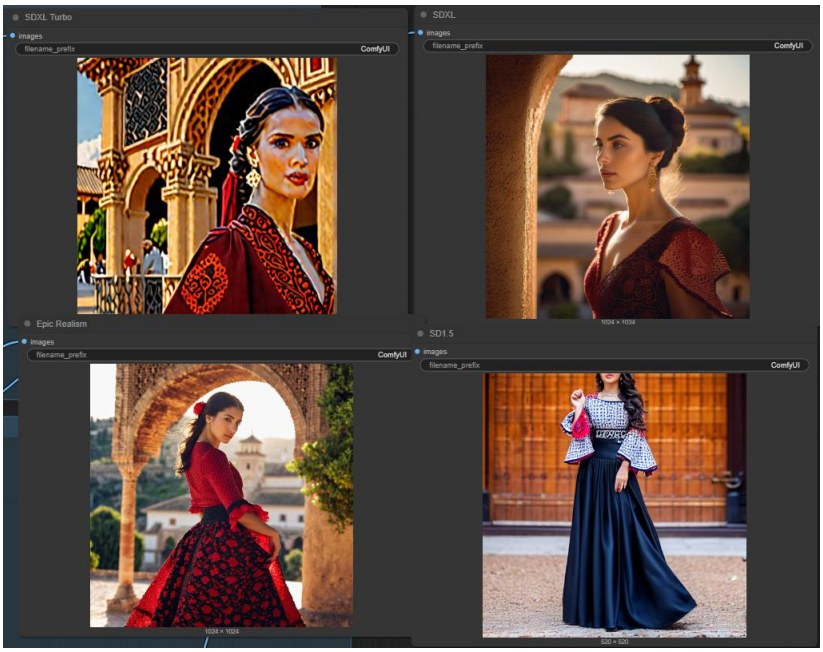



**Figura 75.** Resultados de imágenes generadas para el mismo prompt con SDXL Turbo (esquina izquierda superior), SDXL (esquina derecha superior), Epic Realism (esquina inferior izquierda), y SD1.5 (esquina inferior derecha)

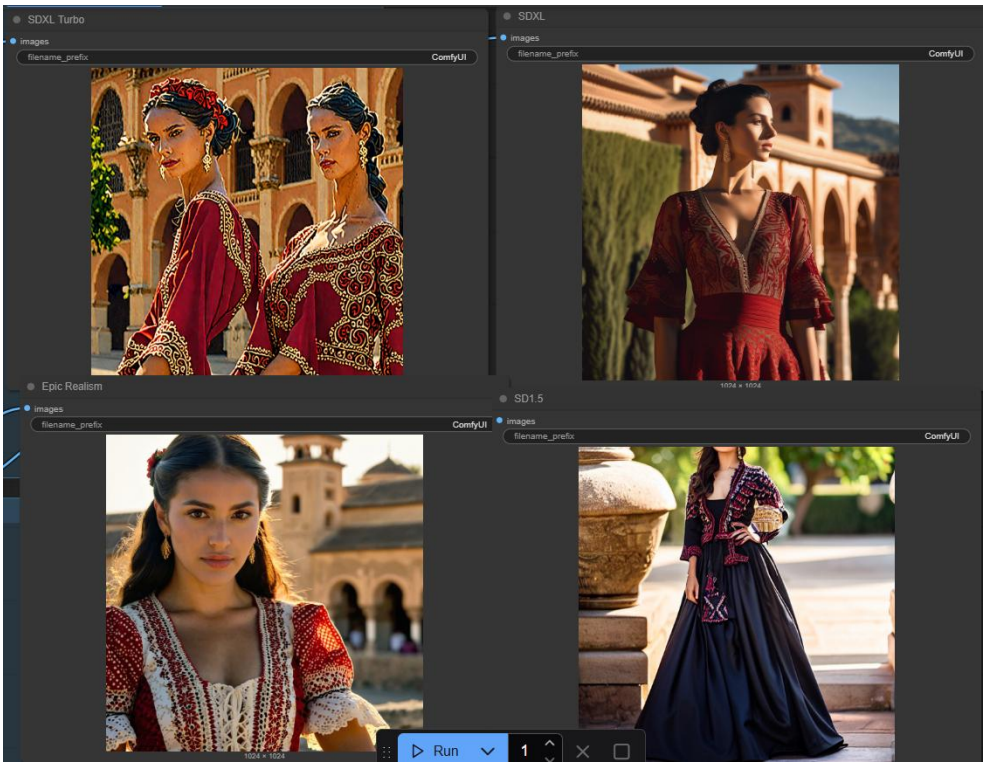



**Figura 76.** Resultados de imágenes generadas para el mismo prompt con SDXL Turbo (esquina izquierda superior), SDXL (esquina derecha superior), Epic Realism (esquina inferior izquierda), y SD1.5 (esquina inferior derecha)

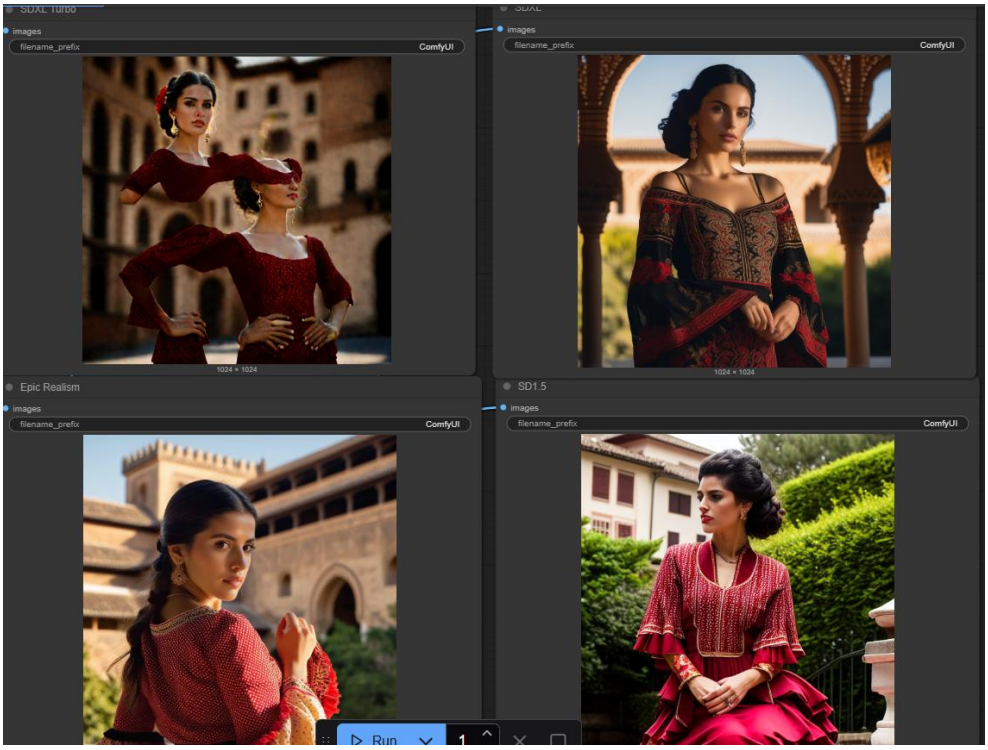



**Figura 77.** Resultados de imágenes generadas para el mismo prompt con SDXL Turbo (esquina izquierda superior), SDXL (esquina derecha superior), Epic Realism (esquina inferior izquierda), y SD1.5 (esquina inferior derecha)

### 6.3.7. Variaciones estilísticas con image-to-image y VAE para epiCRealism XL

Este caso práctico permite crear diferentes versiones estilísticas de una misma imagen, modificando únicamente el estilo en el prompt.

Se probó el mismo prompt con diferentes estilos para el mismo modelo epiCRealism XL:

*"A majestic castle perched atop a lush green hill, surrounded by vibrant blooming gardens and a sparkling river, under a clear blue sky with fluffy clouds, sunlight casting dramatic shadows, intricate architectural details, ornate towers and spires, (in {style} style:1.3), ultra detailed, highly realistic textures, cinematic lighting, masterpiece, 8k, trending on artstation"*

**Modelo usado:**

- epiCRealism XL

**Nodos usados y conexiones:**

- CheckpointLoaderSimple: carga el modelo, el CLIP y el VAE.
- CLIPTextEncode: codifica el prompt positivo y negativo (con variaciones de estilo).
- VAEEncode: convierte la imagen base a un espacio latente.
- KSampler: genera la imagen modificada desde la representación latente.
- VAEDecode: decodifica la imagen latente para obtener la imagen final.
- Preview/SaveImage: previsualiza o guarda la imagen generada.

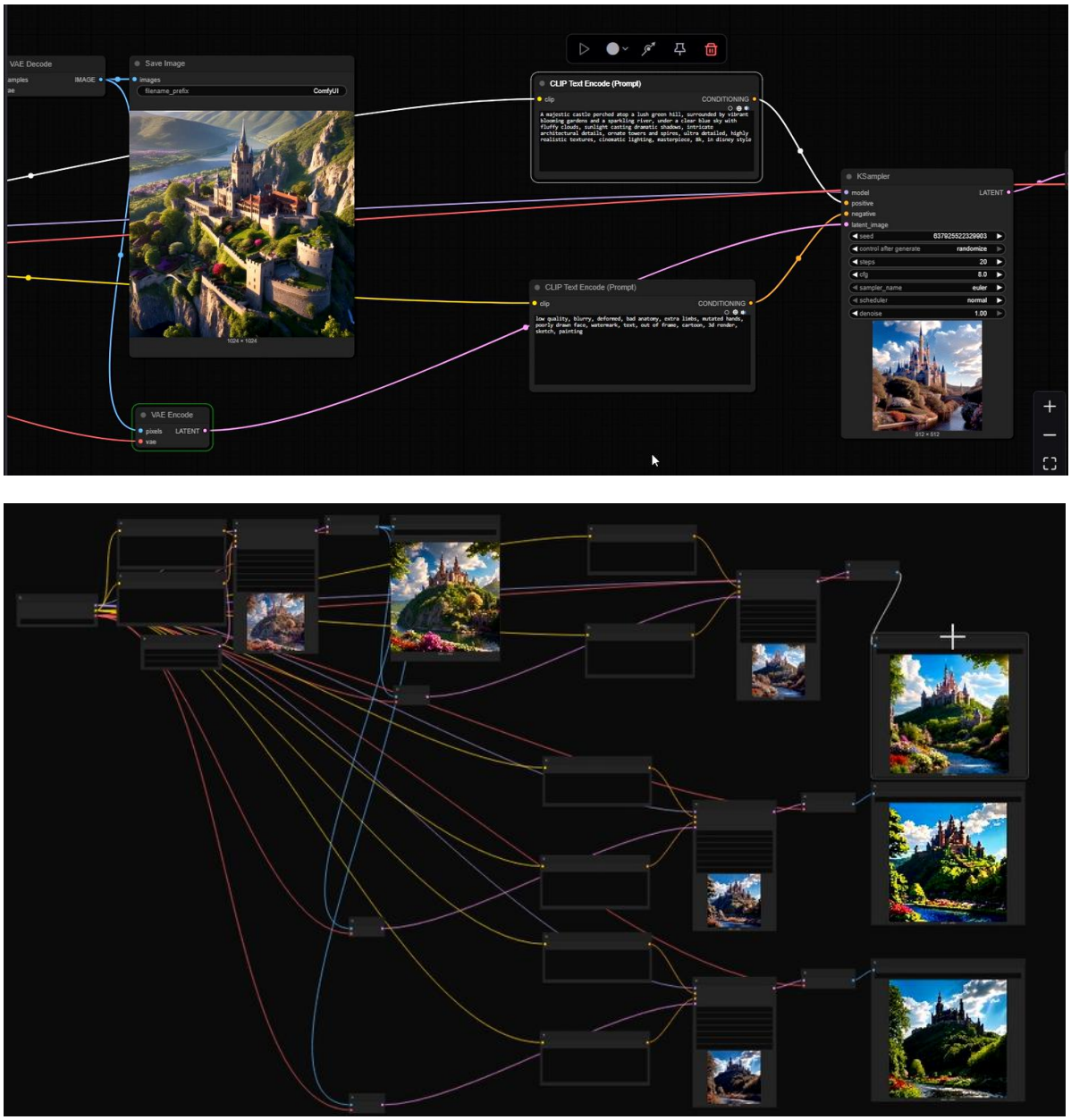



**Figura 78.** Ejemplo de workflow de combinación de Text to Image y Image to Image - Modelo epiCRealism XL

## Proceso paso a paso:

- Se genera una imagen base desde un prompt detallado utilizando el modelo epiCRealism XL.
- Esta imagen se codifica en el espacio latente con VAEEncode.
- Se modifican los prompts para incluir distintos estilos como: in disney style, in pixel art style, in gothic style.
- Se vuelve a usar KSampler para regenerar versiones estilizadas, manteniendo como entrada la codificación latente de la imagen original.

- Las imágenes estilizadas se decodifican con VAEDecode y se guardan con SaveImage.

**Análisis de resultados:**

- Las imágenes generadas mantienen la composición de la imagen original obtenida mediante Text-to-Image, pero presentan un cambio notable en el estilo visual gracias a la incorporación de variaciones estilísticas en el prompt. El modelo epiCRealism XL permite conservar un alto nivel de realismo incluso al aplicar diferentes estilos artísticos, lo que resulta en versiones coherentes y visualmente ricas de la misma escena.

**Figura 79.** Resultados de imagen base generada para usada para diferentes estilos con el modelo epiCRealism XL

**Figura 80.** Resultados de imagen generada añadiendo al prompt “Disney style” en base a la imagen base generada (Figura 79)

**Figura 81.** Resultados de imagen generada añadiendo al prompt “Pixel Art style” en base a la imagen base generada (Figura 79)

**Figura 82.** Resultados de imagen generada añadiendo al prompt “Gothic style” en base a la imagen base generada (Figura 79)

- El modelo epiCRealism XL demuestra que los modelos comunitarios pueden ofrecer resultados sobresalientes en tareas específicas, como el realismo en rostros o detalles arquitectónicos. Existen muchos otros modelos entrenados por la comunidad que pueden explorarse según las necesidades del proyecto y herramientas como ComfyUI permiten probar, ajustar e incluso entrenar modelos propios, sin necesidad de ser expertos en IA.

### 6.3.8. Análisis comparativo de parámetros de control en ComfyUI

La calidad y estilo de las imágenes generadas mediante modelos como Stable Diffusion están altamente influenciados por los parámetros que el usuario define en el proceso. Para comprender el impacto real de cada uno, se ha realizado una exploración práctica variando los principales parámetros uno a uno, mientras se mantiene constante el resto. Los parámetros analizados son:

- Sampler: Algoritmo de muestreo.
- CFG (Classifier-Free Guidance): Nivel de adherencia al prompt.
- Steps: Número de pasos de inferencia.
- Seed: Valor de semilla aleatoria.
- Denoise strength: Solo para flujos de image-to-image.

El objetivo es observar cómo cada uno de estos parámetros modifica el resultado visual, utilizando dos productos como referencia:

- Producto 1: Zapato stiletto rojo (moda/producto de lujo)
- Producto 2: Hamburguesa con queso, tomate y lechuga (alimentación/publicidad creativa)

**Análisis del parámetro Steps**

El parámetro Steps indica el número de pasos que el modelo ejecuta para transformar ruido en una imagen coherente. A más pasos, más refinamiento; sin embargo, un exceso puede provocar imágenes sobreprocesadas. Se evaluaron imágenes de dos productos reales con distintos valores de Steps, manteniendo constantes los demás parámetros:

- **Prompt Producto 1:** *"Glossy red stiletto high heels, product photography, isolated white background, soft studio lighting, realistic shadows, high detail, professional commercial style, 4K resolution"*
- **Prompt Producto 2:** *"A juicy cheeseburger with lettuce and tomato on a wooden plate, high detail, studio lighting, realistic texture, 85mm lens, depth of field"*
- **Seed Producto 1:** 456 | **Producto 2**: 123
- **CFG fijo**: 7
- **Modelo**: SDXL 1.0
- **Resolución**: 1024x1024
- **Sampler**: Euler

**Tabla 2.** Cuadro Comparativo entre distintos valores del parámetro steps

| Steps | Producto 1 | Producto 2 | Observaciones |
|---|---|---|---|
| **5** |  |  | **Stiletto**: Imagen sin detalles. **Hamburguesa**: Imagen poco refinada, con solo manchas. |
| **10** |  |  | **Stiletto**: Hay más estructura pero pocos detalles. **Hamburguesa**: Se ve la estructura pero poco realismo. |
| **20** |  |  | **Stiletto**: Detalles generados correctamente. **Hamburguesa**: Mayor realismo y fidelidad al prompt. |
| **30** |  |  | **Stiletto**: Mayor nitidez pero poco realismo. **Hamburguesa**: Mucho mayor detalle que disminuye el realismo. |

Aunque aumentar el número de Steps mejora la calidad visual hasta cierto punto, el exceso (como con 30 steps) puede producir un efecto artificial, especialmente visible en brillos o texturas sobredefinidas. El valor de 20 steps fue el óptimo en ambos casos, logrando una representación realista sin caer en el sobreprocesamiento.

**Análisis del parámetro CFG**

El parámetro CFG controla el grado en que la generación de la imagen se ajusta al prompt proporcionado. En valores bajos, el modelo tiene mayor libertad creativa y puede alejarse del prompt; en valores altos, fuerza la fidelidad al prompt, lo que puede provocar artefactos o resultados poco naturales si se exagera. Se evaluaron imágenes de dos productos reales con distintos valores de CFG, manteniendo constantes los demás parámetros:

- **Prompt Producto 1:** *"Glossy red stiletto high heels, product photography, isolated white background, soft studio lighting, realistic shadows, high detail, professional commercial style, 4K resolution"*
- **Prompt Producto 2:** *"A juicy cheeseburger with lettuce and tomato on a wooden plate, high detail, studio lighting, realistic texture, 85mm lens, depth of field"*
- **Seed Producto 1:** 456 | **Producto 2**: 123
- **Steps**: 20
- **Modelo**: SDXL 1.0
- **Resolución**: 1024x1024
- **Sampler**: Euler

**Tabla 3.** Cuadro Comparativo entre distintos valores del parámetro CFG

| CFG | Producto 1 | Producto 2 | Observaciones |
|---|---|---|---|
| 3 |  |  | Poca fidelidad al prompt. Más libertad creativa pero menos utilidad para productos. |
| 5 |  |  | Mejor alineación con el prompt. Nivel aceptable para borradores. |

| | | | |
|---|---|---|---|
| **7** |  |  | Buen equilibrio entre fidelidad y naturalidad. Ideal para productos comerciales. |
| **10** |  |  | Se pierde realismo por exceso de precisión. |
| **15** |  |  | Apariencia animada. Sobreajuste al prompt. Pierde valor publicitario realista. |

El parámetro CFG influye directamente en cuán literalmente se interpreta el prompt. Un valor medio como 7 ofrece resultados realistas, detallados y adecuados para contextos profesionales o publicitarios. Valores bajos (3–5) tienden a generar resultados vagos, mientras que valores altos (10–15) sobrecargan la imagen, comprometiendo el realismo.

**Análisis del parámetro Seed**

La seed o semilla es un número que determina el punto de partida del generador de números pseudoaleatorios que interviene en la generación de imágenes. Cambiar la seed permite obtener distintas variaciones de imagen a partir del mismo prompt y configuración, mientras que mantenerla fija garantiza reproducibilidad.

En este experimento se utilizaron seeds fijas (123, 456 y 789) para observar las diferencias visuales generadas, manteniendo constantes todos los demás parámetros. Aunque la seed no influye directamente en la calidad de forma determinista, en ciertos

casos puede dar lugar a resultados más o menos adecuados según el tipo de producto o estilo buscado.

- **Prompt Producto 1:** *"Glossy red stiletto high heels, product photography, isolated white background, soft studio lighting, realistic shadows, high detail, professional commercial style, 4K resolution"*
- **Prompt Producto 2:** *"A juicy cheeseburger with lettuce and tomato on a wooden plate, high detail, studio lighting, realistic texture, 85mm lens, depth of field"*
- **Seed Producto 1:** 456 | **Producto 2**: 123
- **Steps**: 20
- **Modelo**: SDXL 1.0
- **Resolución**: 1024x1024
- **Sampler**: Euler

**Tabla 4.** Cuadro Comparativo entre distintos valores del parámetro seed

| Seed | Producto 1 | Producto 2 | Observaciones |
|---|---|---|---|
| **123** |  |  | Seed ideal para alimentos, pero no para objetos estilizados como calzado. |
| **456** |  |  | Seed más adecuada para productos de moda o diseño preciso. |

| | | | |
|---|---|---|---|
| **789** |  |  | Resultados poco fieles en ambos casos. Seed menos útil para propósitos comerciales. |

La elección de la seed no afecta directamente la calidad visual o fidelidad al prompt de forma determinista, ya que su función es simplemente establecer el punto de partida del proceso aleatorio. Sin embargo, en contextos como la presentación de productos o diseño publicitario, puede ser útil explorar diferentes seeds hasta encontrar una variante visual más adecuada al objetivo deseado. En este experimento, las seeds 123 y 456 generaron imágenes que, subjetivamente, resultaron más útiles según según el tipo de producto o estilo buscado.

**Análisis del parámetro Sampler**

El sampler o algoritmo de muestreo define la forma en que se genera una imagen paso a paso a partir del ruido inicial. No modifica el contenido del prompt, pero sí influye en el estilo visual, nivel de detalle, realismo y estabilidad del resultado. Cada sampler tiene un comportamiento particular y puede ser más o menos adecuado según el tipo de imagen que se desea generar. Se evaluaron imágenes de dos productos reales con los samplers más comunes, manteniendo constantes los demás parámetros:

- **Prompt Producto 1:** *"Glossy red stiletto high heels, product photography, isolated white background, soft studio lighting, realistic shadows, high detail, professional commercial style, 4K resolution"*
- **Prompt Producto 2:** *"A juicy cheeseburger with lettuce and tomato on a wooden plate, high detail, studio lighting, realistic texture, 85mm lens, depth of field"*
- **Seed Producto 1:** 456 | **Producto 2**: 123
- **Steps**: 20
- **CFG**: 7
- **Modelo**: SDXL 1.0
- **Resolución**: 1024x1024

**Tabla 5.** Cuadro Comparativo entre distintos valores del parámetro sampler

| Sampler | Producto 1 | Producto 2 | Observaciones |
|---|---|---|---|
| **DPM++ 2M Karras** |  |  | **Stiletto**: Buena representación del prompt, alto nivel de detalle, especialmente en costuras. **Hamburguesa**: Buena representación del prompt, pero aún faltan algunos detalles. |
| **Euler** |  |  | **Stiletto**: Buen resultado, y representación del prompt, pero detalles menos nítidos. **Hamburguesa**: Excelente representación del prompt, incluyendo muchos detalles. |
| **DDIM** |  |  | **Stiletto**: Mejora en el realismo y silueta, aunque aún presenta detalles menores. **Hamburguesa**: Excelente nitidez, profundidad y texturas. |
| **UNIPC** |  |  | **Stiletto**: Calidad inconsistente, silueta no estilizada. **Hamburguesa**: Poco realismo, no sigue el prompt. |

La elección del sampler tiene un impacto directo sobre la calidad visual y la interpretación estilística del resultado. No todos los samplers funcionan igual de bien para todos los tipos de productos. Aunque algunos samplers son más populares con

ciertos modelos, los resultados pueden variar según el prompt y el tipo de objeto generado.

**Análisis del parámetro Denoise**

El parámetro Denoise controla el grado de influencia que tiene la imagen base o la semilla inicial sobre la imagen generada. Un valor bajo mantiene la imagen cercana a la base original, mientras que valores altos permiten que la generación se aleje más, siguiendo con mayor libertad el prompt textual.

En este análisis se usaron los mismos productos, algunas imágenes de referencia o de base, seed fija, CFG, sampler, steps y modelo para aislar el efecto del denoise sobre la generación:

**Figura 83.** Imágenes base para generación img2img para análisis de denoise

- **Prompt Producto 1:** *"Glossy red stiletto high heels, product photography, isolated white background, soft studio lighting, realistic shadows, high detail, professional commercial style, 4K resolution"*
- **Prompt Producto 2:** *"A juicy cheeseburger with lettuce and tomato on a wooden plate, high detail, studio lighting, realistic texture, 85mm lens, depth of field"*
- **Seed Producto 1:** 456 | **Producto 2**: 123
- **Steps**: 20
- **CFG**: 7
- **Sampler**: Euler
- **Modelo**: SDXL 1.0
- **Resolución**: 1024x1024

**Tabla 6.** Cuadro Comparativo entre distintos valores del parámetro denoise en img2img

| Denoise | Producto 1 | Producto 2 | Observaciones |
|---|---|---|---|
| **0.30** |  |  | Imágenes muy parecidas a las bases original, casi sin diferencias visibles. Las imágenes base domina el resultado. |
| **0.50** |  |  | Variaciones leves respecto a la imagen base, la fidelidad al prompt comienza a notarse pero es limitada. |
| **0.70** |  |  | Cambio más notorio en los productos generados, con mejor representación del prompt, aunque la imagen base aún influye bastante. |
| **0.80** |  |  | Fidelidad alta al prompt manteniendo inspiración de la imagen base. Se agregan elementos nuevos derivados del texto. |

| | | | |
|---|---|---|---|
| **0.90** |  |  | La influencia de la imagen base es menor, predominando el prompt, el realismo se ve afectado y la imagen parece menos natural. |

El valor de Denoise se ajusta según el objetivo: valores bajos (0.3-0.5) conservan la imagen base, útiles para ediciones leves; valores intermedios (0.7-0.8) equilibran la inspiración original y la fidelidad al prompt, ideales para variaciones controladas; y valores altos (0.9) potencian creatividad y apego al texto, pero pueden reducir realismo. Para generar productos con imágenes de referencia, como zapatos o comida, se recomiendan valores entre 0.7 y 0.8 para un buen balance entre originalidad y precisión.

# 7. CONCLUSIONES Y TRABAJO FUTURO

## 7.1. Objetivos cumplidos

A lo largo de este trabajo se han alcanzado los objetivos planteados en la fase inicial:

- Se revisaron los fundamentos de la inteligencia artificial y los modelos de Stable Diffusion, profundizando en su arquitectura modular y funcionamiento interno. Este análisis teórico permitió comprender cómo operan los algoritmos generativos y sentó las bases para el uso efectivo de ComfyUI, permitiendo aprovechar al máximo sus opciones de personalización y control creativo en las pruebas desarrolladas.

- Se han analizado y comparado las principales plataformas del mercado, tanto open-source como propietarias, evaluando aspectos clave como funcionalidad, facilidad de uso, personalización y control técnico. Esta comparativa ha permitido situar a ComfyUI frente a alternativas como MidJourney, DALL·E 3, Leonardo AI y AUTOMATIC1111.

- Se ha elaborado una guía accesible para usuarios sin experiencia técnica, cubriendo desde la instalación y configuración hasta la integración de modelos oficiales como SD1.5, SDXL, SDXL Turbo, LoRA y ControlNet. Los casos prácticos han demostrado cómo adaptar flujos de trabajo a distintos contextos creativos, evidenciando la flexibilidad y el potencial de ComfyUI para facilitar el acceso a la creación visual con IA.

- Los casos prácticos han permitido observar cómo la elección de modelos, parámetros y técnicas influyen en la calidad, rapidez y personalización de los resultados, ofreciendo una visión clara y replicable para futuros usuarios.

- Se ha realizado un análisis comparativo detallado de parámetros clave (Sampler, Steps, CFG, Seed y Denoise) aplicados a productos concretos, identificando el impacto de cada parámetro en la fidelidad al prompt, nivel de detalle, realismo y creatividad. Esto aportó una guía práctica para optimizar la generación de imágenes según las características del producto y los objetivos creativos.

## 7.2. Líneas futuras de trabajo

A partir de los resultados obtenidos y la experiencia adquirida, se proponen varias líneas de trabajo para el futuro:

- Dado el ritmo acelerado de innovación en inteligencia artificial, resulta importante considerar versiones futuras de Stable Diffusion, así como arquitecturas avanzadas y técnicas de control y personalización. La evaluación de estas tecnologías dentro del entorno de ComfyUI podría ofrecer mejoras significativas en calidad visual, coherencia semántica y velocidad de generación explorarando nuevas posibilidades creativas
- Investigar y aplicar flujos de trabajo más complejos, como inpainting, outpainting, transferencia de estilos, y edición avanzada, estos flujos permitirían abordar proyectos profesionales con un mayor nivel de complejidad y exigencia creativa.

- Ampliar el análisis a otros tipos de productos o contextos visuales, con el fin de generalizar las recomendaciones y optimizar flujos creativos para distintas áreas del diseño y la publicidad digital.

- Actualmente, la valoración de los resultados generados se realiza mayoritariamente desde una perspectiva subjetiva; sin embargo, incorporar métricas objetivas como el CLIP Score o el FID, junto con estudios de percepción realizados a través de encuestas a usuarios expertos, permitiría validar de forma más rigurosa los resultados obtenidos.

- La aplicación práctica de ComfyUI en contextos reales, tanto en el ámbito educativo como en entornos profesionales permitiría validar la utilidad práctica de los flujos de trabajo desarrollados, y generar nuevos aprendizajes que alimenten futuras investigaciones o desarrollos técnicos.

# REFERENCIAS BIBLIOGRÁFICAS


Aggarwal, C. C. (2018). Neural Networks and Deep Learning. Springer. https://doi.org/10.1007/978-3-319-94463-0

Amate Bonachera, A. (2024). La IA generativa en el diseño gráfico y la publicidad: Herramientas y estrategias emergentes.Universidad de Diseño, Tecnología e Innovación (UDIT), Grupo de Investigación Genius. https://cidinova.org/wp-content/uploads/2025/01/La-IA-generativa-en-el-diseno-grafico-y-la-publicidad-Herramientas-y-estrategias-emergentes.pdf

Angra, S., & Ahuja, S. (2017). Machine learning and its applications: A review. Proceedings of the 2017 International Conference on Big Data Analysis and Computational Intelligence (ICBDACI 2017), 300-304. https://doi.org/10.55524/ijircst.2022.10.3.61

Atkinson-Abutridy, J. (2024). Large Language Models: Concepts, Techniques and Applications (1st ed.). CRC Press. https://doi.org/10.1201/9781003517245

AUTOMATIC1111. (2022). Stable Diffusion Web UI [Computer software]. https://github.com/AUTOMATIC1111/stable-diffusion-webui

AUTOMATIC1111. (2023a). Home · AUTOMATIC1111/stable-diffusion-webui Wiki. GitHub. https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki

AUTOMATIC1111. (2023b). Features · AUTOMATIC1111/stable-diffusion-webui Wiki. GitHub. https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Features

Black Forest Labs. (2024a). Announcing SDXL 1.0. https://bfl.ai/announcements/24-08-01-bfl

Black Forest Labs. (2024b). FLUX.1 Model Card. Hugging Face. https://huggingface.co/black-forest-labs/FLUX.1-dev

Black Forest Labs. (2025a). FLUX.1 Kontext Technical Report. https://bfl.ai/announcements/flux-1-kontext

Black Forest Labs. (2025b). FLUX API Documentation. https://docs.bfl.ml/quick_start/introduction

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the Opportunities and Risks of Foundation Models. arXiv preprint arXiv:2108.07258. https://doi.org/10.48550/arXiv.2108.07258

Cabanelas Omil, J. (2019). Inteligencia artificial ¿Dr. Jekyll o Mr. Hyde? Mercados y Negocios, 1(40), páginas. https://doi.org/10.32870/myn.v0i40.7403

Chollet, F. (2018). Deep Learning with Python. Manning Publications.

ComfyUI. (2024a). Documentation. https://github.com/comfyanonymous/ComfyUI/wiki

ComfyUI. (2024b). Features. https://github.com/comfyanonymous/ComfyUI

ComfyUI. (2025). Documentación oficial de ComfyUI. https://docs.comfy.org/ Cortés Hernández, A., Hernández Hernández, C. A., García Torres, A. B., & Mata

Quezadas, M. (2024). La inteligencia artificial generativa como un asistente estratégico en la era del aprendizaje digital. Ciencia Latina Revista Científica Multidisciplinar, 8(4), 2158-2171. https://doi.org/10.37811/cl_rcm.v8i4.12456

Cárdenas, J. (2023). Inteligencia artificial, investigación y revisión por pares. Revista Española de Sociología, 32(4), a184. https://doi.org/10.22325/fes/res.2023.184

DataScientest. (2022). Inteligencia artificial: definición, historia, usos, peligros. Recuperado de https://datascientest.com/es/inteligencia-artificial-definicion

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT 2019, 4171–4186. https://doi.org/10.48550/arXiv.1810.04805

Erfani Jahanbakhsh, A. (2024). Generative AI for Node-Based Shaders [Master's thesis, Aalto University]. Aaltodoc. https://aaltodoc.aalto.fi/items/5674e456-3156-4c97-a50d-341816023690

Gao, Z., Peromingo Peromingo, D., & Cubillo Romero, J. (2024). Generación de imágenes mediante modelos de difusión [Trabajo de Fin de Grado, Universidad Complutense de Madrid]. Repositorio Institucional UCM. https://docta.ucm.es/entities/publication/66fa4de2-e0b2-4d2b-88cf-3885523dc16e

Gil Viñarás, A. (2023). Acercamiento filosófico a las imágenes generadas por inteligencia artificial. Análisis y experimentación técnica de imágenes con el programa

Midjourney [Trabajo de Fin de Grado, Universidad de Zaragoza]. Repositorio Universidad Zaragoza. https://zaguan.unizar.es/record/134464#

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural

Information Processing Systems, 27, 2672–2680. https://doi.org/10.48550/arXiv.1406.2661

Herencia Campaña, Ó. (2024). La IA en la profesión del creativo [Trabajo de Fin de Grado, Universitat Politècnica de Catalunya, Centre de la Imatge i la Tecnologia Multimèdia]. Repositorio UPC. http://hdl.handle.net/2117/414133

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. Advances in Neural Information Processing Systems, 33, 6840–6851. https://arxiv.org/abs/2006.11239

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. arXiv preprint arXiv:2106.09685. https://doi.org/10.48550/arXiv.2106.09685

Huang, O., Zhang, Y., Wang, Y., & Wang, Y. (2025). ComfyGPT: A self-optimizing multi-agent system for comprehensive ComfyUI workflow generation (arXiv preprint arXiv:2503.17671). https://doi.org/10.48550/arXiv.2503.17671

Hugging Face. (s.f.). Stable diffusion pipelines. Recuperado de https://huggingface.co/docs/diffusers/v0.5.1/en/api/pipelines/stable_diffusion

Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes. Proceedings of the 2nd International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1312.6114

Krichen, M. (2023). Convolutional Neural Networks: A Survey. Computers, 12(8), 151. https://doi.org/10.3390/computers12080151

Lara Artolazábal, J. (2023). Generación de imágenes con Inteligencia Artificial: revisión de la situación actual y aplicaciones prácticas [Trabajo de Fin de Grado,

Universidad Miguel Hernández de Elche]. Repositorio Institucional UMH. https://hdl.handle.net/11000/30268

Leonardo.AI. (2025). Leonardo.AI Official Website. https://leonardo.ai/

Luo, Z., et al. (2025). LLM4SR: A Survey on Large Language Models for Scientific Research. arXiv preprint arXiv:2501.04306. https://doi.org/10.48550/arXiv.2501.04306

Mayol, J. (2024). Impacto de la Inteligencia Artificial generativa en la publicación científica. Enfermería Nefrológica, 27(3), 187-188. Epub 14 de noviembre de 2024. https://dx.doi.org/10.37551/s2254-28842024019

Microsoft. (2025). DALL-E 3 ahora está disponible en Bing Chat y Bing.com/create gratis. https://news.microsoft.com/source/latam/ia/dall-e-3-ahora-esta-disponible-en-bing-chat-y-bing-com-create-gratis/

MidJourney. (2025a). Documentation. https://docs.midjourney.com/hc/en-us

MidJourney. (2025b). Terms of Service. https://docs.midjourney.com/hc/en-us/articles/32083055291277-Terms-of-Service

Muñoz Murillo, V. (2023). Inteligencia artificial aplicada al análisis de imágenes [Trabajo de Fin de Grado, Universidad Politécnica de Madrid]. Archivo Digital UPM. https://oa.upm.es/80451/

OpenAI. (2024a). DALL·E 3. https://platform.openai.com/docs/guides/images

OpenAI. (2024b). DALL·E 3 API Documentation. https://platform.openai.com/docs/guides/images

OpenAI. (2025a). Investigación y publicaciones. https://openai.com/es-ES/research/index/publication/

OpenAI. (2025b). Investigación. https://openai.com/es-ES/research/

OpenAI. (2025c). Lanzamientos y actualizaciones técnicas. https://openai.com/es-ES/research/index/release/?page=5

OpenVINO™ Documentation. (2023a). Image Generation with Stable Diffusion and IP-Adapter. Recuperado de https://docs.openvino.ai/2023.3/notebooks/278-stable-diffusion-ip-adapter-with-output.html

OpenVINO™ Documentation. (2023b). Text-to-Image Generation with ControlNet Conditioning. Recuperado de https://docs.openvino.ai/2023.3/notebooks/235-controlnet-stable-diffusion-with-output.html

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., &

Rombach, R. (2023). SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. arXiv preprint arXiv:2307.01952. https://doi.org/10.48550/arXiv.2307.01952

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10684-10695. https://arxiv.org/abs/2112.10752

Russell, S., & Norvig, P. (2016). Artificial Intelligence: A Modern Approach (3rd ed.). Pearson

Sauer, A., Lorenz, D., Blattmann, A., & Rombach, R. (2023). Adversarial Diffusion Distillation. arXiv preprint arXiv:2311.17042. https://arxiv.org/abs/2311.17042

Stability AI. (2022). Stable Diffusion v1-5 Model Card. Hugging Face. https://huggingface.co/stable-diffusion-v1-5

Stability AI. (2023). Announcing SDXL 1.0. Stability AI News. https://stability.ai/news/stable-diffusion-sdxl-1-announcement

Stability AI. (2023). SDXL-Turbo Model Card. Hugging Face. https://huggingface.co/stabilityai/sdxl-turbo

Stability AI. (2023). SDXL Turbo. Stability AI News. https://stability.ai/news/stability-ai-sdxl-turbo

Uc Castillo, J. L., Marín Celestino, A. E., Martínez Cruz, D. A., Tuxpan Vargas, J., Ramos Leal, J. A., & Morán Ramírez, J. (2025). A systematic review of Machine Learning and Deep Learning approaches in Mexico: challenges and opportunities. Frontiers in Artificial Intelligence, 7, Article 1479855. https://doi.org/10.3389/frai.2024.1479855

Wang, X. J. (2023). La inteligencia artificial en la industria publicitaria [Trabajo de Fin de Grado, Facultad de Ciencias Sociales, Jurídicas y de la Comunicación, Universidad de Valladolid]. https://uvadoc.uva.es/handle/10324/61666

Wang, X., Zhang, Y., Chen, H., Liu, J., & Zhao, T. (2025). ComfyGPT: A Self-Optimizing Multi-Agent System for Comprehensive ComfyUI Workflow Generation. arXiv. https://doi.org/10.48550/arXiv.2503.17671

Wikimedia Commons. (2006). XOR_perceptron_net.png [Imagen]. Dominio público. https://commons.wikimedia.org/wiki/File:XOR_perceptron_net.png

Zhang, H., Xu, T., Li, H., Zhang, S., Huang, X., Wang, X., & Metaxas, D. N. (2023). ControlNet: Adding Conditional Control to Text-to-Image Diffusion Models. arXiv preprint arXiv:2302.05543. https://doi.org/10.48550/arXiv.2302.05543


UNIVERSIDAD DE GRANADA

MÁSTERES ugr