ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.2236

Diseño de un envase de aceite de oliva empleando técnicas de inteligencia artificial

Gambero del Cid, O. [1]; Domínguez-Merino, E. [2]; Trujillo-Aguilera, F.D.[3]; Blázquez- Parra, E. B.[4]
[1,2,3,4] Universidad de Málaga, España, *oscargdc20000@uma.es*, *enriqued@lcc.uma.es, fdtrujillo@uma.es, ebeatriz@uma.es*

Resumen*– Desde los inicios del diseño, el diseñador ha utilizado sus manos, su ingenio y su lápiz para transformar una idea en un producto funcional que cumpla una función. Con los avances en el aprendizaje profundo y las redes neuronales convolucionales, ha surgido una nueva herramienta: los modelos generativos de imágenes a partir de texto. Con el propósito de probar, evaluar y orientar estos modelos para aplicarlos al diseño de nuevos productos, nos hemos centrado en un campo de productos específico, los envases de aceite de oliva, un campo bastante amplio y perfecto para el uso de estos modelos, ya que solo son capaces de generar una imagen, no pueden proporcionar valores de medición ni volúmenes del objeto. Por lo tanto, los envases de aceite son perfectos como punto de partida, un producto cotidiano cuya forma y estética varían en función de la marca y la forma deseada. El objetivo es utilizar estas nuevas herramientas y darles su lugar de aplicación en el proceso de diseño conceptual de productos en el campo de vanguardia del desarrollo de productos. En este trabajo, estudiaremos qué modelos existen en el mercado y qué técnicas se pueden utilizar para orientar estos modelos. Evaluaremos los resultados obtenidos a través de diferentes métodos de evaluación de imágenes generadas por inteligencia artificial y las valoraciones obtenidas para proponer una metodología y un modelo para la ejecución de nuevos diseños conceptuales.

***Palabras clave*—diseño, envase de aceite, técnicas de inteligencia artificial, desarrollo de productos, estética**

I. Introducción

Actualmente, existen pocas integraciones de inteligencia artificial (IA) en el ámbito específico del diseño industrial; hay ejemplos de uso en los que se integra la arquitectura DALL-E 2 al software Fusion 360 para comenzar a diseñar a partir de una imagen generada por ese modelo [1]. Sin embargo, aún no se han encontrado ejemplos prácticos de entrenamiento de un modelo generativo para desarrollar diseños específicos para un sector, una imagen o un producto.

En los últimos años se han producido numerosos avances en el campo de la inteligencia artificial, el aprendizaje profundo y la creación de diferentes modelos generativos que buscan ofrecer nuevas herramientas para facilitar la realización de diferentes tareas que requieren esfuerzo humano. Estos avances abarcan diferentes campos, desde el análisis y la optimización de datos, la conducción autónoma y los modelos de procesamiento del lenguaje natural como GPT-3, capaces de escribir textos con un nivel de realismo casi humano, hasta la generación de imágenes. Estas herramientas sirven para mejorar y ampliar la ejecución de diferentes tareas que suelen estar limitadas por las capacidades humanas.

Con la aplicación de modelos generativos de imágenes, el objetivo es ampliar las opciones de diseño que puede generar un diseñador humano; en lugar de partir únicamente de la construcción manual de diez diseños conceptuales, la IA permite generar cientos de diseños al mismo tiempo, pudiendo completar estos diseños o partir de los diseños construidos y desarrollar variaciones o productos similares a los imaginados.

Hoy en día, la fase de desarrollo conceptual de un diseño comienza con una lluvia de ideas para satisfacer una necesidad. Se seleccionan las ideas más aceptables y se realizan diseños conceptuales para cumplir con los requisitos de diseño. Este proceso puede llevar más o menos tiempo dependiendo del diseñador y está condicionado por la creatividad de la persona o equipo que crea la idea, con sus conocimientos, experiencia y prejuicios asociados. Estas características también están asociadas al funcionamiento de estos modelos generativos, que están condicionados por la información de entrenamiento si ciertos conceptos son limitados o si un concepto está asociado a una definición limitada, lo que solo puede disminuirse aumentando la información y los conocimientos de entrenamiento.

El objetivo de este estudio es aplicar una nueva herramienta al proceso de diseño que pueda facilitar o ampliar las ideas para nuevos diseños, comprobar su viabilidad y llegar a un proceso de diseño en conjunción con estos nuevos modelos generativos.

II. Envases

Los envases que se encuentran actualmente en el mercado (Fig. 1) son muy variados y creativos, tanto en forma como en diseño y estética, y van desde envases con estructuras orgánicas y naturales hasta estructuras estilizadas compuestas por líneas rectas y acabados minimalistas. Los materiales utilizados son bastante flexibles, desde el plástico para los diseños más económicos, pasando por la cerámica para los más tradicionales, hasta el vidrio, el más utilizado entre las botellas de aceite. Esta flexibilidad en la forma y los materiales hace que sea un producto que se desarrolla rápidamente mediante un modelo de inteligencia artificial generativa.

Fig. 1 Ejemplo de productos del mercado.

El aceite de oliva es un producto muy solicitado que suele ocupar una gran parte del espacio comercial en los grandes centros comerciales. Es uno de los productos de consumo más demandados junto con productos básicos como el vino o la leche, y se encuentra entre los productores locales de aceite. Este producto tiene demanda tanto como alimento básico como producto de regalo y, dependiendo del diseño del producto, como objeto decorativo.

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

• Envases de hojalata: Estos envases están muy extendidos por su funcionalidad y durabilidad. En cuanto a su forma y tamaño, son recipientes cuadrados que van desde los 200 ml hasta grandes recipientes de 5 litros.

• *Recipientes de vidrio:* son los más comunes en el mercado por su durabilidad y aspecto estético. Estos recipientes suelen ser transparentes y permiten observar el contenido para apreciar el color y la calidad del aceite.

Aunque la mayoría de los productos cubren el envase con una tapa para vender la marca y proteger el producto del sol. Desde 25 ml para botellas de regalo hasta 1 litro, siendo los más comunes los de 500 ml.

• *Envases de cerámica:* Estos envases tienen una estética más rústica y natural y se distinguen de otros envases, ya que no permiten ver el contenido del producto, lo que hace que su estética destaque más que la de otros envases, lo que los hace perfectos como regalos o productos gourmet y funcionan como objeto decorativo en el hogar. Los volúmenes de estos envases oscilan entre 200 ml y 750 ml.

• *Envases de PET* (polietileno tereftalato): Estos envases de plástico están pensados para un producto de menor coste, son fáciles de transportar y más económicos que otros envases. Son ligeros y fáciles de transportar, con capacidades de entre 1 y 5 litros, pensados para un consumo a gran escala.

• *Envases aerosoles*: Estos envases suelen estar fabricados en aluminio o vidrio con acero inoxidable. Son envases que permiten dosificar el aceite en forma de spray para tener un mayor control de la cantidad consumida. Suelen venir en botellas de 200 ml, ya que gran parte del contenido es gas a presión que permite pulverizar el aceite fácilmente.

• *Envases bag-in-box:* Estos envases pretenden ser los más sostenibles desde el punto de vista medioambiental, ya que almacenan el aceite en una bolsa reciclable y lo contienen en una caja de cartón con un grifo para extraer el contenido.

El tipo o variedad de aceite que contiene cada botella determina el precio de venta del producto, y la botella 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, dependiendo del olivo del que se extraen las aceitunas.

Para llevar a cabo este estudio, se va a partir como premisa aquellos envases que llegan a una gran parte de los consumidores, como son los envases de vidrio o cerámica, buscando una estética más clásica y limpia centrada en el envasado de aceite de calidad. Preferiblemente, envases rellenables y sostenibles desde el punto de vista medioambiental y destinados a una variedad de aceites de alta calidad. Sobre todo, se buscarán productos con la mayor variedad de formas y estilos que puedan representar una botella de aceite de oliva.

III. Metodología

A. Modelos y Métodos de entrenamiento

Para el desarrollo del estudio, se tendrán en cuenta los diferentes modelos de síntesis de imágenes a partir de texto. Todos ellos comparten conceptos como la difusión y el uso de herramientas como CLIP y redes generativas adversarias.

Todos estos modelos generativos funcionan utilizando un sistema de difusión, en el que se añade iterativamente ruido gaussiano a una imagen, y se entrena un modelo para reconstruir el ruido aplicado a la imagen [2] utilizando el sistema U-Net para codificar y decodificar cada imagen.

CLIP es un modelo de aprendizaje profundo que, utilizando redes convolucionales y transformadores, puede realizar tareas de visión artificial y procesamiento del lenguaje natural [3] [4].

Las GAN (redes generativas adversarias) 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 y evalúa los datos. La red generadora crea nuevas muestras a partir de ruido aleatorio, y la red discriminadora evalúa si estas muestras son reales o falsas. El objetivo del generador es crear muestras lo suficientemente realistas como para engañar al discriminador.

1) Modelos de Código abierto.

Estos modelos están totalmente disponibles para ser ejecutados, editados y transformados; su código es abierto, lo que permite realizar mejoras en conceptos específicos.

• GLIDE: uno de los primeros modelos de código abierto creados tras el desarrollo de la arquitectura VQGAN+Clip por la empresa OpenAI, entrenado con 3,5 millones de parámetros, que ofrece la capacidad de generar imágenes a partir de texto, realizar rellenos, rellenar huecos delimitados con una máscara de una imagen y la capacidad de entrenar el modelo con su banco de imágenes.

• Stable Diffusion: es el modelo que ha tenido un impacto más significativo entre estos sistemas generativos y el que ha contado con un mayor apoyo para su desarrollo por parte de la comunidad. Con este modelo, ha sido posible adaptar numerosas aplicaciones y adaptaciones para mejorar su funcionamiento y ofrecer diferentes formas de generar imágenes, ya sea a partir de texto, de otra imagen, de una estructura o de mapas de profundidad.

2) Modelos Cerrados

Estos modelos no son de código abierto y solo se pueden utilizar a través de los servicios de la empresa que los controla; aunque son cerrados, pueden ofrecer opciones para condicionar el modelo y generar imágenes de alta calidad, y vale la pena considerar sus posibilidades, ya que se pueden combinar con los otros modelos y editar una imagen creada en ellos, dándole el estilo de un modelo preentrenado.

• DALL-E 2: el modelo DALL-E 2 es la evolución de GLIDE a un modelo a mayor escala. Puede generar imágenes de 1024 píxeles, pero a diferencia de GLIDE, su uso está cerrado a través de los servidores de OpenAI y permite generar variaciones de una imagen, rellenar espacios en una imagen o ampliar su contenido, y generar a partir de texto. Es un modelo entrenado con 12 000 millones de parámetros y es multimodal con GPT-3. Este modelo está entrenado en imágenes más artísticas y, en general, tiende a producir imágenes más relacionadas con pinturas o dibujos que con imágenes fotorrealistas.

• Midjourney versión 4: este modelo fue creado por la empresa del mismo nombre, funciona a través de un servidor Discord, ofrece numerosos parámetros para ajustar las dimensiones de la imagen y ofrece diferentes modelos de estilo, pero ninguno centrado en la creación de productos.

3) Métodos y Arquitectura de entrenamiento

• LoRA se basa en un sistema de entrenamiento que utiliza cambios en las características locales para reproducir cambios globales en la imagen. El método consiste en simplificar los pesos del modelo utilizando una técnica de descomposición. Esto implica expresar cada peso como el producto de dos matrices más pequeñas, una que se mantiene constante y otra que se entrena con los datos de destino. Esta estrategia reduce el número de parámetros que se deben entrenar y evita que el modelo se ajuste excesivamente a los datos. Además, se añade una conexión especial entre la entrada y la salida de cada capa para preservar la información original y facilitar el aprendizaje de los cambios locales [5].

• Dreambooth es otro método de adaptación de modelos de texto a imagen, basado en el uso de una capa adicional llamada DreamBooth Layer (DBL) situada entre la entrada de texto y la salida de imagen del modelo preentrenado. La 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 [6].

• Fine-tunnning es una técnica que consiste en adaptar un modelo preentrenado a un problema específico, aprovechando el conocimiento del modelo sobre el dominio general. El ajuste fino se realiza modificando algunos parámetros del modelo, como los pesos de las capas finales o la tasa de aprendizaje, y entrenando el modelo con un conjunto de datos del problema objetivo. De esta manera, se puede mejorar el rendimiento del modelo y obtener resultados más precisos y personalizados.

Para llevar a cabo el entrenamiento, es necesario crear una base de imágenes que indiquen lo que se desea crear. Para ello, las imágenes deben ser claras y nítidas; cuanto mayor sea el detalle y la calidad de la imagen, mejor será el entrenamiento y mejores serán los resultados (Fig. 2). Para comprender mejor lo que se desea entrenar, se recomienda que el objeto se encuentre en un fondo aislado y que solo se vea el objeto. Si se desea entrenar un objeto específico, es aconsejable variar el fondo de la imagen tanto como sea posible. Para la búsqueda de imágenes, se han consultado grandes bases de datos de imágenes como LAION, COCO, Imaginet y Open Image. En estas bases de datos se han buscado imágenes de botellas de aceite de oliva y se han encontrado pocas o ninguna, lo cual es interesante, ya que modelos como Stable Diffusion se entrenan con las bases de datos LAION, lo que indica que es necesario un refinamiento para obtener imágenes cercanas a un recipiente de aceite. Para llevar a cabo el entrenamiento, se establecieron estándares que se utilizarían a lo largo del mismo, y se comenzó utilizando el modelo que mejor reciben los usuarios y más estable: el modelo Stable Diffusion en su versión 1.5. El entrenamiento también se llevó a cabo a través de GLIDE y LoRA.

Fig. 2 Conjunto de imágenes de la base de datos.

B. 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,

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 [7].

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 [8].

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 puntuaje 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 [9].

• 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 [10].

• 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 [11].

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, atractivo y mejor estética visual del diseño.

IV. Resultados

A. Resultados de la compilación

Una vez entrenados los modelos necesarios, se llevó a cabo una recopilación de muestras, realizando inferencias para cada fuente abierta. Se cerró el modelo de código fuente para realizar un estudio comparativo y determinar qué modelo sería más favorable para la realización de este proyecto.

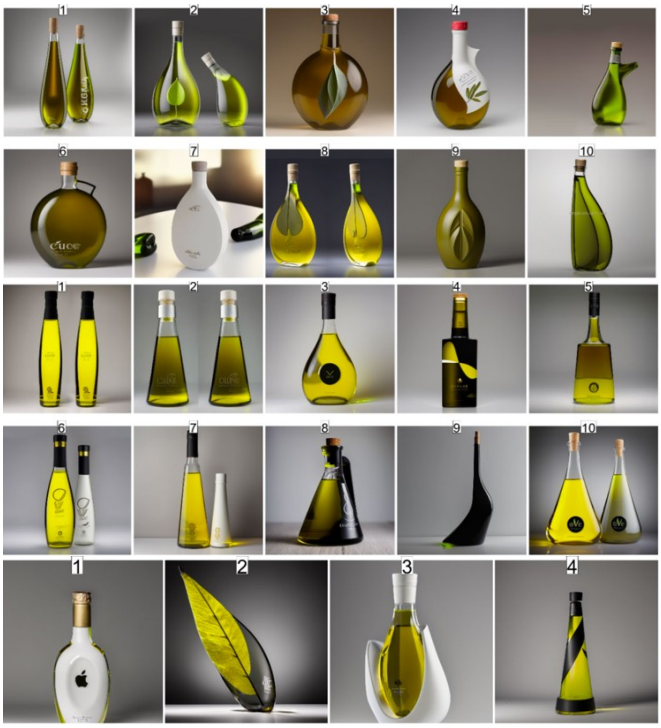

Fig. 3. Diferentes modelos de envase obtenidos por inteligencia artificial.

Se ha obtenido un gran número de imágenes (Fig. 3), que deben compararse utilizando una métrica con la idea de establecer qué modelo es, en general, el que produce la mejor imagen o el que tiene la mejor calidad de imagen, así como decidir qué modelo es más favorable para el desarrollo de un envase de aceite de última generación, en este caso concreto.

Para ello, en los textos de entrada de todos los modelos, se han establecido los mismos conceptos, adaptando cada modelo según sus necesidades y esperando obtener un resultado mediante la descripción; estos textos de entrada, en su forma más simple, son los siguientes:

• Una botella de aceite de oliva al estilo de un ordenador Apple
• Una botella de aceite de oliva con forma de hoja
• Una botella de aceite de oliva con forma curvada
• Una botella de aceite de oliva de estilo futurista

B. Resultados de la 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.

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.

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

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 (Fig.4 y tabla 2).

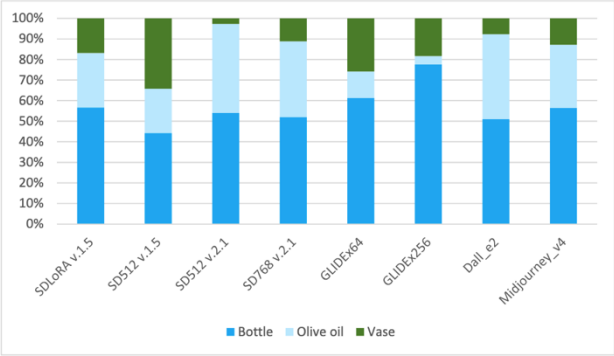

Fig. 4 Distribución porcentual de la Repetición de Palabras.

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

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

Los resultados de la figura 5 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.

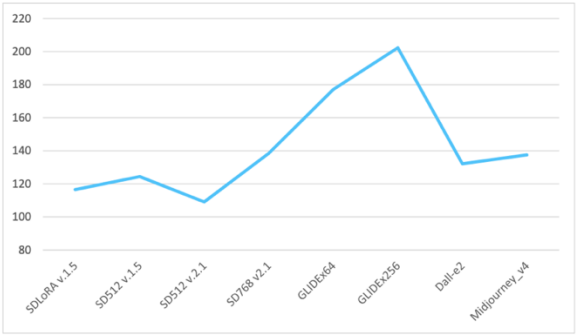

Fig. 5. Distribución de los valores de FID por modelo.

Los resultados mediante la métrica de Brisque de la Tabla 4 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 4: 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**	5,601
	Midjourney_v4	54,973	7,357

En la métrica NIQE (tabla 5), 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 5: 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

La métrica de PIQE (tabla 6) 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 6: 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

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 (Fig. 6).

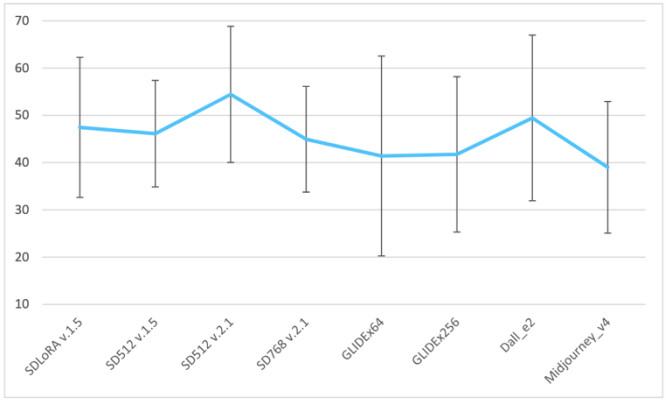

Fig. 6. Distribución de los valores del promedio de contraste.

En la encuesta realizada 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 imagenes 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 (Fig.7).

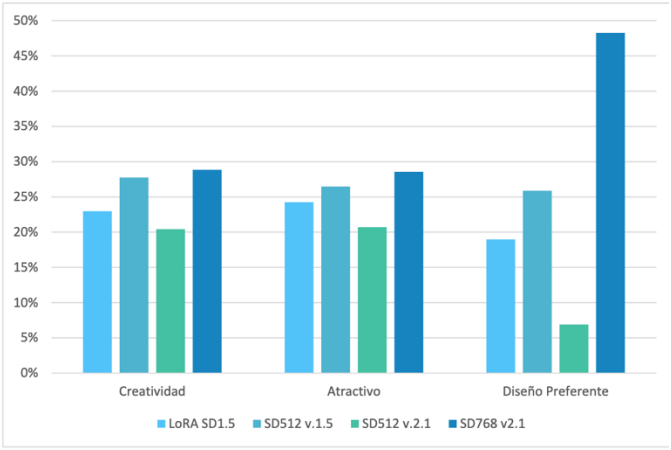

Fig. 7. Distribución porcentual de las preferencias del encuestado por modelo.

B. Análisis de los 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 7).

Tabla 7: Comparativa 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
	GLIDEx64	46	56,49%	0,4018	177,1	35,22	18,87	40,56	41,41	7,98
	GLIDEx256	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 8 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.

Fig. 8. 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. Una vez recogidos todos los datos se ha realizado una ponderación de los criterios de la Tabla 8, 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 8: Resultado del conjunto de evaluaciones.

	ME						MNE	
Criterio	SDLoRA v.1.5	SD512 v.1.5	SD512 v.2.1	SD768 v.2.1	GLIDEx64	GLIDEx256	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

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 izquierda de la figura 9, diseño creado por el modelo de SD en su versión 2.1 a 768 píxeles y reflejada.

Fig. 9. Imagen generada por SD en su versión 2.1 a 768 píxeles seleccionada para caso práctico e imágenes realizadas de diseño en Fusion 360 y estilo realista en Blender

V. Conclusiones

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 siendo 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 valido para el desarrollo de un envase de aceite de oliva, preferentemente el usuario tiene que elegir con cual se siente más cómodo a lo hora de usar una de estas herramientas de trabajo.

Por su parte, Stable Difussion en su versión 1.5 es con el que mejores resultados se ha obtenido a largo plazo.

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. 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.

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 humano, 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.

Como conclusión final, 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.

AGRADECIMIENTO/RECONOCIMIENTO

En esta sección puede agradecer a las personas e instituciones que contribuyeron con el desarrollo del estudio.

REFERENCIAS

[1] Z. Lu, R. H. Kazi, L. Y. Wei, M. Dontcheva, and K. Karahalios, "StreamSketch: Exploring Multimodal Interactions in Creative Live Streams," Proceedings of the ACM on Human-Computer Interaction, 5 (CSCW1), 2021.

[2] J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi, and T. Salimans, "Cascaded Diffusion Models for High Fidelity Image Generation," Computer Vision and Pattern Recognition, pp. 1–33, 2021.

[3] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning Transferable Visual Models From Natural Language Supervision," Computer Vision and Pattern Recognition, pp. 1–48, 2021.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Computer Vision and Pattern Recognition, pp. 1–8, 2015.

[5] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-Rank Adaptation of Large Language Models," Computation and Language, pp. 1–26, 2021.

[6] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, K., "DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation," Computer Vision and Pattern Recognition, pp. 1–25, 2.

[7] 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

[8] 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

[9] 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

[10] Mittal, A., Soundararajan, R., & Bovik, A. C. (2012). Making a "Completely Blind" Image Quality Analyzer. http://live.ece.utexas.edu/research/quality/niqe

[11] Wang, L. (2021). A survey on IQA. http://arxiv.org/abs/2109.00347