

UNIVERSIDAD MIGUEL HERNÁNDEZ DE ELCHE

ESCUELA POLITÉCNICA SUPERIOR DE ELCHE

GRADO EN INGENIERÍA INFORMÁTICA EN  
TECNOLOGÍAS DE LA INFORMACIÓN



**UNIVERSITAS**  
*Miguel Hernández*

UNIVERSITAS Miguel Hernández

"Generación de imágenes con Inteligencia Artificial: revisión de la situación actual y aplicaciones prácticas"

TRABAJO FIN DE GRADO

Junio - 2023

AUTOR: Jorge Lara Artolazábal

DIRECTOR/ES: Antonio Peñalver Benavent



# AGRADECIMIENTOS

Me complace expresar mis más sinceros agradecimientos a las personas que han contribuido de manera significativa a la realización de este Trabajo de Fin de Grado.

En primer lugar, me gustaría agradecer a mi tutor, Antonio Peñalver Benavent, por su invaluable ayuda y dedicación a lo largo de este proyecto. Su guía y consejos han sido fundamentales para su desarrollo y han enriquecido mi experiencia de aprendizaje.

Asimismo, quiero expresar mi gratitud a Federico Botella Beviá por haberme acogido en el proyecto "SMARTUNI" de la UMH, bajo su dirección. Actualmente, trabajar a jornada completa en este proyecto ha sido una gran oportunidad para mi crecimiento académico y profesional.

También quiero reconocer y agradecer a Manuel García Cotes y Eugenio Canals Orts, quienes no solo son amigos cercanos, sino también compañeros de carrera. Su apoyo constante y su colaboración han sido fundamentales para el éxito alcanzado en este grado.

Por último, pero no menos importante, quiero expresar mi más profundo agradecimiento a mi círculo más cercano. A mi pareja, María Rosa Varó Laín, y a mis padres, Pedro Lara Jiménez y María Sacramento Artolazábal Mínguez por su apoyo incondicional, paciencia y comprensión.



# RESUMEN

Este Trabajo de Fin de Grado constituye una investigación exhaustiva y un análisis profundo sobre la generación de imágenes mediante el uso de Inteligencia Artificial. En primer lugar, se examinará detalladamente el estado actual de esta tecnología y se abordarán los aspectos de nuestra vida que pueden verse afectados por ella, además de presentar diversas alternativas disponibles en el mercado.

El enfoque práctico de este trabajo se centrará en la evaluación a fondo de los modelos DALL-E, Midjourney y Stable Diffusion, con el objetivo de comparar y analizar de manera rigurosa tanto sus ventajas como sus limitaciones. En particular, se dedicará especial atención al último modelo mencionado, Stable Diffusion, debido a su naturaleza de código abierto. Asimismo, se llevarán a cabo pruebas con varias herramientas desarrolladas por la comunidad en relación con este modelo.

El análisis realizado y la información expuesta en la presente memoria reflejan el estado de la cuestión hasta junio de 2023. En los últimos tiempos, se han experimentado notables avances en este campo, lo que hace plausible la aparición inminente de nuevas herramientas o versiones mejoradas de las existentes, con funcionalidades adicionales o mejoras respecto a lo analizado en este trabajo. Por tanto, es importante tener en cuenta que el panorama tecnológico puede evolucionar rápidamente, y es probable que surjan nuevos desarrollos que amplíen o modifiquen los conocimientos aquí expuestos.



# ÍNDICE

<b>1 INTRODUCCIÓN</b> .....	<b>17</b>
1.1 INTELIGENCIA ARTIFICIAL EN LA ACTUALIDAD .....	18
1.2 GENERACIÓN DE IMÁGENES MEDIANTE INTELIGENCIA ARTIFICIAL EN LA ACTUALIDAD.....	19
1.3 CONTROVERSIA CON LOS ARTISTAS.....	21
1.4 TRANSFORMERS.....	23
1.5 REDES ADVERSARIAS GENERATIVAS.....	25
<b>2 ESTADO DE LA CUESTIÓN</b> .....	<b>28</b>
2.1 MODELOS EN LA ACTUALIDAD .....	29
2.2 DALL-E.....	30
2.3 MIDJOURNEY.....	32
2.4 STABLE DIFFUSION.....	33
2.5 CONTROLNET .....	34
2.6 OTROS MODELOS DE GENERACIÓN DE IMÁGENES .....	35
<b>3 OBJETIVOS DEL PROYECTO</b> .....	<b>39</b>
3.1 OBJETIVOS GENERALES .....	40
3.2 OBJETIVOS TÉCNICOS.....	40
3.3 OBJETIVOS DIDÁCTICOS.....	41
3.4 OBJETIVOS PERSONALES.....	42
3.5 NO OBJETIVOS.....	42
<b>4 HIPÓTESIS DE TRABAJO</b> .....	<b>45</b>
4.1 USO DE LOS MODELOS.....	46
4.1.1 <i>DALL-E</i> .....	<b>46</b>
4.1.2 <i>MIDJOURNEY</i> .....	<b>46</b>
4.1.3 <i>STABLE DIFFUSION</i> .....	<b>48</b>
4.3 ANÁLISIS DEL PROMPT .....	48
4.4 HERRAMIENTAS COMPLEMENTARIAS STABLE DIFFUSION .....	49
4.4.1 <i>GOOGLE COLAB / PAPERSPACE</i> .....	<b>49</b>

4.4.2	<i>AUTOMATIC1111</i> .....	50
4.4.3	<i>DREAMBOOTH</i> .....	52
4.4.4	<i>CONTROLNET</i> .....	53
4.5	EQUIPOS FÍSICOS.....	54
<b>5</b>	<b>METODOLOGÍA Y RESULTADOS</b> .....	<b>57</b>
5.1	PLANIFICACIÓN DEL PROYECTO.....	58
5.1.1	<i>DESGLOSE DE TAREAS</i> .....	<b>58</b>
5.1.2	<i>DIAGRAMA DE GANTT</i> .....	<b>59</b>
5.1.3	<i>MÉTRICAS</i> .....	<b>60</b>
5.2	HERRAMIENTAS DE LOS MODELOS.....	62
5.2.1	<i>DALL-E</i> .....	<b>62</b>
5.2.2	<i>MIDJOURNEY</i> .....	<b>69</b>
5.2.3	<i>STABLE DIFFUSION</i> .....	<b>76</b>
5.3	ANÁLISIS DEL PROMPT.....	77
5.4	HERRAMIENTAS COMPLEMENTARIAS STABLE DIFFUSION.....	85
5.5	RESUMEN COMPARATIVA.....	119
<b>6</b>	<b>CONCLUSIONES Y PROPUESTAS</b> .....	<b>121</b>
6.1	CONCLUSIONES.....	122
6.2	PROPUESTAS DE FUTURO.....	123
<b>7</b>	<b>BIBLIOGRAFÍA</b> .....	<b>127</b>
7.1	FUENTES BIBLIOGRÁFICAS.....	128





# ÍNDICE DE FIGURAS

Ilustración 1: Página DALL-E 2 de OpenAI.....	46
Ilustración 2: Login de la página de Midjourney .....	47
Ilustración 3: Plan de suscripciones de Midjourney .....	47
Ilustración 4: WebUI AUTOMATIC1111 para Stable Diffusion .....	48
Ilustración 5: GitHub Fast Stable Diffusion de TheLastBen.....	50
Ilustración 6: Intefaz de la WebUI AUTOMATIC111 .....	51
Ilustración 7: Apartado Extensions de AUTOMATIC1111 .....	51
Ilustración 8: Notebook de google colab para Dreambooth .....	52
Ilustración 9: Ejemplo de filtros en ControlNet .....	53
Ilustración 10: Ejemplo del filtro OpenPOSE con ControlNet .....	54
Ilustración 11: Listado de especificaciones del Hardware usado para el plan gratuito de Gradient Paperspace .....	55
Ilustración 12: GPU para el plan gratuito de Google Colab.....	56
Ilustración 13: RAM, RAM de la GPU y espacio de almacenamiento para el plan gratuito de Google Colab.....	56
Ilustración 14: Desglose de tareas para la planificación del proyecto.....	59
Ilustración 15: Diagrama de Gantt para la planificación del proyecto .....	60
Ilustración 16: Gráfico de barras general de las métricas recogidas a lo largo del proyecto .....	61
Ilustración 17: Gráfico circular por Tareas de las métricas recogidas a lo largo del proyecto .....	62
Ilustración 18: Gráfico circular por Categorías de las métricas recogidas a lo largo del proyecto .....	62
Ilustración 19: Ejemplo de uso de la plataforma de DALL-E en el que se escribe un prompt y se muestran los resultados.....	63
Ilustración 20: Ejemplo de uso de DALL-E para generar variaciones de una imagen ..	64
Ilustración 21: Resultados de las variaciones generadas por DALL-E .....	64
Ilustración 22: Ejemplo de uso de DALL-E para editar una imagen .....	65
Ilustración 23: Ejemplo de uso de la herramienta de edición de DALL-E.....	65

Ilustración 24: Ejemplo de uso de la herramienta de edición donde se borra parte de la imagen para que sea editada .....	66
Ilustración 25: Resultado de la edición de la imagen por partes .....	66
Ilustración 26: herramienta para subir imágenes propias de DALL-E donde permite editar y generar variaciones .....	67
Ilustración 27: Resultado de las variaciones con la imagen propia .....	67
Ilustración 28: Ejemplo de edición de una parte de la imagen, la parte delantera de la camiseta .....	68
Ilustración 29: Resultados de la edición de la imagen con distintos logos en la camiseta .....	68
Ilustración 30: Ejemplo de uso de edición para cambiar todo el escenario menos la cara de la persona .....	68
Ilustración 31: Resultados de la edición de la imagen cambiando todo el escenario .....	68
Ilustración 32: Discord de Midjourney .....	69
Ilustración 33: Ejemplo de uso del comando /imagine para introducir el prompt en Midjourney .....	70
Ilustración 34: Ejemplo generación de imágenes a partir del prompt anterior .....	71
Ilustración 35: Uso del botón U2 para generar la segunda imagen de las cuatro motradas .....	71
Ilustración 36: Resultado de la generación de la imagen con toda la resolución .....	72
Ilustración 37: Resultado del uso del botón "Make Variations" para generar variaciones .....	73
Ilustración 38: Imagen propia usada para probar las variaciones con Midjourney .....	73
Ilustración 39: Ejemplo del prompt necesario con el link de la imagen al principio y la descripción después .....	74
Ilustración 40: Resultados de las variaciones generadas con la imagen y el prompt anterior .....	74
Ilustración 41: Imagen completa de las variaciones anteriores .....	75
Ilustración 42: Interfaz de la webUI AUTOMATIC1111 para Stable Diffusion .....	76
Ilustración 43: Resultados para el prompt "Strawberry" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	78

Ilustración 44: Resultados para el prompt "Robot" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	78
Ilustración 45: Resultados para el prompt "Beach" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	78
Ilustración 46: Resultados para el prompt "Happiness" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	79
Ilustración 47: Resultados para el prompt "Skater" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	79
Ilustración 48: Resultados para el prompt "A dog eating strawberries" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente.....	80
Ilustración 49: Resultados para el prompt "Robot arm building a car" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente.....	80
Ilustración 50: Resultados para el prompt "Multiple tourists sunbathing in the beach" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	80
Ilustración 51: Resultados para el prompt "Happy person giving flowers away" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	81
Ilustración 52: Resultados para el prompt "Two Skaters training a new trick" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	81
Ilustración 53: Resultados para el prompt "Cartoon picture of a funny morel mushroom waving to his friends, . Background Setting is a town with half-timbered houses. 8k UHD. Vibrant colors. 3D Hyper realistic." para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	82
Ilustración 54: Resultados para el prompt "Photogenic woman holding a magical book, magical glowing trails, light dust, fantasy robes, highly detailed fantasy art, cinematic lighting, sharp focus, unreal, render, 4k" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	82
Ilustración 55: Resultados para el prompt "Full body photo of a scary alien creature that looks like a humanoid spider with venom coming out from its mouth emerging from a dark cave, green deadly eyes, horror, dark, intricate design and details, dramatic lighting, photorealistic, cinematic, 8k, cinematic noise" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	83

Ilustración 56: Resultados para el prompt "Human colony on unknown planet with retro futuristic buildings, bright, vibrant, colored, vegetation, hyperrealism, highly detailed, insanely detailed, lush detail" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	83
Ilustración 57: Resultados para el prompt "Man wearing a fullbody suit in a disco pub party, his face in the foreground, happy, hyper realistic, detailed touches, neon lighting, sharp focus, 4k " para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente .....	83
Ilustración 58: Ventana de bienvenida de Google Colab .....	85
Ilustración 59: cuaderno Jupyter de google colab en blanco .....	86
Ilustración 60: Github Fast Stable Diffusion de TheLastBen.....	86
Ilustración 61: Pantalla de login para Paperspace .....	87
Ilustración 62: Selección de producto en Paperspace, con las opciones Gradient y CORE .....	87
Ilustración 63: Opciones previas para el inicio en Gradient Paperspace.....	88
Ilustración 64: Pantalla nuevo proyecto para el inicio de Gradient Paperspace.....	88
Ilustración 65: Opción de nuevo proyecto desde la plataforma .....	88
Ilustración 66: Proyecto de Gradient Paperspace para Fast Stable Diffusion .....	89
Ilustración 67: Cuaderno Jupyter de Gradient Paperspace para AUTOMATIC1111 ....	90
Ilustración 68: Instalación de dependencias y repo de AUTOMATIC1111 .....	90
Ilustración 69: Descarga del modelo a utilizar para AUTOMATIC1111 .....	91
Ilustración 70: Instalación de la extensión de ControlNet para AUTOMATIC1111 .....	91
Ilustración 71: Despliegue de la interfaz AUTOMATIC1111 .....	92
Ilustración 72: Interfaz de inicio de AUTOMATIC1111 .....	92
Ilustración 73: Ejemplo de uso de text2img desde AUTOMATIC1111 .....	93
Ilustración 74: Ejemplo de uso de img2img desde AUTOMATIC1111 .....	93
Ilustración 75: Apartado Settings de AUTOMATIC1111 .....	94
Ilustración 76: Apartado Extensions -> Installed de AUTOMATIC1111 .....	95
Ilustración 77: Apartado Extensions -> Available de AUTOMATIC1111.....	95
Ilustración 78: Dataset utilizado para el entrenamiento con Dreambooth.....	97
Ilustración 79: Ejemplo de Dataset incorrecto para el entrenamiento con Dreambooth	97
Ilustración 80: Conexión del Cuaderno Jupyter de Google Colab con Google Drive ...	98

Ilustración 81: Instalación de dependencias de Dreambooth .....	98
Ilustración 82: Elección del modelo para entrenar .....	98
Ilustración 83: Creación de la sesión .....	99
Ilustración 84: Subida del Dataset mostrado anteriormente .....	99
Ilustración 85: Primera parte de los parámetros del entrenamiento con Dreambooth .	100
Ilustración 86: Segunda parte de los parámetros del entrenamiento con Dreambooth	101
Ilustración 87: Modelo entrenado y modelos guardados en el proceso.....	101
Ilustración 88: Instalación de dependencias de AUTOMATIC1111 .....	102
Ilustración 89: Instalación de AUTOMATIC1111 .....	103
Ilustración 90: Descarga del modelo a usar en AUTOMATIC1111 .....	103
Ilustración 91: Instalación de la Extensión ControlNet.....	104
Ilustración 92: Despliegue de la webUI AUTOMATIC1111 con el modelo entrenado .....	104
Ilustración 93: Imágenes generadas con el modelo entrenado para el estilo "Van Gogh" .....	105
Ilustración 94: Imágenes generadas con el modelo entrenado para el estilo "Carne Griffiths" .....	105
Ilustración 95: Imágenes generadas con el modelo entrenado para el estilo "Leonid Afremov" .....	106
Ilustración 96: Imágenes generadas con el modelo entrenado para el estilo "Jean Baptiste Monge" .....	106
Ilustración 97: Imágenes generadas con el modelo entrenado para el estilo "Erin Hanson" .....	106
Ilustración 98: Imágenes generadas con el modelo entrenado para el estilo "Watercolor drawing" .....	107
Ilustración 99: Imágenes generadas con el modelo entrenado para el estilo "Bokeh Cinematic Photography" .....	107
Ilustración 100: Imágenes generadas con el modelo entrenado para el estilo "Alasdair McLellan" .....	107
Ilustración 101: Imágenes generadas con el modelo entrenado para el estilo "Funko Pop" .....	108

Ilustración 102: Imágenes generadas con el modelo entrenado para el estilo "Escultura Griega" .....	108
Ilustración 103: Ejemplo de uso ControlNet para la creación del filtro depth2img con una imagen de Elvis Presley .....	109
Ilustración 104: Ejemplo de uso ControlNet con el resultado de depth2img .....	110
Ilustración 105: Otro ejemplo de filtro depth2img de un hombre haciendo un mate en una canasta.....	111
Ilustración 106: Resultados para el ejemplo de la canasta con el token jlarart creado en el entrenamiento .....	111
Ilustración 107: Ejemplo para el filtro depth2img con un fotograma de una película de guerra .....	112
Ilustración 108: Resultados para el filtro de la película de guerra con el token jlarart del modelo entrenado .....	112
Ilustración 109: Ejemplo para el filtro "Hed"de ControlNet con la imagen de la famosa escena de El Resplandor .....	113
Ilustración 110: Resultados para el filtro Hed con el token jlarart del modelo entrenado .....	114
Ilustración 111: Ejemplo de creación de un filtro OpenPose para la lectura de posturas, en este caso de yoga.....	114
Ilustración 112: Resultados para el filtro OpenPose de yoga .....	115
Ilustración 113: Ejemplo de creación de un filtro Scribble con la imagen de un perro dibujado .....	115
Ilustración 114: Resultado filtro Scribble con el Control Weight a 2 .....	116
Ilustración 115: Resultado filtro Scribble con el Control Weight a 1 .....	117
Ilustración 116: Resultado filtro Scribble con el Control Weight a 0.8 .....	117
Ilustración 117: Resultado filtro Scribble con el Control Weight a 0.5 .....	118
Ilustración 118: Tabla comparativa de los modelos DALL-E, Midjourney y Stable Diffusion.....	120





# **CAPÍTULO 1:**

## **INTRODUCCIÓN**



## 1.1 INTELIGENCIA ARTIFICIAL EN LA ACTUALIDAD

La Inteligencia Artificial ha sido un tema cada vez más relevante en los últimos años y ha logrado avances significativos en una variedad de áreas. Algunas de estas áreas son Sistemas Expertos [1], Redes Neuronales Artificiales [2], Deep Learning [3], Robótica [4] y Agentes Inteligentes [5].

- **Sistemas expertos:** Estos sistemas emulan la toma de decisiones de expertos humanos y se aplican en diversas áreas. Un ejemplo es Dendral [6], un sistema experto creado para ayudar a químicos a resolver la estructura de compuestos desconocidos. Dendral les proporciona información valiosa sobre la composición química de una sustancia desconocida en función de datos experimentales.
- **Redes neuronales artificiales:** Estas son sistemas inspirados en las redes neuronales biológicas que procesan la información en el cerebro humano. Un ejemplo de uso es la tecnología de conducción autónoma de compañías como Tesla (AutoPilot) [7], que utiliza redes neuronales para procesar la enorme cantidad de información sensorial requerida para navegar y controlar un vehículo de forma segura y eficiente. Una variante de estas serían las redes neuronales generativas adversarias, que son la mayoría de modelos de generación de imágenes.
- **Deep learning:** El Deep Learning es una rama del aprendizaje automático que utiliza redes neuronales profundas para abordar problemas complejos. Los asistentes de voz, como Alexa [8] de Amazon, emplean el Deep Learning para procesar lenguaje natural y entender las solicitudes de los usuarios. Estos asistentes son capaces de responder preguntas, realizar tareas y brindar recomendaciones basadas en la información recopilada de la interacción con los usuarios.

- **Robótica:** La robótica abarca el diseño, construcción y operación de robots, así como su programación e integración con otros sistemas. Un ejemplo notable es Sophia [9], el primer robot con Inteligencia Artificial desarrollado por la empresa Hanson Robotics. Sophia puede interactuar con las personas, reconocer rostros y expresiones faciales y mantener una conversación coherente.
- **Agentes inteligentes:** Estos son sistemas que imitan de forma racional el comportamiento humano y se adaptan a diferentes situaciones y entornos. Un ejemplo de agente inteligente es ChatGPT [10], un modelo de lenguaje desarrollado por OpenAI que puede mantener conversaciones coherentes y útiles con los usuarios. ChatGPT puede responder preguntas, brindar información y ayudar en la interacción con otras herramientas y sistemas, adaptándose a las necesidades de cada persona.

Estos ejemplos demuestran que la IA juega un papel fundamental en muchas áreas de nuestras vidas, abordando problemas complejos y mejorando continuamente nuestra cotidianidad a medida que la tecnología avanza.

## **1.2 GENERACIÓN DE IMÁGENES MEDIANTE INTELIGENCIA ARTIFICIAL EN LA ACTUALIDAD**

La generación de imágenes mediante Inteligencia Artificial es un subcampo dentro del Deep Learning y la Visión Artificial [11], que se basa en el uso de modelos de aprendizaje automático para crear imágenes realistas o artísticas a partir de descripciones textuales, transformaciones de imágenes existentes o incluso generaciones aleatorias.

Estos modelos generalmente se basan en Redes Neuronales Artificiales, en particular, las Redes Adversarias Generativas (GAN) [12] y los modelos Transformer [13]. Las GAN consisten en dos redes neuronales, el generador y el discriminador, que trabajan en conjunto para producir imágenes realistas. El generador crea imágenes, mientras que el discriminador evalúa qué tan realistas son estas imágenes. Ambos componentes de la GAN aprenden iterativamente a mejorar su rendimiento en función del desempeño del otro componente.

También usan otras herramientas secundarias para refinar tareas más simples y así conseguir un mejor rendimiento, como es el caso del tratamiento inicial del texto para conseguir un lenguaje que se pueda entender mejor como input a nivel conceptual.

El impacto de la Inteligencia Artificial en la generación de imágenes en la sociedad es amplio y diverso. Por un lado, estos modelos han revolucionado la creatividad y el diseño, permitiendo a artistas y profesionales generar imágenes y arte visual a partir de simples descripciones. Esto ha llevado a un aumento en la creación de contenido digital y un nuevo enfoque en el campo del arte.

Además, en la industria del entretenimiento y la publicidad, la generación de imágenes mediante Inteligencia Artificial ayuda a crear entornos, objetos y personajes realistas o imaginativos para películas, videojuegos y campañas de marketing.

Sin embargo, también existen preocupaciones respecto al uso indebido de estas tecnologías. Por ejemplo, la creación de imágenes falsas o Deepfakes, que pueden usarse con fines maliciosos, como manipulación de información y noticias, difamación y acoso. Por lo tanto, podría ser necesario desarrollar políticas, regulaciones adecuadas para abordar estos desafíos éticos y garantizar que la Inteligencia Artificial en la generación de imágenes no se usa de forma irresponsable o al menos poder limitarla por otros medios, por ejemplo verificaciones externas.

## 1.3 CONTROVERSIA CON LOS ARTISTAS

La irrupción de la Inteligencia Artificial en la generación de imágenes ha generado un debate sobre el impacto en el trabajo de artistas y diseñadores. Algunos argumentan que estas tecnologías pueden desplazar a los profesionales al automatizar tareas creativas y de diseño, lo que podría amenazar la forma en que ganan su sustento y reconocimiento.

Desde una perspectiva negativa, algunos de los puntos importantes son:

Reducir la demanda de artistas y diseñadores humanos: Algunos temen que la IA pueda llegar a ser tan avanzada que sus creaciones sean indistinguibles de las obras de arte humanas, lo que podría conducir a una disminución en la demanda de artistas y diseñadores humanos, causando un impacto económico en el sector creativo.

Desvalorización del arte y del oficio: Existe la preocupación de que la IA, al tener la capacidad de generar imágenes rápidas y en gran volumen, pueda dar lugar a una saturación del mercado, lo que a su vez haría que el valor de las obras de arte disminuya y se pierda la valoración por el proceso creativo y el talento humano.

Pérdida de la autoría y la originalidad: La Inteligencia Artificial en la generación de imágenes podría desdibujar las líneas entre el trabajo humano y el generado por la máquina, lo que podría generar problemas de autoría y originalidad a artistas y diseñadores. Esto podría plantear preguntas sobre la protección de los derechos de autor y el reconocimiento artístico.

Debido a estos puntos negativos, hay artistas librando una batalla tanto en internet como en los tribunales en defensa de los derechos de autor para esta nueva situación [14] [15].

No obstante, hay quienes argumentan que la IA en la generación de imágenes también puede ser beneficiosa en algunos aspectos:

**Estimulación de la creatividad y la colaboración:** La IA podría considerarse una herramienta que ayuda a ampliar la creatividad humana al proporcionar nuevos medios e inspiraciones para que los artistas puedan explorar. Los artistas pueden colaborar con la IA para generar obras de arte únicas que antes no eran posibles.

**Aumento de la eficiencia y la productividad:** Al automatizar ciertas tareas, la IA puede permitir a los artistas y diseñadores centrarse en aspectos más creativos y desafiantes de su trabajo, aumentando así su eficiencia y productividad.

**Democratización del arte:** La IA en la generación de imágenes puede hacer que el arte y el diseño sean más accesibles a personas con habilidades artísticas limitadas. También puede impulsar la formación y el aprendizaje en el sector, haciendo que más personas se involucren en las prácticas creativas.

Es fundamental tener en cuenta que el impacto de la IA en la generación de imágenes en el campo del arte y el diseño dependerá de cómo los artistas, diseñadores y responsables de políticas aborden y se adapten a estos cambios en la tecnología y el mercado en constante evolución.

## 1.4 TRANSFORMERS

Los Transformers son una clase de modelos de aprendizaje profundo basados en la arquitectura de redes neuronales que han revolucionado el procesamiento del lenguaje natural (NLP, por sus siglas en inglés) y también se han utilizado en otras áreas, como la generación de imágenes y la visión por computadora. La arquitectura Transformer permite a los modelos aprender relaciones complejas y estructuras de datos secuenciales.

Los Transformers se basan en la autoatención y la mecanización del proceso de atención a diferentes partes de una secuencia en función de su relevancia en un contexto específico. Esto permite a los Transformers capturar dependencias a larga distancia y relaciones entre elementos de una secuencia, como palabras en un texto.

A continuación, se presenta un resumen general del funcionamiento de los Transformers:

1. Codificación / Embedding: En primer lugar, los datos de entrada, como un texto, se convierten en representaciones vectoriales (Embeddings) que preservan la información semántica de las palabras y su posición en la secuencia.
2. Capas de autoatención: La clave de la arquitectura Transformer es la autoatención. La autoatención permite que el modelo asigne diferentes pesos a cada palabra en función de su relevancia en el contexto. Cada palabra se compara con las demás palabras en la secuencia y se calcula un puntaje de atención para cada par de palabras. Los puntajes de atención se utilizan para calcular un vector de contexto que combina información de todas las palabras en la secuencia. Este proceso se realiza varias veces en paralelo mediante diferentes cabezales de

atención, lo que permite a los Transformers aprender diferentes patrones y relaciones en los datos.

3. Capas de Feed-Forward: Las capas de autoatención se conectan a una serie de redes de perceptrones multicapa conocidas como capas de Feed-Forward. Estas capas procesan la información de los vectores de contexto y agregan representaciones más abstractas.

4. Normalización de capa y conexiones residuales: Para evitar el sobreajuste y mejorar el aprendizaje, los Transformers utilizan técnicas como la normalización de capa y las conexiones residuales. La normalización de capa ayuda a mantener la activación dentro de un rango controlado, mientras que las conexiones residuales facilitan el flujo de información y el proceso de retroalimentación de errores en la red neuronal.

5. Decodificación y predicción: En la parte final del proceso, los Transformers pueden usarse para decodificar información en tareas como la traducción automática, la generación de resúmenes o incluso la generación de imágenes. La decodificación generalmente involucra la generación de una secuencia paso a paso utilizando la información codificada.

En resumen, los modelos Transformer son arquitecturas de redes neuronales basadas en autoatención que han demostrado un gran desempeño en el procesamiento del lenguaje natural y también se han aplicado con éxito en otras áreas. Capturan de manera eficiente las relaciones y las dependencias a larga distancia en los datos secuenciales y han sido fundamentales en el estado actual de la Inteligencia Artificial.



## 1.5 REDES ADVERSARIAS GENERATIVAS

Las Redes Adversarias Generativas (GAN, por sus siglas en inglés) son un tipo de modelo de aprendizaje profundo que utiliza dos redes neuronales en competencia para generar datos artificiales realistas, como imágenes, sonidos o textos. Las GAN constan de dos componentes principales: un generador y un discriminador.

**Generador:** La función del generador es crear datos artificiales (por ejemplo, imágenes) a partir de un espacio de vectores latentes, que puede considerarse como una representación compacta de cierto aspecto del mundo real. El generador toma como entrada un vector aleatorio (ruido) y lo transforma en datos que se parezcan a los datos reales del dominio de interés.

**Discriminador:** La función del discriminador es determinar si un dato (por ejemplo, una imagen) proviene de los datos reales o si ha sido creado por el generador. Para entrenar al discriminador, se le proporciona tanto datos reales como datos generados por el generador, y su objetivo es diferenciar entre ambos.

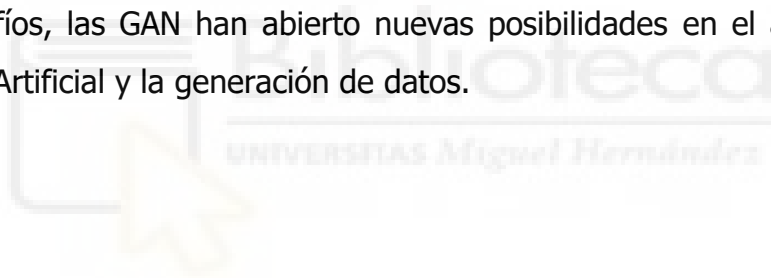
El proceso de entrenamiento de las GAN es un juego iterativo de suma cero entre el generador y el discriminador. En cada iteración, el generador intenta mejorar su capacidad para crear datos artificiales que engañen al discriminador, mientras que el discriminador intenta mejorar su capacidad para distinguir entre datos reales y generados. Este proceso continuaría hasta que el generador sea lo suficientemente bueno como para crear datos artificiales que el discriminador no pueda distinguir de los datos reales.

El entrenamiento se realiza mediante la minimización de una función de costo o pérdida, que cuantifica cuán bien lo está haciendo el generador y el discriminador. En teoría, cuando el proceso de entrenamiento converja, la

distribución de los datos generados por el generador sería indistinguible de la distribución de los datos reales, lo que resulta en una GAN exitosa.

Las GAN tienen numerosas aplicaciones y se han utilizado exitosamente en generación de imágenes, síntesis de voz, creación de muestras de texto, mejoramiento de la resolución de imágenes y videos, entre otras tareas. Cabe mencionar que la capacitación de GAN puede ser desafiante y a veces inestable, debido a la naturaleza competitiva de las dos redes neuronales. A lo largo de los años, se han propuesto varias mejoras y variantes para abordar estos desafíos y mejorar la formación de las GAN.

En resumen, las Redes Adversarias Generativas son un enfoque poderoso en el aprendizaje profundo para la generación de datos realistas a través de la competencia entre dos redes neuronales, el generador y el discriminador. A pesar de sus desafíos, las GAN han abierto nuevas posibilidades en el ámbito de la Inteligencia Artificial y la generación de datos.





# **CAPÍTULO 2:**

## **ESTADO DE LA CUESTIÓN**



## 2.1 MODELOS EN LA ACTUALIDAD

La Inteligencia Artificial ha avanzado significativamente en los últimos años, especialmente en el campo de la generación de imágenes. Los modelos de IA de generación de imágenes actuales han demostrado ser capaces de crear imágenes altamente realistas y detalladas a partir de descripciones en lenguaje natural o parámetros específicos. Estos modelos están abriendo nuevas posibilidades en la creación de contenido visual y están transformando la forma en que los artistas, diseñadores y profesionales de la industria generan imágenes.

Entre los modelos de IA de generación de imágenes más relevantes en la actualidad, tres destacan en particular: "Stable Diffusion" desarrollado por Stability AI [16], "DALL-E" por OpenAI [17] y "Midjourney" [18] creado por su empresa homónima.

### 1. DALL-E:

DALL-E es un modelo desarrollado por OpenAI que se basa en la arquitectura Transformer para generar imágenes realistas y artísticas a partir de descripciones en lenguaje natural. DALL-E captura detalles y relaciones complejas en el texto y utiliza esa información para crear imágenes visualmente coherentes y diversas que reflejen el contenido semántico del texto. El modelo ofrece flexibilidad y diversidad en sus imágenes generadas, con alta coherencia y detalle. Los usuarios pueden acceder a DALL-E a través de una interfaz proporcionada por OpenAI y proporcionar descripciones en lenguaje natural para generar imágenes.

### 2. Midjourney:

Midjourney toma entradas a través de indicaciones de texto y parámetros, utilizando algoritmos de Machine Learning para generar imágenes a partir de las descripciones proporcionadas. Aunque los detalles precisos de la arquitectura y las tecnologías subyacentes no están claramente disponibles, Midjourney parece compartir similitudes con otros modelos de generación de imágenes basados en

texto como DALL-E y Stable Diffusion. Los usuarios finales pueden acceder a Midjourney a través de su plataforma en línea y proporcionar descripciones en lenguaje natural para generar imágenes únicas y personalizadas.

### 3. Stable Diffusion, de Stability AI:

Stable Diffusion está basado en Modelos de Difusión Denoising Score Matching (DSM). El proceso de generación de imágenes se realiza en dos fases, la de difusión y la de difusión inversa o denoising.

Stable Diffusion proporciona una calidad y diversidad de imágenes impresionante, generando imágenes de alta resolución y detalle. Presenta robustez y estabilidad en su proceso de aprendizaje y generación de imágenes en comparación con otros modelos como las GAN. Los usuarios pueden acceder a la plataforma en línea de Stability AI y DreamStudio Lite para utilizar este modelo a través de descripciones en lenguaje natural.

## 2.2 DALL-E

DALL-E, actualmente en su versión 2, es un modelo desarrollado por OpenAI que se basa en la arquitectura Transformer para generar imágenes realistas y artísticas a partir de descripciones en lenguaje natural. Este modelo logra capturar detalles y relaciones complejas en el texto y utilizar esa información para crear imágenes visualmente coherentes y diversas que reflejen el contenido semántico del texto.

DALL-E utiliza una arquitectura Transformer similar a la de modelos de generación de texto, como GPT-3, pero adaptada para manejar tanto texto como información de imágenes. El modelo recibe una secuencia de tokens que combina texto e información de imágenes, y aprende a generar imágenes a partir de las

descripciones en el texto. Durante el entrenamiento, el modelo aprende a ajustar sus parámetros para generar imágenes más realistas y coherentes.

DALL-E presenta algunas ventajas en comparación con otros modelos de generación de imágenes:

**Flexibilidad:** DALL-E es capaz de generar una amplia gama de imágenes, tanto realistas como artísticas, a partir de descripciones en lenguaje natural, o incluso a partir de otras imágenes incompletas para construir en base estas.

**Coherencia y detalle:** A diferencia de otros modelos, DALL-E puede capturar detalles y relaciones complejas en el texto con mayor coherencia y precisión, lo que resulta en imágenes que reflejan más fielmente los elementos descritos en el texto.

**Ampliación a otras modalidades:** La arquitectura Transformer en la que se basa DALL-E también es adecuada para manejar otras modalidades, como sonido o video, lo que podría permitir el desarrollo de modelos capaces de generar contenido multimedia a partir de descripciones de texto.

El usuario final puede acceder a DALL-E a través de una interfaz proporcionada por OpenAI. Para usar DALL-E, los usuarios pueden ingresar descripciones en lenguaje natural, como "un paisaje montañoso al atardecer con un lago" o "una casa futurista en forma de huevo", y el modelo generará imágenes que reflejen esas descripciones. Los usuarios pueden modificar sus descripciones o proporcionar instrucciones adicionales para afinar aún más las imágenes generadas según sus preferencias o necesidades creativas.

## 2.3 MIDJOURNEY

Midjourney es un programa y servicio de Inteligencia Artificial creado por el laboratorio de investigación independiente Midjourney con sede en San Francisco. Se trata de una herramienta de generación de imágenes de IA que toma entradas a través de indicaciones de texto y parámetros, utilizando un algoritmo de Machine Learning para generar imágenes a partir de las descripciones proporcionadas.

Aunque los detalles específicos de la arquitectura y las tecnologías subyacentes no están claramente disponibles, Midjourney comparte ciertas similitudes con otros generadores de imágenes basados en texto como DALL-E y Stable Diffusion. Esto sugiere que se basa en tecnologías de aprendizaje profundo y posiblemente Transformers para generar imágenes que coincidan con las descripciones proporcionadas por los usuarios.

Midjourney V5 ofrece calidad y coherencia mejoradas en la generación de imágenes en comparación con versiones anteriores de su propio modelo, como Midjourney V4. Si bien no se especifican ventajas en su funcionamiento, a nivel de resultado se puede apreciar que Midjourney consigue imágenes más impresionantes y artísticas que la competencia.

El usuario final puede acceder al modelo de Midjourney a través de su plataforma en línea. Los usuarios ingresan descripciones en lenguaje natural de las imágenes que desean generar, y el modelo crea imágenes basadas en estas descripciones. Esto permite a usuarios y artistas generar contenido visual único y personalizado a partir de sus ideas y preferencias.



## 2.4 STABLE DIFFUSION

Stable Diffusion es un modelo de generación de imágenes desarrollado por Stability AI. A diferencia de las Redes Adversarias Generativas (GAN) y los Transformers, Stable Diffusion se basa en un enfoque llamado Modelos de Difusión "Denoising Score Matching" (DSM) [19]. Esta tecnología utiliza un proceso de difusión reversible para generar imágenes realistas y de alta resolución a partir de descripciones textuales y ruido aleatorio.

El proceso de generación de imágenes en Stable Diffusion se divide en dos fases:

Fase de difusión: En esta etapa, se toma una imagen real y se corrompe gradualmente con ruido hasta llegar a un estado puramente ruidoso.

Fase de denoising (desruido): En esta etapa, el modelo de difusión aprende a eliminar el ruido agregado en la fase de difusión. El proceso de desruido se realiza de manera iterativa y reversa desde el estado ruidoso hasta llegar a una imagen reconstruida que se parece a la imagen original. Durante el proceso de entrenamiento, el modelo DSM aprende a ajustar sus parámetros para mejorar su capacidad de desruido y reconstruir las imágenes.

Stable Diffusion ofrece algunas ventajas en comparación con otros modelos de generación de imágenes como GAN y Transformers:

Calidad y de imágenes: Stable Diffusion es capaz de generar imágenes de alta resolución y detalle. Además, este enfoque puede generar imágenes con menos artefactos y distorsiones.

Robustez y estabilidad: A diferencia de las GAN, que pueden ser difíciles de entrenar debido a la competencia entre el generador y el discriminador, los

modelos de difusión pueden ser más estables y robustos en términos de entrenamiento y generación de imágenes.

Código abierto: todo el código de este modelo es completamente libre, lo que viene con muchas ventajas como poder ejecutarlo cómo quieras, modificarlo y poder aprovechar todo tipo de modificaciones que ha hecho la comunidad.

El usuario final puede utilizar Stable Diffusion a través de la plataforma en línea proporcionada por Stability AI, como DreamStudio Lite o Stable Diffusion 2.1 Demo en Hugging Face Spaces o ejecutando el código de forma local gracias a ser un proyecto de código abierto. Los usuarios pueden ingresar descripciones en lenguaje natural de las imágenes que desean generar, y el modelo crea imágenes basadas en estas descripciones. Esto permite a usuarios y artistas generar contenido visual único y personalizado a partir de sus ideas y preferencias.

## 2.5 CONTROLNET

ControlNet [20] es una herramienta creada también por Stability AI, consiste en una estructura de red neuronal diseñada para agregar un control adicional a los modelos de difusión, como el Stable Diffusion, al incluir condiciones adicionales. ControlNet ha revolucionado la generación de imágenes de IA proporcionando un control sin precedentes sobre la generación de imágenes en términos de estilo y estructura.

En el caso de Stable Diffusion, los modelos de texto a imagen utilizan las indicaciones de texto como única información de entrada. ControlNet mejora este enfoque al agregar entradas condicionales como mapas de bordes, máscaras de

segmentación y mapas de características, aportando mayor control sobre el aspecto final de las imágenes generadas.

ControlNet se puede combinar con la difusión enmascarada predeterminada de Stable Diffusion. Por ejemplo, el modelo Canny Edge puede ser utilizado para controlar la generación de imágenes en función de los bordes en una imagen base.

En resumen, ControlNet es una estructura de red neuronal diseñada para mejorar los modelos de difusión como Stable Diffusion de Stability AI, permitiendo un control más preciso y artístico sobre las imágenes generadas. Esta innovadora tecnología ha ampliado aún más las posibilidades en el ámbito de la generación de imágenes y ha proporcionado a los usuarios una mayor flexibilidad para personalizar y controlar sus creaciones visuales.

## **2.6 OTROS MODELOS DE GENERACIÓN DE IMÁGENES**

Además de todo lo mencionado anteriormente, existen una gran variedad de IAs generadoras de imágenes, muchas de ellas buscan especializarse en algún formato para poder destacar entre tanta abundancia de opciones. Algunas de ellas son:

Craiyon (anteriormente llamado DALL-E mini) [21]: Esta plataforma AI de generación de imágenes crea impresionantes imágenes y arte basadas en descripciones de texto proporcionadas por los usuarios. Craiyon es un generador de arte en línea que utiliza algoritmos de aprendizaje profundo y es popular por su capacidad de producir resultados sorprendentes y únicos en segundos. Las imágenes creadas por Craiyon están siendo utilizadas por diversos sectores, como

el diseño gráfico, la publicidad y las redes sociales. El generador es fácil de usar y permite a los usuarios experimentar con diferentes solicitudes para obtener resultados diversos y emocionantes.

Adobe Firefly [22]: Adobe, líder en la industria del software creativo, ha desarrollado Firefly, una nueva línea de modelos generativos de Inteligencia Artificial diseñados para crear imágenes y efectos de texto. Firefly busca revolucionar la generación automática de imágenes y ofrecer una solución integrada y sofisticada para lograr resultados artísticos de alta calidad en productos Adobe, como Photoshop y Illustrator. Firefly promete acelerar y mejorar la experiencia de los profesionales creativos al generar automáticamente contenido visual a partir de descripciones de texto.

Hotpot (AI Art Maker) [23]: Hotpot AI es un generador de ilustraciones e imágenes basado en Inteligencia Artificial que crea rápidamente imágenes sorprendentes y personalizadas. Hotpot AI admite la transformación de imágenes y la modificación de imágenes existentes, lo que facilita a los usuarios experimentar con diferentes estilos y temas.

NightCafe Creator [24]: NightCafe Creator es una aplicación de generación de arte en la que los usuarios pueden cargar sus propias imágenes como "base" y aplicar estilos artísticos seleccionando imágenes de estilo o utilizando descripciones de texto. La plataforma también cuenta con una comunidad en línea, donde los artistas de AI pueden compartir, colaborar y discutir sobre AI Art.

Deep Dream Generator [25]: Desarrollado por Google, el Deep Dream es un programa de visión por computadora que utiliza redes neuronales convolucionales para detectar y mejorar patrones en imágenes digitales. El resultado es una imagen de aspecto "onírico" con formas, colores y patrones únicos. Deep Dream Generator es un lugar en línea donde los usuarios pueden

experimentar con esta tecnología y transformar sus imágenes en obras de arte sorprendentes y llamativas. Los usuarios pueden controlar fácilmente diferentes configuraciones y parámetros para personalizar aún más sus resultados.

Starry AI [26]: Starry AI es una aplicación de generación de arte basada en Inteligencia Artificial que utiliza una combinación de VQGAN-CLIP y CLIP-Guided Diffusion como modelos avanzados para producir imágenes a partir de textos. Starry AI permite a los usuarios explorar estilos innovadores y crear obras de arte únicas simplemente ingresando descripciones de texto en un cuadro de texto. La aplicación ofrece granularidad en su control y es conocida por producir imágenes de alta calidad.

Generador de imágenes de Bing con tecnología de Dall-e [27]: Bing Image Creator emplea la tecnología Dall-e 2 de OpenAI para generar imágenes basadas en palabras clave proporcionadas por usuarios. Este generador permite a los usuarios obtener imágenes, su punto fuerte es estar vinculado al buscador de Bing, gracias a esto lo usarán usuarios que en un principio no se plantearían usar una herramienta como esta, además de que combinándolo con su asistente por chat basado en tecnología de OpenAI que además puede consultar resultados de internet, tendrá un mayor potencial a la hora de conseguir lo que quiere el usuario.

Modelos basados en Stable Diffusion: Gracias a ser un proyecto de código abierto, la comunidad puede aportar de múltiples maneras, y una de ellas es entrenando el modelo base que ofrece Stability AI de Stable Diffusion para añadirle estilos concretos o conceptos nuevos que poder plasmar, se pueden encontrar estos modelos en la página Hugging face [28].

Un ejemplo sería OpenJourney [29], un modelo completamente gratuito basado en Stable Diffusion entrenado con las imágenes y prompts que muestran los usuarios de Midjourney.



# **CAPÍTULO 3:**

## **OBJETIVOS DEL PROYECTO**



En este apartado se van a exponer los diferentes objetivos y no objetivos de de este proyecto, estos se dividen en los siguientes:

- Objetivos generales: se refieren al resultado que se espera lograr del sistema final del proyecto.
- Objetivos técnicos: se refieren a las metas que se deben alcanzar en la metodología de implementación del proyecto durante su desarrollo.
- Objetivos didácticos: son aquellos que se han investigado y que pueden ser útiles para el desarrollo de proyectos relacionados con el mismo tema.
- Objetivos personales: son los desafíos que el autor del proyecto ha tenido que enfrentar.

### **3.1 OBJETIVOS GENERALES**

El principal objetivo de este Trabajo de Fin de Grado es llevar a cabo una exploración exhaustiva de los modelos más relevantes en la generación de imágenes mediante el uso de Inteligencia Artificial. A través de este estudio, se busca adquirir conocimiento sobre los avances y las capacidades actuales de dichos modelos.

Se pretende analizar y evaluar las virtudes y defectos de cada modelo investigado, con el fin de comprender su funcionamiento interno y su efectividad en la generación de imágenes.

### **3.2 OBJETIVOS TÉCNICOS**

En primer lugar, se busca probar las herramientas de los modelos de generación de imágenes DALL-E, Midjourney y Stable Diffusion y realizar un análisis tanto de



las herramientas como de los resultados generados por los distintos modelos teniendo en cuenta los prompts utilizados. Se compararán y evaluarán los resultados obtenidos con el fin de comprender cómo los diferentes prompts influyen en la calidad y coherencia de las imágenes generadas por cada modelo.

Además, se plantea el objetivo de entrenar el modelo de Stable Diffusion utilizando la herramienta Dreambooth. Se llevará a cabo un proceso de entrenamiento para añadir un nuevo concepto al modelo, por ejemplo un individuo, y para probar la complejidad de esta operación.

Por último, se pretende probar y evaluar herramientas complementarias asociadas al modelo Stable Diffusion. Se explorarán diferentes herramientas creadas por la comunidad que amplían las funcionalidades y capacidades del modelo, con el propósito de analizar su efectividad, utilidad y aplicabilidad..

### **3.3 OBJETIVOS DIDÁCTICOS**

Uno de los objetivos didácticos sería concluir de manera fundamentada los métodos más óptimos a nivel de prompt para utilizar los distintos modelos de generación de imágenes.

Además, se busca determinar la forma correcta de entrenar el modelo Stable Diffusion, considerando aspectos como la estructura presente en las fotografías y las distintas resoluciones utilizadas. A través de la realización de experimentos y pruebas, se buscará establecer los parámetros y técnicas adecuadas que optimicen el rendimiento y la precisión del modelo durante el proceso de entrenamiento.

Asimismo, se tiene como objetivo facilitar el acceso y comprensión de estas potentes herramientas de generación de imágenes a un público menos experto.

## 3.4 OBJETIVOS PERSONALES

En primer lugar, deseo familiarizarme con el campo de la Inteligencia Artificial en general, centrándome especialmente en las IAs generativas de imágenes.

Además, mi objetivo es aprender a utilizar los diferentes modelos y herramientas de generación de imágenes. Quiero adquirir habilidades prácticas en la implementación y aplicación de estos modelos, aprovechando las herramientas disponibles para generar imágenes.

Otro objetivo sería comprender el funcionamiento conceptual de los distintos modelos y herramientas. Me interesa obtener un conocimiento sólido sobre los principios en la generación de imágenes mediante Inteligencia Artificial.

## 3.5 NO OBJETIVOS

En este Trabajo de Fin de Grado, es importante establecer también los No Objetivos, es decir, aquellas actividades o enfoques que no serán abordados en el desarrollo del proyecto. A continuación, se detallan los No Objetivos:

1. Desarrollar un nuevo modelo en base a otros: El objetivo principal es analizar y evaluar los modelos existentes de generación de imágenes por Inteligencia Artificial, por lo tanto, no se contempla el desarrollo de un nuevo modelo a partir de otros existentes.
2. Experimentar con todos y cada uno de los modelos disponibles actualmente: Si bien se explorarán modelos relevantes en la generación de imágenes, no se realizará un experimento exhaustivo con todos los modelos disponibles en la actualidad. El enfoque se centrará en aquellos

modelos que sean relevantes y estén relacionados directamente con los objetivos del trabajo.

3. Experimentar con todas las herramientas complementarias de los modelos con los que sí se va a experimentar: Si bien se probarán herramientas complementarias asociadas al modelo Stable Diffusion, no se experimentará con todas las herramientas disponibles para cada modelo investigado. El enfoque se centrará en aquellas herramientas que sean relevantes y aporten valor al análisis y evaluación de los modelos.
4. Profundizar en los modelos de DALL-E y MidJourney al mismo nivel que con el de Stable Diffusion: Si bien se analizarán los modelos de DALL-E y MidJourney, el enfoque principal del trabajo será en el modelo Stable Diffusion. No se profundizará en estos modelos al mismo nivel que en el caso de Stable Diffusion, aunque se realizará un análisis y evaluación adecuados para comprender sus características y capacidades.



# CAPÍTULO 4

## HIPÓTESIS DE TRABAJO

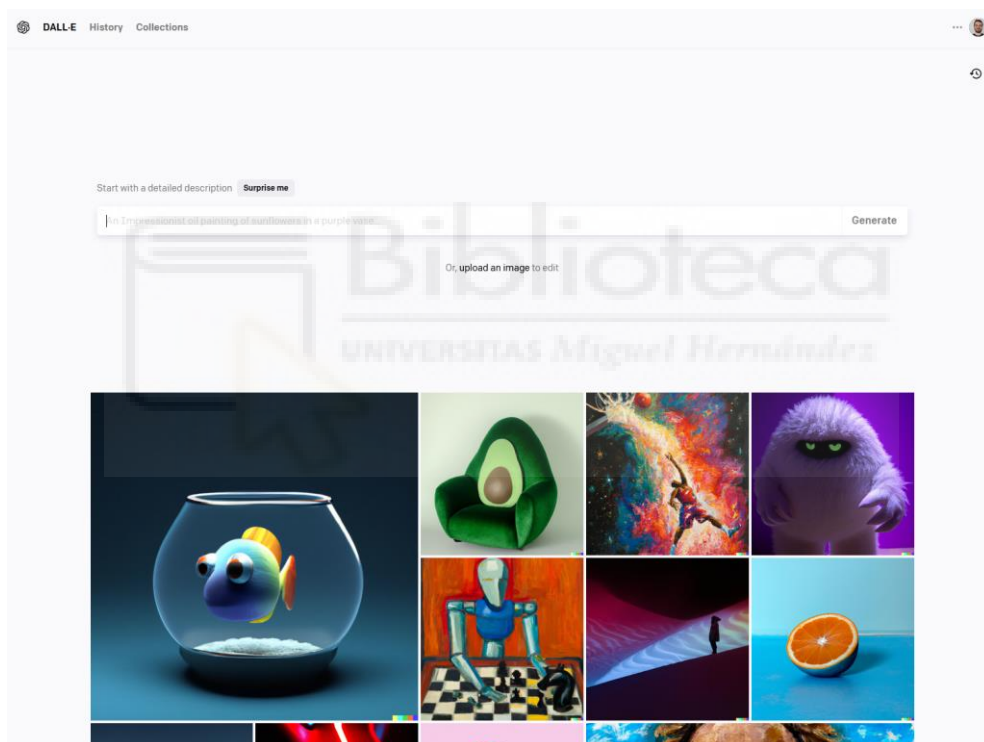


## 4.1 USO DE LOS MODELOS

### 4.1.1 DALL-E

Se probará la plataforma que ofrece OpenAI en su página web para usar su modelo de generación de imágenes DALL-E 2.

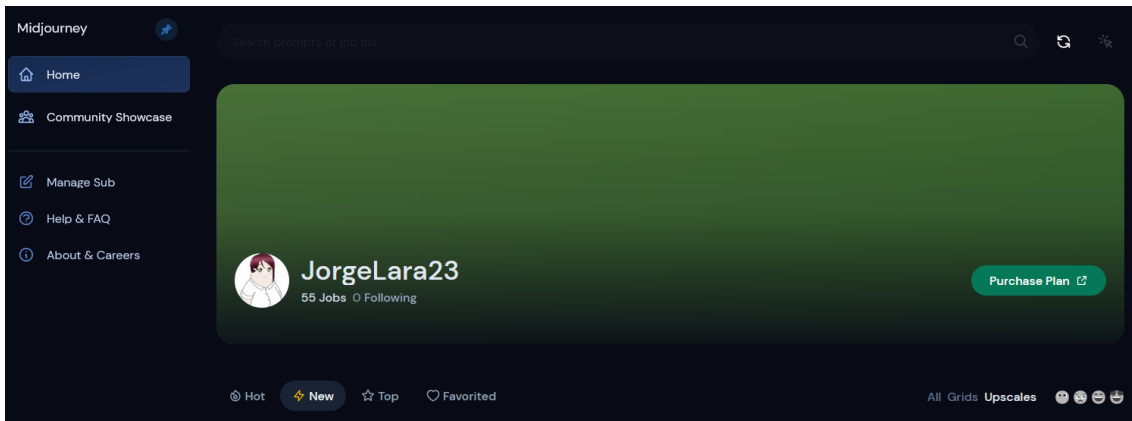
En esta se probarán las distintas funcionalidades que ofrece.



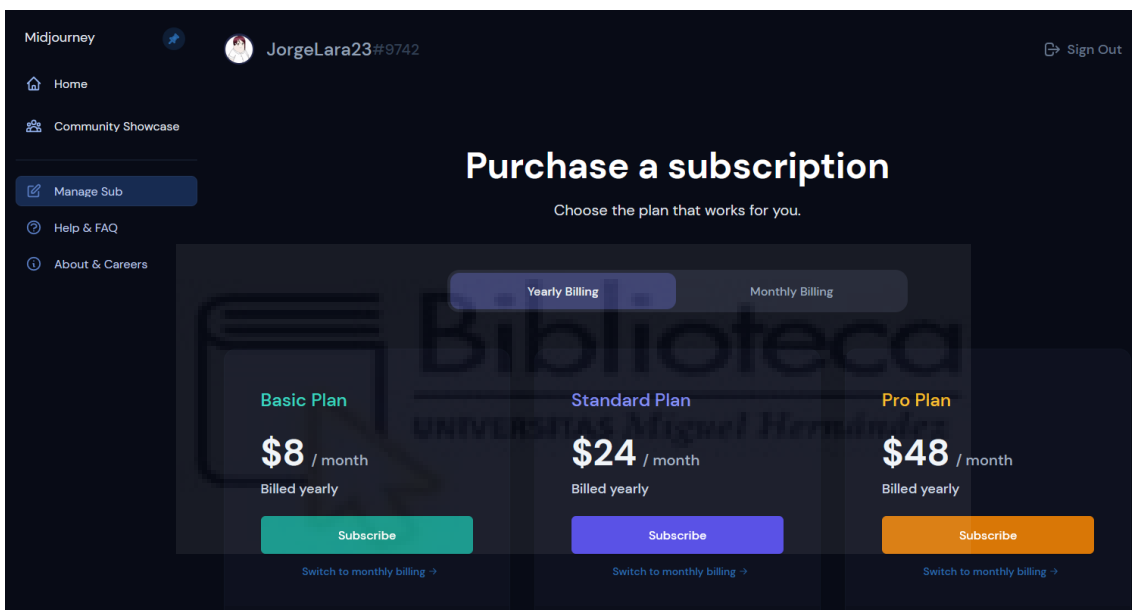
*Ilustración 1: Página DALL-E 2 de OpenAI*

### 4.1.2 MIDJOURNEY

Se probará también el modelo de MidJourney a través de su canal de discord, previamente se necesitará vincular una cuenta de discord desde su página y pagar la suscripción, ya que MidJourney no ofrece ningún plan gratuito.



*Ilustración 2: Login de la página de Midjourney*



*Ilustración 3: Plan de suscripciones de Midjourney*

A partir de aquí se conseguiría acceso al discord y desde este se podrían utilizar las diferentes funcionalidades de Midjourney a través de comandos en el chat, se podrá hacer tanto en alguno de los chats públicos como por el chat privado con el bot de Midjourney, aunque todas las imágenes generadas son igualmente accesibles por los usuarios con suscripción desde su página web.

### 4.1.3 STABLE DIFFUSION

Cómo Stable Diffusion es un modelo de código abierto no tiene una plataforma centralizada, se podría utilizar como alternativa la plataforma que ofrece hugging face, pero para aprovechar todo el potencial del modelo es mejor utilizar una interfaz también de código abierto llamada AUTOMATIC1111 [30], y esta se podría utilizar a su vez desde un servicio de computación en la nube que nos desplegaría una página con su interfaz para el uso de las distintas herramientas.

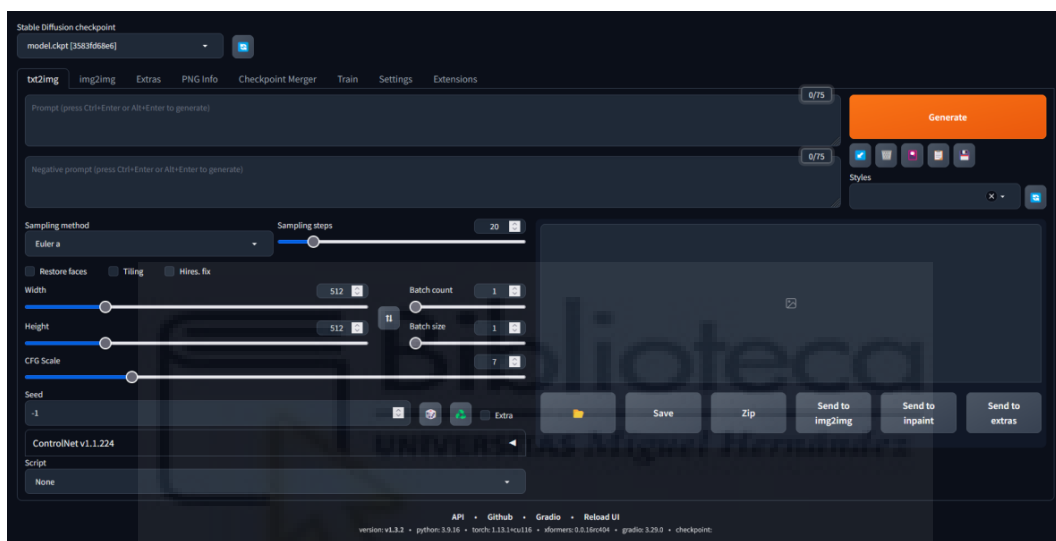


Ilustración 4: WebUI AUTOMATIC1111 para Stable Diffusion

## 4.3 ANÁLISIS DEL PROMPT

Haciendo uso de las plataformas mencionadas en el punto anterior se compararán los diferentes modelos utilizando el mismo prompt para estos tres modelos, se probarán de tres formas para ver los distintos comportamientos que tienen según la complejidad del prompt y el nivel de detalle que le pidamos, con esto será posible hacernos una idea de los puntos fuertes y débiles de cada uno de estos modelos y saber cual se puede adaptar más a un caso concreto en función de las necesidades.



## 4.4 HERRAMIENTAS COMPLEMENTARIAS

### STABLE DIFFUSION

Stable Diffusion es un modelo de código abierto y por su naturaleza existen muchas herramientas que ha creado la comunidad, en este apartado se explorarán varias de estas aunque existen muchas más, también se hablará de las plataformas de computación en la nube que van muy de la mano de este modelo, gracias a estas será posible usar las herramientas aun sin disponer de un hardware suficientemente potente.

#### 4.4.1 GOOGLE COLAB / PAPERSPACE

Para la utilización de las diversas herramientas asociadas a Stable Diffusion, se emplearán las plataformas de computación conocidas como Google Colab [31] y Gradient PaperSpace [32], ambas en su versión gratuita. Inicialmente, se había previsto el uso exclusivo de Google Colab. No obstante, el 10 de abril de 2023, Google Colab modificó sus Términos de Servicio de manera tal que restringía la posibilidad de utilizar interfaces externas con URL independientes, lo cual resulta necesario para el correcto funcionamiento de AUTOMATIC1111. Por consiguiente, en tales casos se recurrirá a Gradient PaperSpace.

Dado que es sumamente común emplear Stable Diffusion en conjunto con estas plataformas, existen numerosos scripts ya desarrollados que permiten un uso aún más conveniente. En el presente trabajo, se hará uso de aquellos ofrecidos por TheLastBen [33] para AUTOMATIC1111 y para DreamBooth [34].

# fast-stable-diffusion Notebooks, AUTOMATIC1111 + DreamBooth

Runpod & Paperspace & Colab pro adaptations AUTOMATIC1111 Webui and Dreambooth.

RunPod



Paperspace



Colab(pro)-AUTOMATIC1111



Colab(pro)-Dreambooth

```
• Minimum 200 steps between each save.  
Start_saving_from_the_step: 200  
• Start saving intermediary checkpoints from this step  
Show code  
#####  
Step: 100 / 1000 [x1.000000, x.000000, 1000x1000, 3x3, Size 4]
```

Step by Step guide by Excalibro1

Dreambooth paper : <https://dreambooth.github.io/>

SD implementation by @XavierXiao : <https://github.com/XavierXiao/Dreambooth-Stable-Diffusion>

Ilustración 5: GitHub Fast Stable Diffusion de TheLastBen

## 4.4.2 AUTOMATIC1111

La interfaz de usuario por excelencia de Stable Diffusion es AUTOMATIC1111, en esta se reúnen las mayorías de herramientas de Stable Diffusion además de por supuesto ofrecer una interfaz amigable en la que poder añadir simplemente un prompt y generar imágenes como en el resto de modelos que sí tienen una plataforma centralizada.

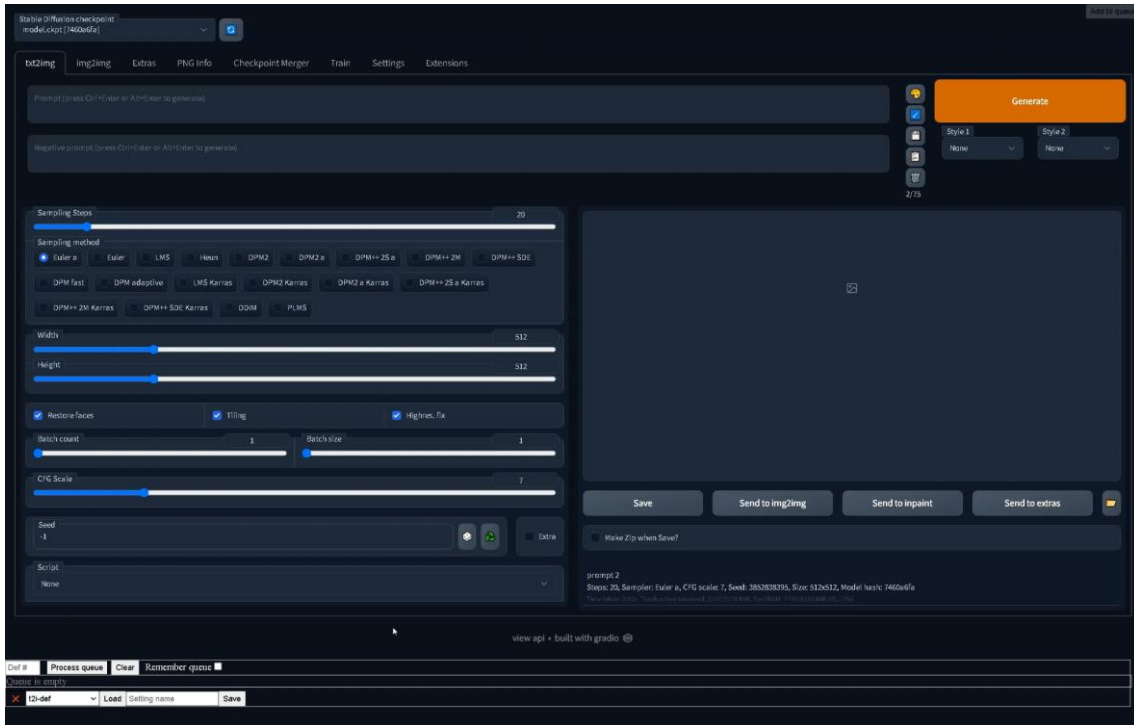


Ilustración 6: Intefaz de la WebUI AUTOMATIC111

Además de las herramientas que se activan por defecto, como img2img, desde el apartado de Extensiones se podrán instalar nuevas extensiones de la comunidad desde el repositorio que le indiquemos.

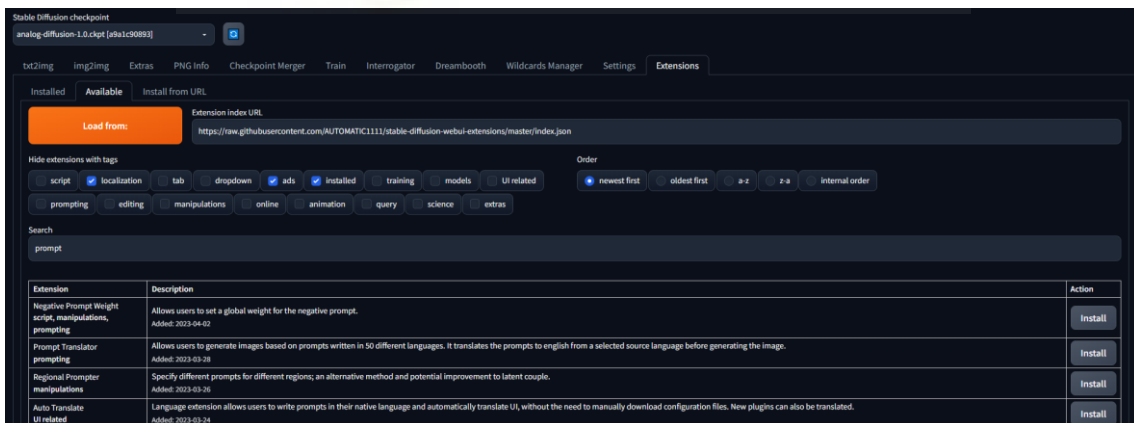


Ilustración 7: Apartado Extensiones de AUTOMATIC111

### 4.4.3 DREAMBOOTH

Para llevar a cabo el entrenamiento del modelo con las imágenes seleccionadas, se empleará la herramienta conocida como DreamBooth. Mediante esta herramienta, se logrará obtener un modelo completo que posteriormente podrá ser utilizado en cualquier instancia de Stable Diffusion, siguiendo el mismo procedimiento que se seguiría con el modelo original. El script desarrollado por TheLastBen se destaca por su alta intuitividad y su exhaustividad, brindando la posibilidad de obtener un modelo entrenado de manera efectiva y satisfactoria.

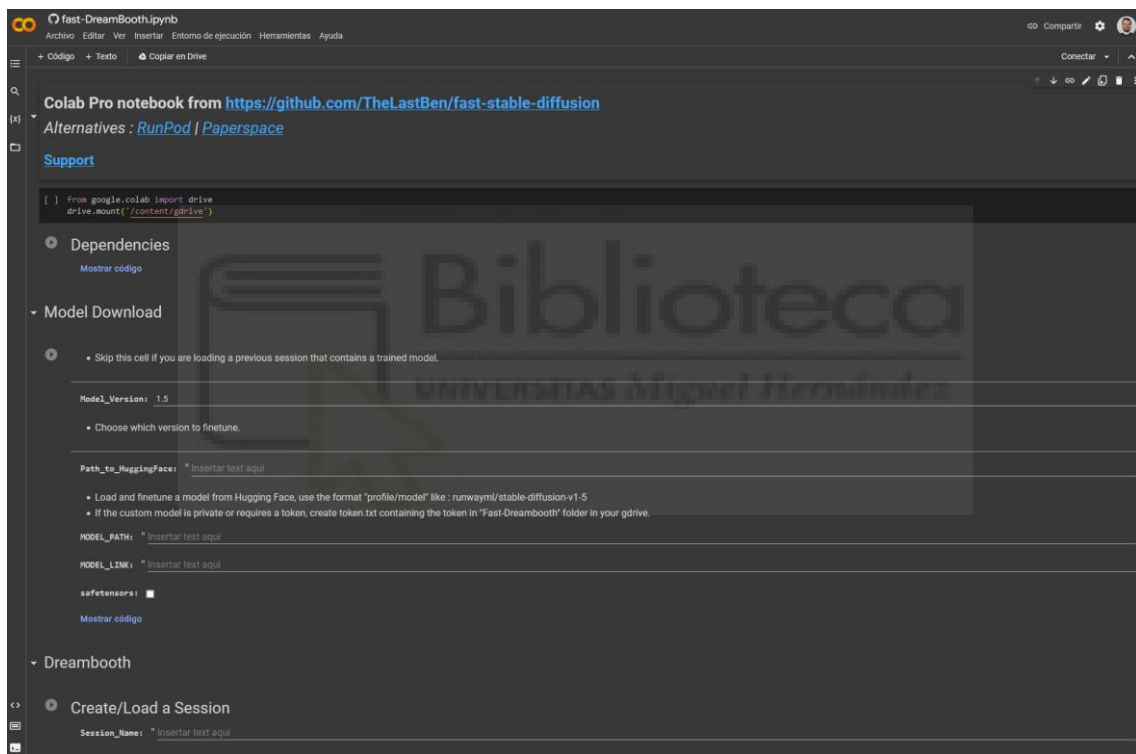


Ilustración 8: Notebook de google colab para Dreambooth

Además del enfoque de entrenamiento mencionado anteriormente, existen otras alternativas disponibles para el proceso de entrenamiento del modelo, como por ejemplo, Lora. Esta alternativa se caracteriza por ofrecer entrenamientos en un tiempo considerablemente reducido, al introducir pequeños cambios en un modelo previamente generado. Si se estuviera trabajando en un entorno local, es posible que Lora resultara la opción más conveniente, especialmente debido

al menor tiempo requerido para la generación de estos modelos auxiliares. Sin embargo, considerando que se utilizarán entornos en la nube, se ha optado por la comodidad de utilizar DreamBooth como herramienta principal. Esta elección se basa en la ventaja de contar con un modelo completo ya generado y la facilidad de incorporarlo directamente en herramientas como AUTOMATIC1111, en lugar de tener que generar un modelo adicional.

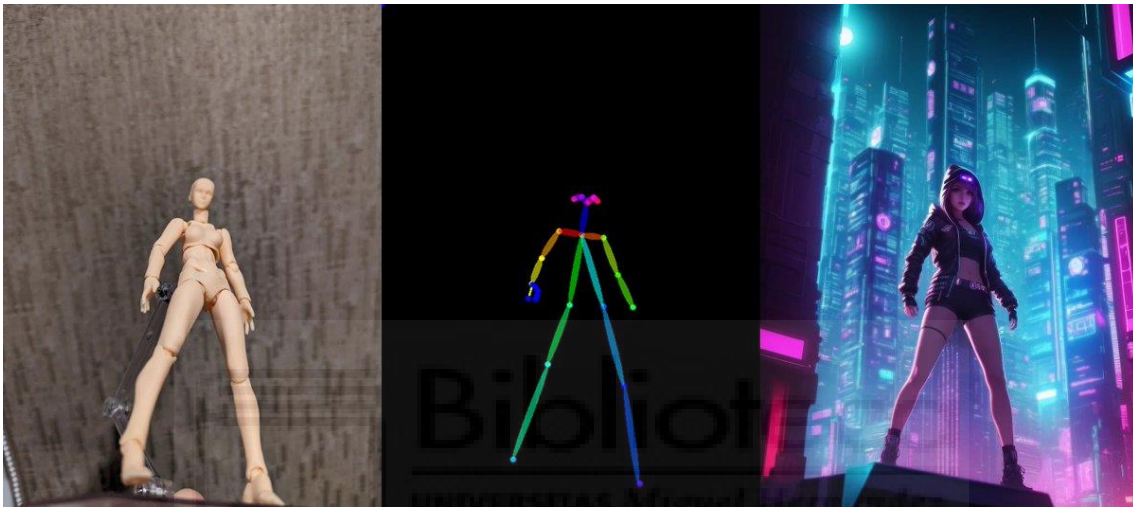
#### 4.4.4 CONTROLNET

ControlNet es una herramienta diseñada para proporcionar un control más preciso sobre los resultados generados por las imágenes, tal como sugiere su nombre. A través de esta herramienta, es posible ajustar parámetros y utilizar mapas de profundidad o esquemas para guiar el proceso de generación de imágenes mediante Stable Diffusion, de modo que las imágenes resultantes cumplan con las directrices establecidas. Además, se brinda la capacidad de especificar el nivel de rigurosidad deseado en la referencia utilizada para la generación de las imágenes. En resumen, ControlNet ofrece la posibilidad de obtener resultados más específicos y personalizados en la generación de imágenes al permitir un mayor control sobre el proceso y la referencia utilizada.



*Ilustración 9: Ejemplo de filtros en ControlNet*

Adicionalmente, es posible utilizar una imagen de referencia que represente la postura deseada de una persona. A partir de esta imagen, el modelo de generación de imágenes es capaz de comprender la postura específica y, en consecuencia, generar imágenes en las que las personas presenten la misma postura. Esta funcionalidad permite obtener imágenes con una mayor coherencia y fidelidad en términos de la pose corporal, brindando la capacidad de generar resultados más realistas y acordes a las directrices establecidas.



*Ilustración 10: Ejemplo del filtro OpenPOSE con ControlNet*

## 4.5 EQUIPOS FÍSICOS

Para el desarrollo de este proyecto se ha hecho uso de un ordenador de sobremesa y de un portátil, aunque la mayor parte de la carga de procesamiento al final recae sobre las plataformas de computación en la nube, en este caso Google Colab y Gradient Paperspace en su versión gratuita, como se explica en el primer apartado de este capítulo.

Las características de los dispositivos usados son:

### EQUIPO DE SOBREMESA

- Sistema Operativo Microsoft Windows 11 (64 bits)
- Procesador AMD RYZEN 5 5700X 3,40GHz
- 32 GB de memoria RAM DDR3
- Disco duro HDD de 1 TB
- Disco duro SSD de 256 GB
- Tarjeta gráfica AMD 6800 16GB

## EQUIPO PORTÁTIL

- Portátil Dell XPS 15
- Sistema Operativo Microsoft Windows 10 (64 bits)
- Intel Core i5-6300H QuadCore Processor 2,30 GHz
- 16 GB de memoria RAM DDR4
- Disco duro SSD de 1 TB
- Tarjeta gráfica NVIDIA® GeForce® GTX 960

El hardware que proporciona Gradient PaperSpace para máquinas con GPU en el plan gratuito es uno de estos dos según disponibilidad:

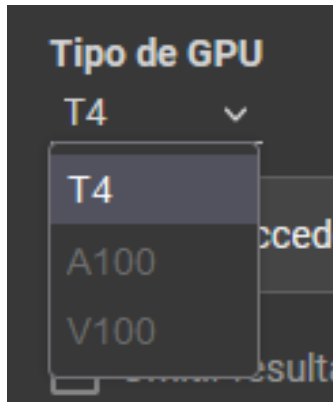
Name	GPU	Memory	vCPUs	TFLOPS (SP)	Price	Multi-GPU	No Plan Required	Pro Plan	Growth Plan
P4000	8 GB	30 GB	8 vCPU	5.3	\$0.51/hr		✓	✓	✓
RTX4000	8 GB	30 GB	8 vCPUs	7.1	\$0.56/hr		✓	✓	✓

*Ilustración 11: Listado de especificaciones del Hardware usado para el plan gratuito de Gradient Paperspace*

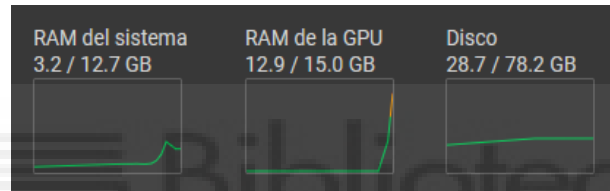
También se ofrecen para el plan de pago, en caso de pagarlo la diferencia sería no tener que entrar en la lista de espera para iniciar tu máquina.

Y en caso del plan gratuito google colab las máquinas con GPU tendrían siempre una gráfica T4 con 15gb de VRAM, una cantidad de RAM de unos 12gb, un disco

de unos 80gb pero estas cantidades podrían variar dependiendo de la disponibilidad.




*Ilustración 12: GPU para el plan gratuito de Google Colab*



*Ilustración 13: RAM, RAM de la GPU y espacio de almacenamiento para el plan gratuito de Google Colab*



# **CAPÍTULO 5: METODOLOGÍA Y RESULTADOS**



## 5.1 PLANIFICACIÓN DEL PROYECTO

La gestión de proyectos implica el uso de diversas técnicas y enfoques para organizar y supervisar las etapas de un proyecto. Un proyecto consiste en una serie de actividades específicas que se llevan a cabo para alcanzar un objetivo determinado, con límites claros en términos de recursos, plazos y alcance. La gestión de proyectos se encarga de planificar, coordinar y controlar todas estas actividades, asegurándose de que se cumplan los objetivos establecidos y se logre el éxito del proyecto.

Aunque podríamos pensar que la gestión de proyectos solo es aplicable para proyectos en los que hace falta gestionar diversas tareas realizadas entre distintas personas y así conseguir un orden, eso no es así, puede ser igualmente útil para organizarse una sola persona y así intentar seguir un orden cronológico coherente que permita la correcta realización de las tareas de una forma eficiente y controlada.

### 5.1.1 DESGLOSE DE TAREAS

En el contexto de un proyecto, las tareas desempeñan un papel fundamental al representar los pasos necesarios para lograr el objetivo establecido. Estas tareas pueden variar en su grado de dependencia entre sí, algunas requiriendo la finalización de otras previas antes de su inicio, mientras que otras pueden llevarse a cabo de forma simultánea o paralela.

En el caso específico de este proyecto, se ha llevado a cabo un desglose de las tareas en diferentes fases, con el propósito de organizar y categorizar adecuadamente el flujo de trabajo. Dichas fases se han identificado como: planificación, estudio, testing, análisis y documentación. Cada una de estas fases

comprende una serie de tareas específicas, las cuales están asociadas con una o varias tareas precedentes.

Esta relación de dependencia entre tareas implica que la finalización exitosa de las tareas anteriores es un requisito previo para el inicio de las tareas siguientes. Es decir, el avance en el cumplimiento de las tareas está directamente vinculado con el progreso y conclusión satisfactoria de las tareas previas. Esta estructura secuencial garantiza una adecuada secuencia de trabajo y una gestión eficiente de los recursos y tiempos disponibles.

A continuación, se muestra el desglose de tareas:

Nº	Categoría	Tarea	Duración (hora)	Requisitos
1	Planificación	Observar el contexto actual del tema y estructurar el trabajo.	12	-
2	Estudio	Inteligencia Artificial en la actualidad	2	1
3	Estudio	Inteligencia Artificial para generación de imágenes	2	1
4	Estudio	Tecnologías de Inteligencia Artificial para generación de imágenes	2	1
5	Estudio	Modelos más relevantes de generación de imágenes	4	2,3,4
6	Estudio	Estudio en profundidad de DALL-E, Midjourney y Stable Diffusion	10	2,3,4
7	Aprendizaje	Usar el modelo de DALL-E	20	5,6
8	Aprendizaje	Usar el modelo de Midjourney	20	5,6
9	Aprendizaje	Usar el modelo de Stable Diffusion	30	5,6
10	Aprendizaje	Usar herramientas complementarias de Stable Diffusion	40	5,6,9
11	Análisis	Evaluar diferencias entre los modelos en cuanto a resultados	20	7,8,9,10
12	Análisis	Evaluar diferencias entre los modelos en cuanto a herramientas que	20	7,8,9,10
13	Documentación	Generar la memoria final	100	1-12

*Ilustración 14: Desglose de tareas para la planificación del proyecto*

### 5.1.2 DIAGRAMA DE GANTT

El diagrama de Gantt es una herramienta ampliamente utilizada en la gestión de proyectos. Proporciona una representación visual de las tareas planificadas, su secuencia temporal, duración y dependencias. Permite programar y dar seguimiento al progreso del proyecto de manera efectiva.

A continuación, se muestra el diagrama de Gantt del proyecto, basado en el desglose de tareas presentado en el apartado 5.1.1. Desglose de tareas y la asignación de horas dedicadas a cada una de ellas.

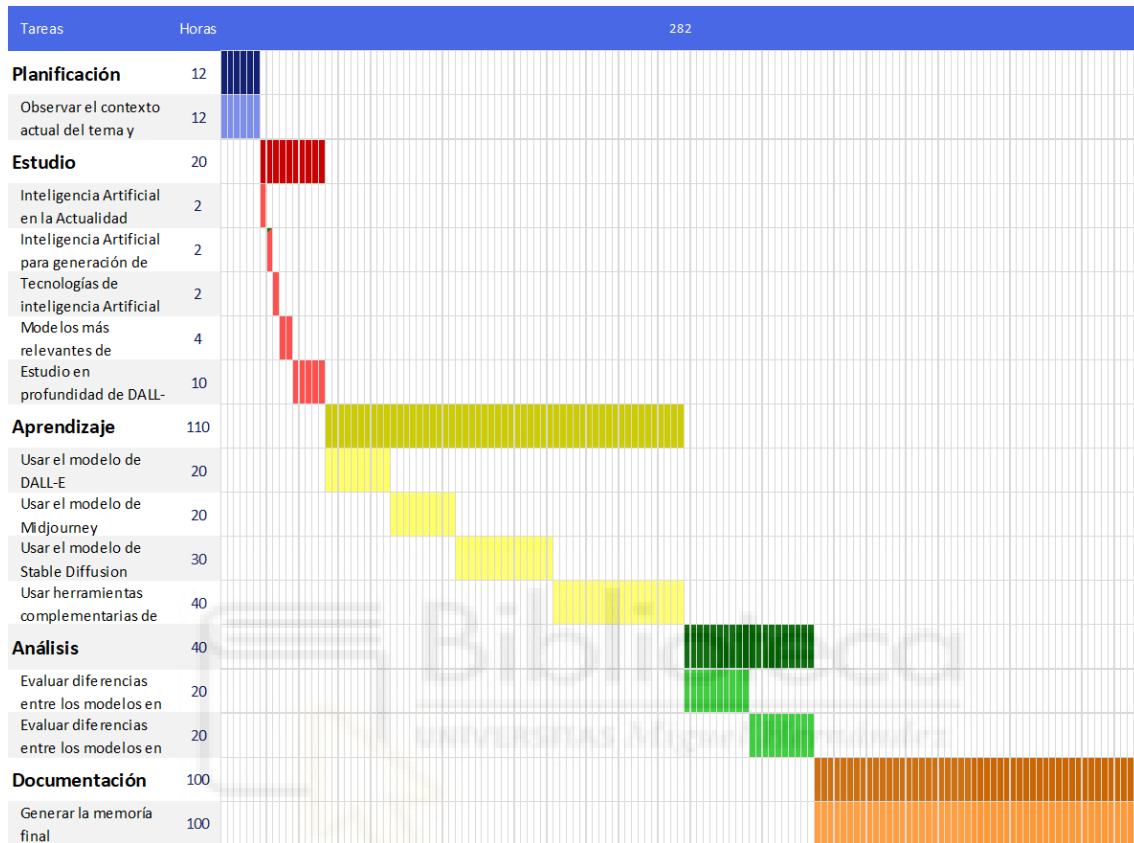


Ilustración 15: Diagrama de Gantt para la planificación del proyecto

### 5.1.3 MÉTRICAS

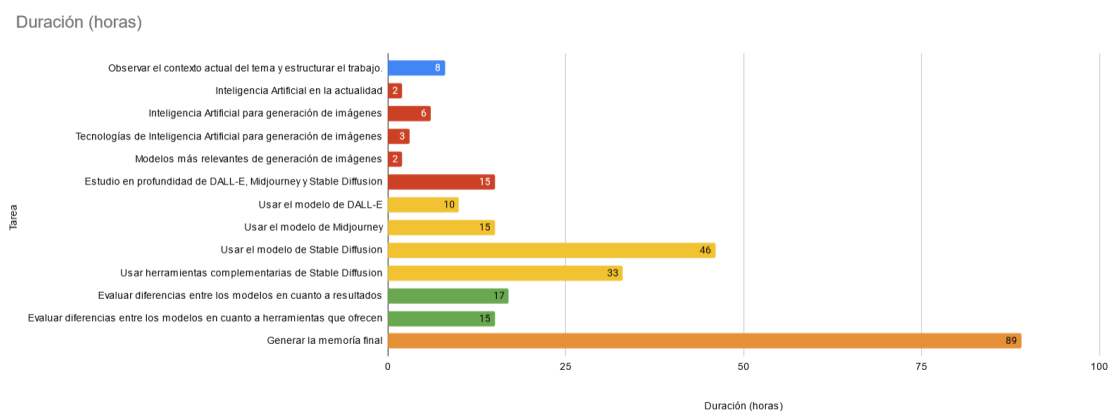
En la gestión de proyectos, las métricas desempeñan un papel crucial al proporcionar variables que permiten medir y controlar el rendimiento de aspectos clave del proyecto. Estas métricas, que pueden tomar diversas formas, son de gran importancia para evaluar y gestionar de manera efectiva el progreso y los resultados del proyecto.

En el caso específico de este proyecto, es necesario establecer algún tipo de control dadas las circunstancias que lo rodean. Estas circunstancias incluyen la

disponibilidad horaria limitada debido a la necesidad de compaginar la realización de la carrera universitaria con un trabajo a jornada completa. Además, existe un tiempo límite para la finalización del proyecto, tanto por la necesidad de completarlo dentro del plazo estimado como por evitar la posible obsolescencia de este debido a los constantes cambios en el ámbito de la Inteligencia Artificial.

En este contexto, las métricas que se tendrán en cuenta se centran principalmente en las horas reales dedicadas a cada una de las tareas. Estas métricas se visualizarán tanto a nivel de tareas individuales como a nivel de categorías, con el objetivo de obtener una visión más amplia del progreso del proyecto. Para llevar a cabo el registro de estas métricas, se utilizará un archivo de Excel en el cual se anotarán las horas empleadas en cada tarea a lo largo del proyecto.

A continuación, se presentan los gráficos correspondientes a estas métricas, los cuales proporcionarán una representación visual de los datos recopilados y permitirán una fácil interpretación de la evolución del tiempo dedicado a las tareas en el proyecto.



*Ilustración 16: Gráfico de barras general de las métricas recogidas a lo largo del proyecto*

### Por Tareas

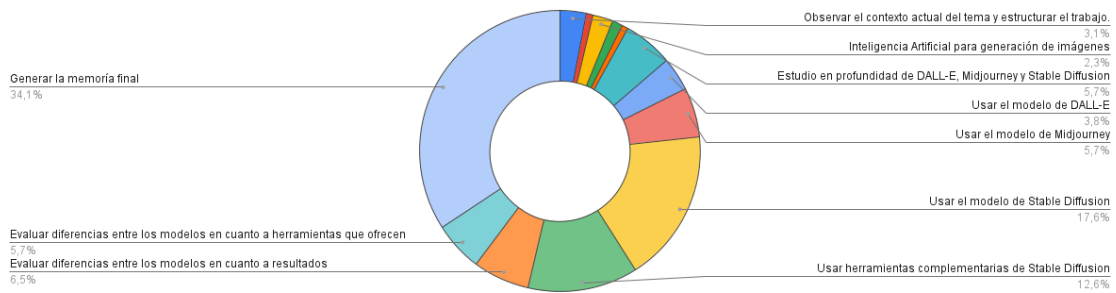


Ilustración 17: Gráfico circular por Tareas de las métricas recogidas a lo largo del proyecto

### Por categorías

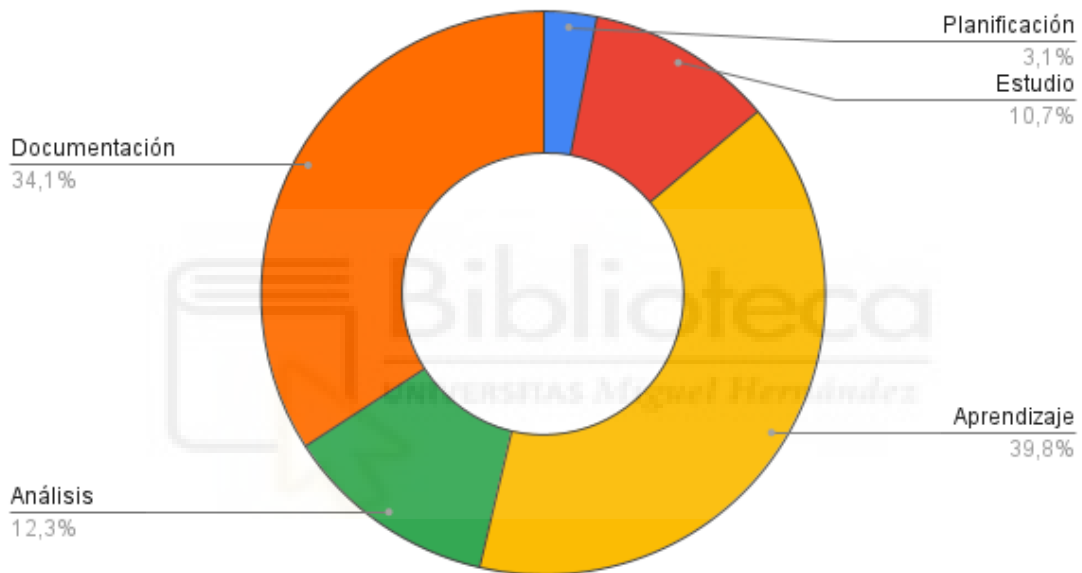


Ilustración 18: Gráfico circular por Categorías de las métricas recogidas a lo largo del proyecto

## 5.2 HERRAMIENTAS DE LOS MODELOS

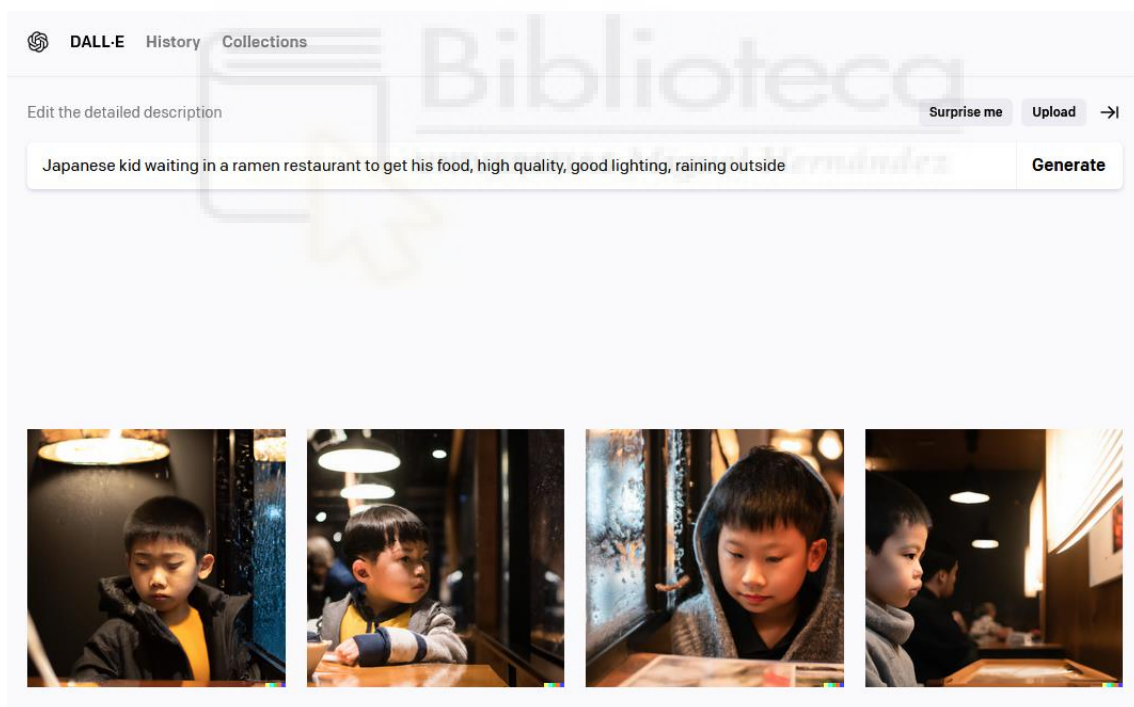
### 5.2.1 DALL-E

Entre los tres modelos analizados, DALL-E destaca por tener una plataforma muy bien desarrollada. En su página web, se encuentra disponible una versión gratuita limitada que permite a los usuarios realizar pruebas y experimentar con el

modelo. La plataforma se caracteriza por contar con una interfaz amigable y proporcionar una serie de herramientas intuitivas que facilitan su uso.

Una vez accedemos a la página de DALL-E, el proceso resulta sumamente sencillo. Basta con ingresar nuestro prompt o consigna en la barra ubicada en la parte superior de la pantalla. Con cada prompt ingresado, se generan cuatro imágenes como resultado, brindando al usuario la posibilidad de elegir aquella que mejor se adapte a sus necesidades o preferencias.

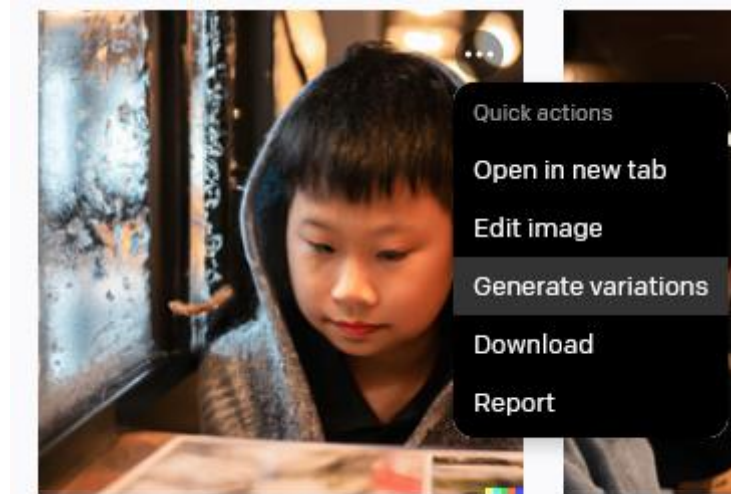
En resumen, DALL-E destaca por su plataforma bien lograda, que incluye una versión gratuita limitada, una interfaz amigable y herramientas intuitivas. La simplicidad del proceso se refleja en la facilidad de ingreso de prompts y en la presentación de múltiples opciones de imágenes generadas para su elección.



*Ilustración 19: Ejemplo de uso de la plataforma de DALL-E en el que se escribe un prompt y se muestran los resultados*

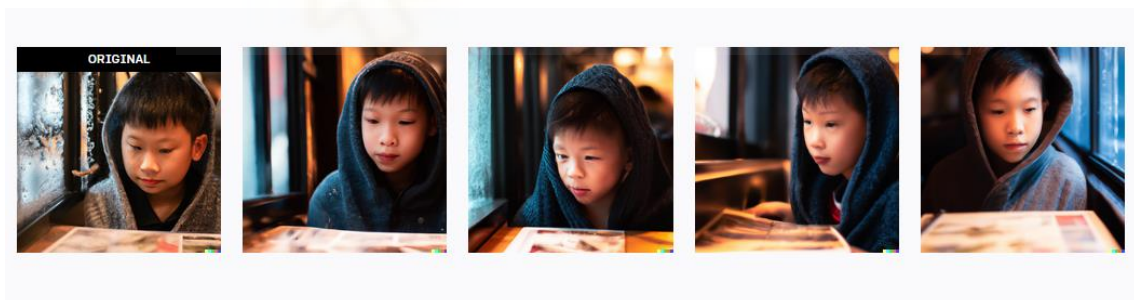
En caso de que una de las imágenes generadas resulte atractiva pero no se ajuste exactamente a lo que se busca, se ofrece la opción de generar variaciones de la

misma. Esto permite explorar diferentes versiones o modificaciones de la imagen inicial, brindando flexibilidad y opciones adicionales al usuario. De este modo, se fomenta la experimentación y la obtención de resultados más acordes a las preferencias y necesidades individuales.



*Ilustración 20: Ejemplo de uso de DALL-E para generar variaciones de una imagen*

De nuevo se muestran 4 imágenes para poder seleccionar la más adecuada.

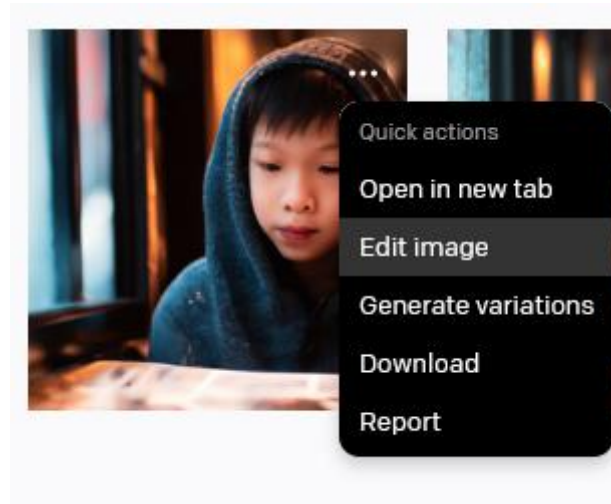


*Ilustración 21: Resultados de las variaciones generadas por DALL-E*

Cabe destacar que, entre los tres modelos considerados, DALL-E es el único que ofrece una funcionalidad de edición de imágenes de una manera tan intuitiva y sencilla. Esta característica brinda la posibilidad de modificar únicamente las partes específicas de una imagen que se deseen cambiar, sin afectar el resto de la composición visual. Esta capacidad de edición selectiva permite un mayor

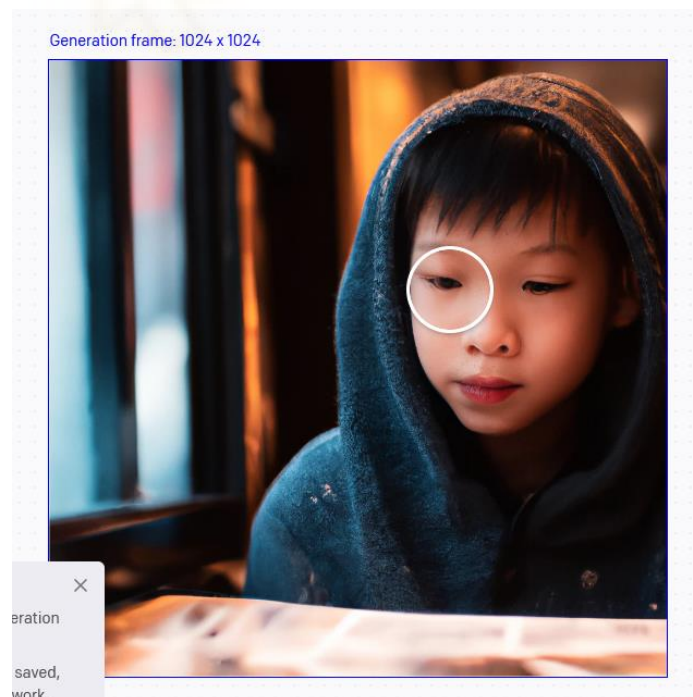


grado de personalización y control sobre los resultados generados, potenciando la adaptabilidad del modelo a las necesidades y preferencias individuales del usuario.



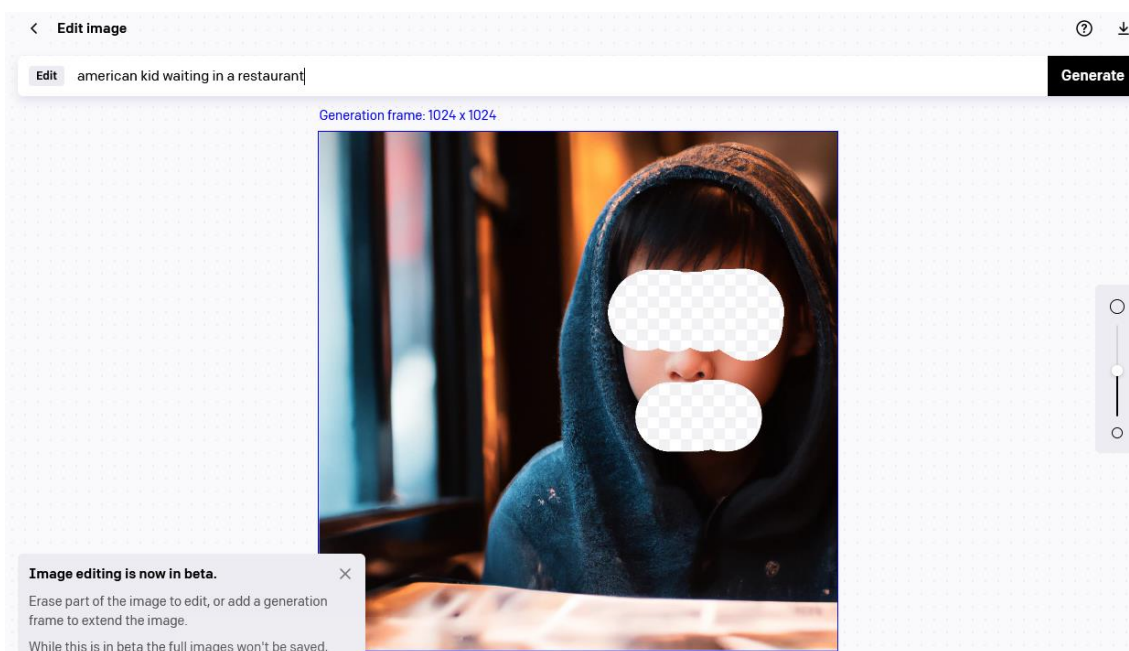
*Ilustración 22: Ejemplo de uso de DALL-E para editar una imagen*

Con su herramienta se podrán "borrar" las partes que queramos cambiar, para poder generar por encima esa parte de la imagen.

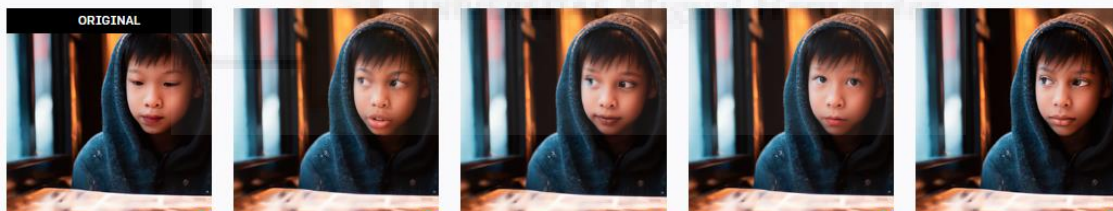


*Ilustración 23: Ejemplo de uso de la herramienta de edición de DALL-E*

Una vez realizada la edición deseada en la imagen, se procede a añadir nuevamente el prompt o consigna deseada.

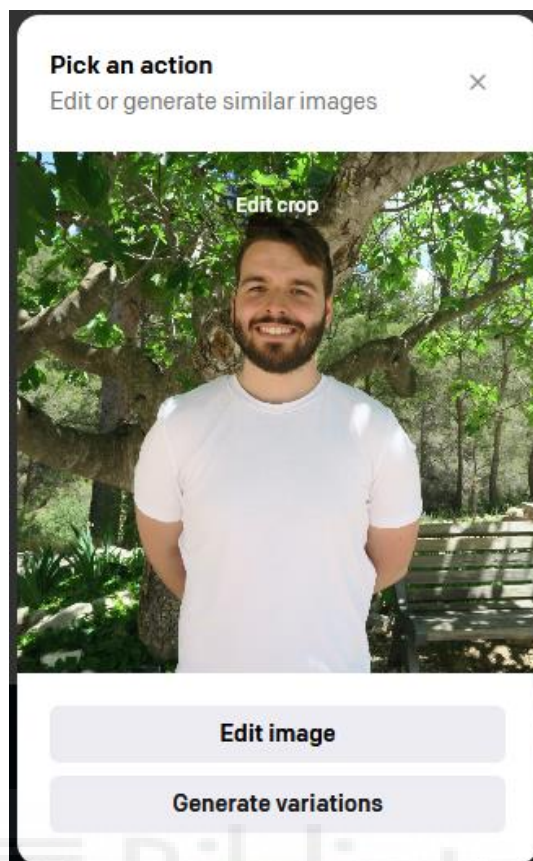


*Ilustración 24: Ejemplo de uso de la herramienta de edición donde se borra parte de la imagen para que sea editada*



*Ilustración 25: Resultado de la edición de la imagen por partes*

Asimismo, DALL-E ofrece la posibilidad de cargar imágenes personalizadas, lo que permite utilizar las mismas herramientas mencionadas anteriormente con fotografías propias. Esta funcionalidad amplía la versatilidad y flexibilidad del modelo, permitiendo a los usuarios utilizar sus propias imágenes como base para la generación y edición de imágenes. Al cargar imágenes personalizadas, se abre un abanico de posibilidades creativas y de adaptación, ya que se pueden aplicar las herramientas y técnicas de generación y edición a contenido visual propio, enriqueciendo aún más la experiencia y los resultados obtenidos.



*Ilustración 26: herramienta para subir imágenes propias de DALL-E donde permite editar y generar variaciones*

De igual manera que en el caso anterior se podrían generar variaciones.



*Ilustración 27: Resultado de las variaciones con la imagen propia*

Además de poder cambiar algún detalle, como el logo de la camiseta en este caso.

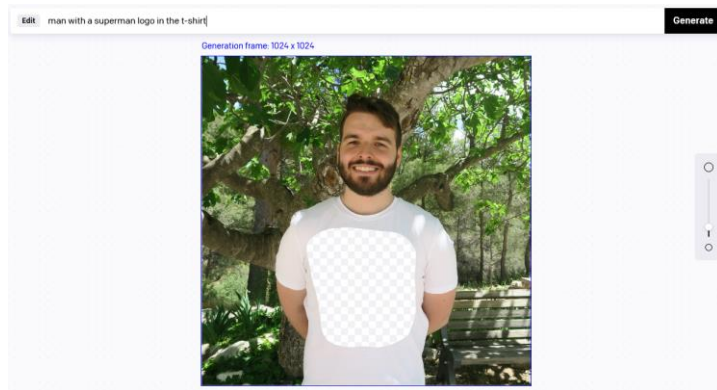


Ilustración 28: Ejemplo de edición de una parte de la imagen, la parte delantera de la camiseta



Ilustración 29: Resultados de la edición de la imagen con distintos logos en la camiseta

O directamente sería posible dejar solo la cara para poder crear el escenario deseado alrededor.

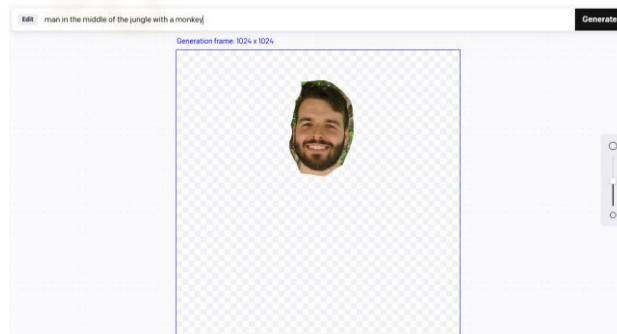
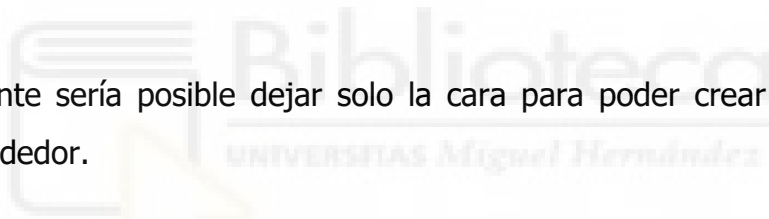


Ilustración 30: Ejemplo de uso de edición para cambiar todo el escenario menos la cara de la persona



Ilustración 31: Resultados de la edición de la imagen cambiando todo el escenario

## 5.2.2 MIDJOURNEY

Con respecto a MidJourney, se presenta una mayor complejidad en comparación a los otros modelos, al menos en lo que respecta al inicio de la generación de las primeras imágenes. Como se expuso en el capítulo de hipótesis de trabajo, se requiere vincular previamente una cuenta de Discord a través de su página web y realizar el correspondiente pago de suscripción, ya que no se ofrece ningún plan gratuito.

Una vez dentro de Discord, el procedimiento consiste en unirse a uno de los grupos públicos disponibles o iniciar una conversación directa con el bot de Discord asociado a MidJourney. Estas interacciones en la plataforma de Discord son necesarias para poder comenzar el proceso de generación de imágenes.

Este enfoque particular de utilización de Discord como plataforma de generación de imágenes añade una capa adicional de interacción y dependencia de servicios externos, lo cual puede influir en la experiencia de uso y la accesibilidad de MidJourney en comparación a otras alternativas analizadas.

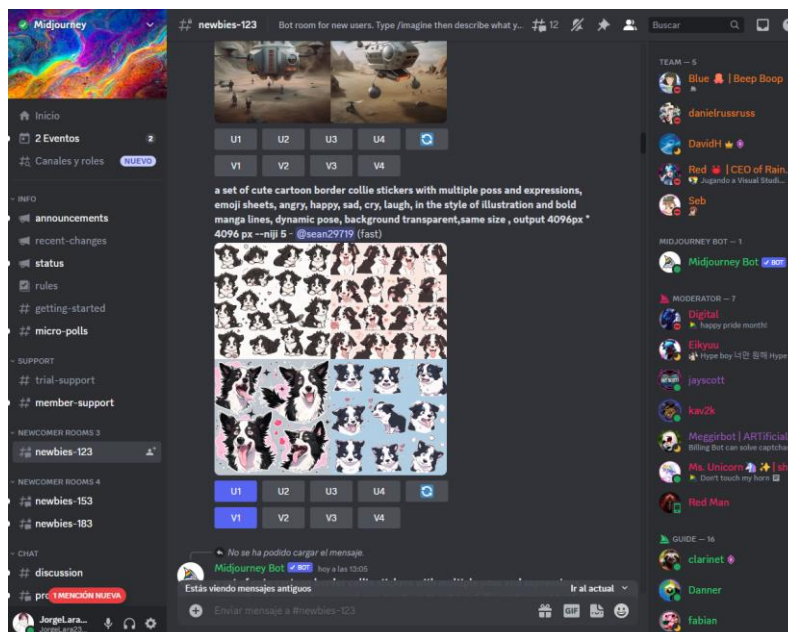
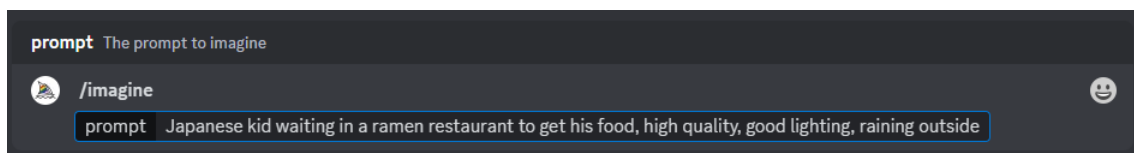


Ilustración 32: Discord de MidJourney

Para ingresar un prompt y generar imágenes utilizando MidJourney, se emplea el comando `"/imagine"`. Este modelo cuenta con una amplia gama de comandos que permiten llevar a cabo diversas tareas. Sin embargo, para los propósitos de este trabajo, se centrará principalmente en la interfaz gráfica proporcionada por MidJourney a través de botones seleccionables en las conversaciones.

A través de esta interfaz gráfica, los usuarios pueden interactuar con MidJourney mediante el uso de botones clickables, lo que facilita la introducción de prompts y la realización de otras acciones relacionadas con la generación de imágenes. Esta modalidad de interacción visual brinda una experiencia más intuitiva y accesible para los usuarios, ya que elimina la necesidad de recordar y escribir comandos específicos, permitiendo una interacción más fluida y eficiente con el modelo.

En resumen, MidJourney ofrece una interfaz gráfica basada en botones clickables en las conversaciones, lo que facilita la introducción de prompts y otras acciones relacionadas con la generación de imágenes. Esta modalidad de interacción mejora la experiencia de uso al proporcionar una opción más visual e intuitiva para los usuarios.



*Ilustración 33: Ejemplo de uso del comando /imagine para introducir el prompt en Midjourney*

Al introducir un prompt aparecerán cuatro imágenes para poder seleccionar alguna, un funcionamiento muy parecido al de DALL-E.

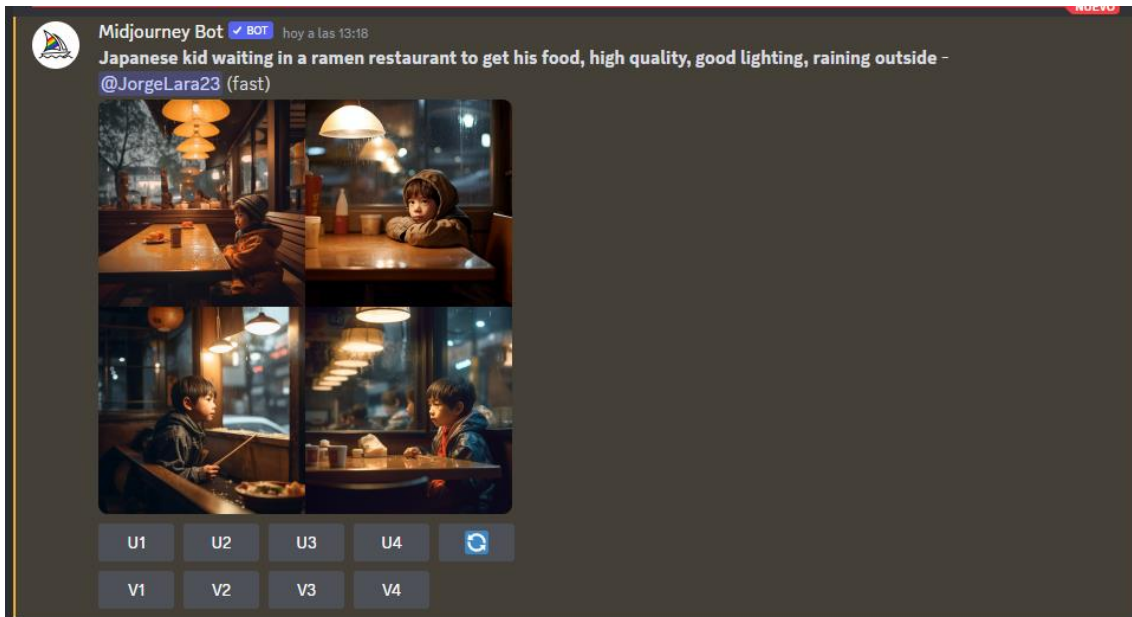


Ilustración 34: Ejemplo generación de imágenes a partir del prompt anterior

Para obtener una imagen individual de entre las cuatro generadas por MidJourney, se emplean los botones U1 - U4. Estos botones permiten seleccionar la imagen deseada de forma independiente y con toda su resolución. Por otro lado, los botones V1 - V4 posibilitan la generación de variaciones directas de una imagen específica.

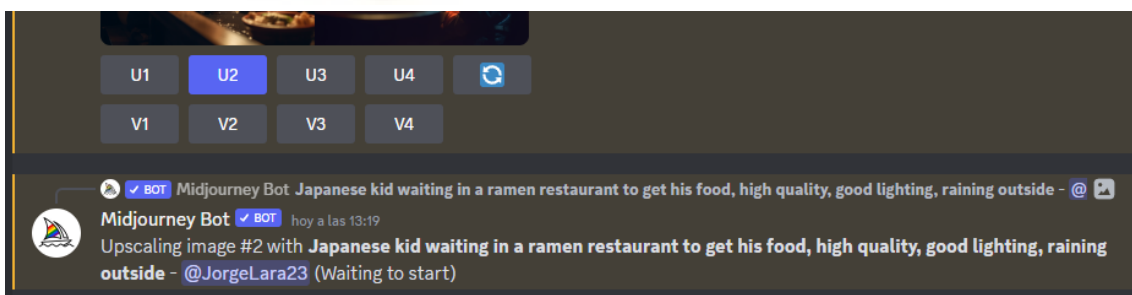
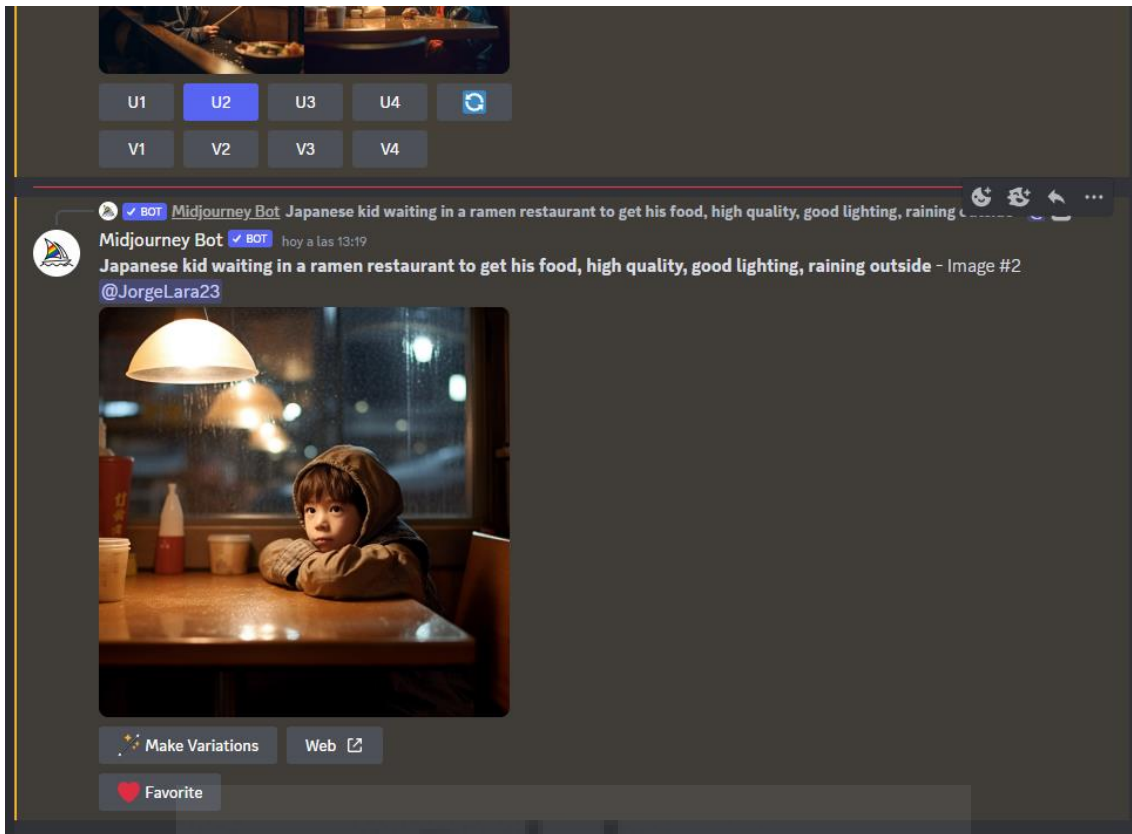


Ilustración 35: Uso del botón U2 para generar la segunda imagen de las cuatro mostradas

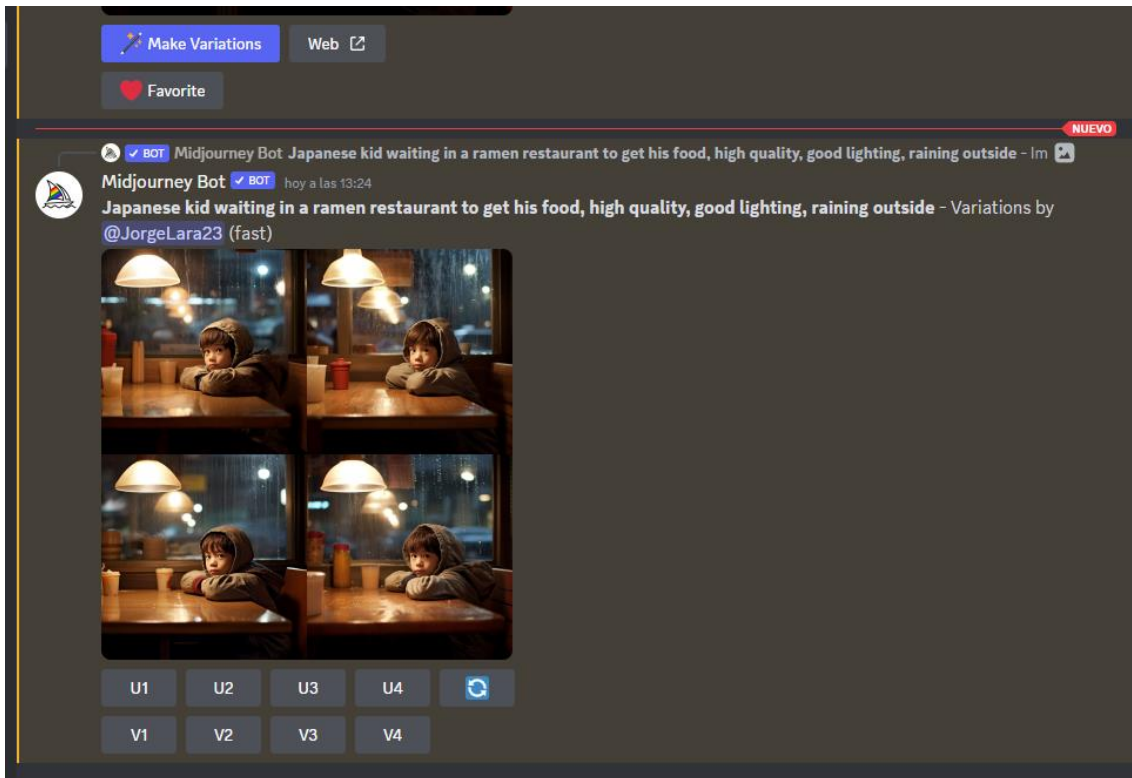


*Ilustración 36: Resultado de la generación de la imagen con toda la resolución*

Además, ofrece la opción de refrescar mediante el uso de un botón designado para esta función. Al seleccionar este botón, se generan cuatro imágenes nuevas, manteniendo el mismo prompt utilizado previamente. Esta funcionalidad brinda la posibilidad de explorar diferentes opciones de imágenes relacionadas con el prompt en cuestión. A partir de esta imagen también se podrían generar variaciones con el botón Make Variations.

Con estas opciones a disposición, el usuario puede acceder nuevamente a la etapa anterior del proceso, donde tiene la capacidad de seleccionar una imagen individual o sus variaciones correspondientes según sus necesidades y preferencias.





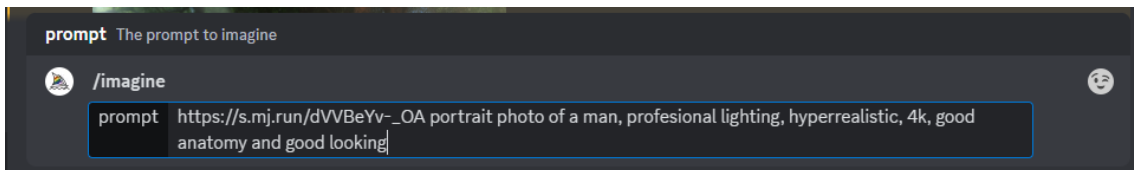
*Ilustración 37: Resultado del uso del botón "Make Variations" para generar variaciones*

MidJourney también ofrece una funcionalidad que permite utilizar imágenes propias, aunque de manera más rudimentaria en comparación con DALL-E. Para utilizar esta opción, se debe cargar la imagen deseada y obtener una URL pública para luego agregarla al prompt correspondiente. Esta característica brinda la posibilidad de utilizar imágenes personalizadas en el proceso de generación, ampliando las opciones creativas disponibles para los usuarios.



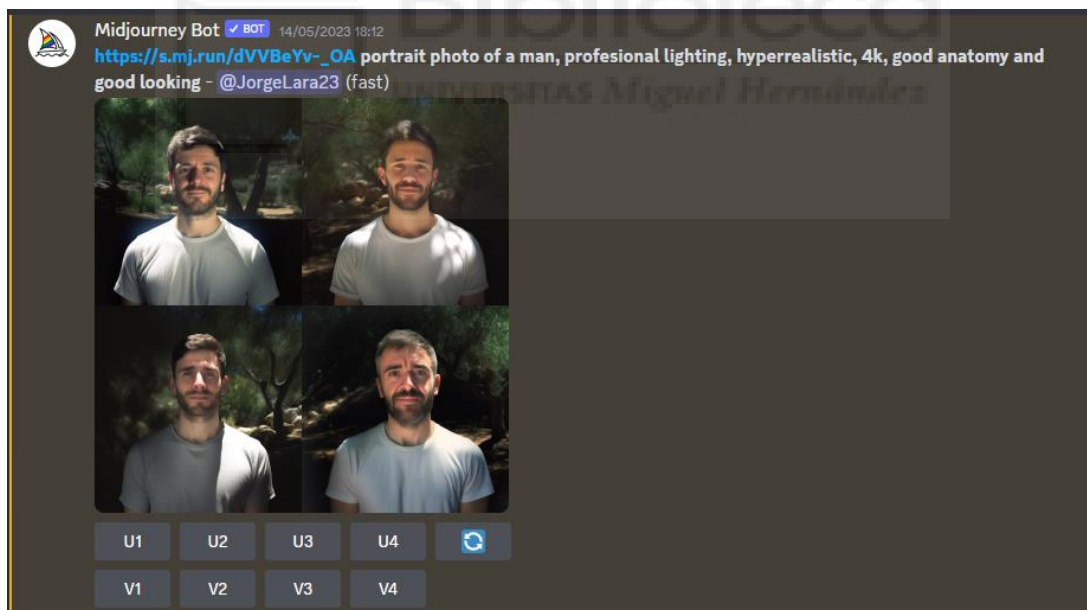
*Ilustración 38: Imagen propia usada para probar las variaciones con Midjourney*

En el caso, se ha optado por subir la imagen directamente a la conversación de Discord para obtener el enlace proporcionado. Posteriormente, dicho enlace se incorporará al prompt junto con las indicaciones necesarias para generar las variaciones deseadas. De esta manera, se aprovecha la funcionalidad de carga de imágenes en Discord como medio para obtener el enlace público requerido para la generación de imágenes personalizadas.



*Ilustración 39: Ejemplo del prompt necesario con el link de la imagen al principio y la descripción después*

Se podría, igual que en el caso anterior, obtener la imagen deseada o generar variaciones de esta.



*Ilustración 40: Resultados de las variaciones generadas con la imagen y el prompt anterior*

Y este sería el resultado:

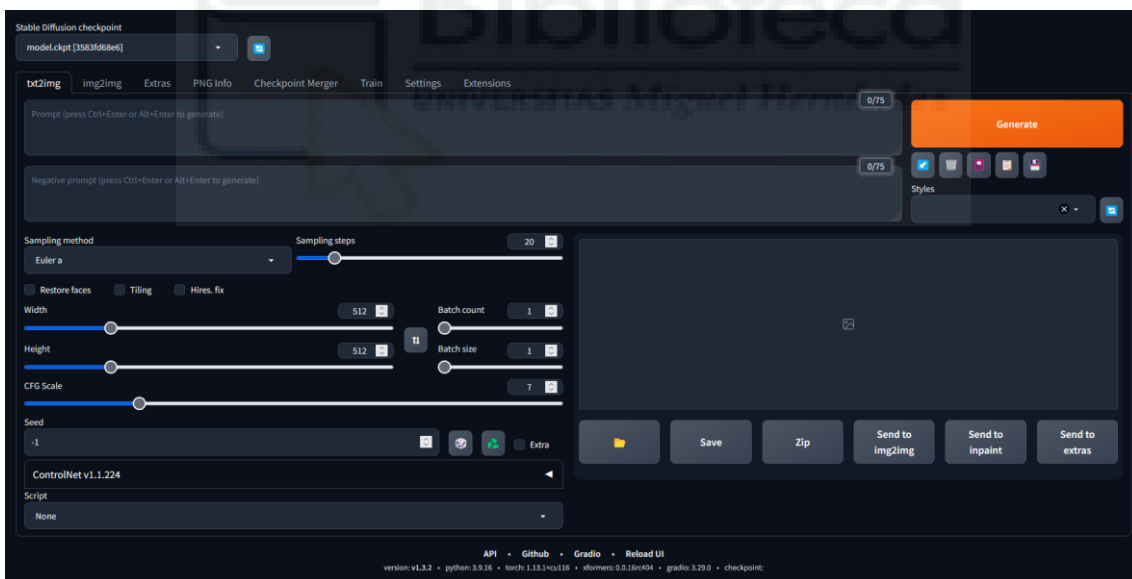


*Ilustración 41: Imagen completa de las variaciones anteriores*

El uso de esta funcionalidad en MidJourney proporciona resultados más refinados en comparación con la opción proporcionada por DALL-E para lograr el objetivo de incluir solo nuestro rostro en la imagen generada. Sin embargo, si el objetivo es obtener una representación realista y precisa de nuestro rostro, este enfoque no sería suficiente. Además, el proceso de utilización de MidJourney resulta más sencillo en comparación con el entrenamiento de una IA personalizada utilizando Stable Diffusion, el cual será abordado posteriormente. Es importante destacar que, en términos de fidelidad con la realidad, no se puede establecer una comparación significativa entre ambos enfoques. La función de uso de imágenes personalizadas en MidJourney se presenta como una herramienta de apoyo para guiar el prompt en lugar de obtener una representación concreta que haya sido previamente entrenada.

## 5.2.3 STABLE DIFFUSION

En relación a Stable Diffusion, como se trata un modelo de generación de imágenes de código abierto, no cuenta con una plataforma centralizada para su uso, aunque existen algunas opciones ofrecidas por otras entidades. En comparación con los otros dos modelos mencionados, este resulta ser el más complejo para comenzar a utilizar, pero a su vez ofrece mayores posibilidades debido a la constante contribución de la comunidad en forma de nuevas herramientas y mejoras en las existentes. Por lo tanto, resulta pertinente realizar un análisis en profundidad de este modelo en una sección independiente, donde se abordarán algunas de las herramientas de manera detallada. Entre las interfaces gráficas más utilizadas se encuentra AUTOMATIC1111, la cual proporciona un contexto adecuado para la comparación de este modelo con los dos anteriores.



*Ilustración 42: Interfaz de la webUI AUTOMATIC1111 para Stable Diffusion*

La interfaz web de Stable Diffusion, si bien es similar a la de DALL-E, se caracteriza por tener una mayor carga visual y una amplia gama de parámetros manipulables. Sin embargo, para poder utilizarla, se requieren varias tareas previas y, una vez en uso, nos enfrentamos a una abrumadora cantidad de

opciones. A nivel básico, encontramos el prompt, presente en las herramientas anteriores, y el prompt negativo, que se utiliza para eliminar características no deseadas en el resultado final, al igual que se hace con el prompt "positivo", que se utiliza para describir la imagen deseada. Se llevará a cabo un análisis detallado de esta herramienta en la sección dedicada a las Herramientas Complementarias de Stable Diffusion.

## 5.3 ANÁLISIS DEL PROMPT

En este apartado, se procederá a probar los mismos prompts en los diferentes modelos, utilizando prompts en inglés, dado que todas las herramientas fueron inicialmente entrenadas con este idioma en mente y su desempeño suele ser mejor en dicho idioma. Dado que el enfoque habitual de estas herramientas consiste en obtener una variedad de resultados y seleccionar el más adecuado según el objetivo, se generarán en promedio 4 imágenes por prompt para cada modelo, y se seleccionará la mejor imagen según los criterios establecidos. En el caso específico de Stable Diffusion 2.1, se requerirá un prompt negativo, el cual se utilizará para indicar que no se desea una imagen borrosa, pixelada, etc. Es importante señalar que dejar este apartado vacío no sería justo, ya que el rendimiento del modelo depende de este aspecto para obtener resultados óptimos. A excepción del prompt negativo, se utilizarán los valores predeterminados de la herramienta de código abierto AUTOMATIC1111 para este modelo.

Prompt negativo genérico usado en SD2.1: ugly, boring, bad anatomy, blurry, pixelated, obscure, unnatural colors, poor lighting, dull, unclear

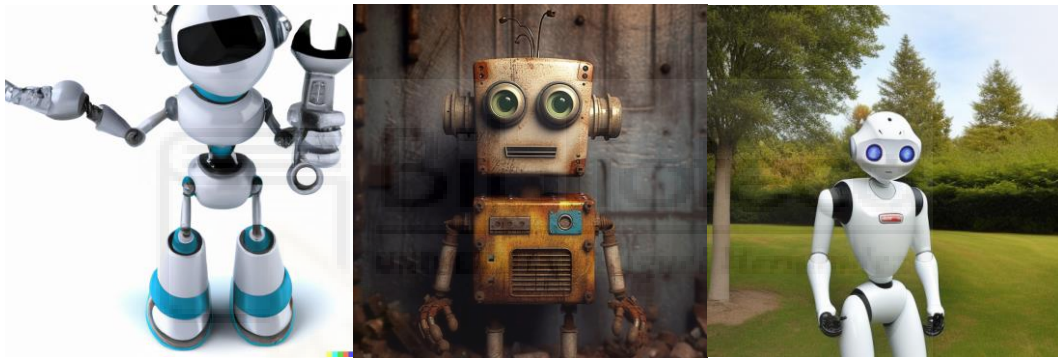
En primer lugar, se realizará una serie de pruebas utilizando palabras simples con el objetivo de observar cómo las distintas herramientas de IA generan imágenes que representen conceptos básicos.

- Strawberry



*Ilustración 43: Resultados para el prompt "Strawberry" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

- Robot



*Ilustración 44: Resultados para el prompt "Robot" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

- Beach



*Ilustración 45: Resultados para el prompt "Beach" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

- Happiness



*Ilustración 46: Resultados para el prompt "Happiness" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

- Skater



*Ilustración 47: Resultados para el prompt "Skater" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

A continuación, se procederá a explorar la generación de imágenes que representen a un sujeto realizando alguna acción específica mediante el uso de las herramientas de IA mencionadas.

- A dog eating strawberries



*Ilustración 48: Resultados para el prompt "A dog eating strawberries" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

- Robot arm building a car



*Ilustración 49: Resultados para el prompt "Robot arm building a car" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

- Multiple tourists sunbathing in the beach



*Ilustración 50: Resultados para el prompt "Multiple tourists sunbathing in the beach" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*



- Happy person giving flowers away



*Ilustración 51: Resultados para el prompt "Happy person giving flowers away" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

- Two skaters training a new trick



*Ilustración 52: Resultados para el prompt "Two Skaters training a new trick" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

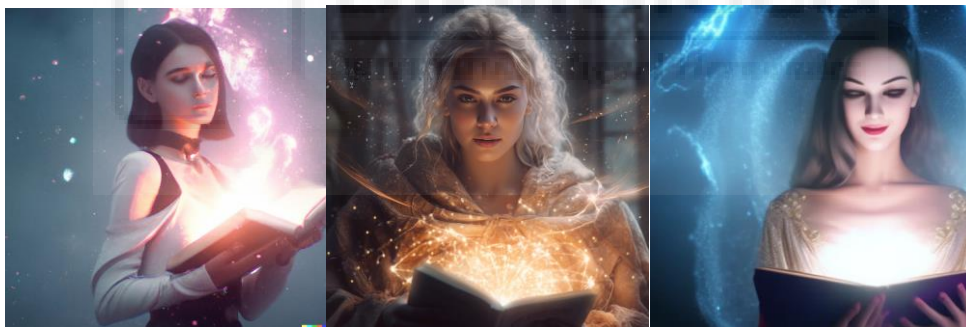
Para concluir, se abordarán los prompts más complejos y específicos con el propósito de obtener imágenes impactantes y con objetivos particulares. Estos prompts han sido recopilados de la página Prompt Hero [35], donde se comparten ejemplos de prompts diseñados para generar resultados llamativos y sorprendentes.

- Cartoon picture of a funny morel mushroom waving to his friends, . Background Setting is a town with half-timbered houses. 8k UHD. Vibrant colors. 3D Hyper realistic.



*Ilustración 53: Resultados para el prompt "Cartoon picture of a funny morel mushroom waving to his friends, . Background Setting is a town with half-timbered houses. 8k UHD. Vibrant colors. 3D Hyper realistic." para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

- Photogenic woman holding a magical book, magical glowing trails, light dust, fantasy robes, highly detailed fantasy art, cinematic lighting, sharp focus, unreal, render, 4k



*Ilustración 54: Resultados para el prompt "Photogenic woman holding a magical book, magical glowing trails, light dust, fantasy robes, highly detailed fantasy art, cinematic lighting, sharp focus, unreal, render, 4k" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

- Full body photo of a scary alien creature that looks like a humanoid spider with venom coming out from its mouth emerging from a dark cave, green deadly eyes, horror, dark, intricate design and details, dramatic lighting, photorealistic, cinematic, 8k, cinematic noise



*Ilustración 55: Resultados para el prompt "Full body photo of a scary alien creature that looks like a humanoid spider with venom coming out from its mouth emerging from a dark cave, green deadly eyes, horror, dark, intricate design and details, dramatic lighting, photorealistic, cinematic, 8k, cinematic noise" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

- Human colony on unknown planet with retro futuristic buildings, bright, vibrant, colored, vegetation, hyperrealism, highly detailed, insanely detailed, lush detail



*Ilustración 56: Resultados para el prompt "Human colony on unknown planet with retro futuristic buildings, bright, vibrant, colored, vegetation, hyperrealism, highly detailed, insanely detailed, lush detail" para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

- Man wearing a fullbody suit in a disco pub party, his face in the foreground, happy, hyper realistic, detailed touches, neon lighting, sharp focus, 4k



*Ilustración 57: Resultados para el prompt "Man wearing a fullbody suit in a disco pub party, his face in the foreground, happy, hyper realistic, detailed touches, neon lighting, sharp focus, 4k " para los modelos DALL-E, Midjourney y Stable Diffusion respectivamente*

Al observar las imágenes anteriores, se puede apreciar claramente que Midjourney supera con creces a los otros dos modelos al generar imágenes espectaculares con una iluminación destacada y abundantes detalles. Dall-e, por otro lado, aunque en ocasiones logra imágenes realistas, tiende a producir ruido e imperfecciones exageradas al ampliar las imágenes, especialmente en aquellas que incluyen rostros humanos. Tanto Stable Diffusion como Midjourney muestran un nivel similar de definición, lo cual tiene sentido considerando que Midjourney utiliza parte del modelo de Stable Diffusion.

De los tres modelos, Stable Diffusion parece ser el menos preciso en base a lo que le pedimos, especialmente cuando se le presenta un prompt complejo. Ejemplos claros incluyen el prompt "Skater" y "Two skaters training a new trick", donde no logra dibujar correctamente una tabla de skate y en su lugar dibuja esquís, así como el prompt más complejo "Cartoon picture of a funny morel mushroom waving to his friends...", donde no representa el elemento principal, que sería el hongo "funny morel".

En cuanto al texto en las imágenes, se puede observar que tanto Dall-e como Stable Diffusion colocan letras de manera incoherente en posiciones y estilos que tendrían sentido en una imagen. Sin embargo, Midjourney no muestra ningún texto en ninguna de las imágenes, incluso al utilizar el prompt "Happiness (using text)". Esto podría deberse a que Midjourney, al ser una tecnología más refinada, parece que ha tratado de abordar directamente este problema conocido en los modelos de generación de imágenes, desactivando esta opción durante el procesamiento de tokens o en algún punto donde se haya eliminado por completo esta limitación.

## 5.4 HERRAMIENTAS COMPLEMENTARIAS

### STABLE DIFFUSION

#### 5.4.1 GOOGLE COLAB / GRADIENT PAPERSPACE

Las plataformas de computación en la nube funcionan de manera similar, principalmente a través de scripts escritos en Python utilizando archivos de cuadernos Jupyter (.ipynb). Estos cuadernos proporcionan un entorno computacional interactivo que se adapta mejor al uso de las plataformas en la nube y al procesamiento de datos. Esto permite a los usuarios programar y ejecutar sus scripts en el lenguaje de programación de su elección, aprovechando las capacidades de la nube para el procesamiento y análisis de datos.

#### Google Colab

Sería necesario estar logueados en Google, y una vez dentro de Google Colab se vería una ventana como esta:

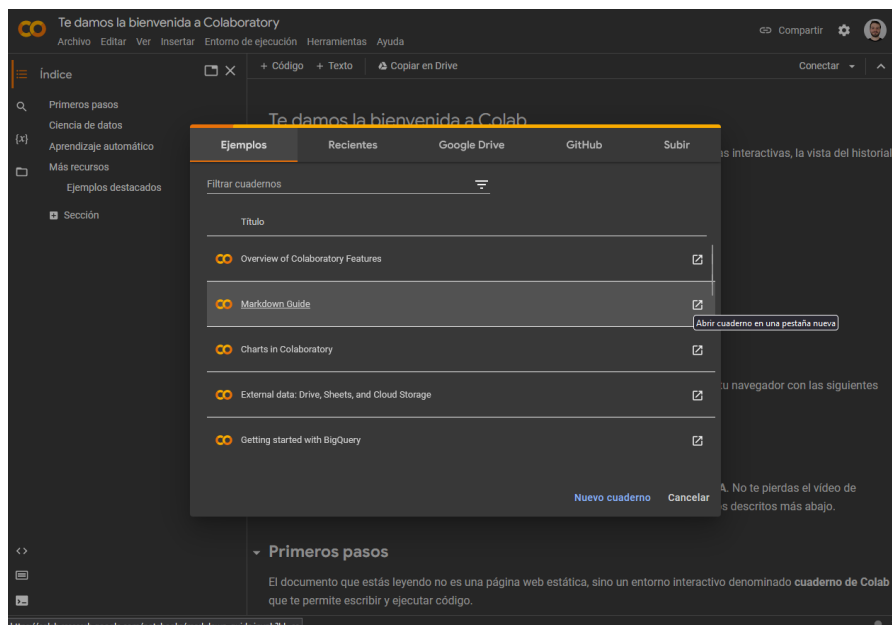
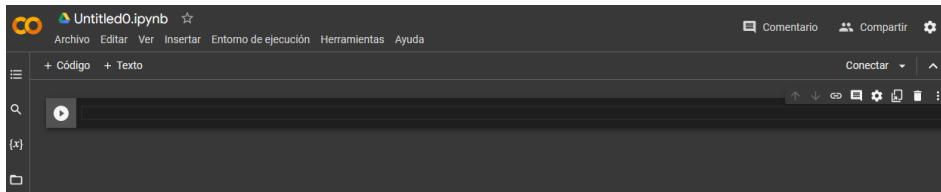


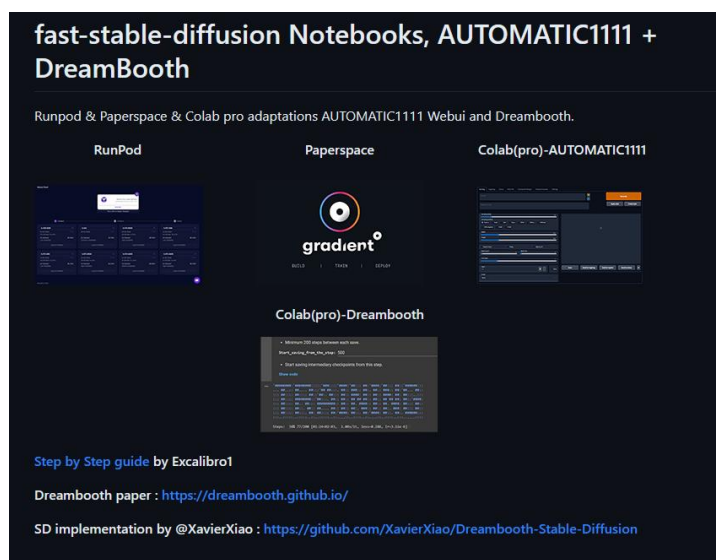
Ilustración 58: Ventana de bienvenida de Google Colab

En esta sección, se encuentran disponibles una serie de tutoriales y ejemplos que permiten familiarizarse con el entorno de programación en la nube. Si se desea utilizar un código propio, se puede abrir un nuevo script seleccionando la opción "Nuevo Cuaderno". Esto proporciona un espacio para crear y desarrollar el código personalizado en el entorno de la plataforma en la nube.



*Ilustración 59: cuaderno Jupyter de google colab en blanco*

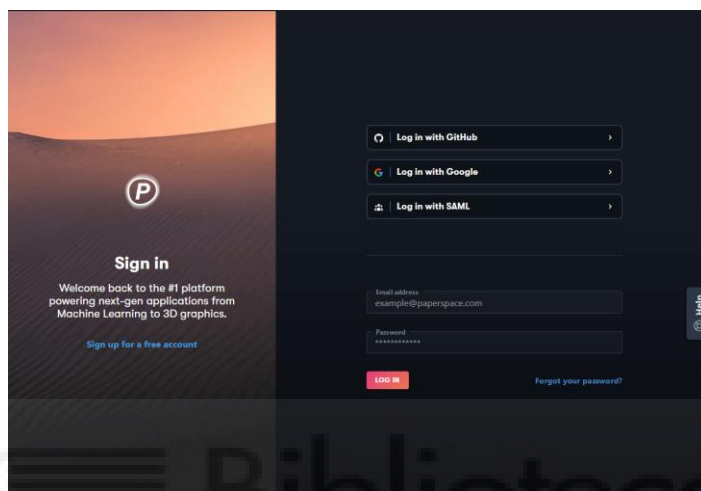
En el contexto de este proyecto, utilizaremos scripts proporcionados por la comunidad para simplificar el uso de Stable Diffusion, específicamente los ofrecidos por TheLastBen. Para acceder a estos scripts, se puede visitar su repositorio en GitHub llamado "fast-stable-diffusion" y acceder a los enlaces de Google Colab disponibles. Inicialmente, tanto DreamBooth como AUTOMATIC1111 estaban disponibles a través de Google Colab, pero debido a cambios en los términos de servicio, actualmente solo es posible utilizar DreamBooth en esta plataforma.



*Ilustración 60: Github Fast Stable Diffusion de TheLastBen*

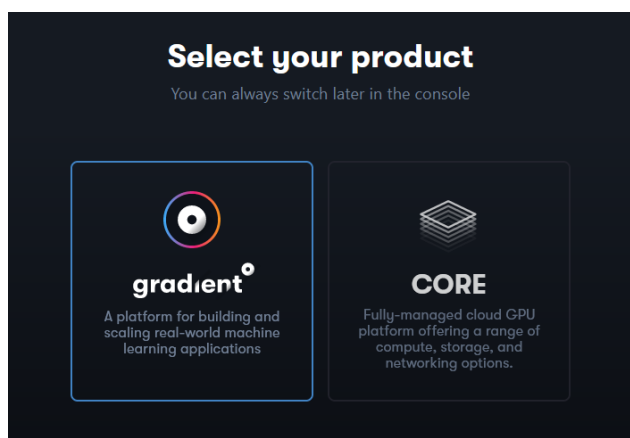
## Gradient PaperSpace

Para utilizar Gradient de Paperspace, también se requerirá iniciar sesión en la plataforma de alguna manera, y en este caso, utilizaremos la opción de iniciar sesión mediante Google.



*Ilustración 61: Pantalla de login para Paperspace*

Al ingresar por primera vez, se nos solicitará seleccionar el producto que deseamos utilizar, y en nuestro caso, nos interesa seleccionar Gradient.



*Ilustración 62: Selección de producto en Paperspace, con las opciones Gradient y CORE*

Rellenamos las opciones que se nos presentan con la información necesaria.

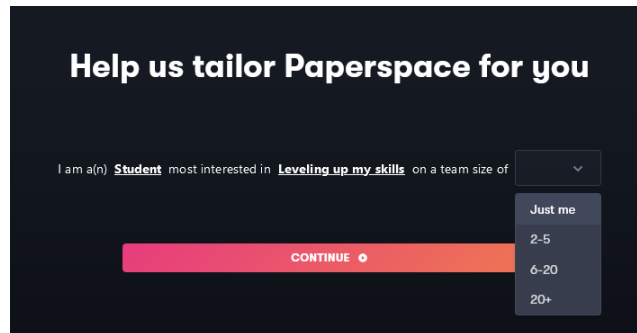


Ilustración 63: Opciones previas para el inicio en Gradient Paperspace

Y ya se podría crear el primer proyecto.

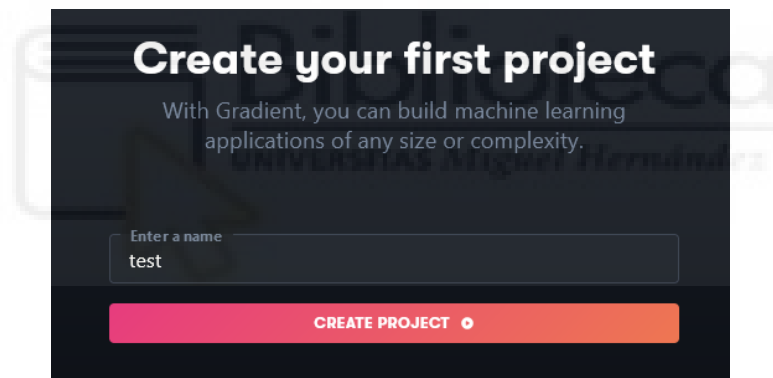


Ilustración 64: Pantalla nuevo proyecto para el inicio de Gradient Paperspace

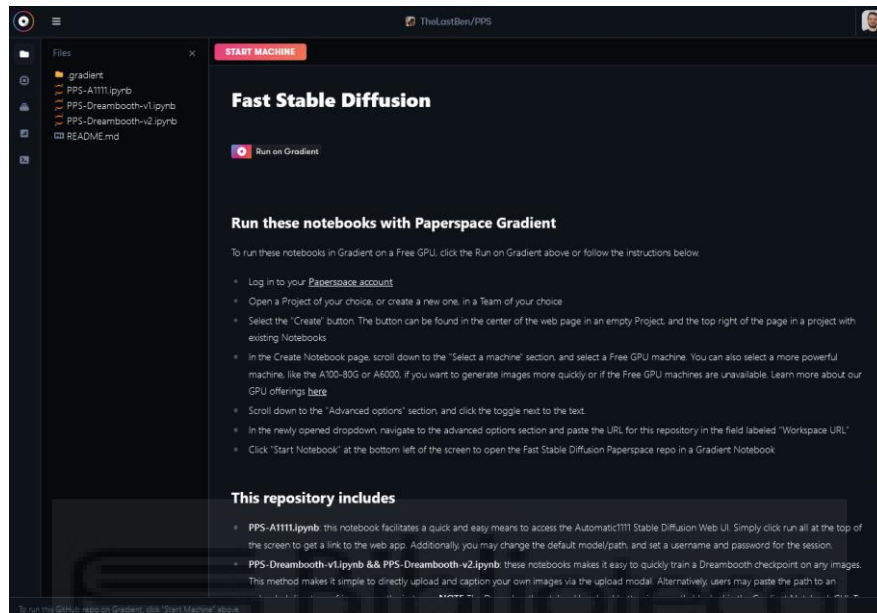
Una vez iniciada la sesión, también es posible crear proyectos directamente desde nuestra página de proyectos.



Ilustración 65: Opción de nuevo proyecto desde la plataforma



Al igual que en Google Colab, se utilizarán los scripts de Fast Stable Diffusion desarrollados por TheLastBen. En este caso, se ha creado un proyecto que incluye las dos versiones de Dreambooth y AUTOMATIC1111. Para este apartado, se utilizará AUTOMATIC1111 como herramienta principal.



*Ilustración 66: Proyecto de Gradient Paperspace para Fast Stable Diffusion*

Será necesario hacer clic en el botón "START MACHINE" y, al utilizar el plan gratuito, esperar que haya poca carga en el sistema para tener éxito en la conexión.

## 5.4.2 AUTOMATIC1111

Para utilizar AUTOMATIC1111, la interfaz web por excelencia de Stable Diffusion, es necesario contar con un entorno de ejecución adecuado. Si bien es posible ejecutarlo localmente, en este caso utilizaremos Gradient de Paperspace siguiendo los pasos previamente descritos. Se accederá al cuaderno Jupyter PPS-

A1111.ipynb, donde se procederá a la ejecución de las diversas secciones que se encuentran disponibles en dicho cuaderno.

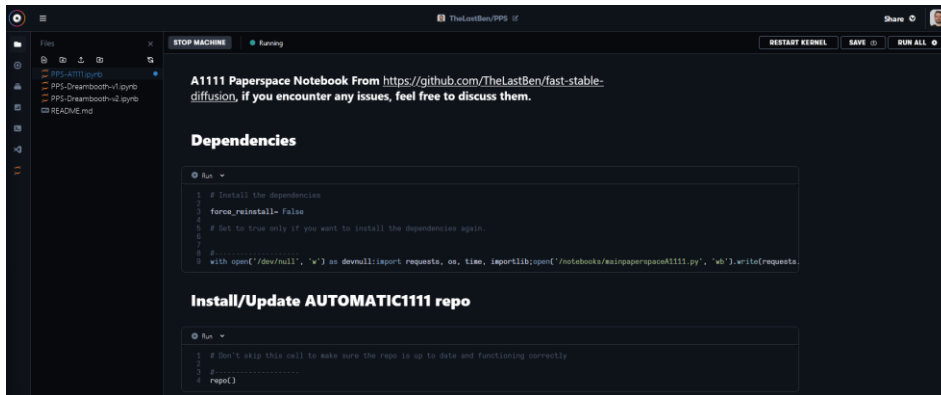


Ilustración 67: Cuaderno Jupyter de Gradient Paperspace para AUTOMATIC1111

Primero se instalarán las dependencias y el repositorio de AUTOMATIC1111:

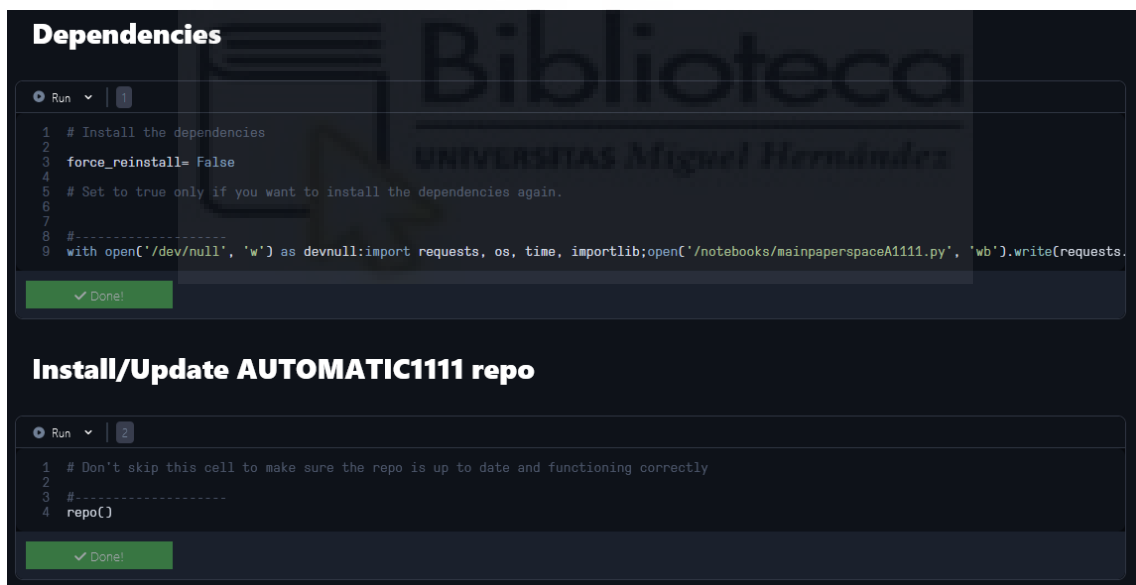
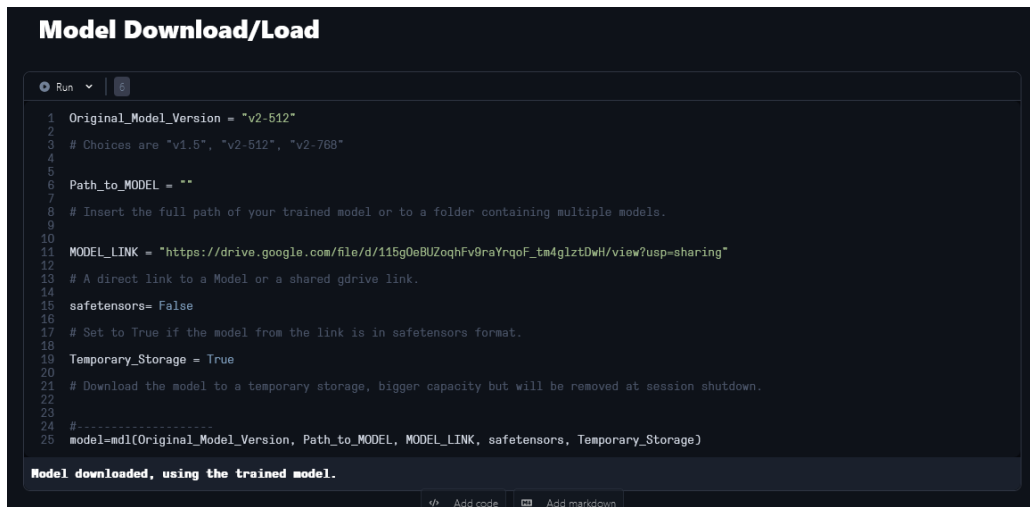


Ilustración 68: Instalación de dependencias y repo de AUTOMATIC1111

El siguiente punto consiste en seleccionar el modelo para nuestro proceso, optando en este caso por utilizar el modelo entrenado previamente en el apartado de Dreambooth. Sin embargo, es más común emplear el modelo original en la versión deseada. En el caso específico de utilizar Gradient, se recomienda evitar

la versión v2-768 debido a su propensión a agotar rápidamente la memoria RAM al agregar extensiones adicionales.

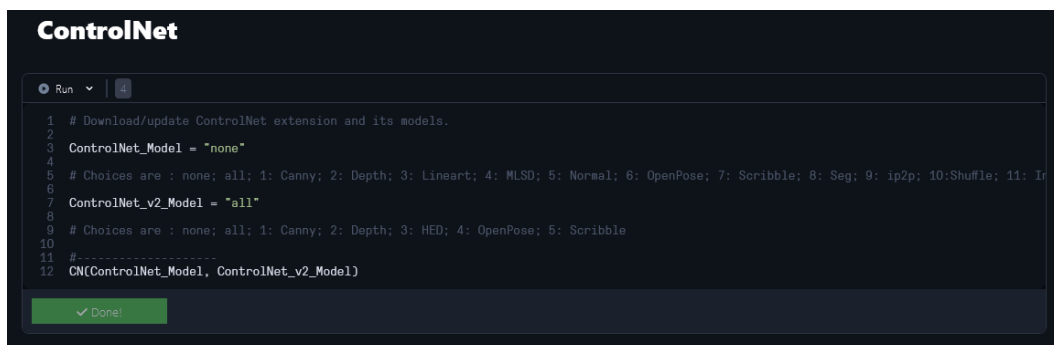


```
1 Original_Model_Version = "v2-512"
2
3 # Choices are "v1.5", "v2-512", "v2-768"
4
5
6 Path_to_MODEL = ""
7
8 # Insert the full path of your trained model or to a folder containing multiple models.
9
10
11 MODEL_LINK = "https://drive.google.com/file/d/115g0eBUZqhfV0raYrqoF_tm4glztdWh/view?usp=sharing"
12
13 # A direct link to a Model or a shared gdrive link.
14
15 safetensors= False
16
17 # Set to True if the model from the link is in safetensors format.
18
19 Temporary_Storage = True
20
21 # Download the model to a temporary storage, bigger capacity but will be removed at session shutdown.
22
23
24 #-----
25 model=mdl(Original_Model_Version, Path_to_MODEL, MODEL_LINK, safetensors, Temporary_Storage)
26
```

Model downloaded, using the trained model.

Ilustración 69: Descarga del modelo a utilizar para AUTOMATIC1111

Procedemos a agregar la extensión de ControlNet a nuestro entorno de trabajo. Esta extensión se puede añadir tanto desde la opción proporcionada en este punto específico, como desde la sección de extensiones en la interfaz de usuario. Ambas opciones nos permitirán incorporar esta funcionalidad adicional al proceso.



```
1 # Download/update ControlNet extension and its models.
2
3 ControlNet_Model = "none"
4
5 # Choices are : none; all; 1: Canny; 2: Depth; 3: Lineart; 4: MLSD; 5: Normal; 6: OpenPose; 7: Scribble; 8: Seg; 9: ip2p; 10: Shuffle; 11: Inpaint
6
7 ControlNet_v2_Model = "all"
8
9 # Choices are : none; all; 1: Canny; 2: Depth; 3: HED; 4: OpenPose; 5: Scribble
10
11 #-----
12 CN(ControlNet_Model, ControlNet_v2_Model)
```

Done!

Ilustración 70: Instalación de la extensión de ControlNet para AUTOMATIC1111

E iniciamos nuestra interfaz de usuario para Stable Diffusion.

### Start Stable-Diffusion

```
1 User = ""
2 Password= ""
3
4 # Add credentials to your Gradio interface (optional).
5
6
7 #-----
8
9 configf=sdUI(User, Password, model) if 'model' in locals() else sdUI(User, Password, "")
10 lpython /notebooks/sd/stable-difusion-webui/webui.py $configf
```

2023-06-14 16:25:59,001 - ControlNet - INFO - ControlNet v1.1.224  
ControlNet preprocessor location: /notebooks/sd/stable-difusion-webui/extensions/sd-webui-controlnet/annotator/downloads  
2023-06-14 16:25:59,207 - ControlNet - INFO - ControlNet v1.1.224  
Calculating sha256 for /models/model1.ckpt: Running on local URL: https://tensorboard-nsyk221127.01g07azj1.paperspacegradient.com  
✓ Connected  
Startup time: 11.1s (import torch: 5.5s, import gradio: 1.1s, import ldm: 0.3s, other imports: 1.2s, load scripts: 1.7s, create ui: 0.9s, gradio launch: 0.2s, scripts app\_started\_callback: 0.2s).

Ilustración 71: Despliegue de la interfaz AUTOMATIC1111

La página principal es la pestaña de txt2img, que es la funcionalidad más usada.

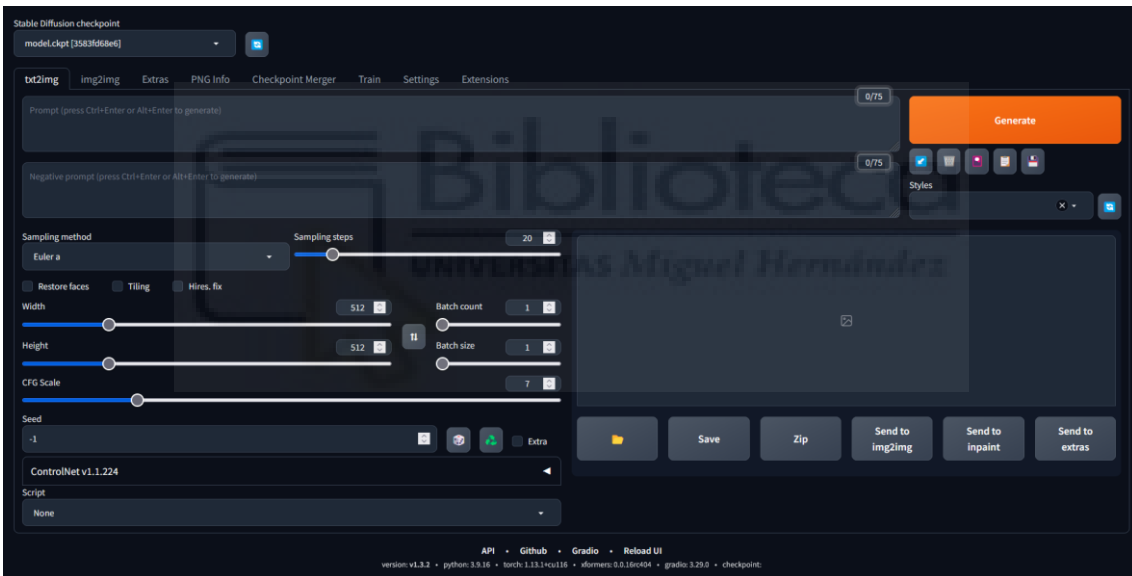


Ilustración 72: Interfaz de inicio de AUTOMATIC1111

Mediante el uso de la extensión txt2img, tendremos la capacidad de seleccionar el prompt positivo y negativo que deseemos para nuestra imagen generada. Además, dispondremos de la opción de ajustar diversos parámetros, como las dimensiones de la imagen, la cantidad de imágenes a generar en cada ejecución, la semilla (con el valor -1 para una semilla aleatoria), el método de muestreo que determinará cómo se inicia la generación de la imagen (siendo importante para obtener imágenes más detalladas en menos pasos, dependiendo del propósito

deseado), y los pasos de muestreo, un parámetro clave para obtener resultados detallados, recomendándose un valor entre 40 y 50 para lograr buenos resultados.

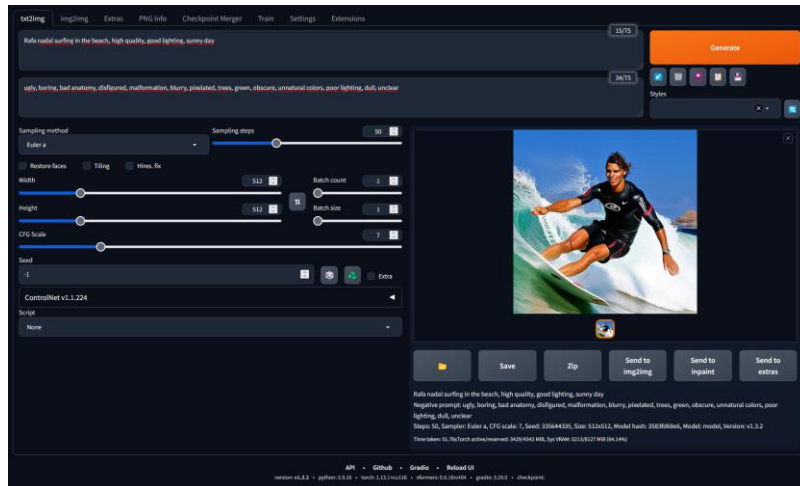


Ilustración 73: Ejemplo de uso de text2img desde AUTOMATIC1111

En la pestaña img2img, se nos brinda la posibilidad de utilizar una imagen como base para generar variaciones mediante la adición de ruido. De esta manera, podemos partir de una imagen inicial y, a través del prompt proporcionado, influir en la generación de la imagen resultante. Esta funcionalidad guarda similitudes con las capacidades de variación presentes en DALL-E y Midjourney.

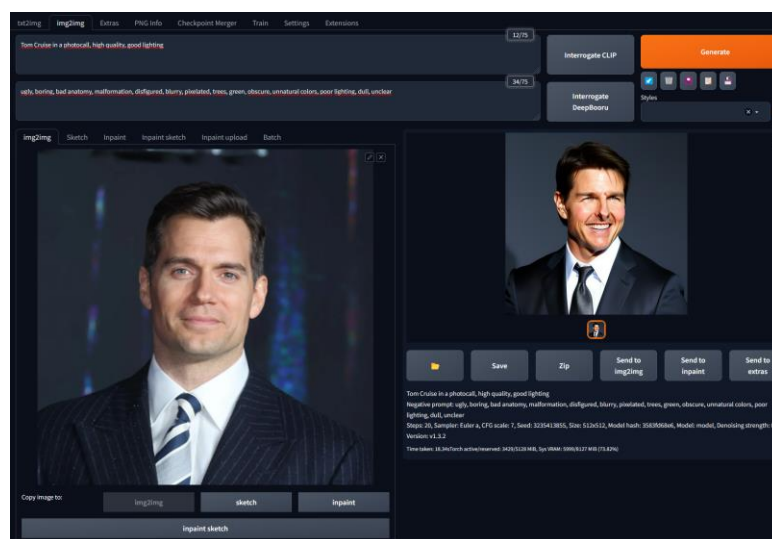
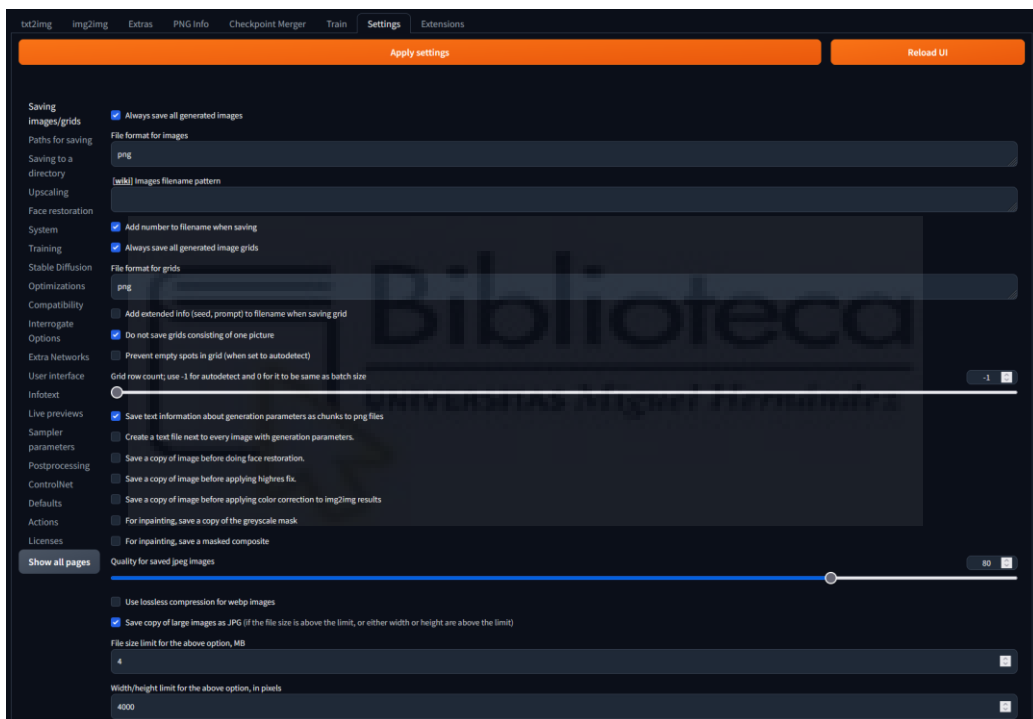


Ilustración 74: Ejemplo de uso de img2img desde AUTOMATIC1111

En la pestaña "Settings", se nos brinda la posibilidad de personalizar diferentes aspectos según nuestras preferencias. Entre las opciones disponibles se encuentran la selección del formato de imágenes que deseamos utilizar, la capacidad de guardar la información de las imágenes generadas tanto en el mismo archivo como en un archivo de texto separado, así como la configuración de la calidad de las imágenes y otras características relacionadas. En definitiva, esta pestaña nos permite ajustar todos los detalles a nuestro gusto y necesidades específicas.



*Ilustración 75: Apartado Settings de AUTOMATIC1111*

En la pestaña "Extensions" de la interfaz, específicamente en la sección "Installed", se nos brinda la capacidad de agregar y seleccionar los módulos que deseamos utilizar de entre los que ya se encuentran instalados. Esta funcionalidad nos permite ampliar y personalizar aún más las capacidades de la herramienta, adaptándola a nuestras necesidades específicas al agregar los módulos que consideremos relevantes para nuestro trabajo.

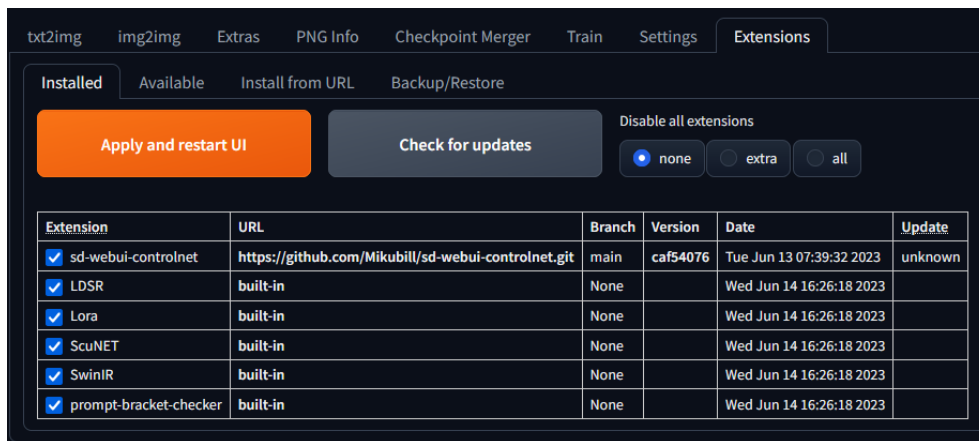


Ilustración 76: Apartado Extensions -> Installed de AUTOMATIC1111

En la pestaña "Extensions" de la interfaz, específicamente en la sección "Available", se nos brinda la posibilidad de instalar una amplia variedad de extensiones desde diferentes repositorios, no necesariamente limitados al repositorio oficial. Esto nos permite explorar y acceder a una amplia gama de funcionalidades adicionales que pueden ser de utilidad en nuestros proyectos. En este contexto, podríamos buscar y seleccionar la extensión de ControlNet, la cual será utilizada más adelante en nuestro trabajo. Sin embargo, es importante destacar que el cuaderno Jupyter utilizado previamente ya proporciona la opción de instalarla por defecto, dada su popularidad y uso frecuente.

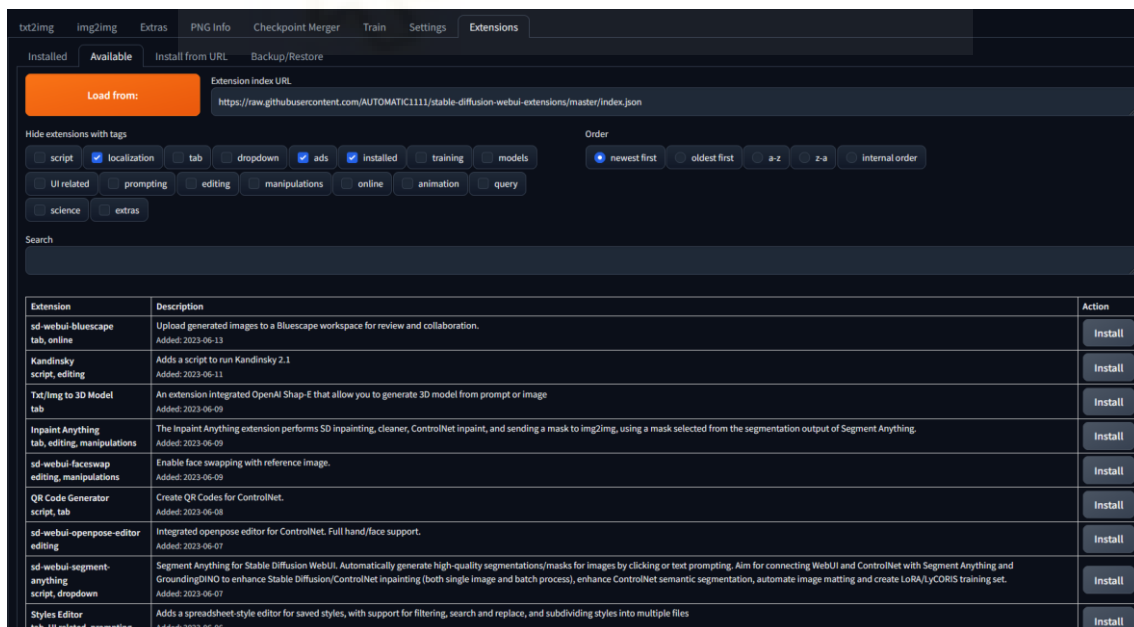


Ilustración 77: Apartado Extensions -> Available de AUTOMATIC1111

AUTOMATIC1111 se destaca como la interfaz más completa entre las tres que hemos explorado. Ofrece una amplia gama de funciones y opciones, brindando a los usuarios capacidades extensas. Sin embargo, como contrapartida, su diseño de interfaz es más complejo y menos intuitivo debido a la abundancia de opciones disponibles. Es más adecuado para usuarios con un nivel más avanzado de experiencia que se sientan cómodos navegando y aprovechando la multitud de opciones proporcionadas por la aplicación.

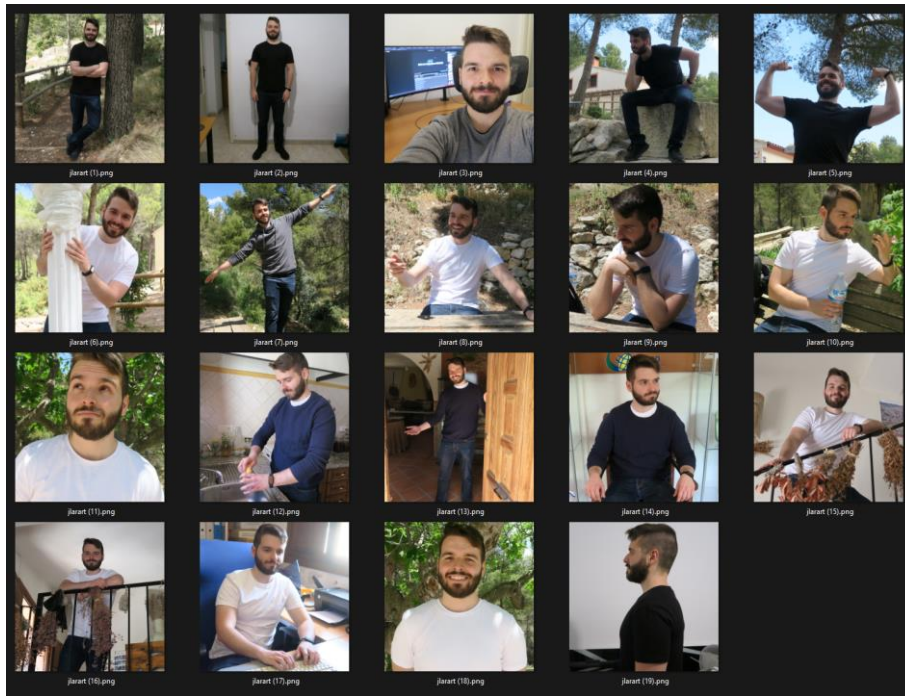
### 5.4.3 DREAMBOOTH

Para llevar a cabo el entrenamiento, es necesario contar con un dataset que incluya imágenes del elemento que se desea entrenar. Se recomienda obtener un dataset diverso que incluya diferentes perfiles, ángulos, expresiones y vestimentas.

Además, es preferible que las imágenes presenten una consistencia en términos de aspecto, especialmente en lo que respecta al peinado y la barba en este caso particular. Por lo tanto, se sugiere tomar todas las fotografías en el mismo día o en días cercanos. Sin embargo, también sería válido seleccionar imágenes que cumplan con esta consistencia independientemente de la fecha en que fueron tomadas.

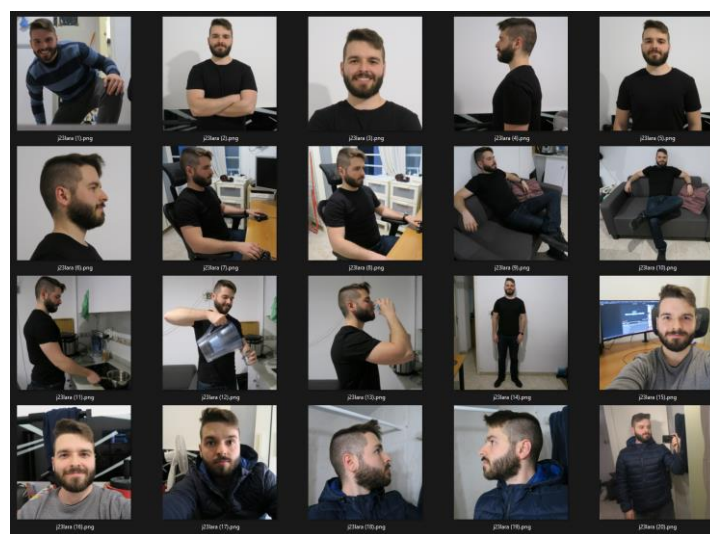
En este caso específico, se utilizarán 19 imágenes con una resolución de 768x768, que coincide con la resolución del modelo Stable Diffusion 2.1 utilizado en el entrenamiento.





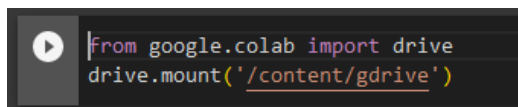
*Ilustración 78: Dataset utilizado para el entrenamiento con Dreambooth*

El primer intento de dataset generó un problema de overfitting debido a la repetición de escenarios, prendas de vestir y posturas. De hecho, este primer intento solo incluía las imágenes con la camiseta negra y ninguna de las demás variantes. Aunque se intentó mejorar la diversidad del dataset, los resultados obtenidos no fueron suficientes. Se muestra este ejemplo como una lección de cómo no abordar la construcción del dataset.



*Ilustración 79: Ejemplo de Dataset incorrecto para el entrenamiento con Dreambooth*

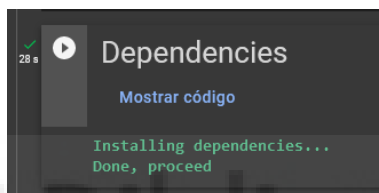
Con el dataset preparado, procederemos al entrenamiento de Dreambooth utilizando la herramienta fast-Dreambooth desarrollada por TheLastBen y aprovechando la plataforma Google Colab. Para ello, seguiremos los siguientes pasos: en primer lugar, estableceremos la conexión con nuestra cuenta de Google Drive para poder guardar el modelo una vez finalizado el entrenamiento.



```
from google.colab import drive
drive.mount('/content/gdrive')
```

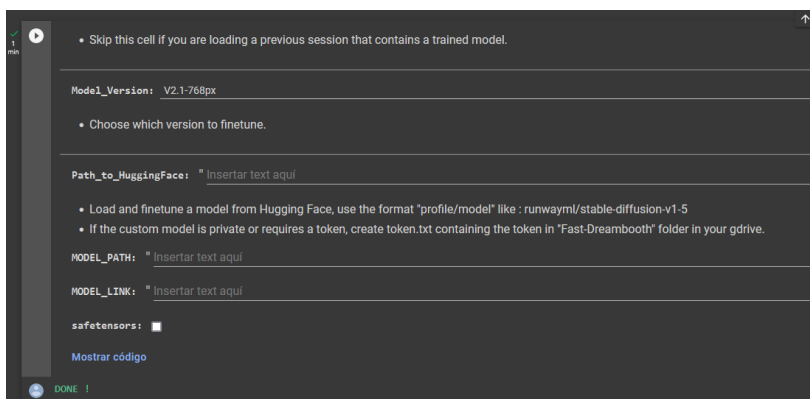
*Ilustración 80: Conexión del Cuaderno Jupyter de Google Colab con Google Drive*

Se instalarán las dependencias:



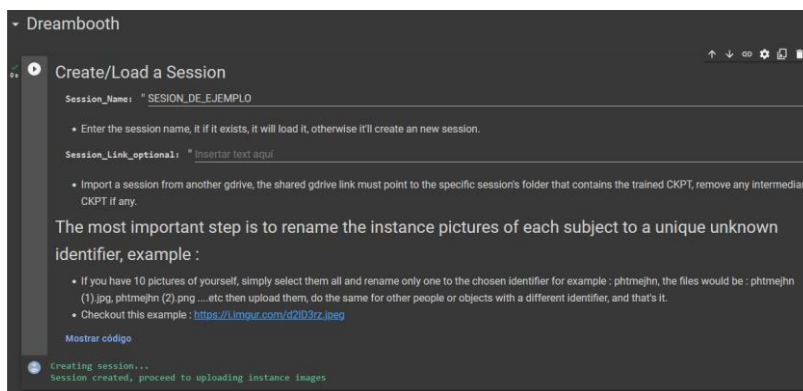
*Ilustración 81: Instalación de dependencias de Dreambooth*

Y se elige el modelo que se quiera entrenar, en este caso el de Stable Diffusion V2.1-768px:



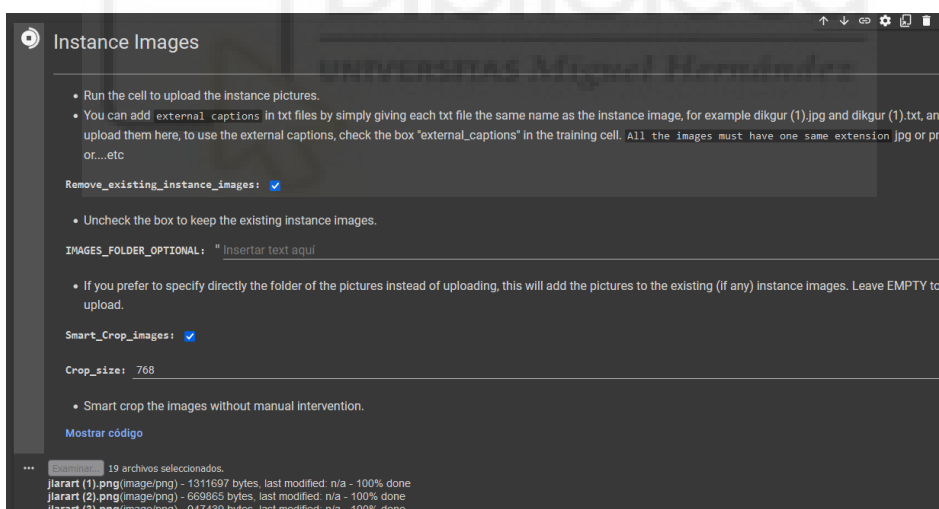
*Ilustración 82: Elección del modelo para entrenar*

Se crea una sesión, este paso es útil para más tarde poder reentrenar un modelo en caso de no ver resultados lo suficientemente afines a nuestro objetivo:



*Ilustración 83: Creación de la sesión*

Añadimos nuestro dataset, es necesario que todas las imágenes tengan el nombre del token que queremos entrenar para mostrar en el futuro. Es importante que este token no sea una palabra común, como "jorge" en mi caso, yo he utilizado "jlarart". También debemos seleccionar la resolución con la que queremos que se utilicen nuestras imágenes. Existe una opción para recortar automáticamente las imágenes a esa resolución (smart\_crop\_images). Sin embargo, si no queremos realizar ese proceso, podemos optar por no utilizarlo.



*Ilustración 84: Subida del Dataset mostrado anteriormente*

Y ahora comenzamos el entrenamiento, donde encontramos varios parámetros que se pueden ajustar, pero nos centraremos en los más importantes. En primer lugar, tenemos "UNet\_Training\_Steps", que representa la cantidad de iteraciones que se realizarán durante el entrenamiento. Cuantas más iteraciones se realicen,

más patrones aprenderá el modelo del elemento que estamos entrenando. Sin embargo, es importante tener en cuenta que si el dataset no es muy grande, como en este caso, el exceso de iteraciones puede llevar al overfitting, incluso si las imágenes seleccionadas son adecuadas. Además, si se realizan demasiadas iteraciones, pueden surgir incoherencias y deformaciones en los resultados. Para estas condiciones, se recomendaría un valor de alrededor de 2500-3000 iteraciones, lo cual dependerá tanto de la cantidad de imágenes como de su diversidad. También es importante tener en cuenta que cuanto más alto sea este valor, más tiempo de procesamiento será necesario.

Otro parámetro interesante es el `Text_Encoder_Training_Steps`, estas serán las iteraciones que se harán para entrenar el codificador de texto para introducir el nuevo token, en mi caso he probado 200, 350 y 450 y no he podido apreciar muchas diferencias, así que finalmente me he quedado con el valor por defecto que es 350.



*Ilustración 85: Primera parte de los parámetros del entrenamiento con Dreambooth*

Otro parámetro importante que debemos ajustar es la resolución. Se recomienda seleccionar la resolución más alta posible sin exceder el tamaño original de las imágenes. En cuanto al apartado "`Save_Checkpoint_Every_n_Steps`", nos será de utilidad para probar diferentes versiones de nuestro modelo entrenado en distintos niveles de avance. En este caso, configuraremos el guardado de checkpoints en 1500 pasos, luego cada 500 pasos y finalmente en 3500 pasos, lo que nos proporcionará tres modelos para probar: uno con 1500 pasos, otro

con 2000 pasos, otro con 2500 pasos, otro con 3000 pasos y otro con 3500 pasos. En el modelo que se ha utilizado, se eligieron 4000 pasos, pero al realizar pruebas no se encontraron diferencias significativas entre los modelos entrenados con 3000, 3500 y 4000 pasos. Por tanto, sería recomendable quedarse con la cantidad de 3000 pasos para este conjunto de imágenes, lo que nos permitiría ahorrar tiempo de procesamiento sin comprometer la calidad del modelo.

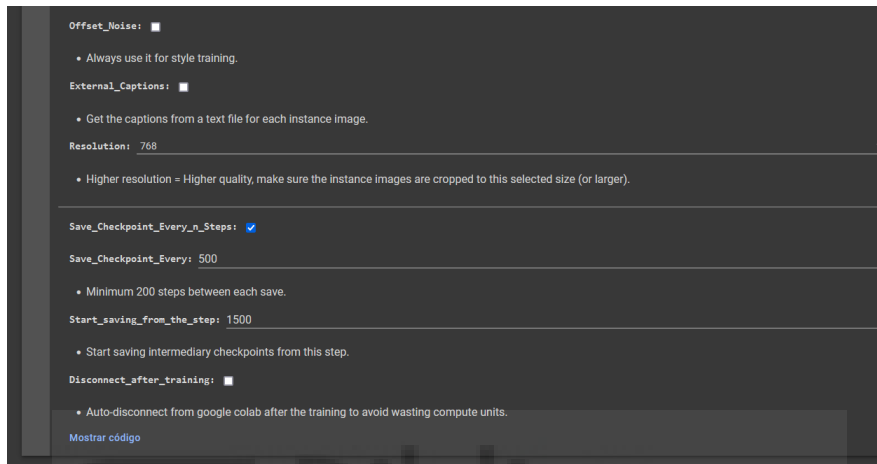


Ilustración 86: Segunda parte de los parámetros del entrenamiento con Dreambooth

El modelo entrenado y sus Checkpoints se encontrarán en nuestra unidad de google drive -> Fast-Dreambooth -> Sessions -> Nombre-de-sesión

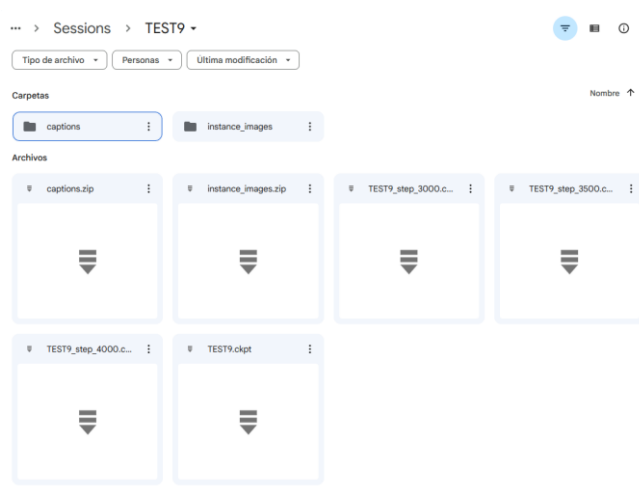



Ilustración 87: Modelo entrenado y modelos guardados en el proceso

Si el resultado obtenido no es satisfactorio, los usuarios tienen la opción de realizar un nuevo entrenamiento del modelo. Esta posibilidad se puede llevar a cabo mediante la utilización de la misma sesión creada anteriormente y utilizando el mismo nombre, o bien cargando un modelo pre-entrenado con diferentes características, como el proporcionado por OpenJourney mencionado previamente.

Anteriormente se mencionó que se podría utilizar la plataforma de computación en la nube Google Colab para llevar a cabo el entrenamiento del modelo. Sin embargo, a partir del 4 de abril de 2023, los términos de servicio de Google Colab han cambiado y ya no permiten la creación de interfaces web fuera de su entorno principal. Esta limitación afecta a la mayoría de las herramientas que ofrecen una interfaz práctica para Stable Diffusion. Por lo tanto, en el presente proyecto se utilizará PaperSpace Gradient en lugar de Google Colab. De este modo, se empleará la aplicación de código abierto AUTOMATIC1111 para llevar a cabo el entrenamiento y la generación de imágenes.

Primero se instalarán las dependencias:



```
Dependencies
Run
1 # Install the dependencies
2
3 force_reinstall= False
4
5 # Set to true only if you want to install the dependencies again.
6
7
8 #-----
9 with open('/dev/null', 'w') as devnull:import requests, os, time, importlib;open('/notebooks/mainpaperspaceA1111.f
```

✓ Done!

Ilustración 88: Instalación de dependencias de AUTOMATIC1111


Se continuará con AUTOMATIC1111:



```
Install/Update AUTOMATIC1111 repo  
Run 2  
1 # Don't skip this cell to make sure the repo is up to date and functioning correctly  
2  
3 #-----  
4 repo()  
✓ Done!
```

*Ilustración 89: Instalación de AUTOMATIC1111*


Se elegirá el modelo entrenado, que en este caso será a través de un link que apunta al fichero en Google Drive:



```
Model Download/Load  
Run 3  
1 Original_Model_Version = ""  
2  
3 # Choices are "v1.5", "v2-512", "v2-768"  
4  
5  
6 Path_to_MODEL = ""  
7  
8 # Insert the full path of your trained model or to a folder containing multiple models.  
9  
10  
11 MODEL_LINK = "https://drive.google.com/file/d/10sqU9ipT4rPRNpw4fRipnvX6eKCGiz95/view?usp=sharing"  
12  
13 # A direct link to a Model or a shared gdrive link.  
14  
15 safetensors= False  
16  
17 # Set to True if the model from the link is in safetensors format.  
18  
19 Temporary_Storage = True  
20  
21 # Download the model to a temporary storage, bigger capacity but will be removed at session shutdown.  
22  
23  
24 #-----  
25 mdl(Original_Model_Version, Path_to_MODEL, MODEL_LINK, safetensors, Temporary_Storage)  
Model downloaded, using the trained model.
```

*Ilustración 90: Descarga del modelo a usar en AUTOMATIC1111*

Se seleccionará el o los modelos de ControlNet deseados, en este caso "all" en el modelo v2 para tener todos disponibles de esta versión, en el siguiente apartado se profundizará sobre esta herramienta.



```
ControlNet

Run

1 # Download/update ControlNet extension and its models.
2
3 ControlNet_Model = "none"
4
5 # Choices are : none; all; 1: Canny; 2: Depth; 3: Lineart; 4: MLSD; 5: Normal; 6: OpenPose; 7: Scribble; 8: Seg; 9:
6
7 ControlNet_v2_Model = "all"
8
9 # Choices are : none; all; 1: Canny; 2: Depth; 3: HED; 4: OpenPose; 5: Scribble
10
11 #-----
12 CN(ControlNet_Model, ControlNet_v2_Model)

Done!
```

Ilustración 91: Instalación de la Extensión ControlNet

Y con el último apartado ya se iniciaría la interfaz de usuario AUTOMATIC1111 con nuestro modelo entrenado:



```
Start Stable-Diffusion

Stop

1 User = ""
2
3 Password= ""
4
5 # Add credentials to your Gradio interface (optional).
6
7
8 #-----
9 configf=sdui(User, Password, model) if 'model' in locals() else sdui(User, Password, "")
10 !python /notebooks/sd/stable-diffusion-webui/webui.py &configf

ControlNet v1.1.183
ControlNet v1.1.183
Calculating sha256 for /models/model.ckpt: Running on local URL: https://tensorboard-
nh7r7tjdmn.clg07azjl.paperspacegradient.com
✓ Connected
Startup time: 15.2s (import torch: 6.5s, import gradio: 1.9s, import ldm: 0.4s, other imports: 1.9s, setup codeformer:
0.2s, load scripts: 2.9s, create ui: 0.9s, gradio launch: 0.2s, scripts app_started_callback: 0.2s).
4e5dcc192370d1d9127e8d206942bc0080bacaa264b295d0b1144c65f2230b
Loading weights [4e5dcc192] from /models/model.ckpt
Creating model from config: /notebooks/sd/stablediffusion/configs/stable-diffusion/v2-inference-v.yaml
LatentDiffusion: Running in v-prediction mode
DiffusionWrapper has 865.91 M params.
Applying cross attention optimization (Doggettx).
Textual inversion embeddings loaded(0):
Model loaded in 27.5s (calculate hash: 12.2s, load weights from disk: 2.4s, find config: 4.5s, create model: 0.5s, apply
weights to model: 2.5s, apply half(): 1.9s, load VAE: 2.7s, move model to device: 0.7s).
```

Ilustración 92: Despliegue de la webUI AUTOMATIC1111 con el modelo entrenado



Aquí se presentan algunos ejemplos de imágenes generadas utilizando el modelo entrenado, donde se le solicitaron estilos de artistas, estilos de pintura, estilos de fotografía y conceptos específicos. Estas imágenes son resultado de la capacidad del modelo para interpretar y combinar diferentes estilos visuales y conceptuales, generando resultados únicos y creativos. Cabe destacar que la calidad y la coherencia de las imágenes generadas dependerán en gran medida de la configuración del modelo, la elección de los estilos solicitados y otros parámetros ajustables durante el proceso de generación.

### Van Gogh (Artista)



*Ilustración 93: Imágenes generadas con el modelo entrenado para el estilo "Van Gogh"*

### Carne Griffiths (Artista)



*Ilustración 94: Imágenes generadas con el modelo entrenado para el estilo "Carne Griffiths"*

Leonid Afremov (Artista)



*Ilustración 95: Imágenes generadas con el modelo entrenado para el estilo "Leonid Afremov"*

Jean Baptiste Monge (Artista)



*Ilustración 96: Imágenes generadas con el modelo entrenado para el estilo "Jean Baptiste Monge"*

Erin Hanson (Artista)



*Ilustración 97: Imágenes generadas con el modelo entrenado para el estilo "Erin Hanson"*

## Watercolor Drawing



*Ilustración 98: Imágenes generadas con el modelo entrenado para el estilo "Watercolor drawing"*

## Bokeh Cinematic Photography



*Ilustración 99: Imágenes generadas con el modelo entrenado para el estilo "Bokeh Cinematic Photography"*

## Alasdair McLellan (fotógrafo)



*Ilustración 100: Imágenes generadas con el modelo entrenado para el estilo "Alasdair McLellan"*

## Funko Pop



*Ilustración 101: Imágenes generadas con el modelo entrenado para el estilo "Funko Pop"*

## Escultura Griega

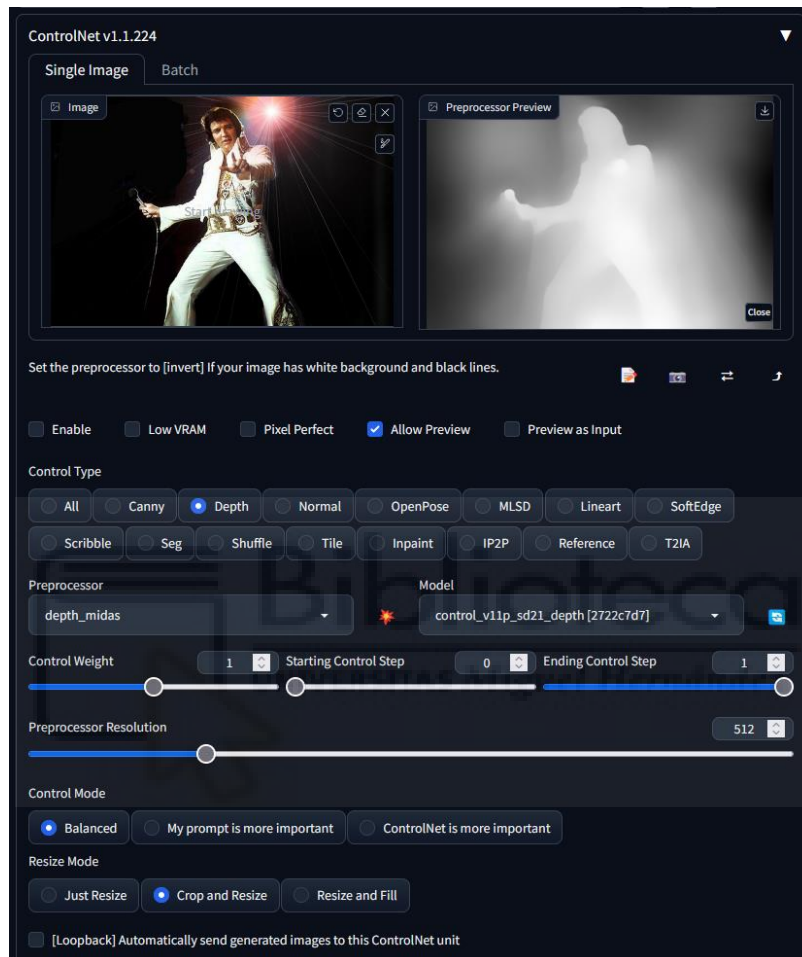


*Ilustración 102: Imágenes generadas con el modelo entrenado para el estilo "Escultura Griega"*

### 5.4.4 CONTROLNET

ControlNet ha emergido como una herramienta fundamental en el contexto de Stable Diffusion, permitiendo alcanzar resultados deseados de una manera más precisa y controlada. Mediante diferentes tipos de control, es posible proporcionar al modelo de Stable Diffusion imágenes con patrones específicos, lo que permite orientar su comportamiento de manera más precisa. Además, ControlNet ofrece

una interfaz que facilita la obtención de imágenes utilizadas por el modelo, generadas a partir de reglas predefinidas aplicadas a imágenes normales. En resumen, ControlNet representa una versión refinada y potente de la funcionalidad inicialmente ofrecida por la herramienta img2img.



*Ilustración 103: Ejemplo de uso ControlNet para la creación del filtro depth2img con una imagen de Elvis Presley*

A continuación, se presenta un ejemplo del tipo de control denominado Depth2img, el cual utiliza la información de profundidad de una imagen para generar un prompt acorde a esta característica. Este enfoque permite ajustar el nivel de restricción con el cual el modelo debe interpretar la guía de profundidad proporcionada. Los parámetros ControlWeight y ControlMode desempeñan un papel crucial en la configuración de los márgenes de actuación del modelo y en la rigidez con la que debe seguir la guía establecida. Estos parámetros permiten

ajustar el grado de flexibilidad y precisión requeridos en la generación de imágenes.

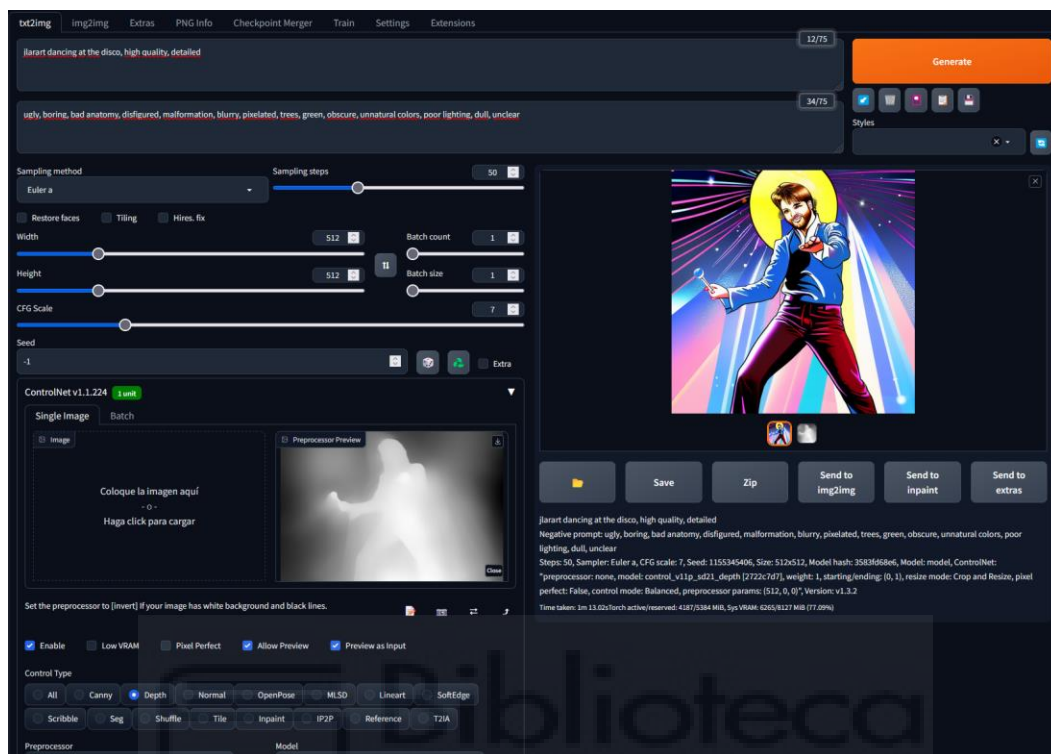
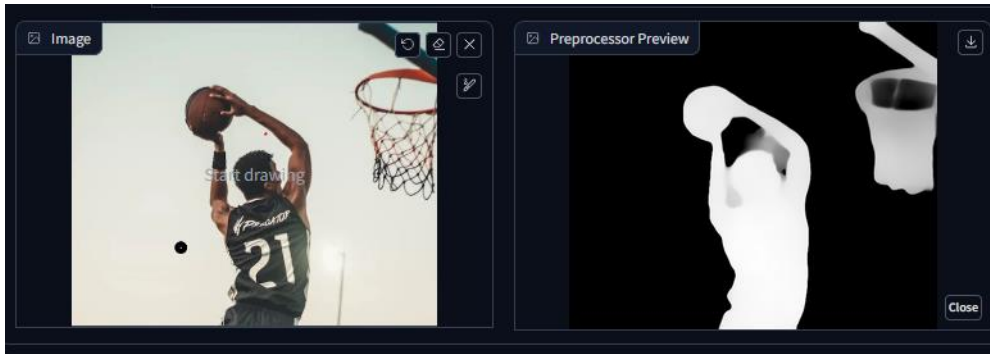


Ilustración 104: Ejemplo de uso ControlNet con el resultado de depth2img

Seguidamente, se presenta otro ejemplo utilizando el tipo de control Depth2img. En esta instancia, se genera una imagen que representa a un hombre realizando un mate en una canasta de baloncesto. Mediante el uso de información de profundidad, se logra una mayor precisión al guiar la generación de la imagen, permitiendo especificar la posición y apariencia del sujeto en relación con su entorno. Esto proporciona un mayor nivel de control y detalle en el proceso de generación de imágenes.



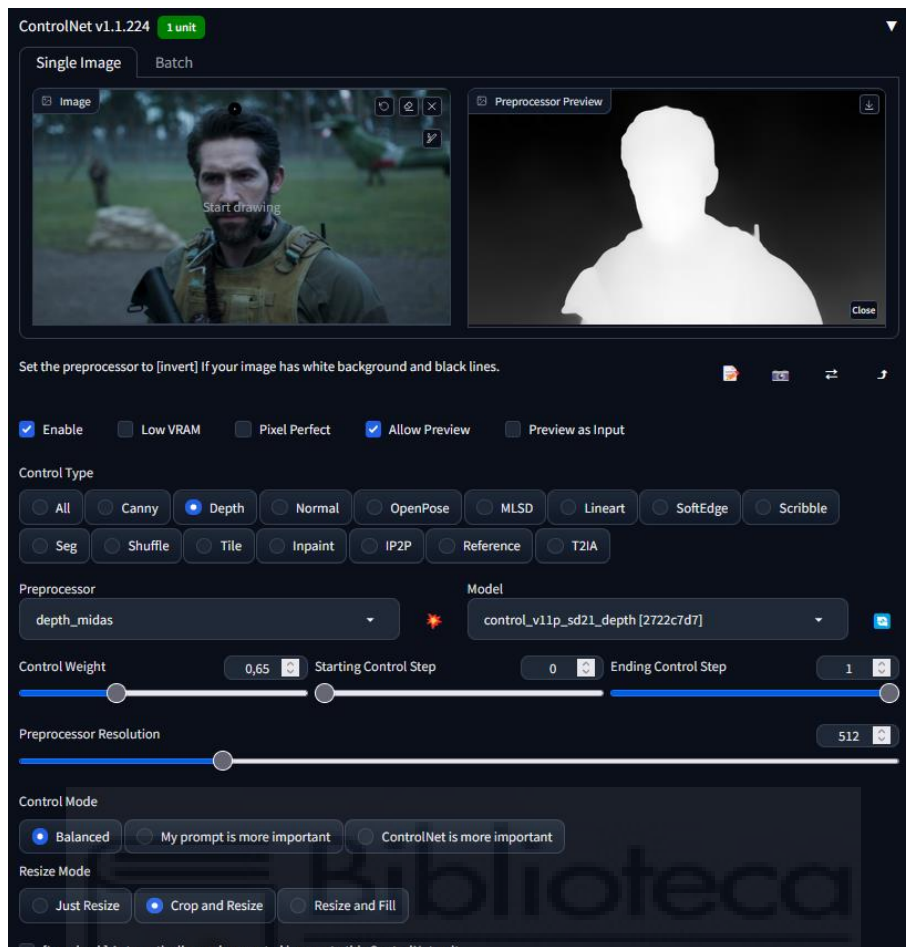
*Ilustración 105: Otro ejemplo de filtro depth2img de un hombre haciendo un mate en una canasta*

Por supuesto, todas estas técnicas son perfectamente compatibles con el entrenamiento de los modelos. En este caso particular, el usuario asume el rol de ejecutar el mate en la imagen generada.



*Ilustración 106: Resultados para el ejemplo de la canasta con el token jlarart creado en el entrenamiento*

En este siguiente ejemplo, se utiliza un plano de una película que muestra a un hombre del ejército como base. Sin embargo, en esta ocasión se reduce el parámetro "Control Weight" con el propósito de brindarle al modelo una mayor libertad para generar el resultado, aunque pueda desviarse ligeramente de la guía inicial. Esta variación en el ajuste permite explorar diferentes posibilidades creativas y obtener resultados sorprendentes, aunque no necesariamente precisos en relación con la guía proporcionada.



*Ilustración 107: Ejemplo para el filtro depth2img con un fotograma de una película de guerra*

Estos serían los resultados:



*Ilustración 108: Resultados para el filtro de la película de guerra con el token jlarart del modelo entrenado*



En esta ocasión, exploraremos otro tipo de control llamado "Hed", el cual utiliza los trazos de una imagen como guía para la generación del resultado. Este enfoque resulta especialmente interesante al trabajar con planos tan distintivos como el de la película "El Resplandor", ya que nos permite capturar con mayor precisión los detalles y expresiones presentes en la imagen original. A diferencia de la técnica "Depth2img", que puede encontrar dificultades al representar estos niveles de detalle, el control basado en trazos ofrece una mayor fidelidad en la reproducción de la imagen guía.

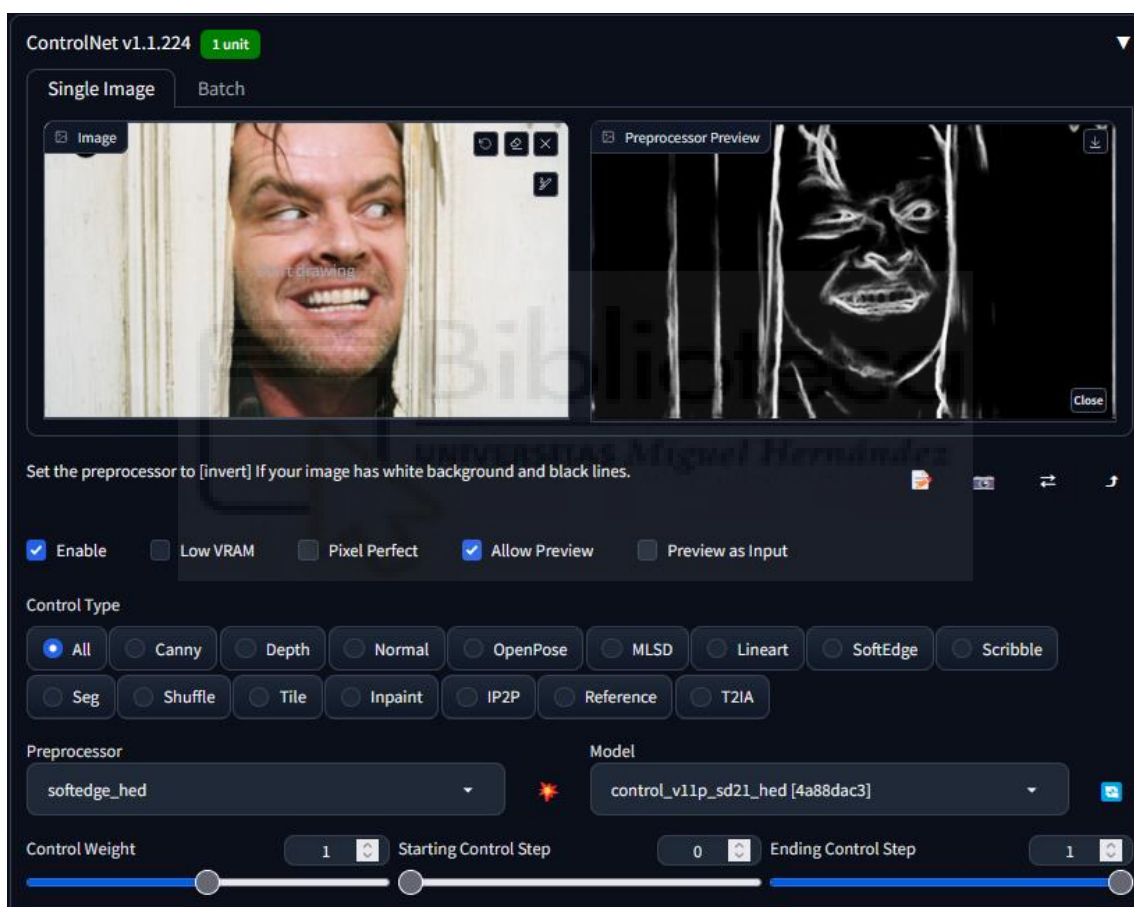


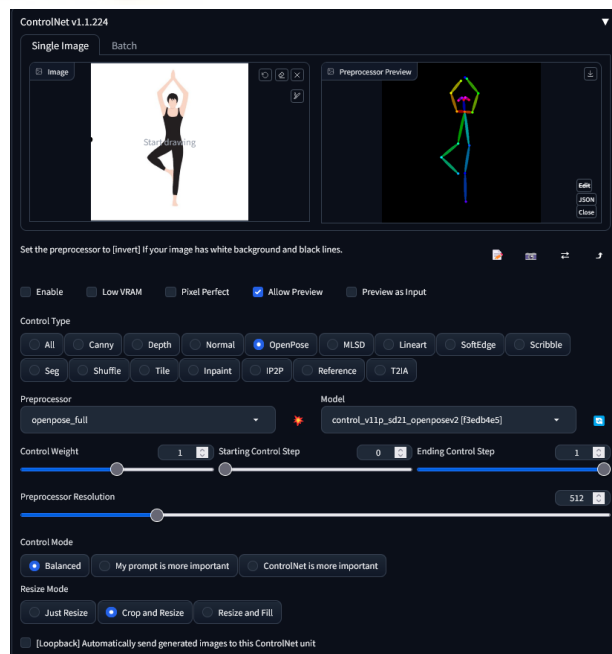
Ilustración 109: Ejemplo para el filtro "Hed" de ControlNet con la imagen de la famosa escena de El Resplandor

Y de nuevo se podría aplicar el filtro de control a nuestro modelo entrenado.



*Ilustración 110: Resultados para el filtro Hed con el token jlarart del modelo entrenado*

Otro tipo de control mucho menos exigente en los detalles sería OpenPose, con él se podrían controlar las posturas de las personas que se quieran representar. Sería posible añadir una imagen para que lea la postura o introducir directamente una postura previamente creada.



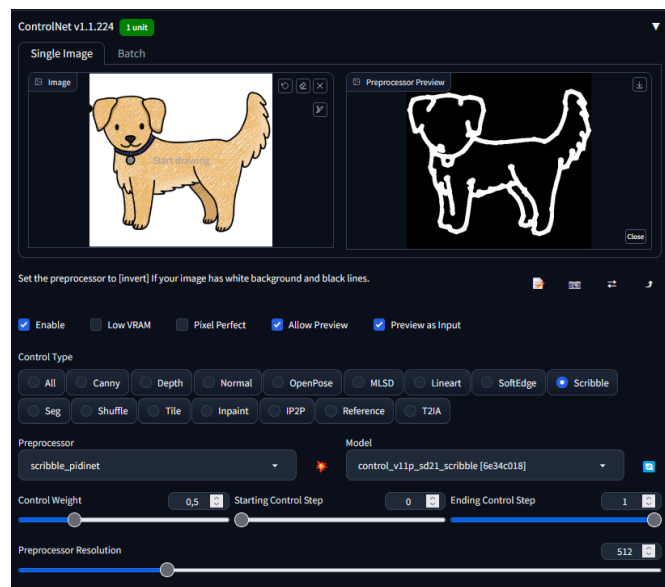
*Ilustración 111: Ejemplo de creación de un filtro OpenPose para la lectura de posturas, en este caso de yoga*

Estos serían los resultados:



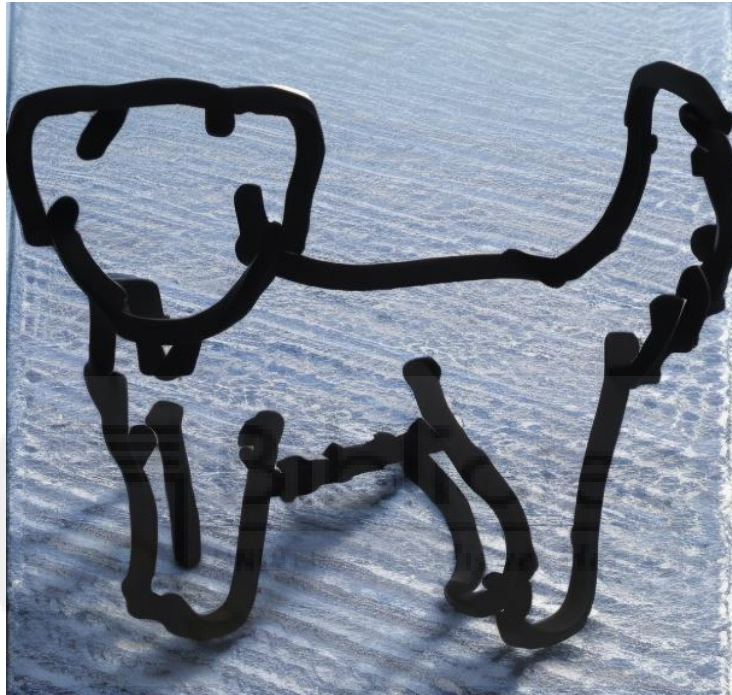
*Ilustración 112: Resultados para el filtro OpenPose de yoga*

Por último, cabe mencionar otro tipo de control disponible en la herramienta, el cual se centra en la generación de imágenes a partir de trazos simples. Esta función resulta especialmente útil para plasmar ideas y bocetos de manera más guiada, permitiendo tanto la introducción directa de trazos como la generación automática de una imagen guía basada en dichos trazos. Con esta opción, los usuarios pueden dar vida a sus ideas y llevarlas a la realidad de forma más precisa y orientada.



*Ilustración 113: Ejemplo de creación de un filtro Scribble con la imagen de un perro dibujado*

Con este tipo de control, se observa claramente la influencia del parámetro "Control Weight" al tener trazos más gruesos y una guía binaria en blanco y negro. En el siguiente ejemplo, se ha establecido el peso máximo, es decir, un valor de 2. Se puede apreciar que el modelo ha replicado fielmente el trazo original, aunque ha agregado un ligero fondo, ya que, al asignar un peso tan alto al control, no se le dará margen de creatividad suficiente en la generación.



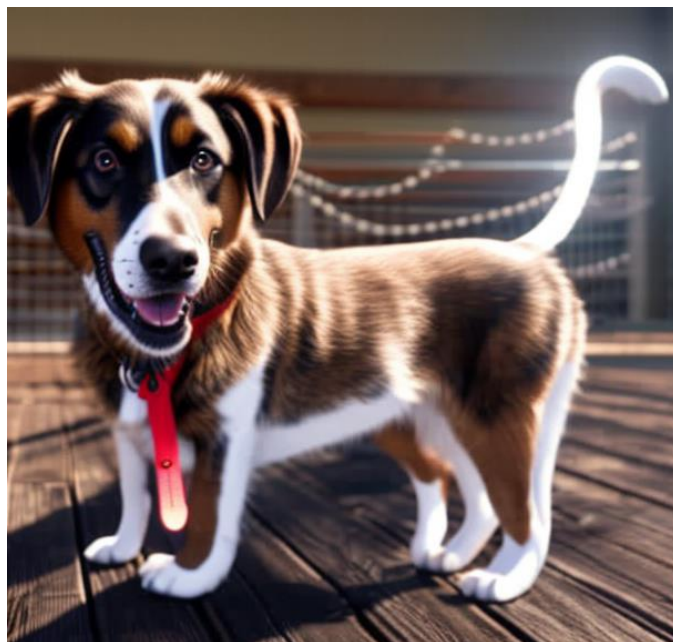
*Ilustración 114: Resultado filtro Scribble con el Control Weight a 2*

Al reducir el parámetro "Control Weight" a 1, se observa que el modelo sigue teniendo dificultades con el trazo grueso utilizado como guía. En este caso, se genera una especie de efecto de iluminación causado por el trazo del perro. Es evidente que el modelo tiene limitaciones para interpretar adecuadamente los trazos más gruesos y sutiles, lo que resulta en resultados menos precisos en comparación con el peso máximo del control.



*Ilustración 115: Resultado filtro Scribble con el Control Weight a 1*

Al ajustar el parámetro "Control Weight" a 0.8, se puede observar que las líneas blancas que recorren el cuerpo del perro son más evidentes y se pueden identificar claramente como resultado del trazo utilizado en la imagen guía. Aunque todavía se pueden apreciar algunas imperfecciones, la imagen generada comienza a ser más aceptable en términos de representar el concepto deseado.



*Ilustración 116: Resultado filtro Scribble con el Control Weight a 0.8*

Al utilizar un valor de "Control Weight" de 0.5, se puede apreciar un resultado bastante satisfactorio en términos de seguir parcialmente la coherencia del modelo de control. Aunque existen algunas dificultades en la representación de la cola del perro, la imagen generada cumple bastante bien considerando que se utilizó un conjunto de trazos crudos como guía.



*Ilustración 117: Resultado filtro Scribble con el Control Weight a 0.5*

## 5.5 RESUMEN COMPARATIVA

A lo largo del apartado 5, se han analizado y comentado las diferencias y similitudes existentes entre los modelos DALL-E, Midjourney y Stable Diffusion. Sin embargo, con el objetivo de sintetizar la información presentada, se procederá a crear una tabla comparativa que incluirá algunos de los puntos más relevantes a considerar para determinar las fortalezas y debilidades de estos tres modelos.

La tabla comparativa permitirá una visualización clara y concisa de las características distintivas de cada modelo, así como de sus puntos destacados y áreas de mejora. Al establecer una estructura tabular, se facilitará la comparación directa entre los modelos, lo que a su vez ayudará a los investigadores y profesionales a evaluar y tomar decisiones informadas sobre cuál de ellos podría ser más adecuado para un determinado contexto o aplicación.

A continuación, se presenta una tabla comparativa de los modelos DALL-E, Midjourney y Stable Diffusion. Es importante destacar que en todos los puntos expuestos, se considera que "más es mejor". Por lo tanto, si se hace referencia al costo de las herramientas, contar con un mayor número de puntos en este apartado indicaría que el costo es inferior o de alguna manera más accesible:

Puntos Relevantes	DALL-E	Midjourney	Stable Diffusion
	Facilidad de uso	█	█
Interfaz amigable	█	█	█
Variaciones de imágenes	█	█	█
Herramientas base	█	█	█
Coste de entrada	█	█	█
Información sobre el modelo	█	█	█
Resultados precisos	█	█	█
Resultados espectaculares	█	█	█
Nivel de detalle	█	█	█
Imágenes más realistas	█	█	█
Menos incoherencias	█	█	█
Nec. menos complej prompt	█	█	█
Nec. menos conocimiento	█	█	█
Libertad para el usuario	█	█	█
Potencial para el usuario	█	█	█
Comunidad detrás	█	█	█

Ilustración 118: Tabla comparativa de los modelos DALL-E, Midjourney y Stable Diffusion





# **CAPÍTULO 6:**

# **CONCLUSIONES Y**

# **PROPUESTAS**



## 6.1 CONCLUSIONES

Este trabajo de fin de grado ha sido desarrollado teniendo en cuenta las tecnologías disponibles al inicio del proyecto, que se remonta aproximadamente a febrero de 2023 cuando se inició de manera más consistente. Aunque algunas tecnologías clave para este trabajo se han añadido a mitad del proceso, resulta imposible adaptarlo continuamente a medida que surgen nuevas tecnologías, dado que el campo de la Inteligencia Artificial, y específicamente el de generación de imágenes, avanza a un ritmo vertiginoso.

Actualmente, contamos con herramientas espectaculares para generar imágenes, especialmente si uno se sumerge en todas las técnicas y se mantiene actualizado. Es posible lograr una imagen completamente a medida según los requisitos establecidos, pero para alcanzar un resultado de calidad a este nivel, seguiría siendo necesario una buena base de conocimiento, un esfuerzo considerable y compromiso con el objetivo. Sin embargo, es importante destacar que este trabajo solo abarca una pequeña parte de la generación de imágenes con Inteligencia Artificial, ya que es imposible abarcar todos los aspectos y profundizar suficientemente en ellos.

La generación de imágenes mediante Inteligencia Artificial es un campo amplio y en constante evolución. A medida que se desarrollan nuevas técnicas y modelos, se abren nuevas posibilidades para la creación de imágenes realistas y personalizadas. Es esencial reconocer la importancia de mantenerse actualizado y explorar continuamente las innovaciones en este ámbito.

A pesar de los avances logrados en este trabajo, aún quedan desafíos por enfrentar. La generación de imágenes que sea completamente indistinguibles y convincentes sigue siendo un objetivo complejo, que requiere un enfoque multidisciplinario y la combinación de diferentes técnicas. Además, los aspectos

éticos y legales relacionados con el uso de Inteligencia Artificial en la generación de imágenes también deben ser considerados y abordados adecuadamente.

En resumen, este trabajo ha contribuido a un campo en constante desarrollo y ha explorado una pequeña porción de las posibilidades de generación de imágenes con Inteligencia Artificial. A medida que se avanza en esta área, es esencial seguir investigando y adaptando nuevas técnicas y tecnologías para lograr resultados aún más impresionantes.

## 6.2 PROPUESTAS DE FUTURO

En el ámbito en constante evolución de la generación de imágenes con Inteligencia Artificial, resulta complejo realizar propuestas concretas debido a la rápida evolución y los continuos avances en esta disciplina. Sin embargo, se pueden identificar algunas áreas que podrían beneficiarse de futuras investigaciones y desarrollos.

En primer lugar, la generación de vídeos puede considerarse una propuesta prometedora, que por supuesto ya está siendo explorada, por ejemplo por parte de Stable Diffusion y ControlNet con sus herramientas text2video y video2video, pero sí que necesita ser pulida. También cabe destacar las herramientas Gen-1 [36] y Gen-2 [37] de Runway [38] de generación de vídeo, la primera en base a vídeo y texto, y la segunda tan solo en base a texto, consiguiendo un nivel de calidad sorprendente. Dado que un vídeo es una secuencia de imágenes en movimiento, el desafío radica en lograr coherencia y fluidez en la sucesión de imágenes. Explorar técnicas y modelos que permitan generar vídeos realistas y consistentes podría abrir nuevas oportunidades en campos como la animación, los efectos visuales y la Realidad Virtual.

En segundo lugar, un área de gran potencial futuro en relación a la generación de imágenes se centra en el reconocimiento de patrones en imágenes como paso previo a su generación. Este enfoque permite identificar relaciones y estructuras entre imágenes, lo que a su vez habilita el uso de la inteligencia artificial para la edición de imágenes. Una herramienta destacada en este ámbito es Drag Your Gan [39], la cual actualmente logra obtener resultados de edición de imágenes altamente controlados mediante la selección de parámetros específicos. Este enfoque ofrece una manera extremadamente precisa de manipular imágenes, brindando oportunidades significativas en el ámbito de la edición visual.

El reconocimiento de patrones en imágenes también abre la puerta a importantes avances en la creación de modelos tridimensionales a partir de texto, haciendo uso de herramientas como ProlificDreamer [40]. Asimismo, se ha observado el desarrollo de técnicas de reconstrucción de modelos 3D a partir de vídeos, entre las cuales destaca Neuralangelo [41]. Estos ejemplos resaltan la existencia actual de diversos proyectos extraordinarios que se centran en la generación de contenido audiovisual con la asistencia de la inteligencia artificial. Estas iniciativas convergen en la búsqueda de soluciones innovadoras y creativas, impulsando así el progreso y la diversificación de las aplicaciones de la inteligencia artificial en el ámbito audiovisual.

Otro punto crucial es acercar esta tecnología al público en general. Al igual que ha sucedido con ChatGPT, donde se ha logrado acercar la generación de texto a la persona promedio, sería deseable mejorar el procesamiento del "prompt" en la generación de imágenes de modo que no sea necesario ser un experto en la forma de expresar conceptos para obtener resultados satisfactorios, esto también está siendo explorado por ejemplo por parte de CLIP, el Text Embedder de Stable Diffusion. Mejorar la usabilidad y la interfaz de usuario de las plataformas y herramientas existentes es un aspecto crucial para ampliar el alcance y la adopción de la generación de imágenes con Inteligencia Artificial.

En conclusión, el futuro de la generación de imágenes con Inteligencia Artificial es prometedor y presenta numerosas oportunidades para la investigación y el desarrollo. Investigar y abordar activamente los avances y desarrollos emergentes en diversas áreas para aprovechar plenamente las oportunidades que están surgiendo y facilitar el acceso a esta tecnología a un público más amplio son propuestas interesantes que podrían impulsar aún más el avance de este campo. Continuar mejorando la usabilidad y la intuitividad de las herramientas existentes, así como fomentar la colaboración interdisciplinaria, serán clave para aprovechar todo el potencial de la generación de imágenes con Inteligencia Artificial en el futuro.





# **CAPÍTULO 7:**

## **BIBLIOGRAFÍA**



## 7.1 FUENTES BIBLIOGRÁFICAS

- [1] Sistemas Expertos «SISTEMAS EXPERTOS: Fundamentos, Metodologías y Aplicaciones» [En línea]. Available: [https://www.researchgate.net/publication/312245490\\_SISTEMAS\\_EXPERTOS\\_Fundamentos\\_Metodologias\\_y\\_Aplicaciones](https://www.researchgate.net/publication/312245490_SISTEMAS_EXPERTOS_Fundamentos_Metodologias_y_Aplicaciones) [Último acceso: 28 Mayo 2023].
- [2] Redes Neuronales Artificiales «Redes Neuronales Artificiales y sus aplicaciones» [En línea]. Available: [https://ocw.ehu.es/pluginfile.php/40137/mod\\_resource/content/1/redes\\_neuro/contenidos/pdf/libro-del-curso.pdf](https://ocw.ehu.es/pluginfile.php/40137/mod_resource/content/1/redes_neuro/contenidos/pdf/libro-del-curso.pdf) [Último acceso: 28 Mayo 2023].
- [3] Deep Learning «Review of deep learning: concepts, CNN architectures, challenges, applications, future directions» [En línea]. Available: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8> [Último acceso: 28 Mayo 2022].
- [4] Robótica, «Robótica e Inteligencia Artificial» [En línea]. Available: <http://www.luisximenez.com/humanidades/pdf/Robotica%20e%20inteligencia%20Artificial.pdf> [Último acceso: 28 Mayo 2023].
- [5] Agentes Inteligentes, «Agentes Inteligentes, definición y tipología» [En línea]. Available: <https://core.ac.uk/download/pdf/11890345.pdf> [Último acceso: 28 Mayo 2023].
- [6] Dendral, «DENDRAL and expert system applications» [En línea]. Available: <https://stacks.stanford.edu/file/druid:pj337tr4694/pj337tr4694.pdf> [Último acceso: 28 Junio 2022].
- [7] AutoPilot «AutoPilot - Tesla» [En línea]. Available: [https://www.tesla.com/es\\_ES/autopilot](https://www.tesla.com/es_ES/autopilot) [Último acceso: 4 Junio 2023].
- [8] Alexa (Amazon) «Amazon scientists applying deep neural networks to custom skills» [En línea]. Available: <https://www.amazon.science/blog/amazon-scientists-applying-deep-neural-networks-to-custom-skills> [Último acceso: 4 Junio 2023].
- [9] Sophia «Sophia - Hanson Robotics» [En línea]. Available: <https://www.hansonrobotics.com/sophia/> [Último acceso: 20 Mayo 2023].



- [10] ChatGPT «Introducing ChatGPT» [En línea]. Available: <https://openai.com/blog/chatgpt> [Último acceso: 12 Junio 2023].
- [11] Visión Artificial «¿Qué es la visión artificial?» [En línea]. Available: <https://www.ibm.com/es-es/topics/computer-vision> [Último acceso: 2 Junio 2023].
- [12] Redes Adversarias Generativas «Generative Adversarial Nets» [En línea]. Available: <https://arxiv.org/abs/1406.2661> [Último acceso: 29 Mayo 2023].
- [13] Transformers «Attention Is All You Need» [En línea]. Available: <https://arxiv.org/abs/1706.03762> [Último acceso: 29 Mayo 2023].
- [14] Controversia Artistas «Artistas en lucha contra los programas de IA que copian sus estilos» [En línea]. Available: <https://www.pagina12.com.ar/535546-artistas-en-lucha-contra-los-programas-de-ia-que-copian-sus-> [Último acceso: 29 Mayo 2023].
- [15] Controversia Artistas «Con los artistas en pie de guerra contra la inteligencia artificial, ArtStation ha decidido censurar sus protestas» [En línea]. Available: <https://www.genbeta.com/actualidad/artistas-pie-guerra-inteligencia-artificial-artstation-ha-decidido-censurar-sus-protestas> [Último acceso: 29 Mayo 2023].
- [16] Stable Diffusion «Stable Diffusion - Stability AI» [En línea]. Available: <https://stability.ai/stablediffusion> [Último acceso: 14 Junio 2023].
- [17] DALL-E 2 «DALL-E 2 - OpenAI» [En línea]. Available: <https://openai.com/dall-e-2> [Último acceso: 14 Junio 2023].
- [18] Midjourney «Midjourney» [En línea]. Available: <https://www.midjourney.com> [Último acceso: 14 Junio 2023].
- [19] DSM «Denoising Score Matching» [En línea]. Available: <https://www.activeloop.ai/resources/glossary/denoising-score-matching/> [Último acceso: 20 Junio 2023].
- [20] ControlNet «Adding Conditional Control to Text-to-Image Diffusion Models» [En línea]. Available: <https://arxiv.org/abs/2302.05543> [Último acceso: 22 Junio 2023].
- [21] Craiyon «Free AI image generator - Craiyon» [En línea]. Available: <https://www.craiyon.com/> [Último acceso: 18 Junio 2023].

- [22] Adobe Firefly «Generador de Arte con IA - Adobe Firefly» [En línea]. Available: <https://www.adobe.com/sensei/generative-ai/firefly.html> [Último acceso: 18 Junio 2023].
- [23] Hotpot «AI Art Generator - Hotpot» [En línea]. Available: <https://hotpot.ai/art-generator> [Último acceso: 18 Junio 2023].
- [24] NightCafe Creator «Create Stunning AI Art - NightCafe» [En línea]. Available: <https://creator.nightcafe.studio/> [Último acceso: 18 Junio 2023].
- [25] Deep Dream Generator «Welcome to Deep Dream Generator» [En línea]. Available: <https://deepdreamgenerator.com/> [Último acceso: 18 Junio 2023].
- [26] Starry AI «Create Art with AI - Starry AI» [En línea]. Available: <https://starryai.com/> [Último acceso: 18 Junio 2023].
- [27] Generador Imágenes Bing Chat «Crear imágenes desde palabras con IA» [En línea]. Available: <https://www.bing.com/create> [Último acceso: 22 Junio 2023].
- [28] Hugging Face «Hugging Face - The AI community building the future» [En línea]. Available: <https://huggingface.co/> [Último acceso: 22 Junio 2023].
- [29] OpenJourney «prompthero/openjourney - Huggin Face» [En línea]. Available: <https://huggingface.co/prompthero/openjourney> [Último acceso: 21 Junio 2023].
- [30] AUTOMATIC1111 «GitHub - AUTOMATIC1111 / WebUI» [En línea]. Available: <https://github.com/AUTOMATIC1111/stable-diffusion-webui> [Último acceso: 21 Junio 2023].
- [31] Google Colab «Te damos la bienvenida a Colaboratory» [En línea]. Available: <https://colab.research.google.com/?hl=es> [Último acceso: 21 Junio 2023].
- [32] Gradient PaperSpace «Machine Learning made simple.» [En línea]. Available: <https://www.paperspace.com/gradient> [Último acceso: 21 Junio 2023].
- [33] TheLastBen Fast Stable Diffusion «Github - TheLastBen/fast-stable-diffusion» 19 Junio 2021. [En línea]. Available: <https://github.com/TheLastBen/fast-stable-diffusion> [Último acceso: 21 Mayo 23].
- [34] Dreambooth «DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation» [En línea]. Available: <https://dreambooth.github.io/> [Último acceso: 22 Junio 2023].

- [35] PromptHero «PromptHero - Search prompts» [En línea]. Available: <https://prompthero.com/> [Último acceso: 24 Junio 2023].
- [36] Gen-1 «Gen-1: The Next Step Forward for Generative AI» [En línea]. Available: <https://research.runwayml.com/gen1> [Último acceso: 29 Junio 2023].
- [37] Gen-2 «Gen-2: The Next Step Forward for Generative AI» [En línea]. Available: <https://research.runwayml.com/gen2> [Último acceso: 29 Junio 2023].
- [38] Runway «Advancing creativity with artificial intelligence.» [En línea]. <https://research.runwayml.com> [Último acceso: 29 Junio 2023].
- [39] Drag Your Gan «Drag Your GAN: Interactive Point-based Manipulation on the Generative Image Manifold» [En línea]. Available: <https://arxiv.org/abs/2305.10973> [Último acceso: 29 Junio 2023].
- [40] ProlificDreamer «ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation» [En línea]. Available: <https://arxiv.org/abs/2305.16213> [Último acceso: 29 Junio 2023].
- [41] Neuralangelo «Neuralangelo: High-Fidelity Neural Surface Reconstruction» [En línea]. Available: <https://arxiv.org/abs/2306.03092> [Último acceso: 29 Junio 2023].

