



UNIVERSIDAD
DE MÁLAGA



ESCUELA DE INGENIERÍAS INDUSTRIALES

Departamento

EXPRESIÓN GRÁFICA, DISEÑO Y PROYECTOS

Áreas de Conocimiento

EXPRESIÓN GRÁFICA EN LA INGENIERÍA

CIENCIA DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL

DIBUJO (B Arte)

TRABAJO FIN DE GRADO

ESTUDIO Y ANÁLISIS DEL DISEÑO DE ENVASES DE ACEITE EMPLEANDO TÉCNICAS DE INTELIGENCIA ARTIFICIAL

Grado en

Ingeniería en Diseño Industrial y D.P.

Autor: **GAMBERO DEL CID, ÓSCAR**

Tutor: **BLAZQUEZ PARRA, ELIDIA BEATRIZ**

Cotutor: **DOMINGUEZ MERINO, ENRIQUE**

MÁLAGA, Junio de 2023

Índice de Contenido

Índice de Contenido	I
Índice de Figuras	II
Índice de Tablas	IV
Índice de Gráficas	V
Resumen	VI
Definiciones y acrónimos	VII
MEMORIA	1
1 Introducción	1
1.1. Objetivos.....	1
1.2. Antecedentes.....	1
1.3. Marco Actual.....	2
1.4. Alcance.....	3
1.5. Fases del Proyecto.....	3
2 Estudio de mercado	4
3 Análisis de soluciones	8
3.1. Descripción de los diferentes modelos y métodos de entrenamiento.....	8
3.2. Costes por modelo	11
3.3. Métodos y Arquitectura de entrenamiento	11
3.4. Recuperación de datos para entrenamiento	13
3.5. Realización de entrenamientos	15
3.6. Recopilación de resultados y primera impresión	18
4 Estudio de resultados y comparativa	28
4.1. Descripción de métodos de análisis.....	28
4.2. Evaluación de cada método	31
4.3. Relación humano-herramienta	40
5 Análisis de Resultados	43
6 Desarrollo de un caso práctico.....	47
7 Conclusión.....	51
8 Referencias	53
ANEXOS.....	58
1 Anexo I	58
1.1. Imágenes Generadas	58
2 Anexo II	66
2.1. Encuesta.....	66
PLANOS	76



Índice de Figuras

Figura 1. Ejemplos de productos en el mercado. ^{III}	3
Figura 2. Ejemplo de envases de lata. ^{III}	4
Figura 3. Ejemplo de envases de cristal. ^{III}	5
Figura 4. Ejemplo de envases de ceramica. ^{III}	5
Figura 5. Ejemplo de envases de plastico. ^{III}	6
Figura 6. Ejemplo de envases en aerosol. ^{III}	6
Figura 7. Ejemplo de envases de bolsa en caja. ^{III}	7
Figura 8. Arquitectura de Modelo de Difusion con U-Net.....	10
Figura 9. Ejemplo de Arquitectura Dreambooth.....	13
Figura 10. Ejemplos de imágenes obtenidas de Aceite de Jaén.....	14
Figura 11. Conjunto de imágenes de la base de datos ^{II}	15
Figura 12. Imágenes generadas en el entrenamieto con Dreambooth.....	16
Figura 13. Imágenes generadas en el entrenamieto con Fine-Tuning.....	17
Figura 14. Imágenes generadas en el entrenamieto con LoRA.....	17
Figura 15. Ejemplo de relación Step-CFG Scale en el modelo Stable Diffusion v2.1 768 píxeles.	18
Figura 16. Ejemplo de generación según criterios de texto de entrada Dreambooth Stable Diffusion V1.5.....	20
Figura 17. Ejemplo de resultado en una remesa Dreambooth Stable Diffusion V1.5.....	20
Figura 18. Ejemplo de resultado en una remesa Dreambooth Stable Diffusion V2.1 a 512 píxeles.....	21
Figura 19. Ejemplo de generación según criterios de texto de entrada Dreambooth Stable Diffusion V2.1 a 512 píxeles.....	21
Figura 20. Ejemplo de resultado en una remesa Dreambooth Stable Diffusion V2.1 a 768 píxeles.....	22
Figura 21. Ejemplo de generación según criterios de texto de entrada Dreambooth Stable Diffusion V2.1 a 768 píxeles.....	22
Figura 22. Ejemplo de generación según criterios de texto de entrada Dreambooth Stable Diffusion V2.1 a 512 píxeles.....	23
Figura 23. Ejemplo de resultado en una remesa LoRA Stable Diffusion V1.5.....	23
Figura 24. Ejemplo de generación según criterios de texto de entrada GLIDE a 64 píxeles.....	24
Figura 25. Ejemplo de resultado en una remesa GLIDE a 64 píxeles.....	24
Figura 26. Ejemplo de generación según criterios de texto de entrada GLIDE a 256 píxeles.....	25
Figura 27. Ejemplo de resultado en una remesa GLIDE a 256 píxeles.....	25
Figura 28. Ejemplo de generación según criterios de texto de entrada.....	26
Figura 29. Ejemplo de resultado en una remesa DALL·E 2.....	26
Figura 30. Ejemplo de generación según criterios de texto de entrada Midjourney v4.....	27
Figura 31. Ejemplo de resultado en una remesa Midjourney v4.....	27
Figura 32. Selección de imágenes para el desarrollo del formulario.....	30
Figura 33. Imágenes generadas mediante el uso de ControlNet.....	41
Figura 34. Conjunto Global de Imágenes para su evaluación.....	44
Figura 35. Imágen genereada por SD en su versión 2.1 a 768 píxeles seleccionada para caso práctico.....	46
Figura 36. Imágenes realizadas de diseño en Fusion 360 y estilo realista en Blender.....	47
Figura 37. Representacion del diseño conceptual a escala con medidas de orientación.....	48
Figura 38. Imágenes de entrenamiento para el caso práctico.....	49
Figura 39. Imágenes Generadas para el caso práctico.....	50
Figura 40. Conjunto de imágenes Dreambooth Stable Diffusion V1.5.....	58
Figura 41. Conjunto de imágenes Dreambooth Stable Diffusion V2.1 a 512 píxeles.....	59
Figura 42. Conjunto de imágenes Dreambooth Stable Diffusion V2.1 a 768 píxeles.....	60
Figura 43. Conjunto de imágenes LoRA Stable Diffusion V1.5.....	61



Figura 44. Conjunto de imágenes GLIDE a 64 píxeles.....	62
Figura 45. Conjunto de imágenes GLIDE a 256 píxeles.....	63
Figura 46. Conjunto de imágenes DALL·E 2	64
Figura 47. Conjunto de imágenes Midjourney v4.....	65
Figura 48. Selección de imágenes como base del formulario	67

Índice de Tablas

Tabla 1: Suma de las Repetición de Palabras por cada modelo.	31
Tabla 2: Promedio de Coincidencia imagen-texto entre las imágenes de los modelos.....	32
Tabla 3: Promedio de Similitud de Coseno Imagen-Texto entre las imágenes de los modelos. 33	
Tabla 4: FID entre las imágenes de los modelos.	34
Tabla 5: Promedio de Brisque entre las imágenes de los modelos.	35
Tabla 6: Promedio de NIQE entre las imágenes de los modelos.....	36
Tabla 7: Promedio de PIQE entre las imágenes de los modelos.....	36
Tabla 8: Promedio de contraste entre las imágenes de los modelos.....	37
Tabla 9: Promedio de Nitidez entre las imágenes de los modelos.	38
Tabla 10: Promedio de Coincidencia imagen-Texto entre las imágenes de los modelos.	39
Tabla 11: Resultado de las Evaluaciones Métricas.	43
Tabla 12: Resultado del conjunto de evaluaciones.	44
Tabla 13: Resultado del conjunto de evaluaciones ponderado.....	45

Índice de Gráficas

Gráfica 1. Distribución porcentual de la Repetición de Palabras.....	31
Gráfica 2. Distribución del los valores del promedio de CIT.	32
Gráfica 3. Distribución del los valores del promedio de SCIT.	33
Gráfica 4. Distribución de los valores de FID por modelo.....	34
Gráfica 5. Distribución del los valores del promedio de Brisque.....	35
Gráfica 6. Distribución del los valores del promedio de NIQE.	36
Gráfica 7. Distribución del los valores del promedio de PIQE.	37
Gráfica 8. Distribución del los valores del promedio de contraste.	38
Gráfica 9. Distribución del los valores del promedio de nitidez entre todos los modelos.	39
Gráfica 10. Distribución porcentual de las preferencias del encuestado por modelo.	40
Gráfica 11. Representación del rango de edad de los encuestados.	67
Gráfica 12. Género del encuestado.	68
Gráfica 13. Experiencia en el sector envase del encuestado.	68
Gráfica 14. Gráfico del porcentaje de las imágenes preferentes por los encuestados.	69
Gráfica 15. Gráfico según conocimientos de las imágenes preferentes por los encuestados. ...	69
Gráfica 16. Gráfico según género de las imágenes preferentes por los encuestados.	70
Gráfica 17. Valoración total de creatividad entre imágenes.....	70
Gráfica 18. Valoración total de creatividad entre modelos.....	71
Gráfica 19. Valoración total de imagen con el objeto más atractivo.	71
Gráfica 20. Valoración total de modelos con el objeto más atractivo	72
Gráfica 21. Pregunta de opinión de comprade todas las imágenes.	72
Gráfica 22. Pregunta de opinión de compra de todas las imágenes.	73
Gráfica 23. Pregunta calidad del diseño seleccionado.	73
Gráfica 24. Pregunta de categoría del diseño del diseño seleccionado.	74
Gráfica 25. Pregunta sobre posible fallos de diseño del producto seleccionado.	74
Gráfica 26. Pregunta de opinión sobre los métodos de diseño utilizados.	75
Gráfica 27. Opinión sobre el uso de la inteligencia artificial.	75

Resumen

Desde el principio del diseño el diseñador se ha valido de sus manos, su ingenio y un lápiz para transformar una idea en un producto útil que cumpliera una función. Con los nuevos avances en Deep Learning, y redes neuronales convolucionales ha surgido una nueva herramienta en la mesa, los modelos generativos de imágenes a partir de texto. Con el propósito de probar, evaluar y dirigir estos modelos para aplicarlos al diseño de nuevos productos se ha centrado en un campo concreto de producto, los envases de aceites de oliva, un campo bastante amplio y perfecto para el uso de estos modelos, ya que estos solamente son capaces de generar una imagen, no pueden aportar valores de medida o volúmenes del objeto, por ello los envases de aceite son perfectos como punto de partida, un producto cotidiano cuya forma y estética varía según la marca y la forma deseada. Con esto se pretende dar uso a estas nuevas herramientas y darles su espacio de aplicación en el proceso de diseño conceptual de productos en el ámbito de estado del arte en el desarrollo de un producto. En este trabajo se va a estudiar qué modelos existen en el mercado, qué técnicas se pueden usar para guiar a estos modelos y se evaluarán los resultados obtenidos mediante diferentes métodos de evaluación de imágenes generadas por inteligencia artificial y valoraciones obtenidas, con el fin de proponer una metodología y un modelo para la ejecución de nuevos diseños conceptuales.¹

¹Palabras Clave: Inteligencia Artificial, Diseño conceptual, Modelo Generativo y Aceite de Oliva.

Definiciones y acrónimos

- **Dataset:** Conjunto de datos ordenados que se usan para el análisis, investigación o aprendizaje de un tema.
- **Open Source:** Referente a un software cuyo código fuente está abierto al público en general, abierto para modificar y distribuir el código de manera libre y gratuita. En castellano se define como código abierto.
- **Fine-tuning:** Técnica de aprendizaje automático del cual se parte de un modelo pre-entrenado y se realiza un entrenamiento con un conjunto de datos específicos. En castellano se define como Ajuste Fino.
- **Step:** Unidad de acción o cálculo dentro de un proceso con el fin de un propósito. En castellano se define como paso.
- **Iteración:** Ciclo repetitivo en el que se realizan una serie de acciones o cálculos. En cada iteración se ejecutan una serie de steps.
- **Epoch:** Pasada completa por el conjunto de iteraciones necesarias.
- **CFG scale:** Técnica que ajusta y escala los pesos de una red neuronal convolucional para preservar la estabilidad durante el entrenamiento o inferencia.
- **Inferencia:** Etapa en la que se utiliza un modelo entrenado para realizar predicciones o tareas específicas basadas en nuevos datos de entrada.
- **Difusión:** Proceso en el que la información se propaga a través de una red neuronal o sistema de aprendizaje automático, transmitiendo los datos de una capa a otra con el fin de realizar cálculos y genera resultados.
- **Modelo Generativo:** Modelo en inteligencia artificial que tiene la capacidad de crear muestras de datos similares a los aprendidos durante el entrenamiento.
- **U-Net:** Arquitectura de red neuronal convolucional usada en el campo de procesamiento de imágenes con una estructura de codificador-descodificador, que permite extraer características y reconstruir la imagen.
- **Prompt:** Instrucción o texto de entrada para una IA con el fin de generar una respuesta o tarea.
- **CLIP:** Contrastive Language-Image Pre-training.
- **GANs:** Generative Adversarial Networks.
- **GLIDE:** Guided Language to Image Diffusion for Generation and Editing.

- LoRA: Local Residual adaptation.
- SD: Stable Difussion.
- BLIP: Bootstrapping Language-Image Pre-training.
- CFG: Classifier-free guidance.
- FID: Fréchet Inception Distance.
- RP: Repetición de palabras.
- CIT: Coincidencia Imagen-Texto.
- SCIT: Similitud del Coseno Imagen-Texto.

MEMORIA



MEMORIA

1 Introducción

“IA es la capacidad de una máquina de realizar análisis e inferencias a partir de información compleja o incompleta”.

1.1. Objetivos

En este trabajo se busca entrenar un modelo generativo de imágenes por inteligencia artificial mediante modelos de difusión y diferentes arquitecturas de entrenamiento, entrenando estos modelos con la creación de una base de imágenes de botellas de aceite, realizar un estudio de estos modelos y compararlos mediante diferentes métricas de control con el fin de crear nuevos diseños de este envase, realizar una encuesta con una muestra de las imágenes generadas con los modelos adecuados y desarrollar una visión previa de un producto final.

Como ejemplo de uso de esta tecnología se ha decidido realizar un entrenamiento en el campo de los diseños de envases primarios de aceite de oliva, ya que es un sector bastante abierto a nuevos cambios, con un producto bastante maleable y que requiere de una fuerte carga estética y crear una comunicación atractiva para el consumidor. Además, existe gran variedad de diseños en el mercado, diversidad que es perfecta para servir como base de datos del entrenamiento de un modelo. Con esto se plantea mejorar el proceso creativo y optimizar el tiempo de diseño de envases, un campo perfecto para explorar y analizar el uso de la inteligencia artificial en los procesos de diseño.

1.2. Antecedentes

Actualmente existen pocas integraciones de esta tecnología en el desarrollo de diseño industriales, existen ejemplos de uso donde integran la arquitectura de DALL-E 2 al software de Fusión 360 para empezar a diseñar en base a una imagen generada por ese modelo (Lu et al., 2021). Sin embargo, hasta el momento no se ha encontrado ningún ejemplo práctico de entrenamiento de un modelo generativo para el desarrollo de diseños específicos de un sector, imagen o producto.

1.3. Marco Actual

1.3.1. Inteligencia Artificial

En los últimos años están desarrollándose numerosos avances en el campo de la inteligencia artificial, el aprendizaje profundo y en la creación de diferentes modelos generativos que buscan ofrecer nuevas herramientas que faciliten la realización de diferentes labores que requieran un esfuerzo humano. Estos avances están abarcando diferentes campos, desde análisis y optimización de datos, conducción autónoma, modelos de procesamiento de lenguaje natural como GPT-3, que son capaces de redactar texto casi a un nivel de realismo humano, hasta la generación de imágenes. Estas herramientas sirven para mejorar y ampliar la ejecución de diferentes tareas que normalmente están limitadas por la capacidad humanas.

Con la aplicación de modelo generativos de imágenes se busca ampliar las opciones de diseños que pueda generar un diseñador humano, en vez de partir solo de la construcción manual de diez diseños conceptuales, la IA permite generar cientos de diseños al mismo tiempo, pudiendo completar esos diseños o partir de los diseños construidos y desarrollar variaciones o productos similares a los imaginados.

1.3.2. Proceso de diseño

Actualmente, la fase de desarrollo conceptual de un diseño parte de una lluvia de ideas que permitan cubrir una necesidad, posteriormente se selecciona las ideas más aceptables y se realizan diseños conceptuales que cumplan con los requisitos del diseño. Este proceso puede llevar más o menos tiempo dependiendo del diseñador y está condicionado con la creatividad del que está creando la idea, con sus conocimientos, experiencia y prejuicios asociados. Estas características también están asociadas a cómo funciona estos modelos generativos, los cuales están condicionados por la información de entrenamiento, si se le limitan ciertos conceptos o se asocia un concepto con una definición limitada y que solo pueden disminuirse aumentando la información y conocimientos de entrenamiento.

1.3.3. Envases

Los envases en el mercado de hoy en día son bastantes variados y creativos, tanto en forma, diseño y estética, formando desde envases con estructura orgánica y natural a estructuras estilizadas, compuestos por líneas rectas y acabados minimalistas. Los materiales usados son bastante flexibles, desde plástico para diseños más económicos, cerámica para los más tradicionales, hasta el vidrio, el más comúnmente usado entre las botellas de aceite. Esta flexibilidad tanto en forma como en materiales hace que sea un producto fácilmente desarrollable por un modelo generativo de inteligencia artificial.



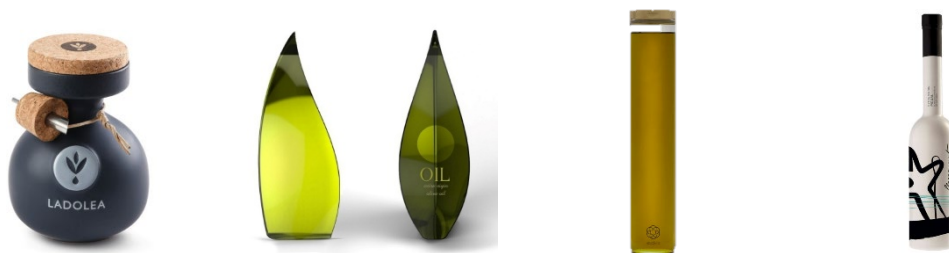


Figura 1. Ejemplos de productos en el mercado.¹¹¹

1.4. Alcance

Este proyecto no busca desarrollar un producto final, busca realizar un estudio y prueba de un nuevo sistema de generación de diseños conceptuales, solo se realizarán pruebas que afecten o intervengan en el estado del arte de la creación de un nuevo envase de aceite de oliva.

Se busca aplicar una nueva herramienta al proceso de diseño que puede facilitar o ampliar las ideas para nuevos diseños, ver su practicidad y llegar a un proceso de diseño en conjunto con estos nuevos modelos generativos.

1.5. Fases del Proyecto

Este proyecto estará separado en diferentes fases, primero un estudio de mercado, seleccionando el contenido que se va a fabricar, un estudio y selección de los diferentes modelos generativos que existen actualmente y se analizarán sus características y distinciones. Después, un estudio y selección de imágenes de envases de aceites de oliva que sirvan para un entrenamiento adecuado y la realización de las condiciones necesarias para un buen entrenamiento. A continuación, se analizarán las diferentes arquitecturas de entrenamiento para los modelos de código abierto y se compararán sus beneficios individuales según diferentes criterios.

Una vez efectuado este estudio y los entrenamientos necesarios y expuestos los resultados, se ejecutará un desarrollo de un caso práctico. Tras realizar los entrenamientos necesarios se tomará una muestra de imágenes de cada modelo, siguiendo criterios que evalúen características similares entre imágenes, y se buscarán diferentes métodos de evaluación que puedan servir para reflejar qué modelo puede ser el más adecuado para un uso práctico. Posteriormente, se desarrollará una encuesta con una muestra de cada modelo que refleje la opinión de los posibles consumidores y muestre su opinión acerca de estos nuevos procesos de diseño, seleccionando en el proceso uno en concreto para ejecutar un modelo visual de una botella de aceite enfocada en la fase del estado del arte del desarrollo de un producto final.



2 Estudio de mercado

El aceite de oliva es un producto altamente demandado y que suelen ocupar gran parte de las ventas de los grandes centros comerciales, siendo uno de los espacios de consumición más demandados junto a productos básicos como el vino o la leche, y entre productores locales de aceite. Este producto es demandado tanto como alimento de consumo básico como producto de regalo y, dependiendo del diseño del producto, como objeto decorativo.

Actualmente en el mercado existen una gran variedad de envases que almacenan aceite de oliva para su venta al público y que están separados según el tipo de aceite que contienen, el volumen que se quiere distribuir, los materiales usados para el envase y función que cumplen. Con estas características podemos encontrar las siguientes categorías de envase de aceite en el mercado según su material:

- **Envases de lata:** Estos envases son populares en términos de funcionalidad y durabilidad. En cuanto a forma y tamaño suelen ser envases cuadriculados y con tamaños de entre los 200 ml hasta grandes envases de 5 litros. Estos envases tienen una forma similar y su estética varía poco entre modelos, siendo un envase prácticamente funcional y con poco énfasis en el diseño visual. Son comunes en la hostelería y restauración ya que son fáciles de apilar y almacenar en grandes cantidades.



Figura 2. Ejemplo de envases de lata.¹¹¹

- **Envases de cristal:** Estos envases son los que más presencia tienen en el mercado, tanto por su durabilidad como por presencia estética. Estos envases suelen ser transparentes y permiten observar el contenido para apreciar el color y calidad del aceite, aunque gran parte de los productos acaban cubriendo el envase con una

cobertura para vender la marca y proteger el producto del sol. Suelen tener formas y diseños variados y se encuentran entre los 25 ml, para botellas de regalo, hasta 1 litro, siendo los más comunes de 500 ml.



Figura 3. Ejemplo de envases de cristal. ^{III}

- Envases de cerámica: Estos envases tienen una estética más rústica y natural, y los hace distinguir del resto de envases no permiten ver el contenido del producto con lo que gran su estética destaca más que en los demás envases, siendo perfecto como regalo o producto gourmet y funcionando como objeto decorativo en los hogares. Estos envases suelen ser rellenables y ser atractivo para aquellos clientes que buscan reutilizar el envase una vez este vacío. Los volúmenes de estos envases están entre los 200 ml y los 750 ml.



Figura 4. Ejemplo de envases de cerámica. ^{III}

- Envases de PET: Estos envases de plástico están pensados para un producto de menor coste, de fácil transporte y más económico que el resto de los envases. Son productos ligeros y fáciles de transportar Existen dos modelos diferentes, en garrafa y en botella de litro, dependiendo del volumen de venta, que se encuentra entre 1 litro y los 5 litros, pensados para el gran consumo.





Figura 5. Ejemplo de envases de plástico.¹¹¹

- Envases en aerosol: Estos envases, normalmente contruidos en aluminio o cristal con acero inoxidable. Son envases que buscan dosificar el aceite en forma de espray para tener un mayor control de la cantidad consumida. Suele venir en envases de 200 ml, ya que gran parte del contenido es gas a presión que permite rociar el aceite con facilidad. Estas botellas son poco comunes en el mercado y su variedad es limitada entre los diferentes productos.



Figura 6. Ejemplo de envases en aerosol.¹¹¹

- Envases de bolsa en caja: Estos envases buscan ser los más sostenibles con el medioambiente, estos almacenan el aceite en una bolsa reciclable y la contienen en una caja de cartón con un grifo para extraer el contenido. Estos envases suelen tener una estética simple y su forma no varía mucho entre productos, siendo poco conocidos en el mercado y centrados a un consumidor preocupado con el medioambiente.



Figura 7. Ejemplo de envases de bolsa en caja. ^{III}

El tipo o variedad de aceite que contiene cada envase determina el precio de venta del producto y el envase suele reflejar la calidad del aceite. Los diseños más atractivos están directamente relacionados con esta calidad y existen diferentes variedades como arbequina, picual, hojiblanca, frantoio, coupage, y muchas más variedades que dependen de la variedad de olivo del que se extrae la aceituna.

Para la realización de este estudio se va a enfocar en envases que lleguen a gran parte de los consumidores como son los envases de cristal o cerámica, buscando una estética más clásica y limpia enfocada a un envase de un aceite de calidad. Preferentemente se buscará un envase rellenable y sostenible con el medioambiente, destinados a una variedad de aceite de alta calidad. Sobre todo se buscarán productos para el entrenamiento con la mayor variedad de forma y estilo puedan representar una botella de aceite de oliva.

3 Análisis de soluciones

Para el desarrollo del estudio se van a tener en cuenta los diferentes modelos de síntesis de imágenes a partir de texto. Para ello hay que tener en cuenta cómo funcionan estos modelos y cuál es la base que comparten, entre estos funcionamientos se comparten varios conceptos como la difusión, el uso de herramientas como CLIP y las redes generativas adversarias.

Todos estos modelos generativos funcionan mediante un sistema de difusión, proceso en el cual a una imagen se agrega ruido gaussiano de forma iterativa y se entrena un modelo para reconstruir el ruido aplicado a la imagen. (Ho et al., n.d.), y usando el sistema U-Net codificando y decodificando cada imagen.

CLIP es un modelo de aprendizaje profundo que mediante redes convolucionales y transformers puede realizar tareas de visión por computadora y procesamiento de lenguaje natural (Radford et al., 2021; Ronneberger et al., 2015).

Las GANs (Generative Adversarial Networks) son una técnica de aprendizaje automático profundo que permite crear nuevos datos a partir de un conjunto existente. Para ello, se utilizan dos redes neuronales: una que genera los datos y otra que los evalúa. La red generadora crea nuevas muestras a partir de un ruido aleatorio y la red discriminadora evalúa si estas muestras son reales o falsas. El objetivo del generador es crear muestras que sean lo suficientemente realistas como para engañar al discriminador.

3.1. Descripción de los diferentes modelos y métodos de entrenamiento

El campo de sintetización de imágenes mediante procesos de inteligencia artificial ha sido procesado mediante diferentes arquitecturas, comenzando por el primer modelo que fue capaz de generar imágenes a través de texto que sería el modelo VQGan+CLIP el cual trabajaba mediante la sintetización del espacio latente hasta coincidir con la descripción guiándose por el modelo de CLIP y sus posteriores modelos como serían GLIDE, modelo que se desarrolló hasta convertirse en la primera versión de DALL·E 2, el modelo de Midjourney, la segunda versión de DALL·E 2, y el modelo de Stable Diffusion.

Estos modelos hay que diferenciarlos según su arquitectura, dimensión de salida, herramientas de control en el proceso de creación y accesibilidad, ya que con modelos de código cerrados no se puede realizar un entrenamiento.



3.1.1. Modelos de Código abierto:

Estos modelos están completamente disponibles para ser ejecutados, editados y transformados, su código es completamente abierto y permite realizar refinamientos a conceptos específicos.

- **GLIDE:** Fue uno de los primeros modelos de código libre que se crearon después del desarrollo de la arquitectura de VQGAN+Clip por la empresa de OpenAI, entrenado en 3.5 millones de parámetros, ofreciendo la capacidad de generar imágenes a partir de texto, realizar inpainting, rellenar huecos demarcados con una máscara de una imagen, y el poder entrenar el modelo con un banco de imágenes propio.

Su resolución es muy limitada, con unas dimensiones de 64x64 píxeles, en el caso de GLIDE se aplica el método de difusión en cascada, en el cual se utiliza varios modelos de difusión en serie para aumentar la resolución de la imagen (Ho et al., 2021), este segundo modelo también se entrena con la misma base de datos y permite agrandar la imagen a una dimensión de 256x256. Este segundo modelo ha sido entrenado con 1.5 billones de parámetros con imágenes de esa resolución, en la práctica no es el modelo más eficiente y en gran parte solo consigue añadir ruido, pero sirve para comparar como estos modelos han ido evolucionando en comparación con los modelos más recientes.

Los desarrolladores de este modelo establecieron dos estrategias de guía en la creación de las imágenes, una usando CLIP como clasificador, atendiendo a etiquetas de clasificación que se usan durante el proceso de muestreo de difusión y otra estrategia sin ningún clasificador. Esta segunda tuvo mayor preferencia entre los evaluadores humanos al tener mayor realismo y mayor similitud con el texto de entrada, versión con lo cual se entrenará el modelo en el estudio (Nichol et al., 2021).

- **Stable Diffusion:** Es el modelo que más repercusión ha tenido entre estos sistemas generativos y el que ha tenido más apoyo de progreso por la comunidad desarrolladora. Con este modelo se ha podido adaptar numerosas aplicaciones y adaptaciones para mejorar su funcionamiento y ofrecer diferentes maneras de generar imágenes, ya sea a partir de texto, partiendo de otra imagen, una estructura o mapas de profundidad.

Existen diferentes modelos base de Stable Diffusion creados por la empresa Stability AI en colaboración con Runway y cada uno con diferencias significativas y con una resolución base de 64x64 píxeles, pasando por su resolución más estable 512x512 píxeles y hasta 1024x1024 píxeles.



La versión 1.4 es la versión original del modelo, fue entrenada usando CLIP ViT-L/14 text y el dataset de código abierto de LION 2.3B, es decir, 2.3 billones de parámetros, usando un 860M U-Net, con una resolución de imágenes de entrenamiento de 512x512 píxeles, y la cual se mejoró con un mayor tiempo de entrenamiento y con un mayor dataset para crear la versión 1.5.

Posteriormente se creó la versión 2.1 que reentrenaba desde cero el modelo, eliminando imágenes de artistas, personas relevantes o imágenes con derechos de autor y usando OpenClip, una versión de CLIP entrenado con un conjunto de datos públicos de LAION 5B, este conjunto presenta un cambio en el proceso de creación del texto de entrada, ya que requiere el uso de más tokens descriptivos. Esta versión ha sido entrenada usando 2B U-Net y realizando dos modelos diferentes, uno entrenado con imágenes de resolución 512x512 y otro con imágenes de una resolución de 768x768 píxeles (Borji, 2022; Ramesh et al., 2022).

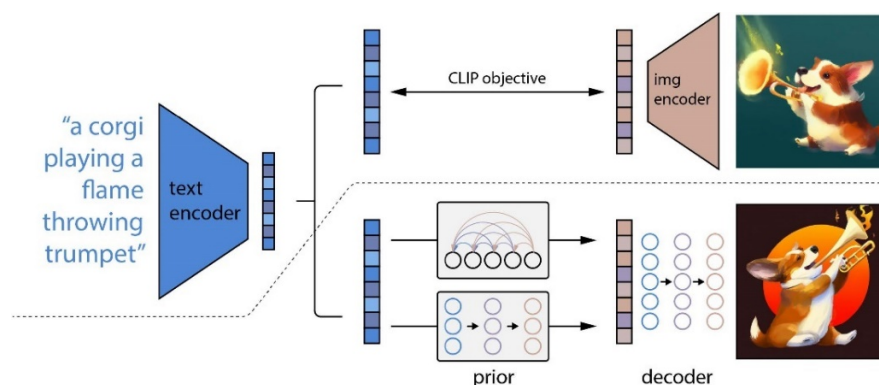


Figura 8. Arquitectura de Modelo de Difusion con U-Net².

3.1.2. Modelos Cerrados:

Estos modelos no son de código libre y sólo pueden ser usados a través de los servicios de la empresa que los controla, aunque son cerrados, pueden llegar a ofrecer opciones para condicionar el modelo y llegar a generar imágenes de alta calidad y merece la pena tener en cuenta sus posibilidades ya que se pueden combinar con los otros modelos y editar una imagen creada en ellos, dándole el estilo de un modelo pre-entrenado. Entre estos modelos destacan Midjourney y DALL·E 2 que están siendo comercializados y están abiertos a usarlos a través de la API de sus empresas correspondientes y otros modelos que aún no están abiertos al público como sería IMAGEN y Parti de Google.

² Adaptado de "Hierarchical Text-Conditional Image Generation with CLIP Latents" (P.3) por A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, M. Chen, (2022). <http://arxiv.org/abs/2204.06125>.

- DALL·E 2: El modelo de DALL·E 2 es la evolución de GLIDE a un modelo de mayor escala, es capaz de generar imágenes de 1024 píxeles, pero a diferencia de GLIDE su uso es cerrado a través de los servidores de OpenAI, y permite generar variaciones de una imagen, rellenar espacios de una imagen o extender su contenido además de generar a partir de texto. Es un modelo entrenado con 12 mil millones de parámetros y realizado de forma multimodal con GPT-3. Este modelo está entrenado en imágenes más artísticas y, en general, suele realizar imágenes más relacionadas con pinturas o dibujos que imágenes fotorrealistas.
- Midjourney versión 4: Este modelo creado por la empresa homónima funciona a través de un servidor de Discord, ofrece numerosos parámetros para ajustar las dimensiones de la imagen y ofrece diferentes modelos de estilo, pero ninguno centrado a la creación de productos.

Por defecto genera 4 imágenes con un tamaño de 512x512 píxeles, con la posibilidad de ampliar una imagen a 1024 píxeles o generar variaciones de la misma imagen, así como darle una imagen base de guía para la generación de la imagen. Sobre la arquitectura y entrenamiento del modelo no se ha ofrecido información por parte de la compañía, pero el modelo de Stable Diffusion fue implementado en el desarrollo de la versión 4 de este modelo.

3.2. Costes por modelo

Cada modelo tiene un coste, tanto de procesamiento como económico, dependiendo del modelo que se esté usando, los modelos cerrados tienen unos costes empresariales por la creación de cada imagen y van ligadas a un coste económico, el modelo de Midjourney tiene como plan de suscripción más barato 8€/mes y permite generar alrededor de 200 imágenes y el modelo de DALL·E 2 tienen como menor coste 15€ que permite obtener 115 créditos, siendo cada crédito equivalente a la generación de 4 imágenes con un mismo texto de entrada o tres variaciones de una imagen.

Los modelos de Stable Diffusion requieren de una cantidad de procesamiento relativamente asequible, requiriendo como mínimo 6 Gb de memoria gráfica dedicada para un funcionamiento adecuado, pudiendo generar una imagen de 512x512 píxeles en un tiempo de alrededor de los 15 segundos.

3.3. Métodos y Arquitectura de entrenamiento

- LoRA: Se basa en un sistema de entrenamiento mediante cambios locales en las características con el fin de reproducir cambios globales en las imágenes. El método consiste en simplificar los pesos del modelo usando una técnica de descomposición.

Esto implica expresar cada peso como el producto de dos matrices más pequeñas, una de las cuales se mantiene constante y la otra se entrena con los datos objetivo. Esta estrategia reduce la cantidad de parámetros que se deben entrenar y evita que el modelo se ajuste demasiado a los datos. Además, se agrega una conexión especial entre la entrada y la salida de cada capa para preservar la información original y facilitar el aprendizaje de cambios locales (Hu et al., 2021). En términos de espacio, el modelo generado por LoRA ocupa entre los 140Mb a los 300Mb, siendo una adaptación relativamente ligera y que permite ser utilizada en diferentes modelos base o entrenados y sus requerimientos de procesamiento pueden llegar a solo necesitar 8GB de VRAM, dependiendo de la adaptación que se esté utilizando.

- Dreambooth: Es otro método de adaptación de modelos de texto a imagen, que se basa en el uso de una capa adicional llamada DreamBooth Layer (DBL) que se coloca entre la entrada de texto y la salida de imagen del modelo pre-entrenado. El DBL se entrena con un conjunto de imágenes objetivo y aprende a modificar la distribución de probabilidad del modelo base para generar imágenes más fieles al objetivo. Para ello, el DBL utiliza dos técnicas: una es el uso de identificadores únicos para cada sujeto, que se añaden al texto de entrada y permiten asociar el sujeto con una región específica del espacio latente; y la otra es el uso de una pérdida llamada Autogenous Class-specific Prior Preservation (ACPP), que consiste en minimizar la divergencia KL, que es la pérdida de datos en la estimación, entre la distribución del modelo base y la del modelo adaptado para cada clase (Ruiz et al., 2022). A diferencia de LoRA, Dreambooth no modifica los pesos del modelo base, sino que añade una capa extra que introduce latencia en la inferencia. Además, Dreambooth requiere más imágenes para entrenar el DBL que LoRA para las matrices de rango bajo. Este entrenamiento necesita como mínimo usar 12 GB de memoria VRAM y 16 GB y memoria RAM.

Este método reescribe el modelo y ocupa el mismo espacio que el modelo original.

- Fine-tuning: El método de fine-tuning es una técnica que consiste en adaptar un modelo pre-entrenado para un problema específico, aprovechando el conocimiento que ya tiene el modelo sobre el dominio general. El fine-tuning se realiza modificando algunos parámetros del modelo, como los pesos de las capas finales o la ratio de aprendizaje, y entrenando el modelo con un conjunto de datos del problema objetivo. De esta forma, se puede mejorar el rendimiento del modelo y obtener resultados más precisos y personalizados.



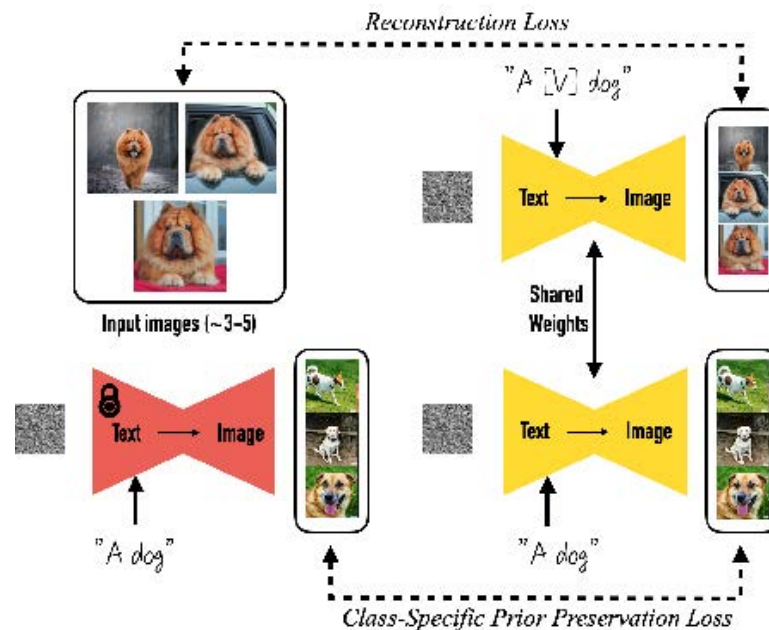


Figura 9. Ejemplo de Arquitectura Dreambooth³.

3.4. Recuperación de datos para entrenamiento

Para realizar un entrenamiento es necesario crear una base de imágenes que sirvan como instrucción de lo que se quiere crear, para ello las imágenes tienen que ser claras y nítidas, cuanto mayor detalle y calidad de la imagen mejor será el entrenamiento y mejores resultados generará. Para poder entender mejor lo que se quiere entrenar es recomendable que el objeto esté en un fondo aislado y que únicamente se vea el objeto. En el caso de querer entrenar un objeto en concreto es más aconsejable variar lo máximo posible el fondo de la imagen.

Se busca encontrar imágenes con un fondo monocromático o transparente, que definiera la forma de la botella. Teniendo en cuenta el estudio de mercado las botellas buscadas tienen que ser de un carácter premium, relacionadas con aceites de alta calidad y compuesto por materiales como el vidrio y la cerámica. El objetivo es obtener la mayor variedad de formas posibles para las botellas de aceite, evitando duplicidades y similitudes entre ellas. Para aumentar la diversidad, se utilizarán imágenes con dos o tres botellas que faciliten la comprensión de lo que se pretende crear, no tanto como elemento único sino como concepto. En cuanto a la cantidad de imágenes necesarias, esto depende del método de entrenamiento, por lo general se requiere un mínimo de 20 imágenes para crear un modelo que se acerque lo suficiente al concepto que

³ Adaptado de "DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation" (p.4), por N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, K. Aberman, (2022). <http://arxiv.org/abs/2208.12242>.

se quiere recrear. En este caso y para ampliar la variedad y diversidad se utilizarán entre 50 y 100 imágenes. Se realizarán pruebas partiendo de la menor cantidad de imágenes posibles y cuando se llegue a un punto de equilibrio, cuando el modelo seleccionado como base de prueba sea capaz de generar las imágenes deseadas, se establecerá el número de imágenes adecuadas para este concepto de envase. Es importante tener en cuenta que a medida que se aumenta la cantidad de imágenes el modelo requerirá más steps de procesamiento, por eso se busca el menor número de imágenes de entrenamiento posible.

Para la búsqueda de imágenes se ha buscado en las grandes bases de imágenes como LAION, COCO, Imagnet y Open Image. En estas bases se ha buscado imágenes de botellas de aceite de oliva y se han encontrado pocas o ninguna imagen, lo cual es interesante, ya que modelos como Stable Diffusion están entrenados con las bases de LAION, lo cual indica que es necesario hacer un refinamiento para poder conseguir imágenes cercanas a un envase de aceite. Ante la falta de una base de imágenes con contenido de aceite de oliva, la otra opción es recopilar manualmente imágenes de botellas de internet. Este conjunto de imágenes ha sido posible gracias a la página de *Aceite de Jaén*, que tiene un gran catálogo con imágenes de las botellas que pone a la venta y con diseños variados, fondos en blanco y productos visualmente bien definidos como se puede ver en la Figura 10. Además, se incluirán imágenes con diseños diferentes para añadir variedad al modelo o incorporar imágenes de envases clásicos que aporten más fondo y contexto.



Figura 10. Ejemplos de imágenes obtenidas de Aceite de Jaén⁴.

⁴ Adaptado de “Aceite de Jaén Online S.L.”, Aceites de oliva virgen extra [Imagen], <https://www.aceitedejaen.es/#premium>.



Figura 11. Conjunto de imágenes de la base de datos //

3.5. Realización de entrenamientos

3.5.1. Entrenamiento con Dreambooth

Para empezar con un entrenamiento y establecer unos estándares para utilizar en el resto de entrenamiento se empezará usando el modelo que mejor recepción tiene entre los usuarios y que es más estable, que es el modelo de Stable Diffusion en su versión 1.5.

Se empezó realizando un entrenamiento con un dataset de 62 imágenes para establecer un baremo y comprobar si es necesario añadir imágenes por falta de precisión o quitar imágenes por sobreajuste. Para ello se ha usado el repositorio de TheLastBen/fast-stable-diffusion el cual esta optimizado para funcionar con bajos recursos de procesamiento y una mejor optimización

que el repositorio original, así como ofrecer una interfaz gráfica y diferentes extensiones y complementos.

Para entrenar el modelo solo es necesario dar la base de imágenes y establecer un token de reconocimiento que definan el modelo a la hora de escribir el input. El token utilizado en este caso ha sido “Aceitfg”. Este token no puede hacer referencia a ninguna palabra existente, ya que reemplazaría ese concepto y generaría imágenes mezclando ambos. En este caso no es necesario añadir la etiqueta de cada imagen y funciona sin la necesidad de definir ninguna característica de lo que se está entrenando. Lo siguiente sería establecer el número de steps necesarios, el creador del repositorio recomienda usar 1000 steps por cada 10 imágenes, con lo cual sería necesario realizar 6200 steps de entrenamiento. Con esta cantidad de entrenamiento, cuyo tiempo de ejecución se realizó en poco más de hora y veinte minutos, se consiguió construir un modelo capaz de generar imágenes de alta calidad y cuyos objetos generados se asemejaban bastante a la idea y concepto de lo que es un envase de aceite, con un gran margen de creatividad generando imágenes muy variadas entre sí y que no reflejan ningún síntoma de sobreentrenamiento. Con esta base se continuó entrenando los modelos de la versión 2.1 y usando las imágenes de entrenamiento con una dimensión de 768 para su correspondiente modelo.



Figura 12. Imágenes generadas en el entrenamiento con Dreambooth.

3.5.2. Entrenamiento Fine-tuning con GLIDE

El entrenamiento para el modelo GLIDE se realizó usando el repositorio de [afiaka87/glide-finetune](#) y estableciendo las dimensiones de las imágenes de entrenamiento a una dimensión de 64 píxeles. En esta situación se ha realizado un entrenamiento con 100 epochs y 40 epochs para el modelo del up sampler. Para este entrenamiento es necesario tener junto a cada imagen la descripción de esta, en concreto a cada imagen se le dio la definición de “An olive oil bottle” para que cuando se le dé el mismo texto de entrada genere imágenes basadas en ese entrenamiento. Éste genera los dos modelos de alrededor de 1.59 Gb cada uno y la calidad obtenida es relativamente correcta en la versión de 64 píxeles, pero al realizar el up sampler con el segundo

modelo se empeora la imagen añadiendo una gran cantidad de ruido, con lo cual cada modelo será evaluado de forma independiente con respecto al resto de modelos.

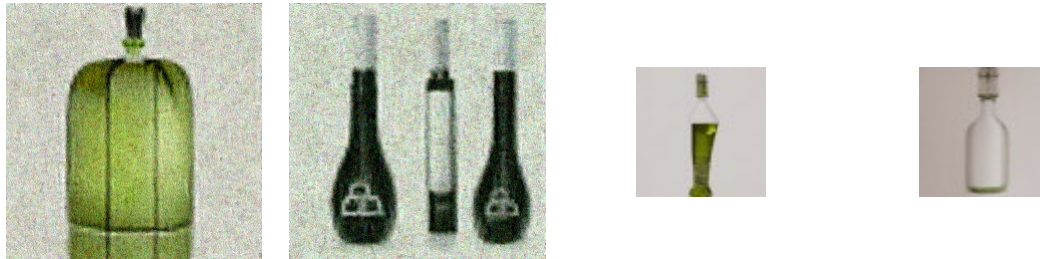


Figura 13. Imágenes generadas en el entrenamiento con Fine-Tuning.

3.5.3. Entrenamiento con LoRA

Para el desarrollo del entrenamiento de LoRA se entrenará el modelo Stable Diffusion V1.5 para comprobar entre sí qué modelo resulta ser más eficiente y con una calidad en principio más cercana a las imágenes de entrenamiento y con mejor entrenamiento. En este caso se realizarán 100 steps por cada imagen, es decir 6200 steps como en el entrenamiento de Dreambooth y se usará el mismo token que define al modelo (Aceitfg) en el nombre de cada imagen, así como archivos .txt asociados a cada imagen con la definición de “Aceitfg an olive oil bottle” El resultado de este entrenamiento es un modelo de 144.51 MB que se puede usar en cualquier modelo de Stable Diffusion.

En total se han realizado 3 entrenamientos diferentes generando 6 modelos entrenados finales preparados para realizar pruebas y evaluaciones y compararse con los modelos de código cerrado.



Figura 14. Imágenes generadas en el entrenamiento con LoRA.

3.6. Recopilación de resultados y primera impresión

Una vez entrenados los modelos necesarios se ha realizado una recolección de muestras, realizando inferencias a cada modelo, tanto los de código libre como los cerrados, para posteriormente realizar un estudio comparativo y determinar qué modelo sería más favorable para la realización de este proyecto.

Para ello se han tenido en cuenta diferentes parámetros: en los modelos entrenados con opciones de control se han mantenido los valores de steps a 40, a partir de los 25 steps el ruido de la imagen ya ha formado una forma y a partir de 50 la imagen no varía mucho con las anteriores, se usarán semillas aleatorias en cada imagen y el CFG scale a un valor de 7, El valor de CFG establece el peso que tiene el texto sobre la imagen, técnicamente a mayores valores más se acerca a la descripción, pero cuanto más alto es, más probable es que la imagen se corrompa. Un ejemplo de cómo funcionan los valores de steps y CFG en conjunto sería en la Figura 15 donde se utiliza una misma semilla como base.

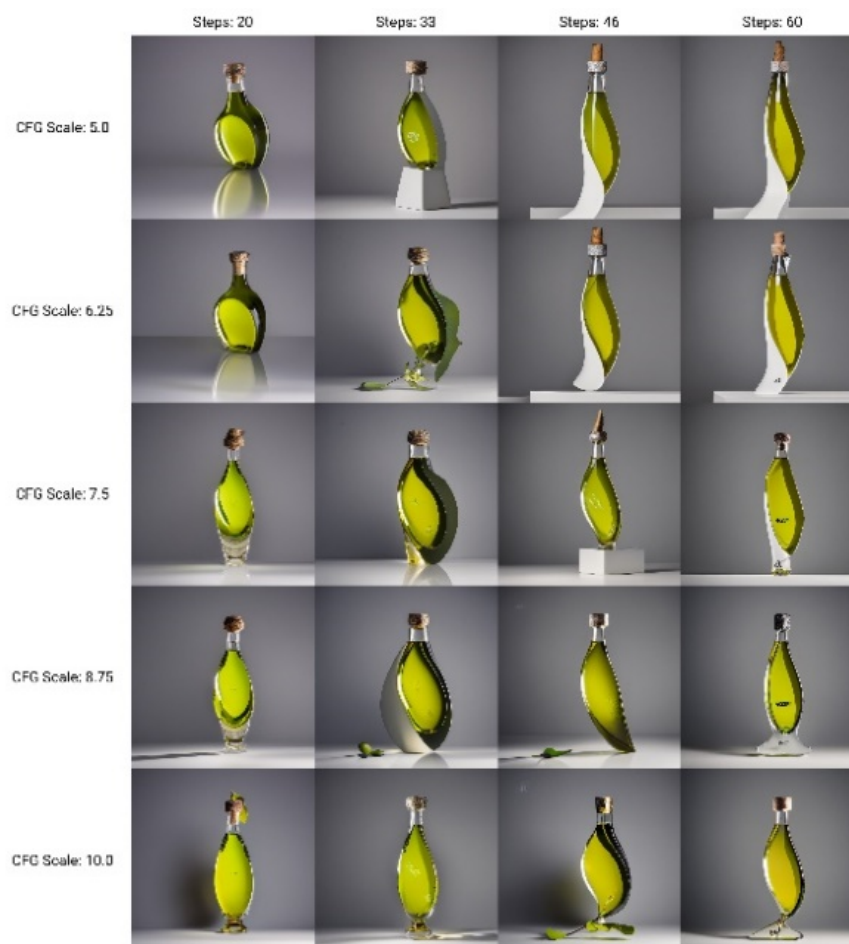


Figura 15. Ejemplo de relación Step-CFG Scale en el modelo Stable Diffusion v2.1 768 píxeles.

En los textos de entrada de todos los modelos se ha buscado establecer los mismos conceptos, adaptando a cada modelo según sus necesidades y esperando obtener un resultado acorde a la descripción, estos textos de entrada, en su forma más simple, los siguientes:

- “An olive oil bottle in the style of Apple computer”.
- “An olive oil bottle in the shape of a leaf”.
- “An olive oil bottle in a curved shape”.
- “An olive oil bottle in a futuristic style”.

Con estas descripciones se busca comprobar que conocimiento tienen acerca de los estilos de diseños existentes y ver si es capaz de adaptar la generación a una forma en concreto, elementos que cualquier diseñador necesita aplicar en la creación de un diseño.

En total se extraerá una muestra de 62 imágenes, misma cantidad de imágenes usadas en el entrenamiento, para evaluar y determinar qué modelo es el más conveniente para desarrollar nuevos diseños conceptuales de envases de aceite y usando el mismo entorno de ejecución, una GPU Tesla T4 con 16 GB de RAM.

3.6.1. Resultados Dreambooth Stable Diffusion V1.5

Este modelo entrenado es sin duda el modelo que mejor interacción tiene con el diseñador, no solo es capaz de adecuarse bastante a la descripción dada, sino que la cantidad de imágenes que son útiles por cada remesa de generación es bastante alta y las imágenes resultantes son de alta calidad, tanto visual como compositivamente hablando, mucha variedad entre imágenes y no presenta síntomas de sobre-entrenamiento. Este modelo funciona con un texto de entrada simple, no necesita de muchos detalles para poder generar una imagen de calidad. En la Figura 16 se puede comprobar como el modelo es capaz de adecuarse al texto de entrada fácilmente y es capaz de desarrollar imágenes con estilo de Apple en la imagen 1, con forma de hoja de olivo en la imagen 2, con un diseño curvo en la imagen 3 y con un diseño futurista en la imagen 4. Este mismo orden se repite en los ejemplos de cada modelo.



Figura 16. Ejemplo de generación según criterios de texto de entrada Dreambooth Stable Diffusion V1.5.

El tiempo de generación por una imagen con 40 steps es de 6 segundos, con una velocidad promedio de 5.85 it/s y usando modificadores como ControlNet es posible guiar la imagen que se quiere crear usando dibujos, formas, mapas de profundidad o imágenes



Figura 17. Ejemplo de resultado en una remesa Dreambooth Stable Diffusion V1.5.

3.6.2. Resultados Dreambooth Stable Diffusion V2.1 a 512 píxeles

La versión 2.1 en 512 píxeles puede generar una imagen en un tiempo medio de 6 segundo, con un promedio de 6.7 it/s, este modelo no consigue generar mucha variedad entre imágenes y tiende a realizar las imágenes con un estilo en concreto, aun especificando no querer ese estilo en los prompts negativos. Consigue generar buenos detalles en los objetos y en ciertos resultados suele generar un buen producto.



Figura 18. Ejemplo de resultado en una remesa Dreambooth Stable Diffusion V2.1 a 512 píxeles.



Figura 19. Ejemplo de generación según criterios de texto de entrada Dreambooth Stable Diffusion V2.1 a 768 píxeles.

3.6.3. Resultados Dreambooth Stable Diffusion V2.1 a 768 píxeles

Este modelo tarda en generar una imagen 17 segundos con una media de 2.30 it/s este modelo genera mejores imágenes que su versión de 512 píxeles, genera con imágenes con mejor coherencia, aunque tiende a realizar imágenes con un contraste fuerte, con generaciones bastante variadas sus tasas de éxito por cada 10 imágenes relativamente bajas, dando resultados que no son válidos para el desarrollo de un producto y que se aleja de la descripción y entrenamiento.



Figura 20. Ejemplo de resultado en una remesa Dreambooth Stable Diffusion V2.1 a 768 píxeles.



Figura 21. Ejemplo de generación según criterios de texto de entrada Dreambooth Stable Diffusion V2.1 a 768 píxeles.

3.6.4. Resultados LoRA Stable Diffusion V1.5

Los resultados del entrenamiento mediante LoRA comparten ciertas similitudes con el de Dreambooth, aunque su precisión es un poco inferior, pero consigue resultados igual de creativos con la misma calidad que el modelo de Dreambooth, respondiendo de una manera igual de precisa claramente la ventaja de este modelo es la portabilidad, la posibilidad de usar este entrenamiento dentro de otros modelos. Esta posibilidad abre la puerta a entrenar un modelo con imágenes de productos de una empresa y combinarlo para generar envases con el estilo de diseño de la empresa.

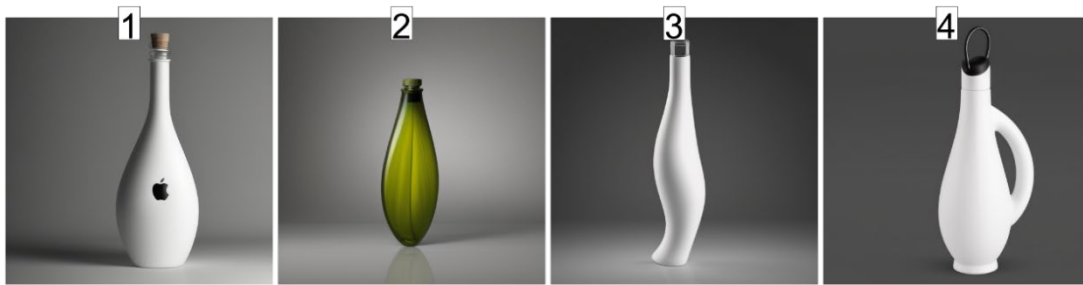


Figura 22. Ejemplo de generación según criterios de texto de entrada Dreambooth Stable Diffusion V2.1 a 512 píxeles.

El tiempo de generación es el mismo que en el modelo entrenado con Dreambooth y para obtener unas imágenes decentes hay que asignarle un peso de 0.7 al modelo, ya que con un peso de 1 las imágenes generadas se empiezan a parecer demasiado a las de entrenamiento.



Figura 23. Ejemplo de resultado en una remesa LoRA Stable Diffusion V1.5.

3.6.5. Resultados GLIDE a 64 píxeles

Los resultados de este entrenamiento son bastante interesantes ya que las imágenes generadas aun siendo de unas dimensiones bastante reducidas consiguen definir ciertas formas, no consiguen adecuarse totalmente a la descripción como se puede ver en la Figura 24 y la cantidad aprovechable de imágenes por cada remesa es aceptable, a pesar de no ser capaz de definir detalles, los tiempos de ejecución por cada 100 pasos es de 8 segundos por imagen alrededor de las 11.87 it/s.

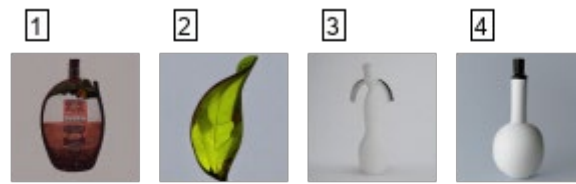


Figura 24. Ejemplo de generación según criterios de texto de entrada GLIDE a 64 píxeles.

El modelo se ha evaluado usando una escala CFG de 7 y 100 pasos.

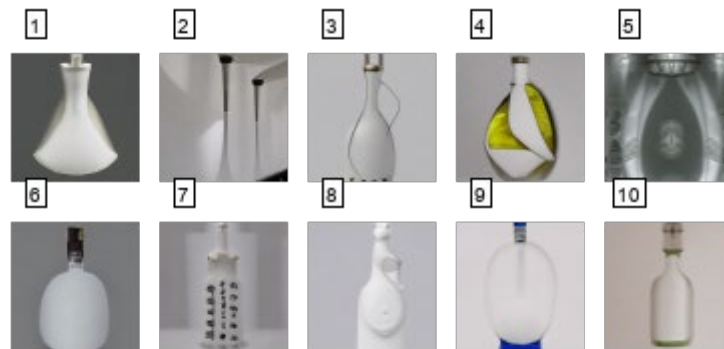


Figura 25. Ejemplo de resultado en una remesa GLIDE a 64 píxeles.

3.6.6. Resultados GLIDE a 256 píxeles

Este modelo, que básicamente aumenta las dimensiones de la imagen generada por la versión de GLIDE a 64 píxeles consigue aumentar el tamaño a unas dimensiones de 256 píxeles, pero los resultados tras el entrenamiento parecen solo añadir ruido sin transformarlo en información nueva, ni aun aumentando el número de pasos, usando 27 steps el tiempo de generación es de 40 segundos por imagen, rondando los 5.89 it/s.

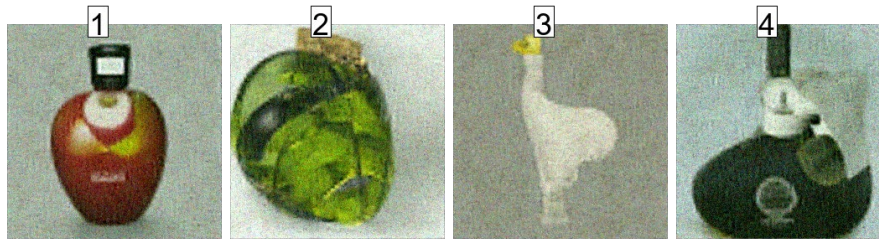


Figura 26. Ejemplo de generación según criterios de texto de entrada GLIDE a 256 píxeles.

El modelo se ha evaluado usando una escala CFG de 7 y 40 pasos.

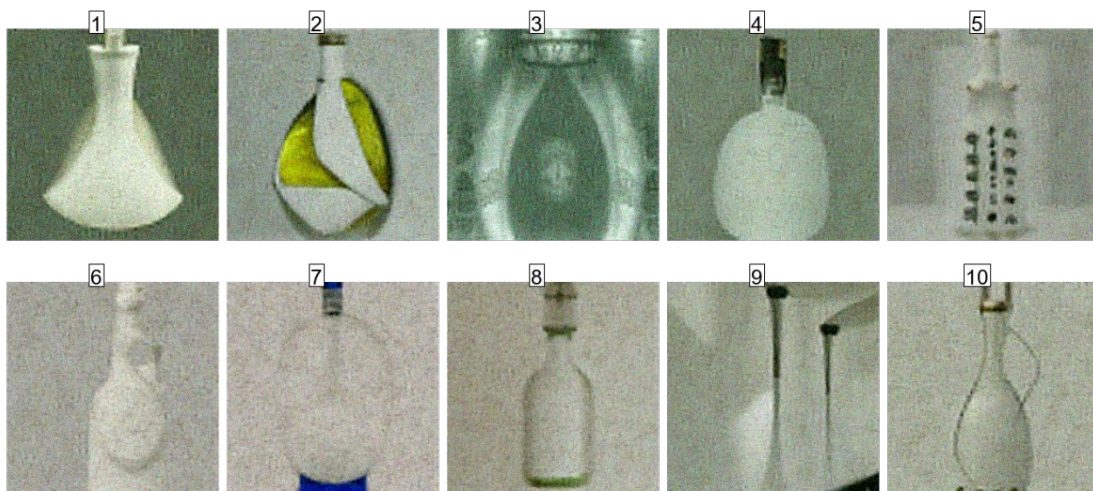


Figura 27. Ejemplo de resultado en una remesa GLIDE a 256 píxeles.

3.6.7. Resultados DALL·E 2

Los resultados de este modelo, a pesar de generar imágenes con mayores dimensiones, no es completamente capaz de generar una imagen correcta de una botella de aceite. Tiene tendencia a generar botellas simples de cristal, aun especificando que tenga cierta cobertura o color en concreto, o genera envases que serían más cercanos a jarrones o elementos decorativos y en caso de la remesa de la Figura 29 se puede ver como no consigue establecer la forma de hoja y simplemente coloca ramas de olivo.



Figura 28. Ejemplo de generación según criterios de texto de entrada

DALL-E 2.

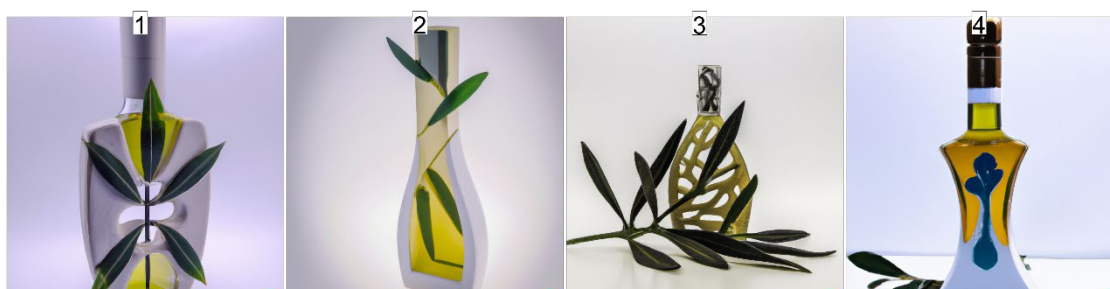


Figura 29. Ejemplo de resultado en una remesa DALL-E 2.

3.6.8. Resultados Midjourney v4

El modelo de Midjourney en su cuarta versión es capaz de generar imágenes bastante creativas y con un estilo bastante marcado. Las descripciones se adaptan en la mayoría de los casos a la imagen, aunque no siempre es capaz de generar una forma o estilo en concreto. En la Figura 31 se puede comprobar cómo le cuesta definir esa forma de hoja y tiende a generar ramas de olivo, como el modelo de DALL-E 2. El tiempo de generación para una remesa de 4 imágenes es de entre 50 y 60 segundos, aunque en caso de tener otro plan de pago los tiempos de duración son menores.



Figura 30. Ejemplo de generación según criterios de texto de entrada Midjourney v4.



Figura 31. Ejemplo de resultado en una remesa Midjourney v4.

4 Estudio de resultados y comparativa

Con este estudio no se busca establecer que modelo es, en general, el que mejor imagen realiza o el que mejor calidad de imagen tiene, sino ver que modelo es más favorable para el desarrollo de un estado del arte de envases de aceite, con lo cual las conclusiones de este estudio pueden ser variables según lo que se pretenda generar y puede haber más de un modelo favorable para el desarrollo de un estado del arte.

4.1. Descripción de métodos de análisis

Se van a describir una serie de métodos de evaluación, de los cuales posteriormente se extraerán una serie de valores representativos de la calidad de cada modelo y se ponderarán para establecer los modelos más favorables,

4.1.1. Relación imagen-texto

En este análisis se va a usar la herramienta de BLIP, que es un sistema de visión por lenguaje, similar a CLIP, pre-entrenado para realizar tareas de comprensión y generación visión-lenguaje. Esta herramienta permite realizar títulos de descripción de imágenes, y realizar respuestas a preguntas visuales y la comprobación de la coincidencia imagen-texto entre un texto de entrada descriptivo y una imagen (Li et al., 2022).

Este sistema permite realizar 3 evaluaciones diferentes:

- La primera realizando la captación de todas las imágenes generadas y comprobando que palabras son las que más se repiten y que están más relacionadas con las botellas de aceite y que palabras hay que se alejen de ese concepto (RP).
- La segunda es calculando la probabilidad de que la descripción “An olive oil bottle” se adecue con cada imagen comprobando la coincidencia entre imagen y texto (CIT).
- La tercera calculando la similitud del coseno entre el vector de texto, usado en la descripción del CIT, y el vector de imagen en el espacio latente, pudiendo evaluar la relación semántica entre ambos vectores (SCIT).

4.1.2. Métrica con referencia

En este apartado de evaluación se va a realizar evaluaciones mediante métricas que tienen en cuenta el dataset original como referencia de comparación, estas métricas funcionan cuando se evalúan con la base de datos que se ha usado de entrenamiento, en los casos donde no es

posible realizar ese entrenamiento sirve simplemente como una posible referencia con la realidad:

- FID: Es una medida de distancia entre dos distribuciones de imágenes, en este caso la distribución del conjunto de entrenamiento y los conjuntos de imágenes generadas por cada modelo, y permite capturar la similitud de las imágenes generadas con las reales. Es la métrica más usada para la evaluación de la calidad de los modelos generativos y las redes adversarias generativas (GANs) y suele dar resultados más precisos cuando los conjuntos de imágenes son de tamaños considerables (Heusel et al., 2017).

4.1.3. Métrica sin referencia

- **Brisque:** Blind/Referenceless image spatial quality evaluator es una métrica de evaluación de calidad de imagen que usa estadísticas de escenas naturales, y permite cuantificar las pérdidas de naturalidad. Estas estadísticas son patrones que se encuentran en las imágenes naturales como los valores de luminancia, intensidad por unidad de área iluminada, o contraste de los píxeles de una imagen, que en una imagen natural sigue una distribución gaussiana y en una imagen distorsionada o de calidad inferior no siguen esa curva. El puntaje de esta métrica se base en la distribución gaussiana estos coeficientes. En el proceso de entrenamiento también se usan valores de opinión, con puntuaciones subjetivas humanas de las imágenes (Mittal et al., 2012).
- **NIQE:** Naturalness image quality evaluator funciona en con una base similar a la de brisque, es una métrica de evaluación que hace uso de desviaciones medibles de regularidad estadísticas que se observan en imágenes naturales, básicamente este modelo esta entrenado de manera que compara las imágenes naturales con imágenes distorsionadas, compara las características y devuelve un valor, cuanto menor sea el valor, más se acerca a una imagen natural. Esta métrica, a diferencia de Brisque no tiene en cuenta la opinión (Mittal, Soundararajan, et al., 2012).
- **PIQE:** Perception-based image quality evaluator Esta al igual que el modelo de NIQE, es una métrica tiene en cuenta valores de luminancia, estructura espacial y contraste de la imagen para determinar la calidad de la imagen, no ha sido entrenada con valores de opinión y estima las distorsiones por bloques y después mide la varianza local de esos bloques para calcular el valor, pudiendo dar medias locales y globales.
- **Contraste y Nitidez:** Se tendrán en cuenta los valores de contraste y nitidez de las imágenes por separado mediante gradientes, leyendo cada imagen y suavizándolo

iterativamente, con el fin de conocer que modelos generan imágenes más claras y con bordes mejores definidos (Wang, 2021).

4.1.4. Encuesta al usuario

Se ha realizado una encuesta en la que se ha buscado tener en cuenta la opinión de los posibles clientes, para ello se han realizado diferentes preguntas para comprobar la creatividad, que diseño es más atractivo y tiene mejor estética visual. Se ha recogido una muestra de los mejores resultados de cada modelo, seleccionando 3 resultados que se dirigen a una misma intención, el diseño de una botella con forma o estética de hoja de olivo. Estas imágenes son representativas de cada modelo y son aquellas imágenes que mejor se adaptaban a un mismo texto de entrada “An olive oil bottle with the shape of a leaf”. También se busca la opinión de cada diseño y la opinión acerca del uso de estas nuevas herramientas de diseño. Con esta encuesta también se obtendrá un diseño elegido por los posibles clientes con el que realizar un diseño preliminar del producto. En total se realizarán 13 preguntas, incluyendo preguntas acerca del perfil del encuestado para conocer el marco de intereses de los posibles productos. Los resultados de esta encuesta se detallan en el Anexo II.



Figura 32. Selección de imágenes para el desarrollo del formulario.

4.2. Evaluación de cada método

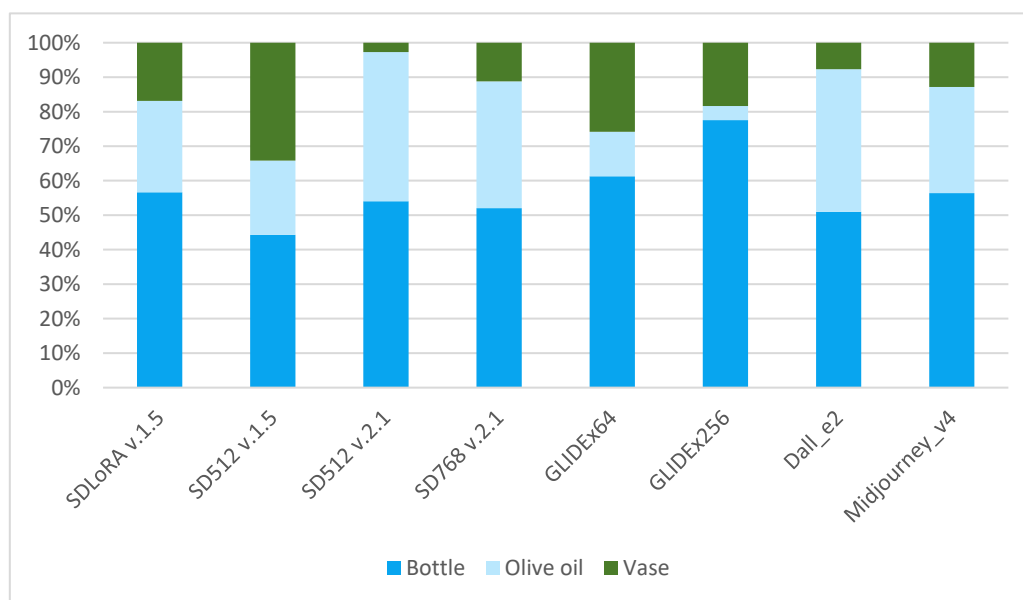
Cada modelo se ha separado entre modelos entrenados (ME) y Modelos no entrenados (MNE).

En la relación imagen texto el apartado de Repetición de palabras, se ha comprobado que las palabras más usadas para describir los elementos de cada imagen han sido, botella (Bottle), aceite de oliva (Olive oil) y jarrón (Vase).

Se ha contabilizado cuantas veces se repite cada palabra en cada modelo y se ha puntuado “Bottle” y “Olive Oil” como palabras positivas para el modelo y “Vase” como palabra menos conveniente para el modelo, mostradas en Tabla 1.

Tabla 1: Suma de las Repetición de Palabras por cada modelo.

	Modelo	Bottle	Olive oil	Vase
ME	SDLoRA v.1.5	47	22	14
	SD512 v.1.5	35	17	27
	SD512 v.2.1	60	48	3
	SD768 v.2.1	51	36	11
	GLIDEx64	38	8	16
	GLIDEx256	38	2	9
MNE	Dall_e2	53	43	8
	Midjourney_v4	44	24	10



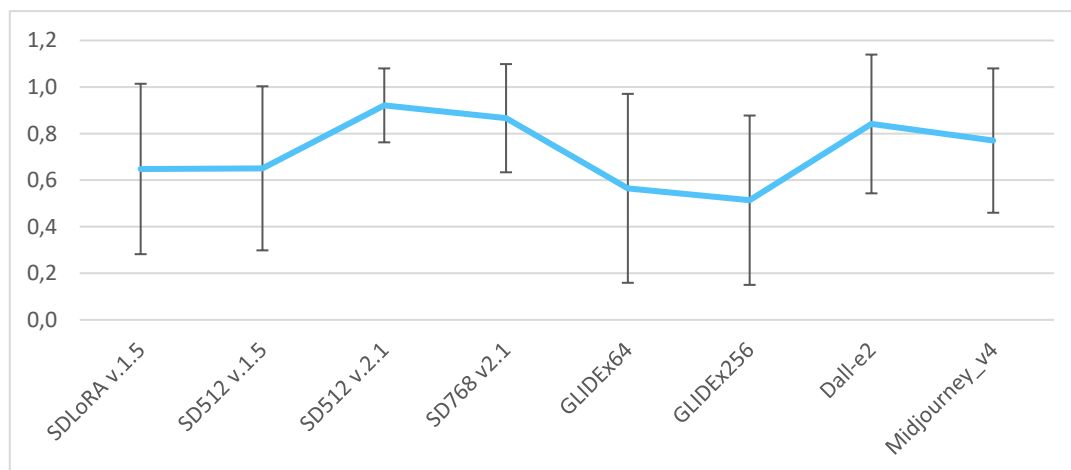
Gráfica 1. Distribución porcentual de la Repetición de Palabras.

En la coincidencia imagen texto se ha realizado el promedio de cada modelo, teniendo en cuenta su desviación típica, en la se representan estos valores y se comprueba que el modelo

de Stable Diffusion 2.1 en su versión de 512 píxeles es el que más porcentaje de probabilidad tiende a acercarse a la descripción de “An olive oil bottle”, un 92,12% y con la menor desviación típica entre los modelos.

Tabla 2: Promedio de Coincidencia imagen-texto entre las imágenes de los modelos.

	Modelo	Promedio	Desviación Típica
ME	SDLoRA v.1.5	0,6481	0,3660
	SD512 v.1.5	0,6508	0,3524
	SD512 v.2.1	0,9212	0,1588
	SD768 v.2.1	0,8662	0,2324
	GLIDEx64	0,5649	0,4058
	GLIDEx256	0,5139	0,3637
MNE	Dall_e2	0,8414	0,2981
	Midjourney_v4	0,7700	0,3100

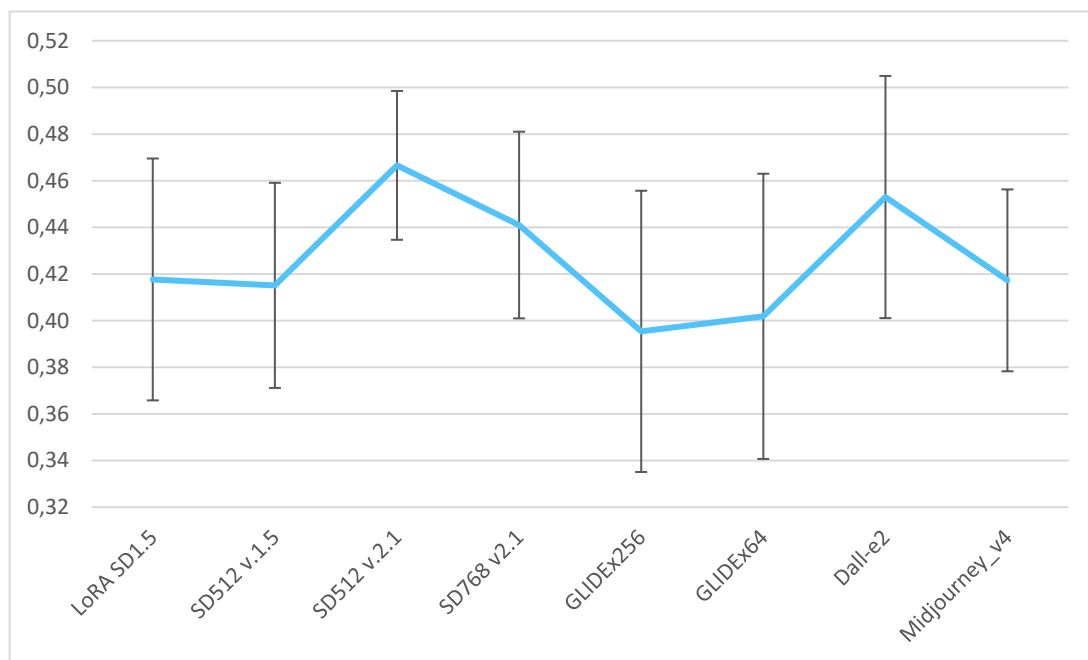


Gráfica 2. Distribución del los valores del promedio de CIT.

En la similitud del coseno imagen-texto se puede comprobar en la Tabla 3 como los valores cuanto más cercanos a 1 mayor es la similitud entre los vectores, todos los resultados parecen estar en un mismo rango de valores, entre 0.4 y 0.5, pero el modelo de Stable Diffusion 2.1 en su versión de 512 píxeles es el que tiene un valor más alto y con una desviación típica entre todas sus imágenes más bajo.

Tabla 3: Promedio de Similitud de Coseno Imagen-Texto entre las imágenes de los modelos.

	Modelo	Promedio	Desviación Típica
ME	SDLoRA v.1.5	0,4177	0,0519
	SD512 v.1.5	0,4151	0,0440
	SD512 v.2.1	0,4666	0,0319
	SD768 v.2.1	0,4410	0,0400
	GLIDEx64	0,3954	0,0603
	GLIDEx256	0,4018	0,0612
MNE	Dall_e2	0,4530	0,0519
	Midjourney_v4	0,4173	0,0390

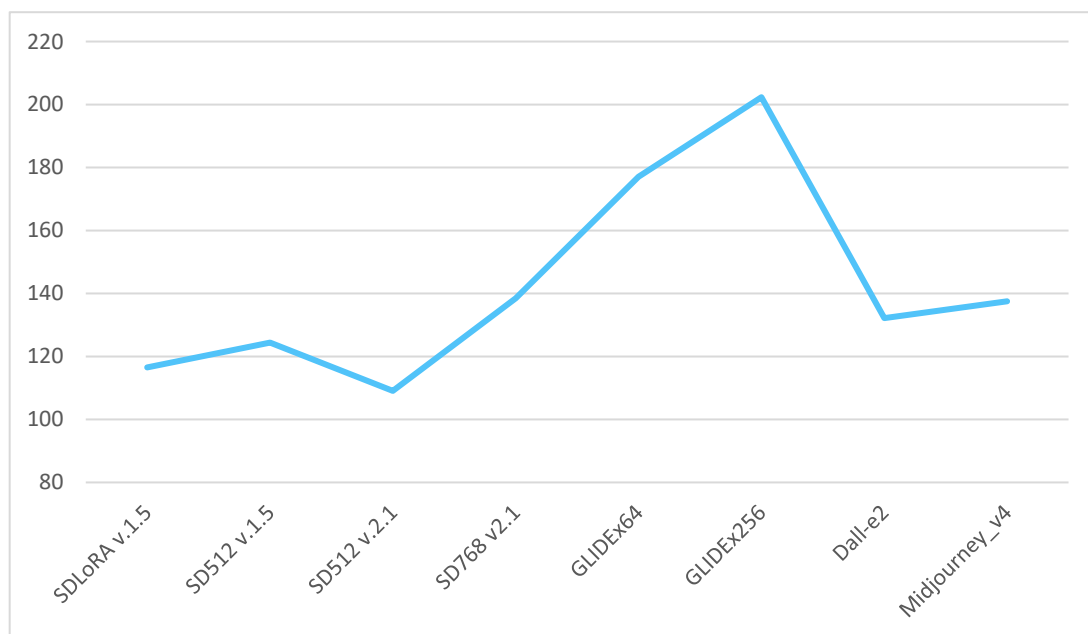


Gráfica 3. Distribución de los valores del promedio de SCIT.

Los resultados de la Tabla 4 muestran como los valores del FID describe más adecuadamente la calidad entre los modelos, acercándose a lo que a primera vista se puede apreciar entre los modelos. Este método establece que el modelo de Stable Diffusion 2.1 en su versión de 512 píxeles es el modelo que más se acerca a las imágenes de entrenamiento, y, por tanto, a las imágenes reales y que los modelos de GLIDE son los que más se alejan.

Tabla 4: FID entre las imágenes de los modelos.

	Modelo	FID
ME	SDLoRA v.1.5	116,51
	SD512 v.1.5	124,41
	SD512 v.2.1	109,07
	SD768 v.2.1	138,46
	GLIDEx64	177,06
	GLIDEx256	202,32
MNE	Dall_e2	132,16
	Midjourney_v4	137,50

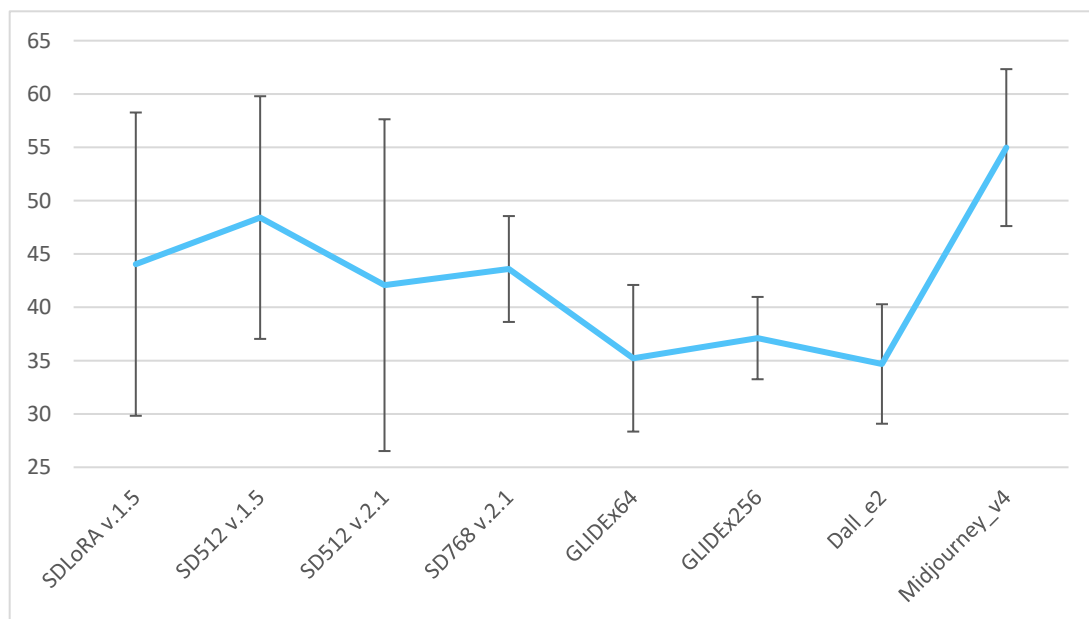


Gráfica 4. Distribución de los valores de FID por modelo.

Los resultados mediante la métrica de Brisque de la Tabla 5 se establece que el modelo con mayor naturalidad es el modelo de Midjourney. Los valores de esta métrica se establecen entre 0 y 100, siendo 0 el valor de una imagen más natural. En este caso se comprueba que el siguiente modelo es el de Stable Diffusion 2.1 en su versión de 512, siendo el que obtiene mejor puntuación entre los modelos entrenados.

Tabla 5: Promedio de Brisque entre las imágenes de los modelos.

	Modelo	Promedio	Desviación Típica
ME	SDLoRA v.1.5	44,045	14,218
	SD512 v.1.5	48,413	11,376
	SD512 v.2.1	42,075	15,554
	SD768 v.2.1	43,590	4,964
	GLIDEx64	35,220	6,874
	GLIDEx256	37,113	3,859
	MNE	Dall_e2	34,683
	Midjourney_v4	54,973	7,357

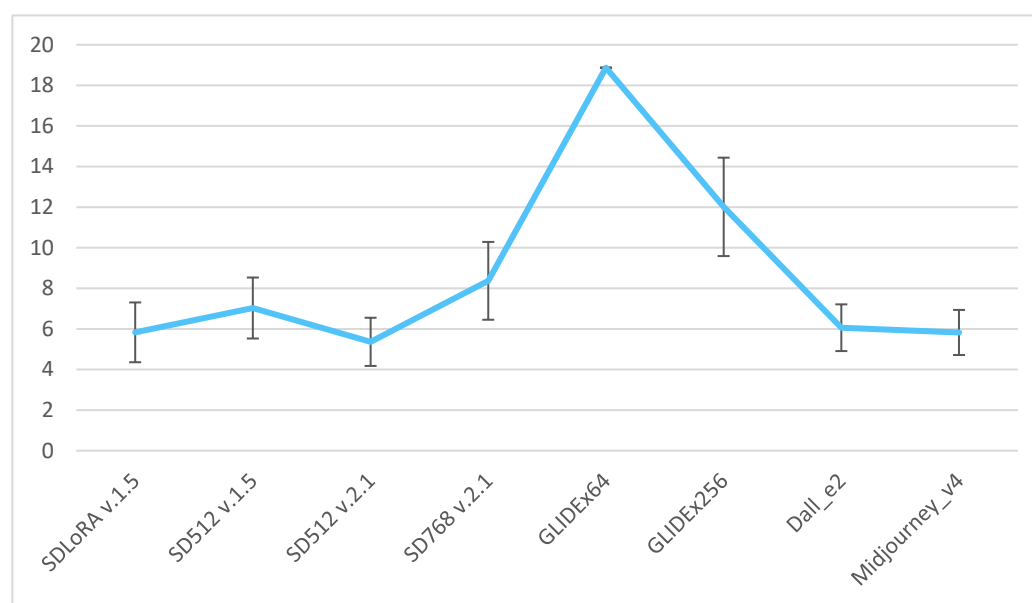


Gráfica 5. Distribución de los valores del promedio de Brisque.

En la métrica NIQE, cuanto menor sea la puntuación mayor calidad visual será percibida por un observador. Los valores de esta métrica se encuentran entre 0 y 20, en este caso el método con menor valor y por tanto mejor calidad visual sería el modelo de Stable Diffusion 2.1 en su versión de 512, seguido por los modelos de Stable Diffusion 1.5 entrenado con LoRA y Midjourney.

Tabla 6: Promedio de NIQE entre las imágenes de los modelos.

	Modelo	Promedio	Desviación Típica
ME	SDLoRA v.1.5	5,831	1,473
	SD512 v.1.5	7,030	1,503
	SD512 v.2.1	5,364	1,187
	SD768 v.2.1	8,369	1,917
	GLIDEx64	18,870	0,001
	GLIDEx256	12,014	2,424
MNE	Dall_e2	6,060	1,151
	Midjourney_v4	5,826	1,111

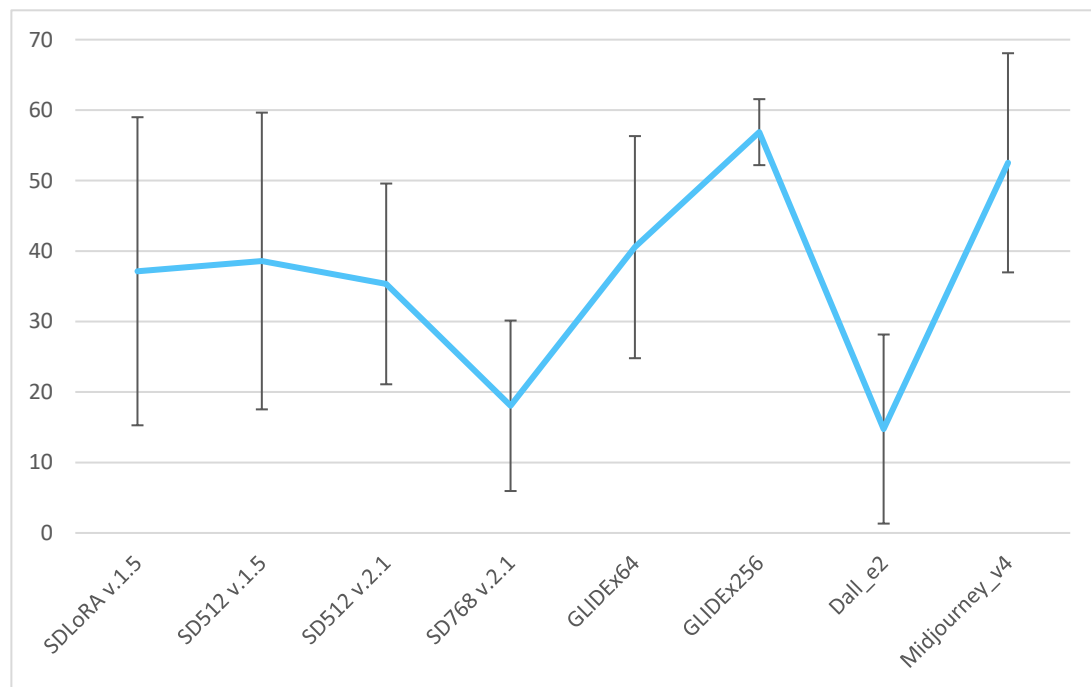


Gráfica 6. Distribución de los valores del promedio de NIQE.

La métrica de PIQE establece su rango entre 0 y 100, donde 0 es la mejor calidad de imagen. En este caso el modelo de DALL·E 2 resulta ser el que mejor promedio tiene.

Tabla 7: Promedio de PIQE entre las imágenes de los modelos.

	Modelo	Promedio	Desviación Típica
ME	SDLoRA v.1.5	37,135	21,859
	SD512 v.1.5	38,591	21,055
	SD512 v.2.1	35,337	14,243
	SD768 v.2.1	18,045	12,094
	GLIDEx64	40,555	15,759
	GLIDEx256	56,875	4,681
MNE	Dall_e2	14,744	13,415
	Midjourney_v4	52,525	15,549

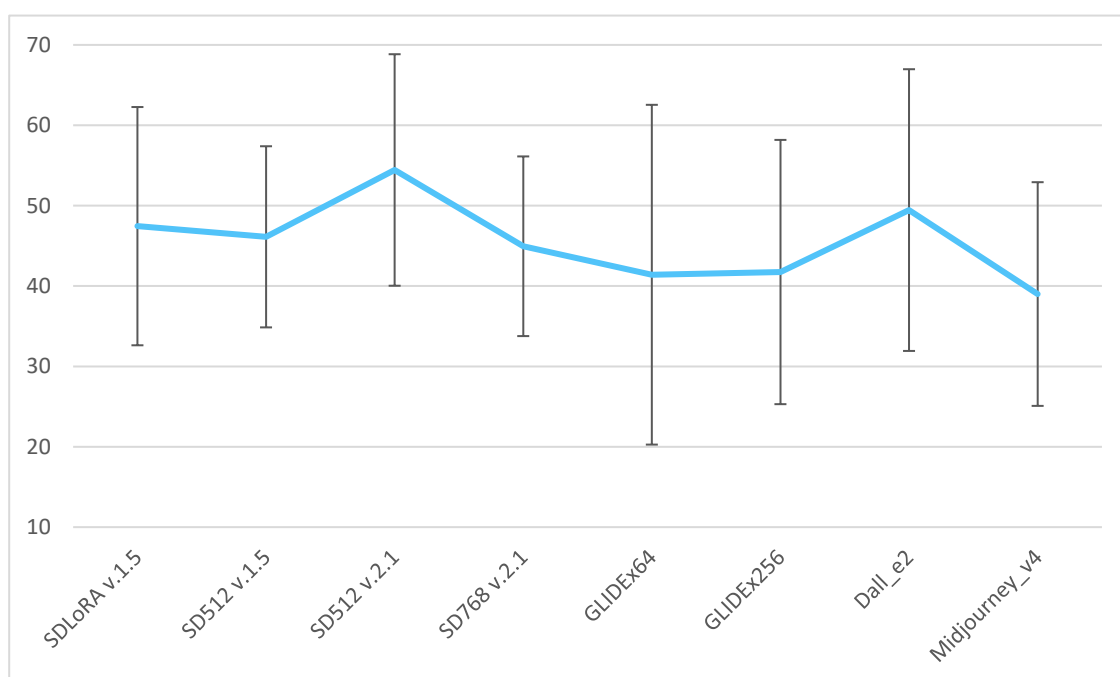


Gráfica 7. Distribución del los valores del promedio de PIQE.

En términos de contraste el modelo de Stable Diffusion 2.1 en su versión de 512 es el que más contraste tiene, haciendo que sea el modelo que más define la forma de los objetos y los destaca más del fondo.

Tabla 8: Promedio de contraste entre las imágenes de los modelos.

	Modelo	Promedio	Desviación Típica
ME	SDLoRA v.1.5	47,449	14,822
	SD512 v.1.5	46,123	11,269
	SD512 v.2.1	54,439	14,403
	SD768 v.2.1	44,948	11,177
	GLIDEx64	41,409	21,137
	GLIDEx256	41,741	16,439
	Dall_e2	49,450	17,526
MNE	Midjourney_v4	39,004	13,916



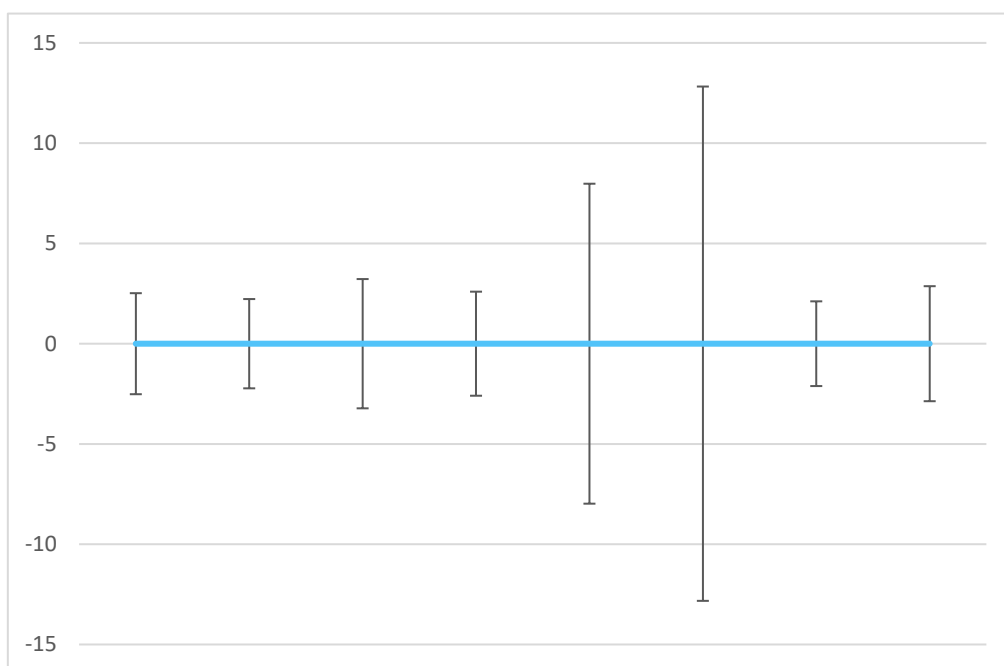
Gráfica 8. Distribución de los valores del promedio de contraste.

Los valores de nitidez están evaluados tal que, a mayores valores, menor es la nitidez de la imagen. Evaluando en todos los modelos comprobamos que el modelo de DALL·E 2 parece ser el modelo con las imágenes más nítidas, seguido del modelo de Dreambooth del modelo de Stable Diffusion 1.5, que tiene valores bastante cercanos y con las desviaciones más bajas.

Tabla 9: Promedio de Nitidez entre las imágenes de los modelos.

	Modelo	Promedio	Desviación Típica
ME	SDLoRA v.1.5	2,52	1,21
	SD512 v.1.5	2,23	0,98
	SD512 v.2.1	3,22	1,12
	SD768 v.2.1	2,59	0,87
	GLIDEx64	7,98	4,12
	GLIDEx256	12,82	1,88
MNE	Dall_e2	2,11	0,58
	Midjourney_v4	2,87	1,45



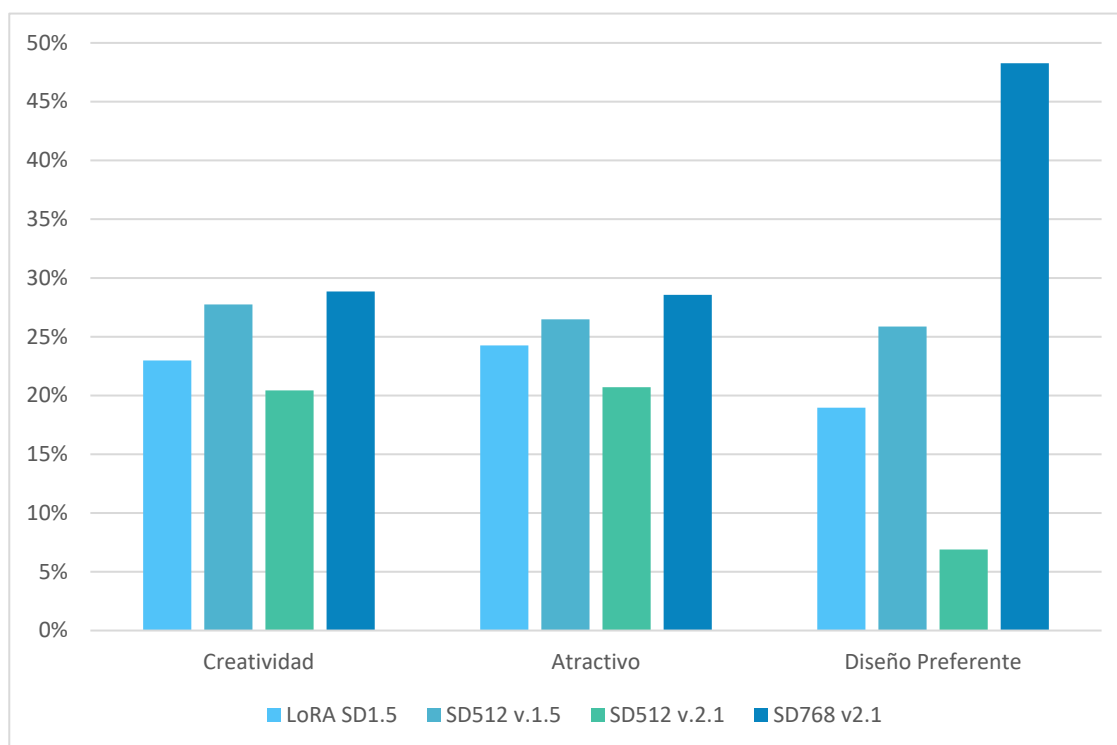


Gráfica 9. Distribución de los valores del promedio de nitidez entre todos los modelos.

En la encuesta, desarrollada en el Anexo 2, se han valorado tres criterios evaluados por los encuestados: creatividad, atractivo visual y diseño preferente por el usuario. Los valores de la creatividad representan la suma de los valores entre el 1 al 10 entre las imágenes de cada modelo, los valores del atractivo visual la suma del 1 al 12 de cada imagen de cada modelo y el diseño preferente cuantas veces ha seleccionado cada encuestado una imagen de cada modelo. Entre los resultados el modelo de Stable Diffusion en su versión 2.1 a 768 píxeles es el que más han elegido los encuestados cada categoría.

Tabla 10: Promedio de Coincidencia imagen-Texto entre las imágenes de los modelos.

Modelo	Creatividad	Atractivo	Diseño Preferente
LoRA SD1.5	496	1105	11
SD512 v.1.5	599	1206	15
SD512 v.2.1	441	943	4
SD768 v2.1	623	1301	28



Gráfica 10. Distribución porcentual de las preferencias del encuestado por modelo.

4.3. Relación humano-herramienta

Uno de los aspectos a evaluar es la interacción entre el usuario y cada modelo, donde hay que tener en cuenta las facilidades que te ofrece cada modelo para crear, editar y generar cada idea conceptual. Cada modelo viene con un entorno de uso que facilita su aplicación y que en ciertos casos genera un flujo de trabajo más eficiente y con resultados más convenientes a lo que se pretende crear.

Los modelos de Stable Diffusion tienen el beneficio de ser de código libre y por lo tanto los desarrolladores interesados en el modelo han creado diferentes interfaces gráficas que establecen una conexión personalizable con el modelo. En el desarrollo de este trabajo se ha estado utilizando la interfaz de AUTOMATIC1111, una interfaz modular basada en Gradio donde el control a la hora de formar imágenes es bastante amplio, no ofrece solo control sobre las características esenciales del modelo como serían la descripción del texto, dimensiones y el número de imágenes que quieres crear, sino también control en las semillas, pasos, escala CFG, método de muestreo. Se puede generar imágenes usando la técnica de texto-a-imagen e imagen-a-imagen, pudiendo partir de una fotografía o dibujo como base de la generación de una nueva imagen o editar partes de la imagen rellenándolas con nuevo contenido. Con estos

controles se puede volver a crear una misma imagen y cambiar ciertos valores esperando alguna variación en la imagen real como se puede ver en la Figura 15.

Adicionalmente se le puede incluir el módulo de Controlnet, que incluye diferentes modelos entrenados de control, que permiten generar una imagen siguiendo un mapa de profundidad, mapas normales, elementos de posición, sketch que permiten guiar a la imagen sin condicionarla tanto como en el método de imagen a imagen y pudiendo combinar varios métodos. Un ejemplo de uso sería el de la Figura 33, donde partiendo de una imagen de una hoja de olivo se extrae su mapa de profundidad y se establece como guía a la hora de crear diferentes imágenes siguiendo el texto de entrada, en este caso se ha usado el modelo de Stable Diffusion en su versión 1.5 entrenada con Dreambooth y asignándole un peso a la imagen de 0.7.



Figura 33. Imágenes generadas mediante el uso de ControlNet.

Estos modelos requieren también un texto de entrada bastante descriptivo, utilizando palabras clave y usando números modificadores de estilo dependiendo de la versión que se use, la versión 2.1 requiere significativamente más palabras que describan la escena que se quiere crear y funciona de manera más efectiva usando textos de entrada negativos que especifiquen aquello que no se quiere ver en la imagen. Un ejemplo de texto de cada modelo serían los siguientes:

Versión 1.5: "A photo of aceitefg, leaf shape, aceitefg, olive leaf form, ergonomic bottle, curved form"

Versión 2.1:

- Prompt: "A photo of aceitefg, shape of a leaf, aceitefg, leaf form, ergonomic bottle, curved form, centered, dramatic lighting, dark, neon, hyperrealistic, highly detailed, intricate, sharp focus, 8k, olive oil bottle, hyper detailed, white background, product photo, white cover"
- Negative Prompt: "distored, unreal, weird, close up, two, deformed, semi-realistic, cgi, 3d, render, sketch, cartoon, drawing, anime:1.4), cropped, out of frame, worst quality, low quality, jpeg artifacts, ugly, duplicate, morbid, mutation, deformed, blurry, dehydrated, bad proportions, gross proportions, melted"

El modelo de GLIDE no cuenta con ninguna interfaz de usuario adaptada hasta el momento, las únicas herramientas de control son el texto de entrada, el número de pasos que va a realizar, la escala CFG y la cantidad de imágenes que se va a generar y se definen mediante código. Los textos de entrada no requieren ser demasiado explícitos y con pocos detalles el modelo puede entender el propósito de la imagen. Un ejemplo sería:

“An olive oil bottle, olive leaf shape, natural shape, leaf form”.

El modelo de DALL·E 2 cuenta con una interfaz de la página de OpenAI que permite generar imágenes a partir de texto, efectuar variaciones de una imagen y realizar inpainting y outpaintings de una imagen, pudiendo cambiar elementos o ampliar las dimensiones de una imagen guiándose por el texto. No ofrece más herramientas de control, es simple y limitado.

Los textos de entrada necesarios serían comparables con los de GLIDE y la versión 1.5 de Stable Diffusion, aunque sí requiere de más modificadores de estilo ya que tiende a realizar cuadros artísticos, un ejemplo sería:

“A product photo of an olive oil bottle, centered image, beautiful presentation, a bottle in the shape of a leaf, white covered”.

El modelo de Midjourney usa la interfaz de la aplicación de Discord para generar sus imágenes, pero ofrece bastantes herramientas, como elegir entre diferentes modelos anteriores, establecer dimensiones, usar imagen como base, juntar dos imágenes diferentes y aplicar estilos. Los textos de entrada necesarios para este modelo son similares a los anteriores:

“Product design concept of olive oil bottle shaped like a olive leaf by Dieter Rams, Frank Lloyd Wright, Achille Castiglioni and Syd Mead --v 4”.

5 Análisis de Resultados

Tras realizar los diferentes métodos de evaluación y viendo en conjunto que modelos son los más favorecidos por las diferentes métricas, siendo en conjunto más favorable el modelo de Stable Diffusion en su versión 2.1 a 512 píxeles, pero estando muy cerca de otros modelos como el de su versión a 768 píxeles, DALL·E 2 y la versión 1.5 de Stable Diffusion entrenada con Dreambooth.

Tabla 11: Resultado de las Evaluaciones Métricas.

	Modelo	RP	CIT	SCIT	FID	Brisque	NIQE	PIQE	Contraste	Nitidez
ME	SDLoRA v.1.5	69	64,81%	0,4177	116,5	44,05	5,83	37,14	47,45	2,52
	SD512 v.1.5	52	65,08%	0,4151	124,4	48,41	7,03	38,59	46,12	2,23
	SD512 v.2.1	108	92,12%	0,4666	109,1	42,08	5,36	35,34	54,44	3,22
	SD768 v.2.1	87	86,62%	0,441	138,5	43,59	8,37	18,04	44,95	2,59
	GLIDEx6 4	46	56,49%	0,4018	177,1	35,22	18,87	40,56	41,41	7,98
	GLIDEx2 56	40	51,39%	0,3954	202,3	37,11	12,01	56,87	41,74	12,82
MNE	Dall_e2	96	84,14%	0,453	132,2	34,68	6,06	14,74	49,45	2,11
	Midjourney_v4	68	77,00%	0,4173	137,5	54,97	5,83	52,53	39,00	2,87

En la Figura 34 se puede observar una comparativa dimensional entre las imágenes usadas en la evaluación de cada modelo, ordenadas de menor tamaño (modelo de GLIDE) a mayor tamaño (modelo de DALL·E 2), incluyendo las imágenes usadas para el entrenamiento y comparadas en la métrica FID, dimensión que hay que tener en cuenta a la hora de valorar.



Figura 34. Conjunto Global de Imágenes para su evaluación.

Teniendo en cuenta apartados como los de la interacción con los modelos y la encuesta realizada se puede establecer que los modelos de Stable Diffusion son los más adecuados para la generación de imágenes y que destacan los modelos con una versión de 2.1 a 768 píxeles por los encuestados y la versión 1.5 por facilidad de uso.

Se ha realizado una evaluación completa todos los criterios, añadiendo una valoración del uso de cada modelo y ponderando el peso de relevancia de cada criterio.

Tabla 12: Resultado del conjunto de evaluaciones.

Criterio	ME						MNE	
	SDLoRA v.1.5	SD512 v.1.5	SD512 v.2.1	SD768 v.2.1	GLIDEx6 4	GLIDEx25 6	Dall_e2	Midjourney_ v4
RP	69	52	108	87	46	40	96	68
CIT	64,81%	65,08%	92,12%	86,62%	56,49%	51,39%	84,14%	77,00%
SCIT	0,4177	0,4151	0,4666	0,441	0,4018	0,3954	0,453	0,4173
FID	116,5	124,4	109,1	138,5	177,1	202,3	132,2	137,5
Brisque	44,05	48,41	42,08	43,59	35,22	37,11	34,68	54,97
NIQE	5,83	7,03	5,36	8,37	18,87	12,01	6,06	5,83
PIQE	37,14	38,59	35,34	18,04	40,56	56,87	14,74	52,53
Contraste	47,45	46,12	54,44	44,95	41,41	41,74	49,45	39
Nitidez	2,52	2,23	3,22	2,59	7,98	12,82	2,11	2,87
Creatividad	496	599	441	623	-	-	-	-
Atractivo	1105	1206	943	1301	-	-	-	-
Diseño Preferente	11	15	4	28	-	-	-	-
Valoración usuario	6	8	7	5	2	1	3	4
Dimensiones	50	50	50	75	6,25	25	100	50

Una vez recogidos todos los datos se ha realizado una ponderación de los criterios de la Tabla 12, teniendo en cuenta que criterios como FID, las encuestas realizadas o la adecuación al texto puede tener más relevancia que otros criterios de calidad visual como la nitidez o el contraste.

Tabla 13: Resultado del conjunto de evaluaciones ponderado.

Modelo	ME						MNE		Peso de cada Criterio
	SDLoRA v.1.5	SD512 v.1.5	SD512 v.2.1	SD768 v.2.1	GLIDEx64	GLIDEx256	Dalle 2	Midjourney v4	
RP	12%	9%	19%	15%	8%	7%	17%	12%	90%
CIT	11%	11%	16%	15%	10%	9%	15%	13%	100%
SCIT	12%	12%	14%	13%	12%	12%	13%	12%	70%
FID	90%	89%	90%	88%	84%	82%	88%	88%	100%
Brisque	87%	86%	88%	87%	90%	89%	90%	84%	50%
NIQE	92%	90%	92%	88%	73%	83%	91%	92%	50%
PIQE	87%	87%	88%	94%	86%	81%	95%	82%	50%
Contraste	13%	13%	15%	12%	11%	11%	14%	11%	40%
Nitidez	93%	94%	91%	93%	78%	65%	94%	92%	40%
Creatividad	23%	28%	20%	29%	0%	0%	0%	0%	80%
Atractivo	24%	26%	21%	29%	0%	0%	0%	0%	80%
Diseño Preferente	19%	26%	7%	48%	0%	0%	0%	0%	100%
Valoración usuario	17%	22%	19%	14%	6%	3%	8%	11%	70%
Dimensiones	12%	12%	12%	18%	2%	6%	25%	12%	60%
Porcentaje ponderado	37,9%	39,1%	37,8%	42,6%	28,0%	27,3%	33,6%	31,2%	
Porcentaje Global	13,67%	14,09%	13,61%	15,34%	10,10%	9,85%	12,10%	11,24%	

Atendiendo al porcentaje global resultante de estas evaluaciones se establece que el modelo de Stable Diffusion en su versión 2.1 a 768 píxeles es el modelo más conveniente y con mejores resultados seguidos de la versión 1.5 entrenada con Dreambooth.

Como diseño de ejemplo a desarrollar los encuestados han elegido el diseño de la imagen número 8 de la Figura 32, diseño creado por el modelo de SD en su versión 2.1 a 768 píxeles y reflejada en la Figura 35, imagen que será tratada para un desarrollo conceptual final en el apartado 6.



Figura 35. Imágen generada por SD en su versión 2.1 a 768 píxeles seleccionada para caso práctico.

6 Desarrollo de un caso práctico

Una vez realizada la encuesta y seleccionado la imagen 8 de la Figura 32 como diseño preferente por los encuestados se realizará un diseño 3D preliminar, donde establecer volúmenes y posibles dimensiones del diseño, estas dimensiones no son las definitivas, solo forman parte del diseño conceptual, las dimensiones reales están demarcadas por el volumen total que en el envase tiene que albergar y la normativa de envases referentes a los grosores que tiene que tener las paredes del diseño.

Este modelo en 3D, realizado en el software Fusion 360 de Autodesk para el modelado y el programa Blender para un renderizado realista, ha tomado de referencia la imagen de la Figura 35 y se ha seguido con toda la información que esa imagen transmite, se han demarcado tres figuras, una que configura el recipiente que contiene el aceite, otro elemento que sirve como base y un tercer modelo que compone el tapón de la botella.

En la imagen se intuyen diferentes materiales, vidrio para el recipiente, corcho para el tapón y un tercer material que puede ser entre cerámica esmaltada, porcelana o vidrio opaco. Para establecer un conjunto de materiales coherentes se usará cerámica, al ser un material comúnmente usado en los envases más tradicionales de aceite. Estos materiales son orientativos para la aplicación de materiales en el procesado de las imágenes realistas necesarias para la compresión del modelo generativo en el nuevo entrenamiento y los materiales definitivos tienen que ser determinados en las siguientes fases del desarrollo del producto.



Figura 36. Imágenes realizadas de diseño en Fusion 360 y estilo realista en Blender.

Para establecer unas dimensiones orientativas del producto se va a partir del volumen final de aceite que se busca conservar, para este producto se ha pensado en un volumen de partida de 200 ml, para establecer un volumen de partida y comprobar que dimensiones debería tener.

Al realizar el modelo se puede ver como el volumen de producto en comparación con el envase es proporcionalmente más pequeño y que para albergar ese volumen sería necesario un producto con una altura de 222 mm, unas dimensiones bastante grandes para un envase de ese volumen. Como grosores de la pared se ha establecido unos valores de 4.25 mm, siguiendo la normativa UNE-EN 12726:2019⁵ para envases con corcho de diámetro de entrada 18.5 mm, como se puede observar en la Figura 37.

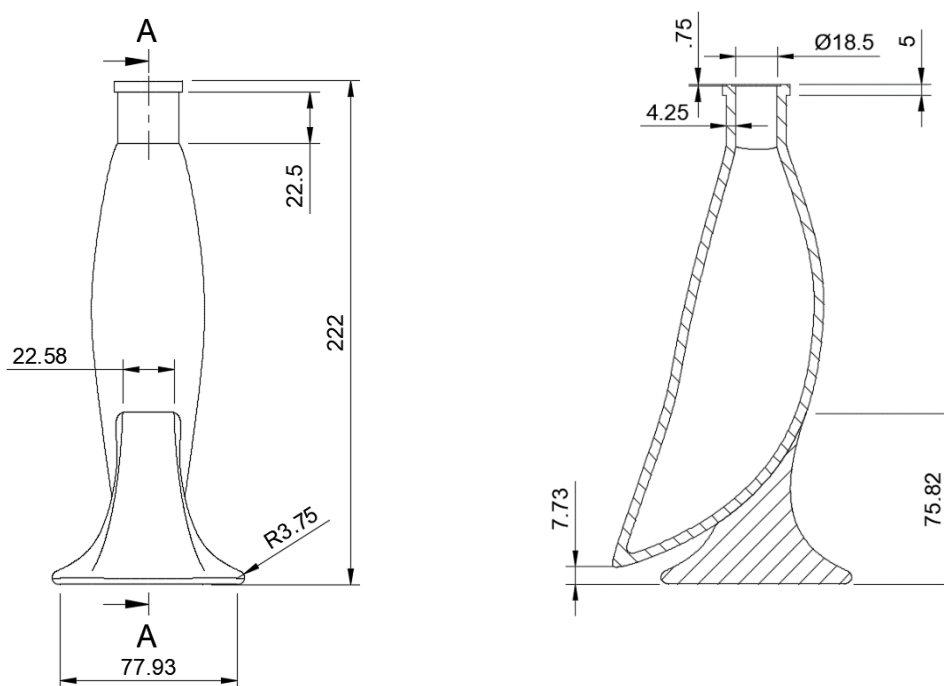


Figura 37. Representación del diseño conceptual a escala con medidas de orientación.

De forma orientativa se puede comprobar las masas relativas de cada parte, el recipiente de vidrio estaría alrededor de los 253 g, el corcho alrededor de los 3 g y la base entorno a los 342 g. El peso orientativo que incluiría los 200 ml de aceite estaría próximo a los 798 g. y los 598 solo del envase. Un peso relativamente aceptable para un envase con una visión de venta de un producto de alta calidad.

⁵ UNE-EN 12726:2019: Envases y embalajes. Boca para tapón de corcho con un diámetro de entrada de 18,5 mm para corchos y cápsulas de seguridad.

Estos pesos se han calculado dentro del programa de Fusion 360 indicando los materiales con sus densidades correspondientes y los planos conceptuales completos se encuentran en la sección de Planos.

Con este diseño se puede establecer un nuevo entrenamiento sobre el modelo de Stable Diffusion 1.5 usando Dreambooth, para ello se realiza una base de 20 imágenes renderizadas del producto, bajo diferentes condiciones de iluminación y fondo, para generar imágenes de presentación, exposición e investigación en caso de querer probar con nuevos materiales o modificar la forma para posteriormente realizar el diseño en detalle del producto del diseño seleccionado. Estas imágenes han sido realizadas de manera realista buscando que el modelo entienda como debería verse en el mundo real.

Para este entrenamiento se han realizado 3200 steps en el modelo con una duración de 40 minutos de entrenamiento y con unos resultados bastante cercanos al diseño establecido.



Figura 38. Imágenes de entrenamiento para el caso práctico.

Ahora es posible generar diferentes diseños partiendo del mismo concepto generado por la inteligencia artificial, un proceso en el cual se vuelve a nutrir al modelo en busca de un refinamiento preciso para un caso en concreto.

En la Figura 39 se puede ver cómo es capaz de generar imágenes del modelo basándose en el mismo diseño pudiendo aplicar nuevos materiales como madera a la base en la imagen 2 o generar nuevos diseños como en la imagen 5 y crear nuevas escenas como en la imagen 4 o la imagen 1 para una presentación de producto. Con estas imágenes ya se puede presentar una idea final del producto que se quiere diseñar.

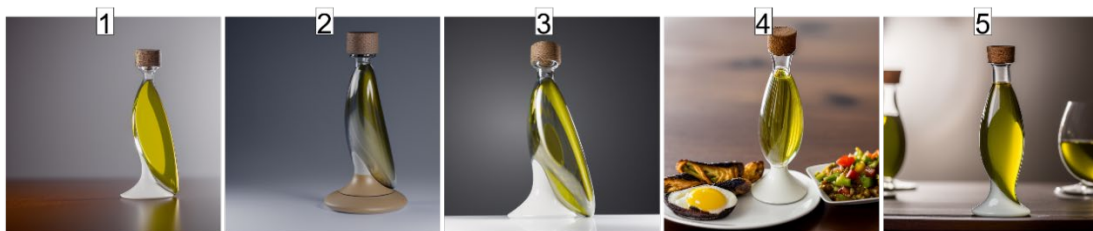


Figura 39. Imágenes Generadas para el caso práctico.

Con esta información y diseños nuevos se podría pasar a la siguiente fase del proceso de diseño, la definición del diseño final, el desarrollo técnico del envase, definiendo las especificaciones finales del diseño mediante un estudio ergonómico, análisis de materiales, optimización del volumen del diseño y necesidades de fabricación que pueda requerir el producto como un dosificador o difusor de aceite y adaptarlo a la ergonomía de la mano para un agarre adecuado.

7 Conclusión

Este sistema de creación de ideas conceptuales es bastante prometedor, en principio el sistema más eficiente y que técnicamente resulta más conveniente, tanto en términos de accesibilidad como de control, personalización, costes, adecuación al texto y calidad de imagen según los métodos de medición ponderados en la Tabla 13 es el modelo entrenado de Stable Diffusion en su versión 2.1 a 768 píxeles. Según las respuestas realizadas a los encuestados el modelo de Stable Diffusion en su versión 2.1 a 768 píxeles es el modelo que más aceptación ha tenido y, tras haber realizado numerables generaciones entre los distintos modelos, se ha observado de forma subjetiva que el modelo de Stable Diffusion en su versión 1.5 tiende a dar un resultado de manera más rápida y precisa que el resto de los modelos, con un menor número de especificaciones.

Cualquiera de estos modelos puede ser capaz de generar un resultado adecuado para la necesidad de un envase de aceite, el modelo de Stable Diffusion en su versión 2.1 a 768 píxeles ha generado los 2 diseños que los encuestados han valorado como más atractivos y diseños preferentes.

Utilizando cualquiera de estos 3 métodos se puede realizar un diseño conceptual válido para el desarrollo de un envase de aceite de oliva, preferentemente el usuario tiene que elegir con cual se siente más cómodo a la hora de usar una de estas herramientas de trabajo.

Como preferencia personal el modelo de Stable Diffusion en su versión 1.5 es con el que mejores resultados obtengo a largo plazo, como alegoría sería interesante tratar con estos modelos como si fueran un diseñador diferente, y a cada uno te va a responder diferente según la información que le facilites.

Hay que tener en cuenta que por mucho que un diseño pueda ser más o menos atractivo, éste puede no ser viable para satisfacer las necesidades físicas del producto. Como se puede comprobar en el apartado 6, donde el producto acaba requiriendo unas dimensiones bastante considerables para conservar una cantidad limitada de aceite.

Es importante encontrar un equilibrio entre el atractivo estético y la funcionalidad práctica. Para ello, es recomendable realizar pruebas y estudios de viabilidad antes de llegar a la etapa de producción. Además, los diseños tienen que seguir ciertas regulaciones técnicas, ergonómicas y normativas que estos modelos no son capaces de contemplar y aplicar a un diseño y que la elección de los materiales puede no ser las más convenientes para el diseño.

Con estos modelos se puede seguir la siguiente arquitectura de trabajo:

1. Selección de campo o producto de diseño.
2. Estudio de mercado del producto.
3. Generación de ideas textuales centradas en un concepto.
4. Introducción de estas ideas adaptadas al modelo seleccionado.
5. Selección de los mejores diseños creados.
6. Evaluación del diseño preferente.
7. Entrenamiento del modelo usando imágenes del diseño para el diseño final o su edición.
8. Comprobación de practicidad del diseño.

El uso de estos modelos permite aportar nuevos posibles diseños para tener en cuenta en el desarrollo del envase y ofrecer diferentes opciones de conceptos donde elegir, si se usara estas herramientas en el proceso de diseño se optimizaría el tiempo y aportaría más variedad a los diseños creados por el diseñador, ayudando en situaciones de bloqueo creativo y ofreciendo ideas frescas adaptadas a un campo del diseño.

Por el momento esta tecnología no llega a suplantar el trabajo del diseñador, su papel sigue siendo clave en el desarrollo del diseño final y su criterio es que decide que modelo realizar y puede tener en cuenta o no las soluciones ofrecidas por estos modelos generativos. El diseñador es el que controla las necesidades del cliente y los aspectos técnicos del diseño del envase y maneja los aspectos estéticos, emocionales y comunicativos que necesita el envase.

En resumen, el uso de modelos generativos en el proceso de diseños conceptuales es una herramienta con gran potencial que puede aumentar la productividad y la variedad de opciones de diseño, en conjunto con el trabajo del diseñador y el poder adaptar estos modelos a diferentes conceptos de diseños acaba añadiendo un valor añadido a un producto que puede favorecer en su resultado.

8 Referencias

I. Referencias Bibliográficas:

- Borji, A. (2022). Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. <http://arxiv.org/abs/2210.00586>
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. <http://arxiv.org/abs/1706.08500>
- Ho, J., Jain, A., & Abbeel, P. (n.d.). Denoising Diffusion Probabilistic Models. <https://github.com/hojonathanho/diffusion>.
- Ho, J., Saharia, C., Chan, W., Fleet, D. J., Norouzi, M., & Salimans, T. (2021). Cascaded Diffusion Models for High Fidelity Image Generation. <http://arxiv.org/abs/2106.15282>
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. <http://arxiv.org/abs/2106.09685>
- Li, J., Li, D., Xiong, C., & Hoi, S. (2022). BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. <http://arxiv.org/abs/2201.12086>
- Lu, Z., Kazi, R. H., Wei, L. Y., Dontcheva, M., & Karahalios, K. (2021). StreamSketch: Exploring Multi-Modal Interactions in Creative Live Streams. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1). <https://doi.org/10.1145/1122445.1122456>
- Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12), 4695–4708. <https://doi.org/10.1109/TIP.2012.2214050>
- Mittal, A., Soundararajan, R., & Bovik, A. C. (2012). Making a “Completely Blind” Image Quality Analyzer. <http://live.ece.utexas.edu/research/quality/niqe>
- Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., & Chen, M. (2021). GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. <http://arxiv.org/abs/2112.10741>
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. <http://arxiv.org/abs/2103.00020>
- Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical Text-Conditional Image Generation with CLIP Latents. <http://arxiv.org/abs/2204.06125>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. <http://arxiv.org/abs/1505.04597>
- Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2022). DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. <http://arxiv.org/abs/2208.12242>
- Wang, L. (2021). A survey on IQA. <http://arxiv.org/abs/2109.00347>



II. Referencias a códigos

- Xiao, Z. (2022, septiembre 20). Dreambooth on Stable Diffusion. GitHub. <https://github.com/XavierXiao/Dreambooth-Stable-Diffusion>
- Mullis, C. (2021). Glide-finetune. GitHub. <https://github.com/afiaka87/glide-finetune>
- Li, J., Li, D., Xiong, C., & Hoi, S. (2022). BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. <https://github.com/salesforce/blip>
- Seitzer, M. (2020, agosto). pytorch-fid: FID Score for PyTorch. GitHub. <https://github.com/mseitzer/pytorch-fid>
- MathWorks. (s.f.). Image quality metrics. <https://www.mathworks.com/help/images/image-quality-metrics.html>
- MathWorks. (s.f.). Puntuación de la calidad de la imagen sin referencia con el Evaluador de la calidad espacial de la imagen sin referencia/a ciegas (BRISQUE). MathWorks. <https://www.mathworks.com/help/images/ref/brisque.html>
- MathWorks. (s.f.). Puntuación de calidad de imagen sin referencia del Evaluador de la calidad de imágenes naturales (NIQE). MathWorks. <https://www.mathworks.com/help/images/ref/nique.html>
- MathWorks. (2018). Perception based Image Quality Evaluator (PIQE) no-reference image quality score. MathWorks. <https://www.mathworks.com/help/images/ref/piqe.html>
- Birdal, T. (2023). Sharpness Estimation From Image Gradients. MATLAB Central File Exchange. <https://www.mathworks.com/matlabcentral/fileexchange/32397-sharpness-estimation-from-image-gradients>

III. Referencias de imágenes del estudio de mercado

- Picualia. (s.f.). Aceite de oliva virgen extra picual lata 2,5 l [Imagen]. Aceite de Jaén. <https://www.aceitedejaen.es/aceite-de-oliva-virgen-extra/gran-formato/picualia-gourmet-picual-lata-25-l>
- Padilla. (s.f.). Aceite de oliva virgen extra ecológico picual lata 5 l [Imagen]. Aceite de Jaén. <https://www.aceitedejaen.es/aceite-de-oliva-virgen-extra/ecologico/padilla-bio-ecologico-picual-lata-5-l>
- Oro Bailén. (s.f.). Aceite de oliva virgen extra reserva familiar arbequina lata 2,5 l [Imagen]. Aceite de Jaén. <https://www.aceitedejaen.es/aceite-de-oliva-virgen-extra/gran-formato/oro-bailen-reserva-familiar-arbequina-lata-25-l>
- Aceites Tapia. (2021). Verde Oleum / Can [Imagen]. En Aceites Tapia. Catalogue de produits (p. 10). https://www.aceitestapia.com/wp-content/uploads/2021/07/AceitesTapia_Catalogo_FRANCES_compressed.pdf
- Regalos Ecology. (s.f.). Aceite detalles boda [Imagen]. Regalos Ecology. <https://www.regalosecology.com/aceite-detalles-boda>
- Verde Esmeralda. (s.f.). Red Diamond Royal 6 botellas 500 ml [Imagen]. Aceite de Jaén. <https://www.aceitedejaen.es/aceite-de-oliva-virgen-extra/premium/verde-esmeralda-red-diamond-royal-6-botellas-500-ml>

- Georgetown Olive Oil Co. (s.f.). Olive oil cruets arabesco one handle [Imagen]. Georgetown Olive Oil Co. <https://www.georgetownoliveoil.com/collections/italian-ceramics/products/olive-oil-cruet-arabesco-one-handle>
- Picualia. (s.f.). Aceite de oliva virgen extra picual garrafa 5 l [Imagen]. Aceite de Jaén. <https://www.aceitedejaen.es/aceite-de-oliva-virgen-extra/gran-formato/picualia-familiar-picual-garrafa-5-l>
- Myst. (s.f.). Myst bio [Imagen]. Myst. <https://www.myst.gr/product/myst-bio/>
- Carbonell. (s.f.). Gran selección aceite de oliva virgen extra [Imagen]. El Corte Inglés. <https://www.elcorteingles.es/supermercado/0110120903205027-carbonell-gran-seleccion-aceite-de-oliva-virgen-extra-botella-250-ml/>
- Maeva. (s.f.). Aceite de oliva virgen extra spray [Imagen]. El Corte Inglés. <https://www.elcorteingles.es/supermercado/0110120903500492-maeva-aceite-de-oliva-virgen-extra-spray-200-ml/>
- Moli Coloma. (s.f.). Condimento de aceite de oliva aromatizado a la trufa negra spray [Imagen]. El Corte Inglés. <https://www.elcorteingles.es/supermercado/0110120903500658-moli-coloma-condimento-de-aceite-de-oliva-aromatizado-a-la-trufa-negra-spray-250-ml/>
- Pago de Espejo. (s.f.). Aceite de oliva virgen extra picual bag in tube 3 l [Imagen]. Aceite de Jaén. <https://www.aceitedejaen.es/aceite-de-oliva-virgen-extra/gran-formato/pago-de-espejo-picual-bag-in-tube-3-l>
- Sierra de Cazorla. (s.f.). AOVE Sierra de Cazorla Royal Verde Bag In Box 3l [Imagen]. Productos Cazorla. https://www.productoscazorla.com/aceites-de-oliva/625-AOVE-Sierra-de-Cazorla-Royal-Verde-Bag-In-Box-3l.html?gclid=CjwKCAjwhJukBhBPEiwAnilcNXV_WIExOZCUeknptbB0tb9wqVcakPzTTvf61DhYXGJ04BzbOcRgOBoC644QAvD_BwE
- García de la Cruz. (s.f.). Aceite de oliva virgen extra 3l bag in box [Imagen]. Aceites García de la Cruz. <https://www.aceitesgarciaidelacruz.com/tienda/clasicos/78-aceite-de-oliva-virgen-extra-3l-bag-in-box.html>

IV. Referencias del Dataset de imágenes

- Aceite de Jaén Online S.L (2019). Aceites de oliva virgen extra-Premium [Imagen]. Aceite de Jaén. <https://www.aceitedejaen.es/#premium>
- Euperion. (2021). Canard 34 oz Stoneware Olive Oil Bottle [Imagen]. Euperion. <https://euperion.co/products/canard-34-oz-stoneware-olive-oil-bottle?variant=44591112749307>
- Pereira O'Dell. (2019). Hero Bottle [Imagen]. AdForum. <https://www.adforum.com/creative-work/ad/player/34600897/callegari-olive-oil-brand-identity/callegari-olive-oil>
- The Dieline. (2010). Lintar Olive Oil [Imagen]. The Dieline. <https://thedieline.com/blog/2010/3/29/lintar-olive-oil.html>
- Ladolea. (2022). Extra Virgin Olive Oil 200ml/600ml Wooden [Imagen]. Ladolea. <https://ladolea.gr/products/olive-oils/extra-virgin-olive-oil-200ml-600ml-wooden/>
- Lukina, A, Ershov, R. (2016). Inizio [Imagen]. Behance. <https://www.behance.net/gallery/35571467/Inizio>



- Aceites Tapia. (2021). Verde Oleum / Can [Imagen]. En Aceites Tapia. Catalogue de produits (p. 10). https://www.aceitestapia.com/wp-content/uploads/2021/07/AceitesTapia_Catalogo_FRANCES_compressed.pdf
- The Bright Angle. (2022). Ceramic Olive Oil Dispenser Bottle [Imagen]. The Bright Angle. <https://thebrightangle.com/products/ceramic-olive-oil-dispenser-bottle>
- González, I. (2018). Aceite de oliva [Imagen]. UDIT. <https://udit.es/incubadora-creativa/del-olivo-al-vidrio/>
- Elenianna. (2022). FIVE Ultra Premium Extra Virgin Olive Oil 500ml [Imagen]. Elenianna. <https://www.elenianna.es/cinco-ultra-premium-aceite-de-oliva-virgen-extra-500-ml>
- Shah, J. (2012). Oliveo Olive Oil [Imagen]. Behance. <https://www.behance.net/gallery/3704170/Oliveo-Olive-Oil>
- Brachia. (2022). Brachia Ceramic Bottle Lemon 250 ml [Imagen]. Uje. <https://www.uje.hr/en/brachia-ceramic-bottle-lemon-250-ml/>
- Sistori, A. (2015). Olio D'Oliva [Imagen]. Behance. <https://www.behance.net/gallery/23633847/OLIO-DOLIVA>
- Artistica. (2022). Oliva: Olive Oil Bottle Deluxe [Imagen]. Artistica. <https://www.artistica.com/collections/olive-oil-vinegar-bottles/products/olive-olive-oil-bottle-deluxe-1041-oli>

ANEXOS



ANEXOS

1 Anexo I

1.1. Imágenes Generadas

Las imágenes recogidas para realizar las evaluaciones de calidad han sido elegidas de forma aleatoria dentro de un conjunto de imágenes significativas de cada modelo. Cada modelo tiene sus dimensiones reales, para establecer las diferencias de dimensiones entre modelos se ha escrito en la parte superior el nombre de cada modelo en el mismo tamaño de letra.

1.1.1. Resultados Dreambooth Stable Diffusion V1.5

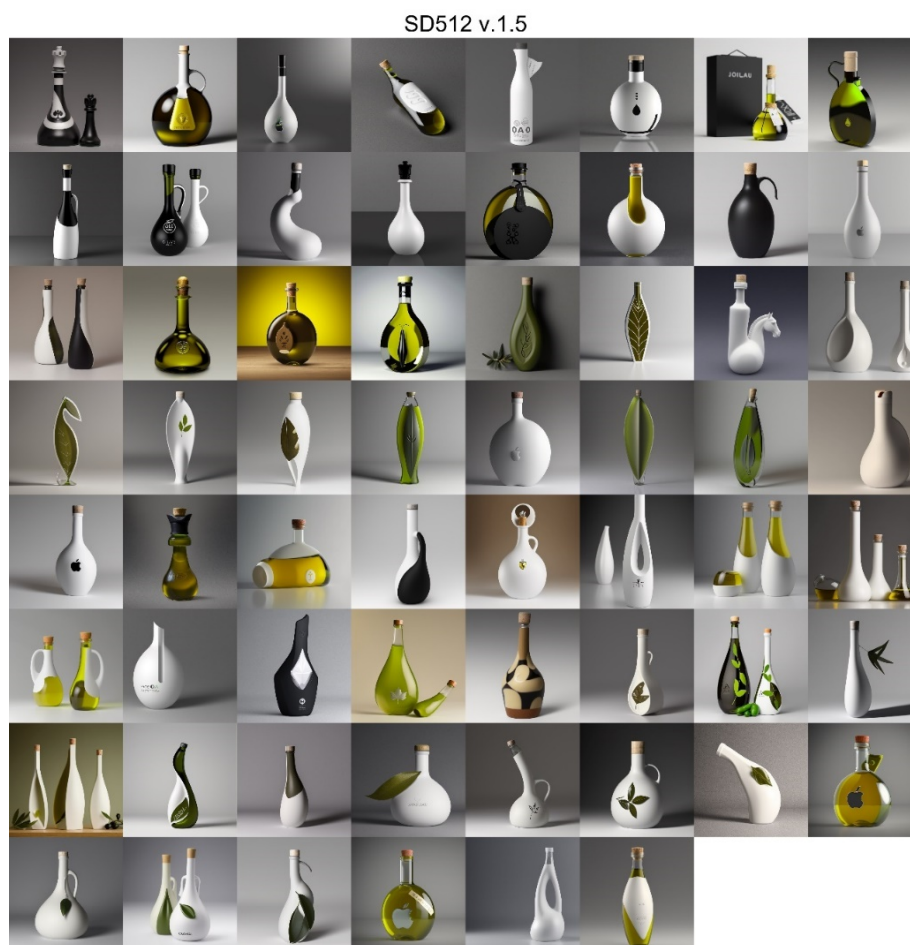


Figura 40. Conjunto de imágenes Dreambooth Stable Diffusion V1.5

1.1.2. Resultados Dreambooth Stable Diffusion V2.1 a 512 píxeles



Figura 41. Conjunto de imágenes Dreambooth Stable Diffusion V2.1 a 512 píxeles

1.1.3. Resultados Dreambooth Stable Diffusion V2.1 a 768 píxeles

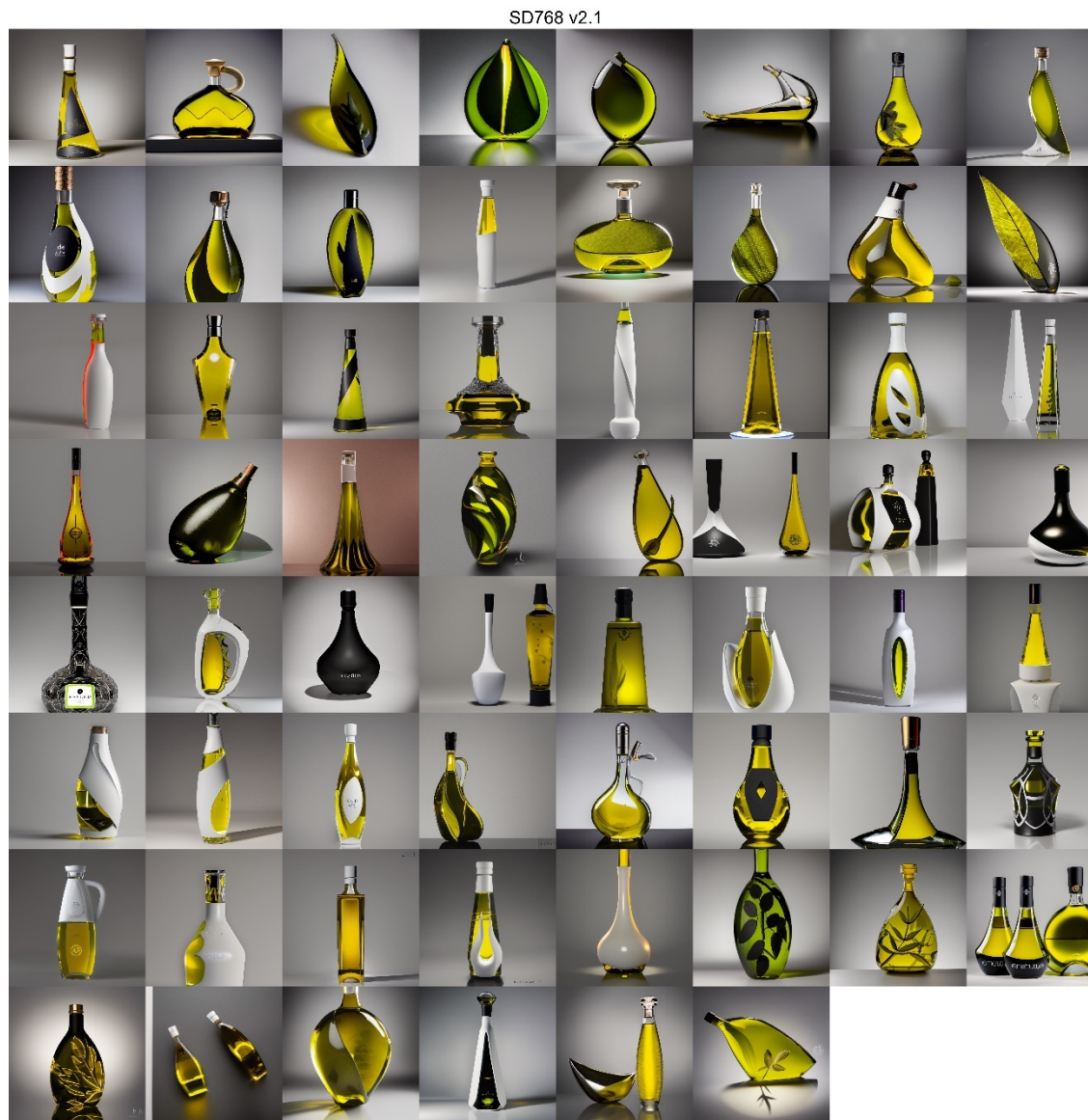


Figura 42. Conjunto de imágenes Dreambooth Stable Diffusion V2.1 a 768 píxeles

1.1.4. Resultados LoRA Stable Diffusion V1.5



Figura 43. Conjunto de imágenes LoRA Stable Diffusion V1.5

1.1.5. Resultados GLIDE a 64 píxeles

GLIDE_{x64}



Figura 44. Conjunto de imágenes GLIDE a 64 píxeles

1.1.6. Resultados GLIDE a 256 píxeles

GLIDEx256



Figura 45. Conjunto de imágenes GLIDE a 256 píxeles

1.1.7. Resultados DALL·E 2



Figura 46. Conjunto de imágenes DALL·E 2

1.1.8. Resultados Midjourney v4

Midjourney 512



Figura 47. Conjunto de imágenes Midjourney v4

2 Anexo II

2.1. Encuesta

En la realización de la encuesta se han desarrollado 13 preguntas, 3 preguntas de información del encuestado, 8 preguntas sobre las imágenes de los modelos entrenados y 2 preguntas acerca de la opinión de las imágenes de la Figura 48 y los nuevos modelos de inteligencia artificial aplicados al diseño. Las preguntas realizadas son las siguientes:

- ¿En qué rango de edad se encuentra?
- ¿Con que género te identificas?
- ¿Tiene usted experiencia o conocimientos en el diseño o producción de envases?
- En una escala del 1 al 5, ¿Qué tan creativo encuentras el diseño de cada producto?
(1= menos creativo, 5=más creativo)
- Seleccione preferencia del 1 al 12 las imágenes que crea que tienen mejor estética visual y le resulta un objeto más atractivo: *Siendo 12 el que más le resulte atractivo y 0 el que menos.*
- Si tuviera preferencia por un diseño, ¿Cuál elegiría? *Las siguientes preguntas irán con relación a esta elección.*
- Como primera impresión, ¿Qué crees que representa el producto?:
Botella de vino, botella de aceite, Jarrón y otro objeto decorativo, escribir otra opción.
- ¿Cree que sería un buen diseño de botella de aceite de oliva? Evalúe del 1 al 5 siendo 1 un pésimo diseño y 5 un excelente diseño.
- ¿Cree que ese diseño representa un envase de aceite de oliva premium? Evalúe del 1 al 5 siendo 1 un envase regular y 5 un envase premium.
- ¿Cree que puede tener algún fallo de diseño?
- Si fuera a comprar aceite, ¿Le llamaría la atención lo suficiente como para considerar comprar el producto?
- ¿Qué proceso crees que se utilizó para crear este diseño?
- ¿Qué opina del uso de la Inteligencia Artificial aplicado a la creación de nuevos diseños?

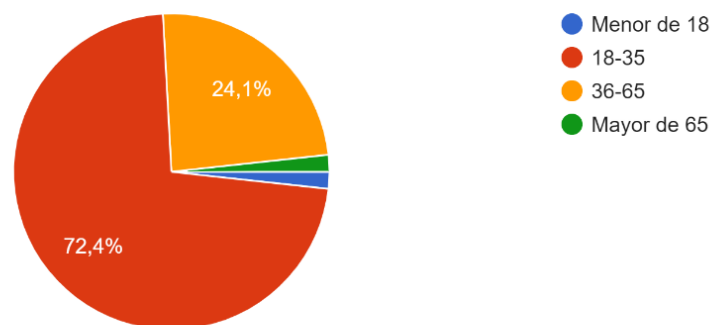


Figura 48. Selección de imágenes como base del formulario

En total se han recibido 58 respuestas. El perfil de los encuestados establece que el 72.4% tienen una edad entre 18-35 años, que la mayor parte de los encuestados son mujeres, un 53.4% y que solo un 12.3% tiene experiencia directa con el sector envase.

¿En qué rango de edad se encuentra?

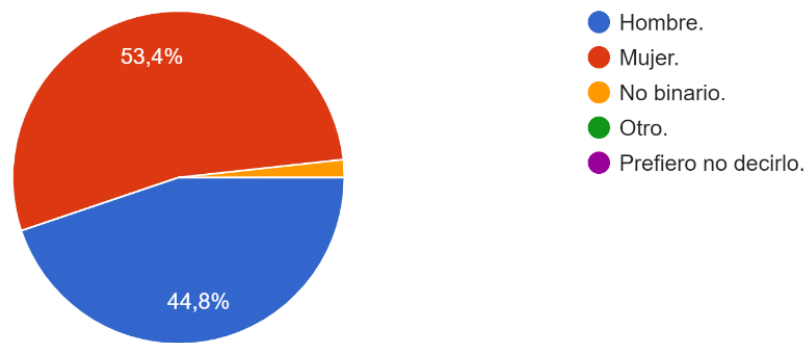
58 respuestas



Gráfica 11. Representación del rango de edad de los encuestados.

¿Con que genero te identificas?

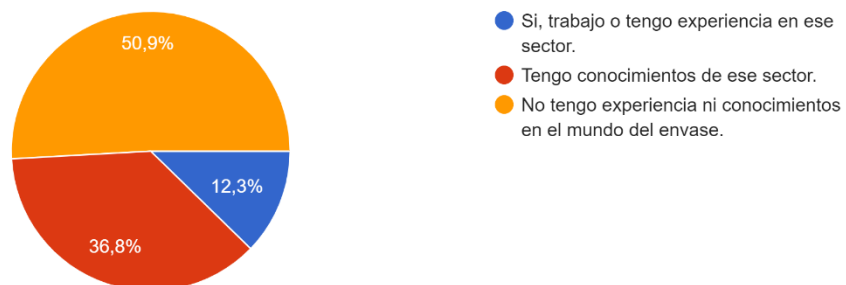
58 respuestas



Gráfica 12. Género del encuestado.

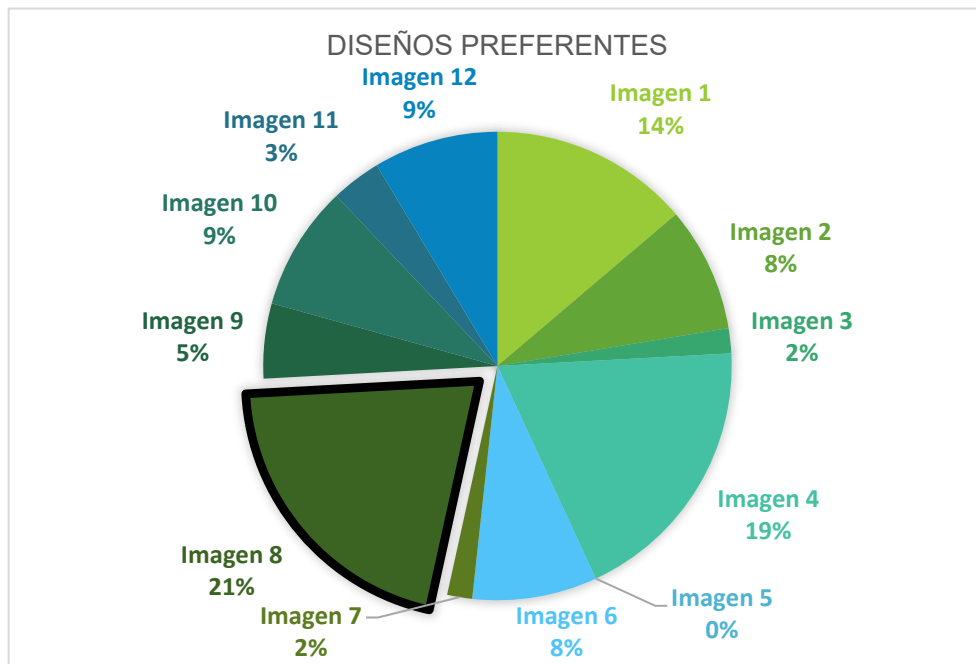
¿Tiene usted experiencia o conocimientos en el diseño o producción de envases?

57 respuestas



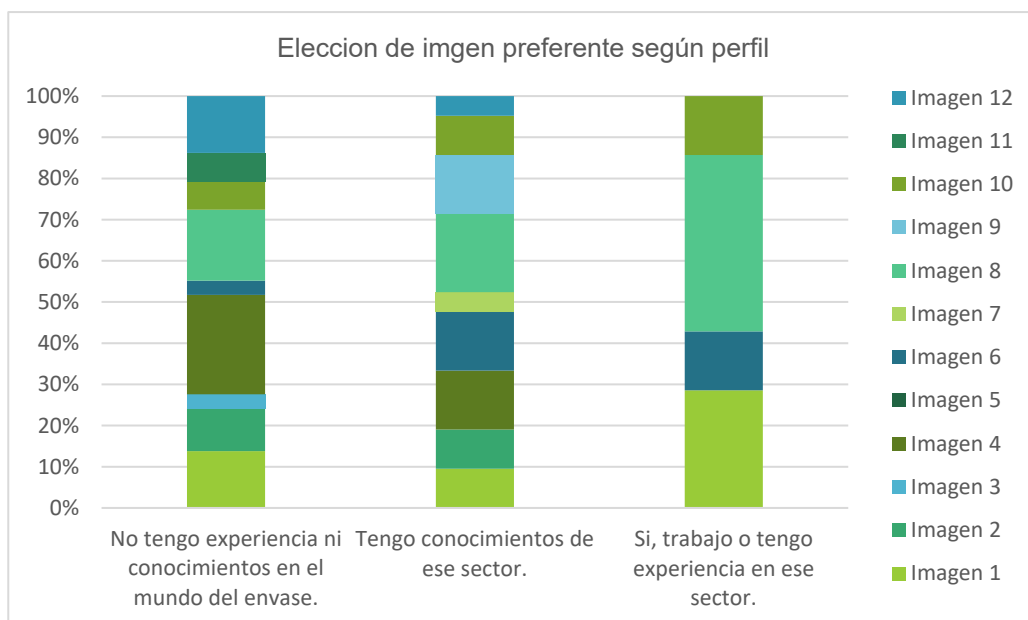
Gráfica 13. Experiencia en el sector envase del encuestado.

De entre todas las imágenes, los encuestados han repartido su valoración de preferencia y entre las imágenes mejor valoradas han sido en primer lugar la imagen 8, en segundo la imagen 4 y en tercer lugar la imagen 1.

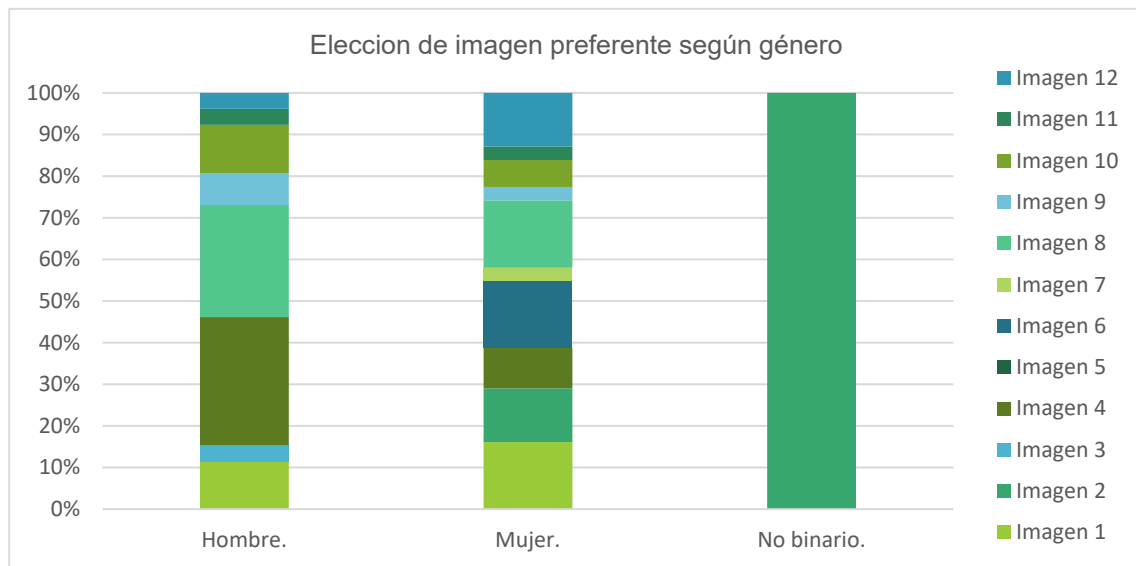


Gráfica 14. Gráfico del porcentaje de las imágenes preferentes por los encuestados.

Según el perfil del encuestado se puede comprobar como aquellos que trabajan en el sector tienen preferencia sobre los diseños de la imagen 8 y la imagen 1.

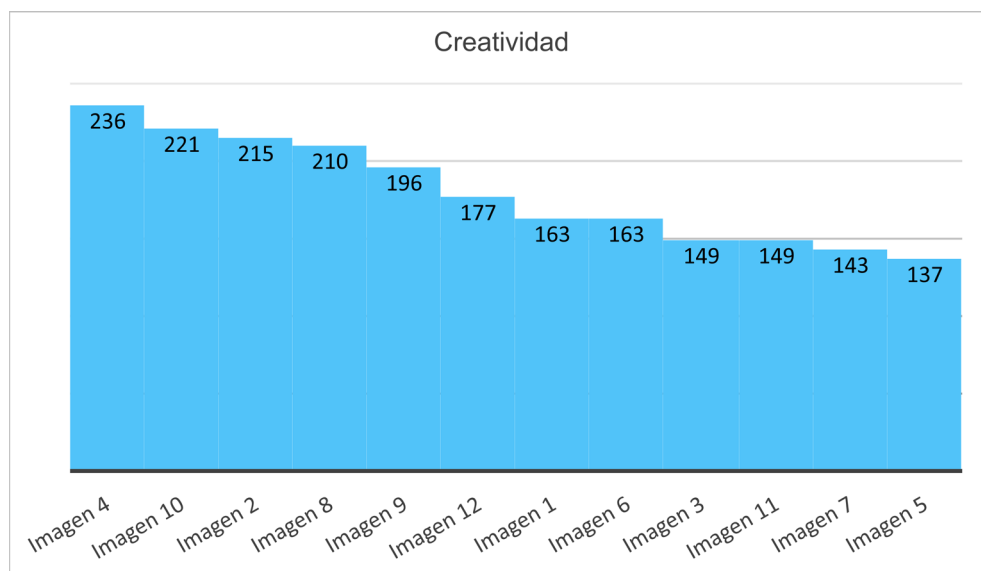


Gráfica 15. Gráfico según conocimientos de las imágenes preferentes por los encuestados.

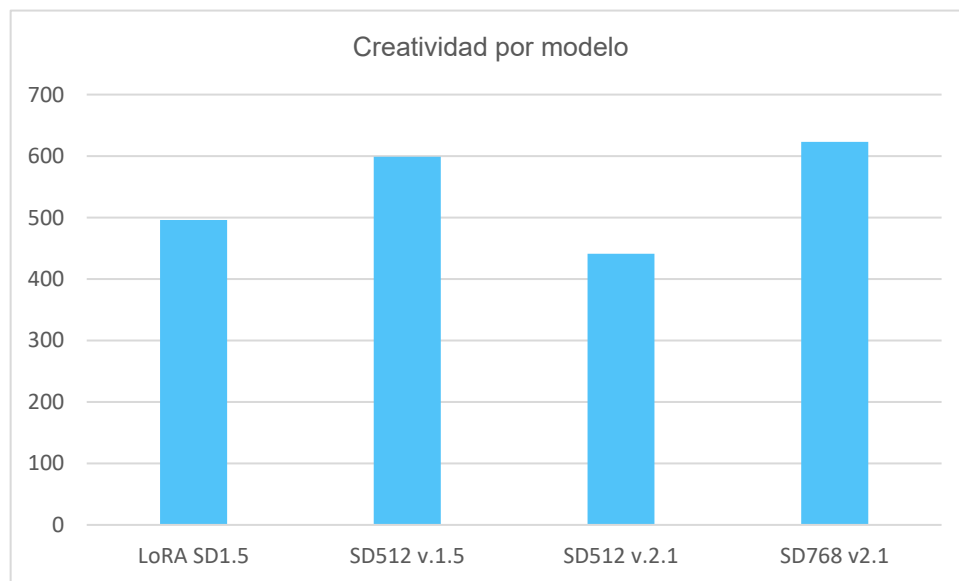


Gráfica 16. Gráfico según género de las imágenes preferentes por los encuestados.

Entre las imágenes que los encuestados encuentran más creativas, valoradas del 1 al 10, destacan la imagen 4, la imagen 10 y la imagen 2, aunque el valor global de cada modelo el modelo de Stable Diffusion en su versión 2.1 a 768 píxeles es que el que ha sido valorado más creativo.

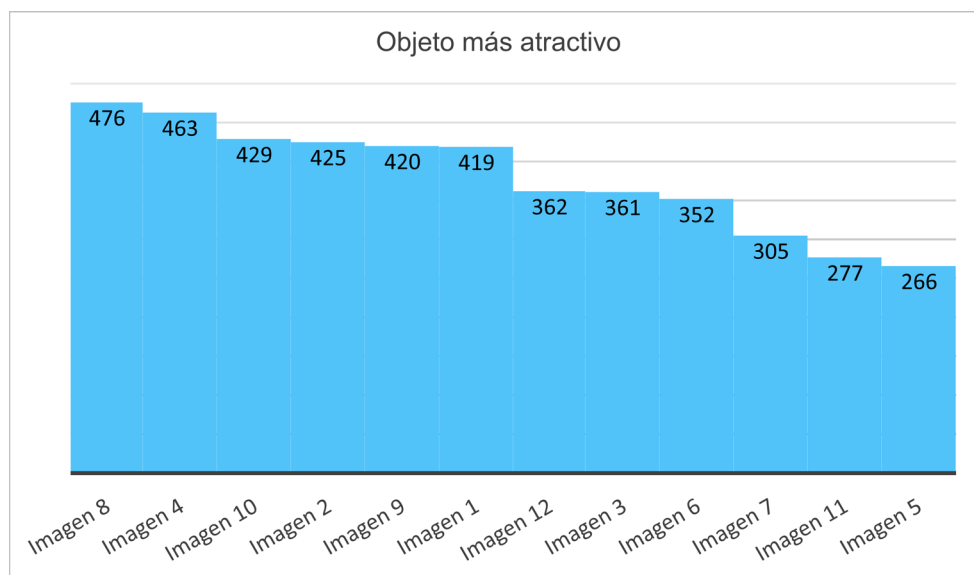


Gráfica 17. Valoración total de creatividad entre imágenes.

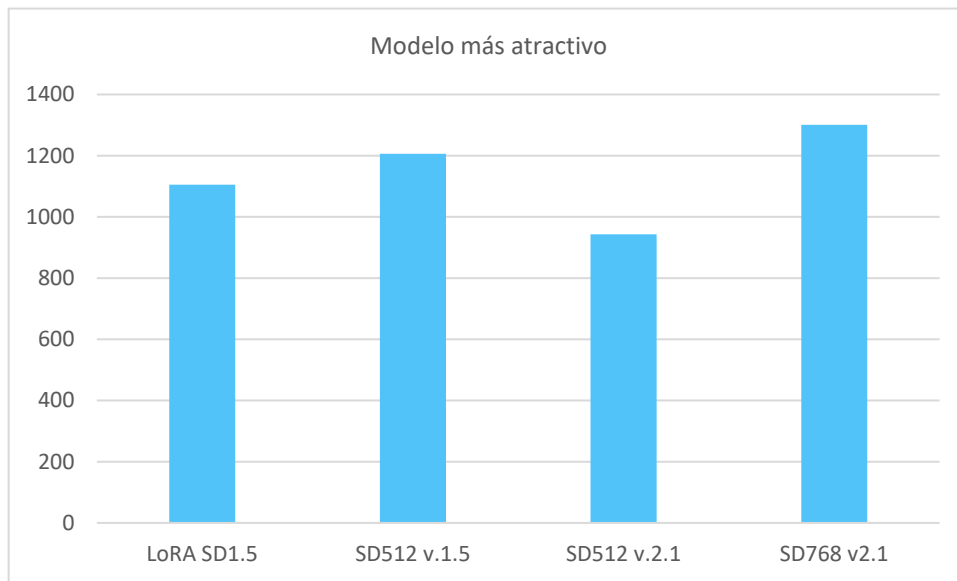


Gráfica 18. Valoración total de creatividad entre modelos.

En cuanto a imagen más atractiva, ordenadas por preferencia del 1 al 12, las imágenes mejor valoradas han sido la imagen 8, la imagen 4 y la imagen 10, en valor global el modelo de Stable Diffusion en su versión 2.1 a 768 píxeles vuelve a ser el preferido.



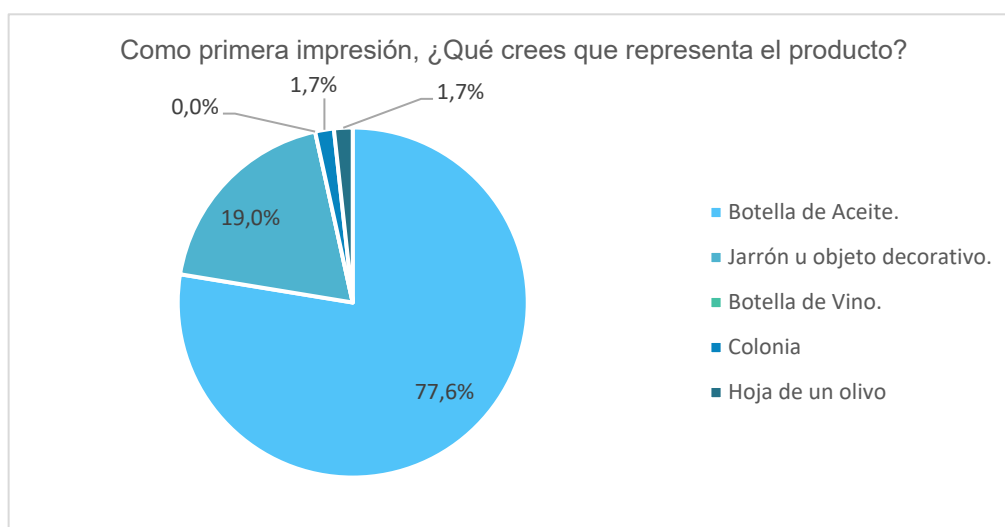
Gráfica 19. Valoración total de imagen con el objeto más atractivo.



Gráfica 20. Valoración total de modelos con el objeto más atractivo

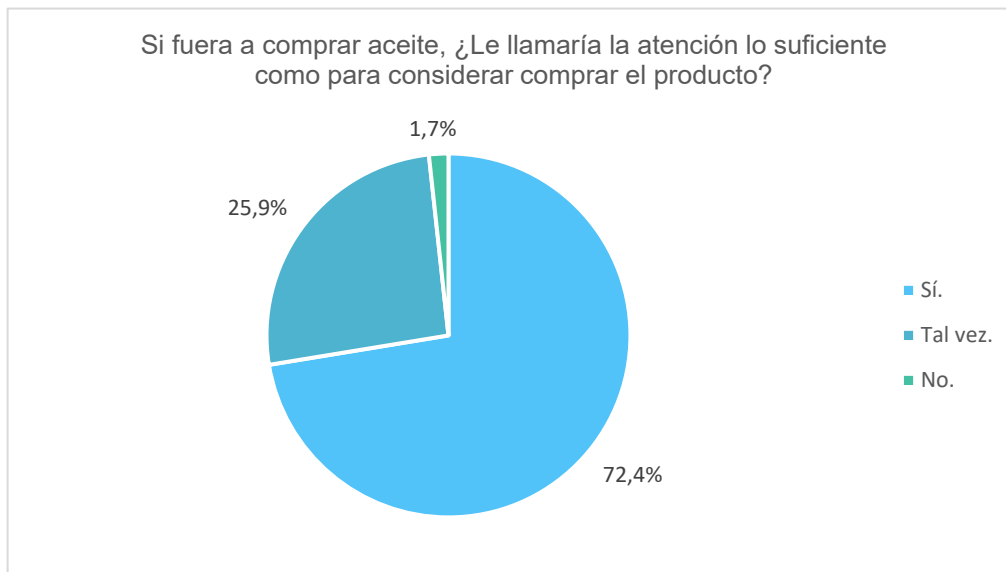
En conclusión, como modelo preferente, los encuestados prefieren las imágenes creadas por el modelo de Stable Diffusion en su versión 2.1 a 768 píxeles, y en concreto destacan la imagen número 8, imagen que será usada para el desarrollo del caso práctico.

También se ha recuperado información acerca de la opinión de la calidad del diseño técnicamente, el 77.6 cree que las imágenes representan una botella de aceite.



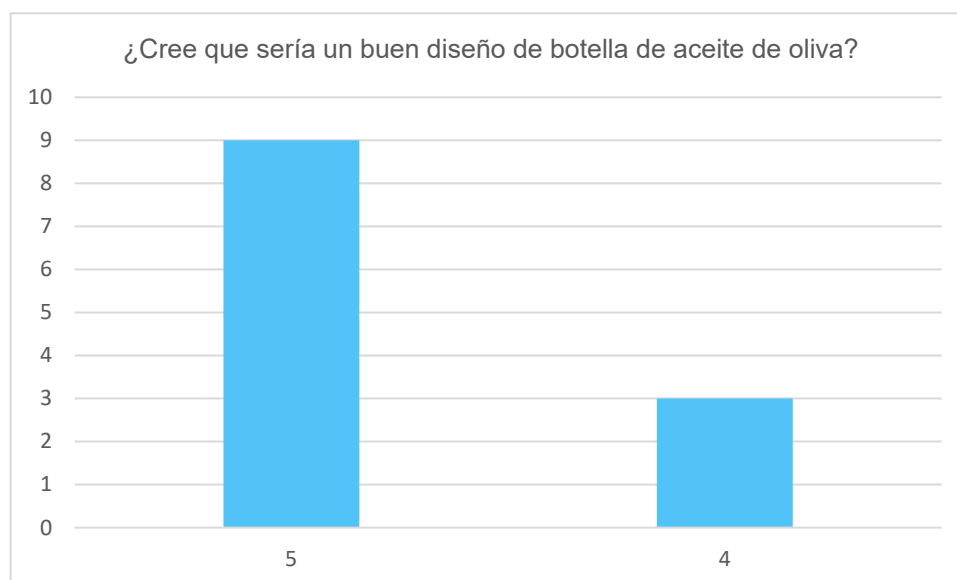
Gráfica 21. Pregunta de opinión de compra de todas las imágenes.

El 72.4 % cree que compraría ese producto si lo viera y un 25,9 pensaría en comprárselo.

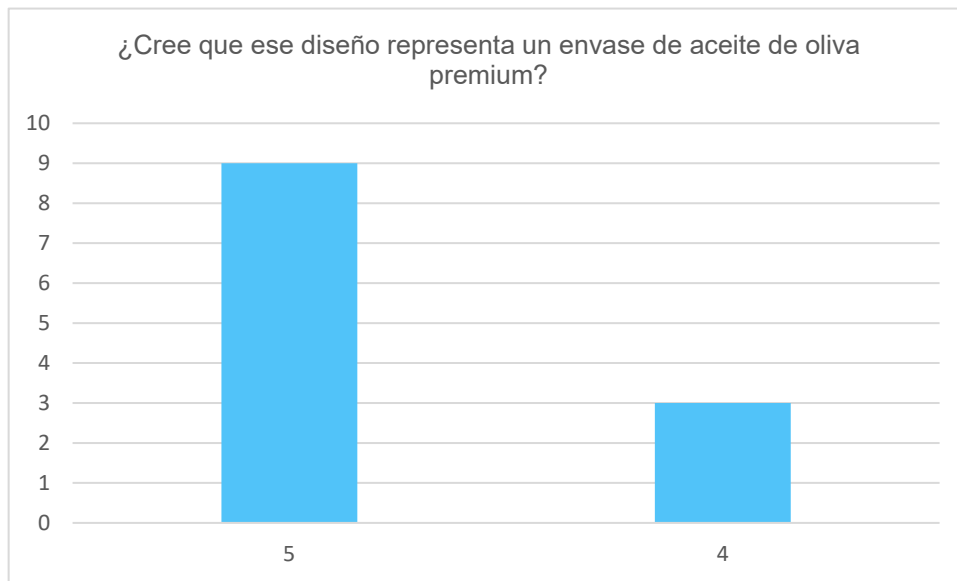


Gráfica 22. Pregunta de opinión de compra de todas las imágenes.

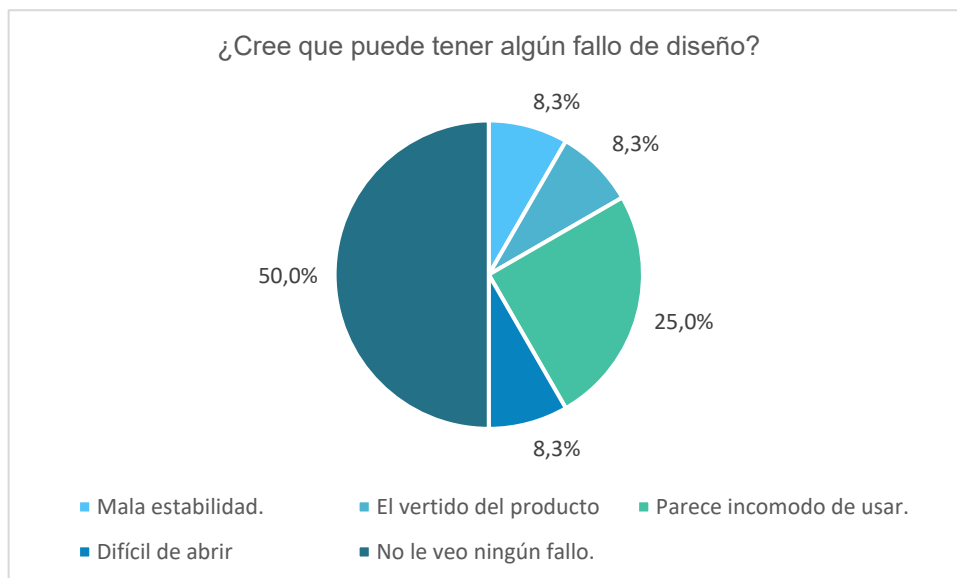
En el caso concreto de la imagen 8, la seleccionada para el desarrollo práctico, se ha recogido la valoración de si cree que sería una buena botella de aceite y si ve algún defecto en el diseño, donde el 50% no ve ningún fallo en el producto o hacen referencia a su posible incomodidad de uso.



Gráfica 23. Pregunta calidad del diseño seleccionado.



Gráfica 24. Pregunta de categoría del diseño del diseño seleccionado.

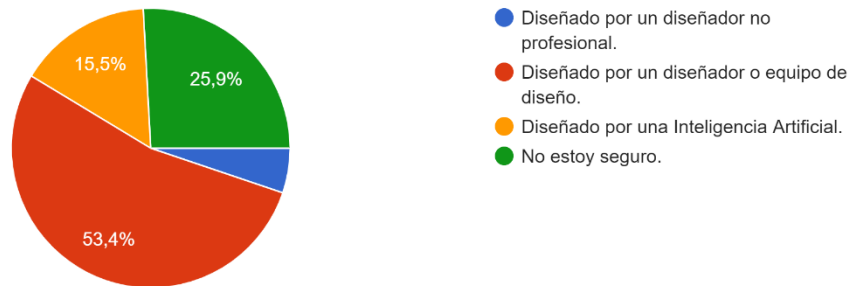


Gráfica 25. Pregunta sobre posible fallos de diseño del producto seleccionado.

En cuanto a la opinión de cómo se han creado las imágenes un 58,6% cree que han sido diseñadas por un humano y solo un 15,5% está seguro de que lo ha generado una IA. Referente a la opinión acerca de estas herramientas solo un 25,9% creen que puede ser perjudicial para el sector del diseñador y el resto cree que puede ser beneficioso de una manera u otra para el desarrollo de nuevos diseños.

¿Qué proceso crees que se utilizó para crear este diseño?

58 respuestas



Gráfica 26. Pregunta de opinión sobre los métodos de diseño utilizados.

¿Qué opina del uso de la Inteligencia Artificial aplicado a la creación de nuevos diseño?

58 respuestas

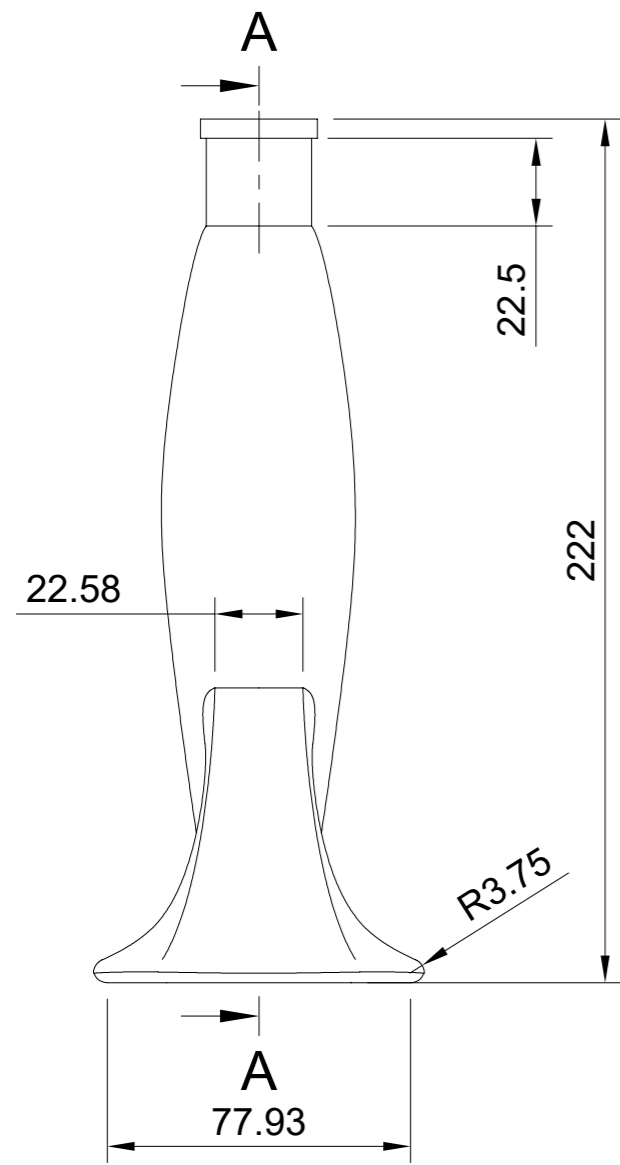


Gráfica 27. Opinión sobre el uso de la inteligencia artificial.

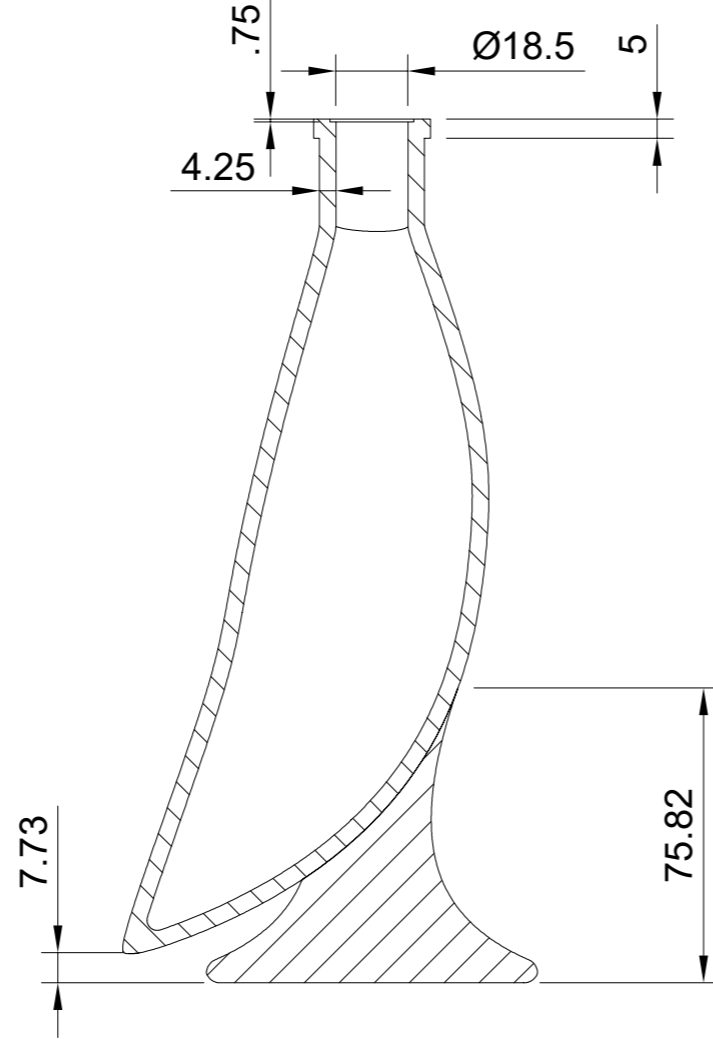
PLANOS



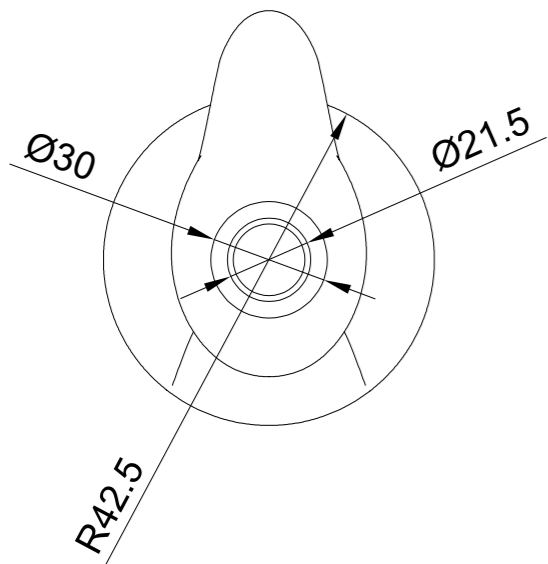
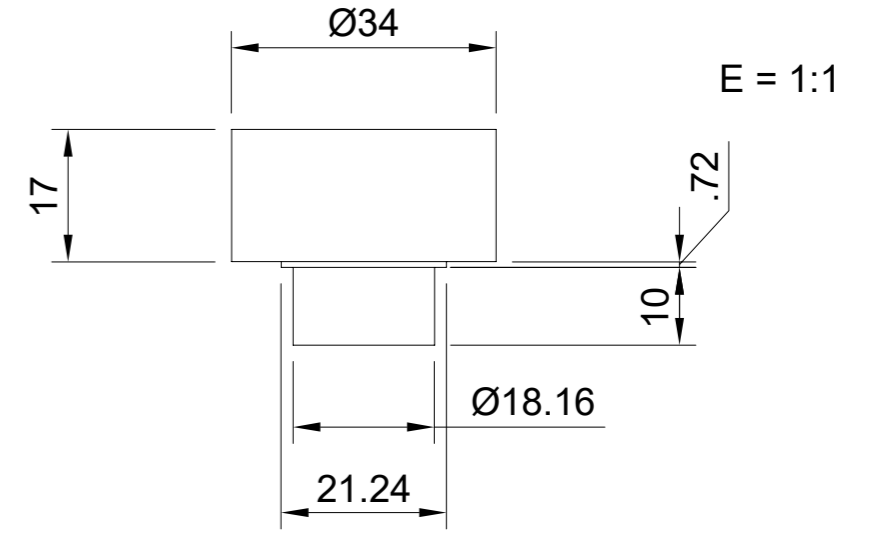
Cuerpo y Base




A-A (1:2)



Tapón



Creado por Óscar Gambero del Cid	
Escala 1/1	
Título Vision Conjunta Base, Tapón y Cuerpo	Numero de Documento 1
	Fecha. 10/06/2023