



INSTITUTO TECNOLÓGICO DE ESTUDIOS SUPERIORES DE MONTERREY  
ESCUELA DE INGENIERÍA Y CIENCIAS  
CONCENTRACIÓN INTELIGENCIA ARTIFICIAL AVANZADA PARA LA CIENCIA DE DATOS  
MONTERREY, NUEVO LEÓN

TC3007C.503. INTELIGENCIA ARTIFICIAL AVANZADA PARA LA CIENCIA  
DE DATOS II

IMPLEMENTACIÓN DE MODELOS GENERATIVOS DE I.A. PARA LA  
CREACIÓN DE PUBLICIDAD

ARCA CONTINENTAL  
**CÉSAR ALEJANDRO CRUZ SALAS A00825747**  
**FRANCISCO JOSÉ JOVEN SÁNCHEZ A00830564**  
**DAVID EMILIANO MIRELES CÁRDENAS A01633729**  
**MARIO JAVIER SORIANO AGUILERA A01384282**  
**MARLON BRANDON ROMO LÓPEZ A00827765**

## Índice

<b>1. Introducción</b>	<b>3</b>
<b>2. Problema/Reto</b>	<b>3</b>
<b>3. Propuesta de Solución</b>	<b>3</b>
<b>4. Licencias</b>	<b>4</b>
4.1. ChatGPT 3.5-Licencia de uso comercial . . . . .	4
4.2. Stable Diffusion (SDXL) Creative ML OpenRAIL-M (Machine Learning Open-Source for Creativity) . . . . .	5
4.3. Automatic1111- GNU AFFERO GENERAL PUBLIC LICENSE . . . . .	5
4.4. Microsoft Designer . . . . .	6
<b>5. ChatGpt-3.5 (Generación de prompts para modelo de imágenes)</b>	<b>6</b>
5.1. Objetivo . . . . .	6
5.2. Entrenamiento . . . . .	7
5.3. Resultados . . . . .	8
<b>6. ChatGpt-3.5 (Generación de ideas y texto para el anuncio)</b>	<b>9</b>
6.1. Objetivo . . . . .	9
6.2. Entrenamiento . . . . .	9
6.3. Resultados . . . . .	10
<b>7. Stable Diffusion XL (SDXL)</b>	<b>11</b>
7.1. Beneficios . . . . .	11
<b>8. Tipos de entrenamiento para SDXL</b>	<b>12</b>
8.1. HyperNetworks . . . . .	12
8.2. Dreambooth . . . . .	13
8.3. LoRA (Low-Rank-Adaptation) . . . . .	13
8.4. Textual Inversion . . . . .	14
<b>9. Entrenamiento de SDXL</b>	<b>14</b>
9.1. AutoTrain . . . . .	15
9.2. Dreambooth-LoRA . . . . .	15
9.3. Resultados . . . . .	15
<b>10. Automatic 1111 como UI</b>	<b>16</b>
10.1. Beneficios . . . . .	18
10.2. Herramientas mas importantes . . . . .	18

<b>11. Microsoft Designer</b>	<b>20</b>
11.1. Beneficios . . . . .	20
11.2. Resultados . . . . .	21
<b>12. STEFANN</b>	<b>22</b>
12.1. Objetivo . . . . .	22
12.2. Uso . . . . .	23
12.3. Resultados . . . . .	23
<b>13. Conclusión</b>	<b>23</b>
<b>14. Anexos</b>	<b>24</b>

## 1. Introducción

En los últimos años, el auge de la inteligencia artificial (IA) ha impactado los procesos de negocio de los participantes de la mayoría de las industrias. La inteligencia artificial generativa tiene la capacidad de generar imágenes, texto e incluso videos en algunas ocasiones, bajo las especificaciones que un usuario puede establecer. Esta rama particular de la inteligencia artificial tiene la capacidad de irrumpir en áreas de trabajo donde se creía que las computadoras no podían competir contra la creatividad humana: artes visuales, diseño, escritura e incluso música. Para las industrias donde esta clase de actividades creativas forman una parte importante, la IA generativa tiene la capacidad de transformar la forma de trabajo, reduciendo el costo y el tiempo que toma hacer diseños, escritos o imágenes. Sin embargo, la IA en general continua en un proceso de adopción por el público general, además de que conlleva retos relacionados a la necesidad de poder computacional.

Como muchos de los desarrollos en las tecnologías de software de las últimas décadas, los avances en la IA generativa son en la mayoría de las ocasiones contribuciones “open source”, es decir, abiertos completamente para que el público general pueda descargar, modificar y utilizar el código de los autores que generan software nuevo. Existen grandes plataformas open source en el mundo de la IA donde el público general puede descargar y utilizar modelos de aprendizaje automático (ML por sus siglas en inglés), incluyendo los más recientes y más poderosos. Muchas de estas contribuciones de código abierto incluso permiten que los desarrollos sean utilizados por compañías con fines de lucro, como lo es la industria de la mercadotecnia, quien puede utilizar los avances en la IA generativa para crear contenido publicitario con poca necesidad de intervención humana.

## 2. Problema/Reto

La problemática se basa en la complejidad y la demanda de recursos asociados con la creación de contenido visual publicitario para marcas como Coca-Cola, Fanta y Sprite. La dificultad reside en la búsqueda de un equilibrio entre la eficiencia del proceso y la personalización del contenido, aspecto crucial en el ámbito publicitario para garantizar la conexión efectiva con la audiencia. La falta de una herramienta integral que combine de manera fluida la inteligencia artificial y el diseño gráfico para generar imágenes publicitarias específicas y personalizadas ha generado la necesidad de una solución que optimice este proceso.

El reto consiste en desarrollar una aplicación lineal que integre herramientas avanzadas en inteligencia artificial y diseño gráfico para facilitar la generación y personalización de imágenes publicitarias. Esto implica la necesidad de superar los desafíos inherentes a la interacción entre la inteligencia artificial y la creatividad humana, así como la integración eficiente de los elementos visuales de las marcas mencionadas en un proceso fluido y accesible para los usuarios.

## 3. Propuesta de Solución

La solución propuesta para abordar la problemática planteada se centra en el desarrollo de una aplicación lineal que integra diversas herramientas avanzadas en inteligencia artificial (IA) y diseño gráfico. A lo largo

de este proceso, se seguirá una secuencia de pasos definidos para facilitar la generación y personalización de imágenes. A continuación, se detallan los componentes clave de esta solución integral:

El usuario inicia el proceso en una interfaz intuitiva, generando prompts a través de la aplicación. ChatGPT-3.5, modificado con una fórmula mejorada para la generación de prompts efectivos, responde proporcionando sugerencias creativas.

En la siguiente etapa, el usuario revisa y selecciona el prompt más adecuado, junto con el token del objeto a formar (por ejemplo, "sks - coca-cola" o "sfk - sprite"). Esta combinación se convierte en la entrada precisa para el modelo de generación de imágenes.

La interfaz de usuario Automatic1111 entra en juego, sirviendo como plataforma para ingresar los prompts y tokens seleccionados. Aquí, se inicia el modelo entrenado específicamente para la generación de imágenes de botellas de refresco, utilizando la potente tecnología SDXL (Stable Diffusion XL).

El resultado de este proceso es una imagen generada, que el usuario tendrá que transferir a Microsoft Designer para su edición y personalización. En este paso, los usuarios pueden realizar modificaciones, agregar texto u otros elementos gráficos según sus necesidades específicas.



Figura 1: Esquema de la propuesta solución

## 4. Licencias

### 4.1. ChatGPT 3.5-Licencia de uso comercial

La licencia comercial de GPT-3.5 Turbo para marketing permite a las empresas utilizar el modelo para una amplia gama de propósitos de marketing [2], incluyendo:

- Generación de contenido creativo, como anuncios, correos electrónicos, publicaciones en redes sociales, etc.
- Personalización de la experiencia del cliente, como la generación de recomendaciones de productos o servicios, o la creación de chatbots.
- Análisis de datos, como la identificación de tendencias o la generación de informes.

La licencia comercial de GPT-3.5 Turbo ofrece las siguientes ventajas para las empresas:

- Mayor flexibilidad y control sobre el uso del modelo.
- Acceso a soporte técnico y recursos adicionales.
- Posibilidad de personalizar el modelo para adaptarse a las necesidades específicas de la empresa.

Las restricciones de la licencia comercial de GPT-3.5 Turbo para marketing incluyen:

- El modelo no puede utilizarse para fines que sean ilegales, dañinos o discriminatorios.
- El modelo no puede utilizarse para generar contenido que sea engañoso o falso.
- El modelo no puede utilizarse para generar contenido que infrinja los derechos de propiedad intelectual de terceros.

Los precios de la licencia comercial de GPT-3.5 Turbo para marketing varían en función de la cantidad de tokens que la empresa necesite utilizar.

#### **4.2. Stable Diffusion (SDXL) Creative ML OpenRAIL-M (Machine Learning Open-Source for Creativity)**

**Tipo de Licencia:** Esta es una licencia específica para proyectos de aprendizaje automático y creatividad.[4]

**Características Clave:** La licencia Creative ML OpenRAIL-M es una licencia de código abierto que se utiliza en proyectos de aprendizaje automático relacionados con la creatividad. Su enfoque es promover la colaboración y el uso creativo de modelos de aprendizaje automático. Puede incluir términos específicos relacionados con la atribución y el uso de los modelos generados.

**Restricciones:**

- Restricciones de uso: La licencia establece restricciones de uso, como no utilizar el Modelo para actividades ilegales, no dañar a menores, no difundir información falsa con el propósito de dañar a otros, entre otras restricciones detalladas en la sección "Attachment A: Use Restrictions" de la licencia. Es importante asegurarse de que el uso del Modelo cumpla con estas restricciones.
- Responsabilidad: La licencia establece que los usuarios son responsables de las salidas generadas por el Modelo. Por lo tanto, si utiliza el Modelo para crear contenido publicitario, debe garantizar que dicho contenido cumpla con todas las leyes y regulaciones aplicables y no sea perjudicial ni engañoso.
- Marcas comerciales: Como se mencionó anteriormente, no está permitido utilizar las marcas comerciales del Licenciatario en su publicidad, según los términos de la licencia.
- Cumplimiento de la licencia: Asegúrese de cumplir con todos los términos y condiciones de la licencia, incluidas las disposiciones sobre distribución y redistribución, y cualquier otro requisito establecido en la licencia.

#### **4.3. Automatic1111- GNU AFFERO GENERAL PUBLIC LICENSE**

Automatic1111 user interface tiene una licencia de uso. La licencia es de tipo MIT, lo que significa que es de código abierto y gratuita para su uso, modificación y redistribución [3]. Los términos y condiciones de la licencia son los siguientes:

- La licencia permite el uso, modificación y redistribución del código de automatic1111 user interface sin restricciones, siempre que se respeten los siguientes términos:
- Se debe incluir una copia de la licencia en todas las copias del código.
- Se debe proporcionar crédito al autor original del código.
- Se pueden realizar modificaciones al código, pero las modificaciones deben publicarse bajo la misma licencia MIT.

La licencia de automatic1111 user interface es una licencia de código abierto muy permisiva que permite a los usuarios utilizar, modificar y redistribuir el código sin restricciones. Esto hace que automatic1111 user interface sea una herramienta muy versátil que puede ser utilizada por una amplia gama de usuarios.

En particular, los términos y condiciones de la licencia permiten a los usuarios utilizar automatic1111 user interface para cualquier propósito, incluso para fines comerciales. Esto hace que automatic1111 user interface sea una herramienta muy atractiva para las empresas que buscan desarrollar aplicaciones de interfaz de usuario automatizadas.

Además, los términos y condiciones de la licencia permiten a los usuarios modificar el código de automatic1111 user interface para adaptarlo a sus necesidades específicas. Esto hace que automatic1111 user interface sea una herramienta muy flexible que puede ser adaptada para una amplia gama de aplicaciones.

En general, la licencia de automatic1111 user interface es una licencia de código abierto muy permisiva que permite a los usuarios utilizar, modificar y redistribuir el código sin restricciones. Esto hace que automatic1111 user interface sea una herramienta muy versátil y atractiva para una amplia gama de usuarios.

#### 4.4. Microsoft Designer

Microsoft Designer tiene una licencia de uso. La licencia es de tipo comercial, lo que significa que se requiere una suscripción de pago para utilizar el software [8]. Los términos y condiciones de la licencia son los siguientes:

**Licencia:** Microsoft Designer se licencia como una suscripción mensual o anual. La suscripción mensual cuesta \$5,99 por usuario y mes, y la suscripción anual cuesta \$49,99 por usuario y año.

**Requisitos:** Microsoft Designer requiere una suscripción activa de Microsoft 365, así como un dispositivo con Windows 10 o 11.

**Uso:** Microsoft Designer se puede utilizar para crear prototipos de interfaces de usuario para aplicaciones web y móviles. El software incluye una variedad de herramientas

### 5. ChatGpt-3.5 (Generación de prompts para modelo de imágenes)

#### 5.1. Objetivo

A pesar de que la mayoría de los modelos de generación de imágenes cuentan con la capacidad de trabajar con base a “prompts” o entradas de texto formuladas por el usuario, muchas veces se presenta la dificultad

de controlar el resultado del modelo con el lenguaje natural de estas. Modelos de generación de imágenes como Stable Diffusion o DALL-E funcionan con base a estos prompts. El usuario redacta un texto breve que describe la imagen que desea obtener como resultado y el modelo de IA hace su mejor esfuerzo para generar una imagen que coincide con el deseo que el usuario expresó. Sin embargo, muchas veces es posible que la computadora omita detalles que para los usuarios humanos son tan obvios que no son necesarios de incluir. En muchas otras ocasiones la computadora también añade detalles que el usuario no pidió, genera imágenes que son mezclas de objetos reales, crea cuerpos humanos y de animales que no corresponden a su anatomía real y en general comete una serie de errores que no sucederían con un diseñador humano.

En el caso de desarrollar imágenes para Coca Cola, muchas veces los modelos de IA generativa incluyeron manos humanas sosteniendo productos de la marca con dedos faltantes, proporciones irreales, distorsionaron los productos, las tipografías de sus textos, utilizaron colores que no son de la marca y utilizaron estilo gráficos que no son utilizables para fines publicitarios.

Con esto en mente, exploramos la posibilidad de utilizar dos tipos de modelos generativos en conjunto para mitigar esta clase de problemas, que en su aplicación a los negocios se traduciría en recursos computacionales y financieros desperdiciados en generar imágenes inutilizables. Partimos de la premisa de que los usuarios no son capaces de aprender a utilizar su lenguaje de la mejor manera para alimentar a los modelos de generación de imágenes, pero que sí existe una herramienta capaz de moldear su lenguaje para adaptarse a patrones: la IA generativa de texto. A continuación utilizamos el modelo gratuito de generación de texto ChatGPT 3.5.

## 5.2. Entrenamiento

Partiendo de la idea de que el modelo Stable Diffusion XL es capaz de comprender palabras relacionadas a ciertos atributos de una imagen, como los objetos que aparecen y el estilo gráfico, se le alimentó una fórmula de texto a ChatGPT 3.5 que consiste en una combinación específica de palabras clave y texto que se debía de sustituir. Esta fórmula existe en forma de una oración con espacios en blanco que corresponden a sujetos, adjetivos, adverbios, acciones, estilos de imagen, texto, descriptores de resolución o calidad de la imagen, entre otros.

Después de explicarle a ChatGPT acerca del objetivo en mente se le introdujo la fórmula y se le pidió que genere una variedad de prompts para el modelo de generación de imágenes. ChatGPT reemplazó espacios en blanco que corresponden con el sujeto de la oración con frases como “botella de Coca Cola” o “mano sosteniendo lata de Sprite”, otros espacios correspondientes a adjetivos con frases como “colorido” o “vibrante”, eligió acciones como “sosteniendo” (para la mano), estilos de iluminación como “claro” o “etéreo” e incluso otros atributos descriptivos como estilos de arte. Las oraciones generadas por ChatGPT fueron introducidas al modelo de generación de imágenes. Cada oración o prompt se procesó con tres a cinco variaciones en los atributos aleatorios del modelo.

Como sucedió cuando los usuarios introducen prompts de lenguaje natural a los modelos de generación de imágenes, las entradas generadas por ChatGPT 3.5 en muchos casos resultaron en imágenes inutilizables para fines publicitarios, como paredes en dos dimensiones de botellas que parecían más una pintura ilusionista que una fotografía o ilustración de publicidad. Sin embargo, a diferencia de cuando el usuario mismo modifica

los prompts ingresados a los modelos como Stable Diffusion, pedirle al modelo de generación de texto que realizara cambios resultó en procesos de afinación mediante prueba y error muchos más cortos, pues elegir acerca de la gran cantidad de descriptores dentro del prompt puede resultar abrumador mientras que pedirle a ChatGPT cosas como “por favor evita que la imagen genere botellas demasiado grandes” resultó más sencillo y práctico. Esto representa que en una aplicación del mundo real donde es necesaria una interfaz amigable utilizar estas dos clases de IA generativa en conjunto puede resultar de gran utilidad.

### 5.3. Resultados

Las primeras imágenes generadas con prompts de ChatGPT 3.5 no fueron mucho más útiles que las imágenes generadas con prompts elaboradas por usuarios, sin embargo, tomó una y como máximo dos instancias en las que se le pidió a ChatGPT realizar correcciones para obtener resultados más deseables, mientras que realizar cambios a los prompts como usuario resulta menos sistemático y más elaborado. En la "Fig.13" se pueden apreciar los resultados obtenidos.

Ejemplos de los mejores prompts utilizados:

#### Sprite Prompts

- “An image of effervescent Sfk bottle pouring over ice, vibrant and dynamic lighting, extremely detailed, ultra realistic, 10k high resolution, in the style of contemporary product illustration, glossy commercial photography, and modern digital art.”
- “An image of classic Sfk bottle with condensation droplets, timeless and iconic lighting, extremely detailed, ultra realistic, 10k high resolution, in the style of vintage soda advertisements, traditional product illustration, and glossy commercial photography.”
- “An image of a cinematic Sfk bottle, dramatic lighting casting deep shadows, extremely detailed, ultra realistic, 10k high resolution, in the style of film noir, classic product photography, and modern digital art.”

#### Coca-cola Prompts

- “An image of effervescent Sks bottle pouring over ice, vibrant and dynamic lighting, extremely detailed, ultra realistic, 10k high resolution, in the style of contemporary product illustration, glossy commercial photography, and modern digital art.”
- “An image of classic Sks bottle with condensation droplets, timeless and iconic lighting, extremely detailed, ultra realistic, 10k high resolution, in the style of vintage soda advertisements, traditional product illustration, and glossy commercial photography.”
- “An image of a cinematic Sks bottle, dramatic lighting casting deep shadows, extremely detailed, ultra realistic, 10k high resolution, in the style of film noir, classic product photography, and modern digital art.”



Figura 2: Imágenes generadas Sprite con prompts ChatGpt-3.5

## 6. ChatGpt-3.5 (Generación de ideas y texto para el anuncio)

### 6.1. Objetivo

El objetivo principal es generar ideas y texto para el anuncio tomando en cuenta un enfoque integral destinado a cultivar la creatividad y potenciar la eficacia comunicativa en el proceso de concepción y redacción de un anuncio publicitario impactante. Se propone establecer un marco metodológico que, mediante una serie de fases iterativas, estimule la generación sistemática de ideas innovadoras y oriente la formulación de textos persuasivos y cautivadores. En este proceso se buscará trascender las convenciones tradicionales de la publicidad, explorando nuevos enfoques y estrategias que vayan más allá de la mera transmisión de información sobre los productos o servicios. La intención es construir una narrativa publicitaria coherente y significativa, en la que cada elemento contribuya a la creación de una experiencia envolvente para la audiencia. Este enfoque iterativo implica la consideración de los requisitos y expectativas específicos de la audiencia objetivo, así como del contexto publicitario en el que se desenvolverá el anuncio. Se aspira a la creación de mensajes que trasciendan lo convencional y resuenen profundamente con el público, generando no solo un interés momentáneo, sino una conexión emocional perdurable. Este proceso abarca la exploración de diversas fuentes de inspiración, desde las tendencias culturales hasta las características únicas de la marca, con el propósito de alimentar un flujo constante de ideas que puedan ser refinadas y transformadas en textos publicitarios con un impacto significativo. Asimismo, se incorporarán herramientas y técnicas específicas de redacción publicitaria para maximizar la persuasión y la retención del mensaje. Por último, el anuncio se erige como una pieza distintiva de comunicación que refleja la esencia de la marca y crea una experiencia inolvidable para el consumidor.

### 6.2. Entrenamiento

El proceso de entrenamiento destinado a la concepción de ideas y la redacción de textos para el anuncio se llevó a cabo meticulosamente, considerando diversos escenarios clave para la presentación efectiva del material

publicitario. Dichos escenarios abarcaban entornos como instituciones educativas, la ciudad de Monterrey, Nuevo León, así como las marcas emblemáticas Coca-Cola y Fanta.

Durante esta fase de entrenamiento, se implementó el modelo de lenguaje ChatGPT en una serie de pruebas iterativas, con el propósito de perfeccionar continuamente la calidad del texto y obtener resultados óptimos para la generación de eslogan. A partir de los insights extraídos de estas pruebas, se desarrolló un texto base para la creación de eslogan publicitarios.

Posteriormente, se ingresaron los parámetros esenciales, tales como "Producto/Servicio", "Características Clave", "Público Objetivo", "Tono y Estilo", "Llamada a la Acción", "Longitud/Formato", "Información Adicional", para contextualizar la generación de eslogan específicamente para Coca-Cola, Fanta, Sprite, entornos escolares y la localidad de Monterrey, Nuevo León.

Este enfoque estratégico permitió optimizar la relevancia y adecuación de los eslogan al contexto previamente identificado. Como resultado de este proceso detallado y orientado, se lograron obtener eslogan mejor articulados y alineados con los objetivos del anuncio publicitario, culminando así en la creación de mensajes impactantes y persuasivos para su presentación.

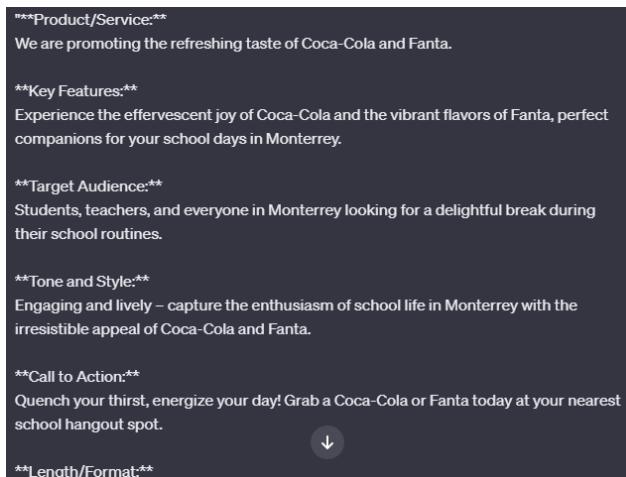


Figura 3: Generación Texto para eslogan

### 6.3. Resultados

Los resultados obtenidos a partir del proceso de texto generativo de eslogan ha culminado en la creación de eslogan publicitarios envolventes y persuasivos para Coca-Cola y Fanta. Estos mensajes, diseñados con un enfoque en la conexión local en el contexto escolar de Monterrey, buscan no solo informar, sino también generar una conexión emocional con la audiencia. La adaptación precisa a las características específicas de la marca y del entorno local asegura la pertinencia y efectividad de los eslogan, ofreció así un anuncio publicitario distintivo y memorable.

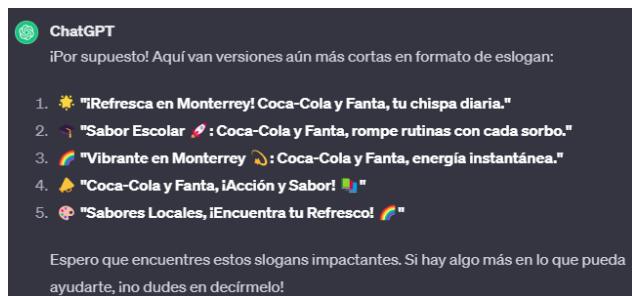


Figura 4: Eslogan

## 7. Stable Diffusion XL (SDXL)

El modelo de Stable Diffusion texto-imagen más reciente y sofisticado creado por Stability AI se llama ***Stable Diffusion XL (SDXL)***. Se trata de un modelo de código abierto que amplía los logros de los modelos de Stable Diffusion v1.5 y v2.

***SDXL*** tiene aplicación en diversos ámbitos, contribuyendo al desarrollo de aplicaciones novedosas e innovadoras. Una aplicación notable es la de los proyectos creativos, donde artistas, diseñadores e ilustradores lo utilizan para generar imágenes de alta calidad. La tecnología resulta inestimable para crear imágenes realistas y detalladas, mejorando la producción creativa de los profesionales de estos campos.

Además, extiende su impacto al ámbito de las redes sociales contribuyendo a la creación de contenidos personalizados. Esto incluye la generación de avatares y fondos personalizados, enriqueciendo la experiencia en las redes sociales al ofrecer a los usuarios elementos a medida y visualmente atractivos. La versatilidad de que este posee es evidente, ya que sigue impulsando la innovación a través de diversas aplicaciones, mostrando su potencial para remodelar y mejorar diversos aspectos de nuestro panorama digital. Para conocer mas acerca del modelo se puede ir a la siguiente refencia [5]

### 7.1. Beneficios

En comparación con sus predecesores ***SDXL*** aporta una serie de mejoras, como por ejemplo:

- **Mayor fidelidad y calidad de los gráficos:** *SDXL* produce visuales más realistas, intrincados y acordes con las sugerencias del texto.
- **Mayor estabilidad y solidez frente al ruido:** *SDXL* produce imágenes más refinadas y libres de artefactos y ruido.
- **Mayor velocidad de inferencia y menor uso de memoria:** funciona de forma más eficiente, lo que lo hace perfecto para su despliegue en dispositivos periféricos y aplicaciones en tiempo real.

A parte de estos avances tecnológicos, ***SDXL*** proporciona a los usuarios muchas ventajas:

- **Disponibilidad de código abierto:** El modelo puede ser utilizado y modificado libremente por cualquiera, lo que fomenta la creatividad y el trabajo en equipo en el ámbito de la generación de texto a imagen.
- **Compatibilidad con las herramientas y procesos de difusión estable actuales:** Los usuarios pueden incluir fácilmente ***SDXL*** en sus flujos de trabajo gracias a su compatibilidad con las herramientas y pipelines de difusión estable actuales.

## 8. Tipos de entrenamiento para SDXL

Se exploran diversas modalidades de entrenamiento para modelos generativos de imágenes de difusión, con el propósito de identificar la más eficaz para abordar el desafío en cuestión. A continuación, se detallan los enfoques considerados, acompañados de breves descripciones de su funcionamiento.

### 8.1. HyperNetworks

El enfoque de las ***redes hipertextuales*** utiliza una red neuronal compacta e independiente para controlar los pesos del modelo de difusión principal, en la "Fig.5", podemos ver un esquema de donde se encuentran las redes en el modelo de difusión modificado. Este enfoque permite una mayor flexibilidad y delicadeza en la manipulación de los resultados. No obstante, es importante recordar que la complejidad de este enfoque puede dificultar el entrenamiento y el ajuste.

La producción de imágenes muy estilizadas o artísticas pone de manifiesto la eficacia de las ***redes hipertextuales***. El mayor grado de control facilita la alteración intencionada del proceso de difusión, lo que permite una expresión artística más compleja. No obstante, es crucial reconocer que este enfoque no está tan extendido o respaldado como otros, lo que dificulta el proceso de obtención de modelos entrenados y la resolución de problemas.

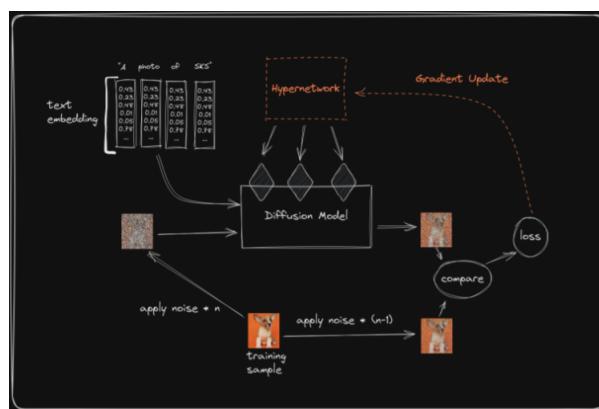


Figura 5: Esquema de Hypernetworks training

*Nota:* Imagen obtenida de [7].

## 8.2. Dreambooth

La técnica **Dreambooth** consiste en afinar el modelo de difusión utilizando un conjunto limitado de imágenes de referencia. Este proceso permite al modelo asimilar conceptos y estilos específicos de las imágenes proporcionadas, lo que lo hace muy adecuado para generar modelos personalizados que produzcan imágenes de individuos, lugares u objetos de significado personal.

Sin embargo, **Dreambooth**, considerado el método más eficaz y versátil, presenta el inconveniente de que requiere mucho espacio de almacenamiento para el modelo personalizado ya que entrena el modelo por completo. Además, cabe destacar que **Dreambooth** puede tener problemas de generalización, que se manifiestan en resultados incoherentes para imágenes que van más allá del ámbito específico en el que se ha entrenado.

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

El método **LoRA** integra un codificador de texto en el proceso de difusión como se puede ver en la imagen "Fig.6", introduciendo así información adicional para controlar con precisión los resultados basados en indicaciones textuales. Aunque **LoRA** es un desarrollo relativamente reciente, resulta prometedor en tareas como la generación de imágenes alineadas con descripciones específicas o la mejora de imágenes existentes con detalles adicionales.

Es importante reconocer que una de las limitaciones de **LoRA** es su coste computacional, sobre todo en el caso de las imágenes de alta resolución. Además, el ajuste fino de un modelo **LoRA** exige a menudo un mayor nivel de conocimientos en comparación con otras metodologías.

LoRA weights,  $W_A$  and  $W_B$ , represent  $\Delta W$

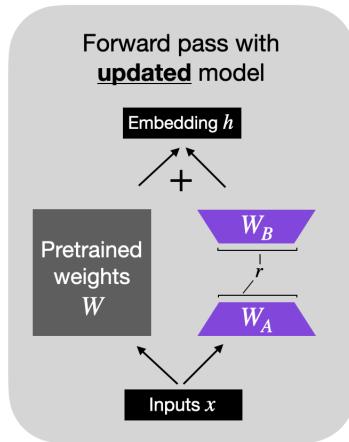


Figura 6: Esquema de LoRA training

*Nota:* Imagen obtenida de [11].

## 8.4. Textual Inversion

El método ***Textual Inversion*** incorpora indicaciones textuales directamente en el espacio latente del modelo de difusión se busca modificar los colores del embedding solamente, como se puede observar en el esquema de la figura "Fig.7", lo que da lugar a representaciones más eficientes y compactas de los resultados deseados. Posicionado como un compromiso equilibrado entre ***Dreambooth*** y ***LoRA***, ***Textual Inversion*** ofrece una calidad y eficiencia encomiables sin los requisitos de almacenamiento de ***Dreambooth*** ni los costes computacionales asociados a ***LoRA***.

Sin embargo, hay que tener en cuenta que ***Textual Inversion*** puede encontrar dificultades cuando se le plantean preguntas complejas o cuando se le pide que genere detalles muy específicos. La codificación de la información textual puede no captar todas las sutilezas inherentes a la imagen deseada, lo que contribuye a ocasionales dificultades para alcanzar el nivel deseado de complejidad o especificidad.

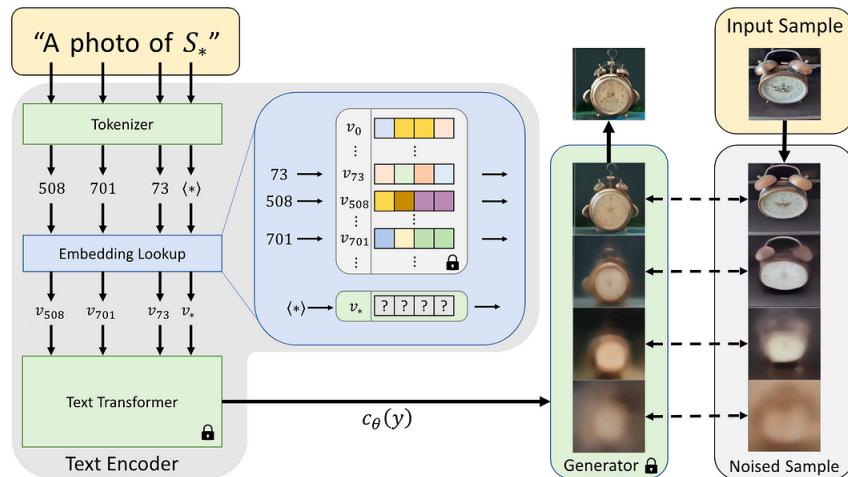


Figura 7: Esquema de Textual Inversion training

Nota: Imagen obtenida de [9].

## 9. Entrenamiento de SDXL

En este proyecto, se implementó el entrenamiento utilizando ***AutoTrain*** y seleccionando la configuración ***Dreambooth-Lora***. Se estableció un **learning rate** de  $1 \times 10^{-4}$ , un **steps number** de 500, basándose en los valores recomendados para objetos según el paper [12]. Además, se aplicó **gradient accumulation** con un valor de 4, manteniendo la **resolución** nativa del modelo SDXL en 1024. El **optimizador** utilizado fue 8-bit Adam.

A pesar de haber utilizado un número relativamente reducido de imágenes gracias a Dreambooth, el proceso de entrenamiento de cada modelo tomó aproximadamente 2 horas y media. El peso de los modelos LoRA resultó ser de 22.3 MB. Se entrenó un modelo específico para la generación de imágenes de Coca-Cola de 355 ml de vidrio, Sprites de 355 ml de vidrio, Fantas de 355 ml de lata y fondos para los anuncios. La

elección de estos productos de la familia Coca-Cola se basó en los datos proporcionados por la empresa Arca Continental, que indicaban que eran los productos más vendidos.

### 9.1. AutoTrain

Además para el entrenamiento este se facilitó con AutoTrain el cual es: "Una herramienta sin código para entrenar modelos de última generación para tareas de Procesamiento del Natural Language Processing (NLP), para tareas de Computer Vision (CV), y para tareas de Habla e incluso para tareas Tabulares. Está construido sobre las impresionantes herramientas desarrolladas por el equipo de Hugging Face, y está diseñado para ser fácil de usar."<sup>[6][1]</sup>

### 9.2. Dreambooth-LoRA

Este tipo de entrenamiento es del repositorio ***Parameter-Efficient Fine-Tuning (PEFT) methods*** [10] Dreambooth-LoRA es una potente síntesis de Dreambooth y LoRA, dos métodos que como se mencionó anteriormente funcionan para optimizar modelos de difusión texto-imagen como Stable Diffusion XL.

Se fusionan lo mejor de ambos modelos: con Dreambooth, puedes entrenar un modelo de difusión utilizando una colección limitada de tus propias fotos, personalizándolo de manera exclusiva para reflejar tus preferencias y estilo específico. Lo cual facilita que el modelo genere imágenes más alineadas con tu visión específica. Por otro lado, LoRA (Low-Rank Adaptation) ofrece una técnica eficaz para el ajuste fino. En lugar de modificar el modelo de difusión en su totalidad, LoRA introduce pequeñas capas flexibles que capturan de manera efectiva la esencia de tus imágenes de entrenamiento. Este enfoque permite mantener la calidad de las fotos de salida, al mismo tiempo que optimiza el modelo para que sea más compacto y rápido en su ejecución.

Además, es relevante destacar que los modelos obtenidos pertenecen al tipo LoRA. Esta distinción es significativa, ya que difiere de los modelos Dreambooth, en los cuales se entrena y modifica el modelo completo. En el caso del SDXL, que es inherentemente un modelo bastante pesado con un tamaño aproximado de 6.94 GB, la implementación de modelos LoRA se presenta como una solución eficiente. Cada modelo LoRA para este proyecto tiene un peso reducido a 22.3 MB, generando un ahorro sustancial de espacio. Esta reducción en el tamaño del modelo no solo optimiza el almacenamiento, sino que también agiliza significativamente la carga del modelo en la interfaz de usuario (UI).

### 9.3. Resultados

El modelo ha generado resultados sobresalientes, como se aprecia en la "Fig.8". Cada variante específica de las imágenes (Coca-Cola, Sprite, Fanta y fondo) fue obtenida utilizando un modelo LoRA distinto. Estos modelos LoRA se incorporan de manera modular sobre la estructura base del modelo Stable-Diffusion XL. Esta característica posibilita la flexibilidad de intercambiar entre modelos, permitiendo la creación de imágenes según las preferencias y requisitos específicos.

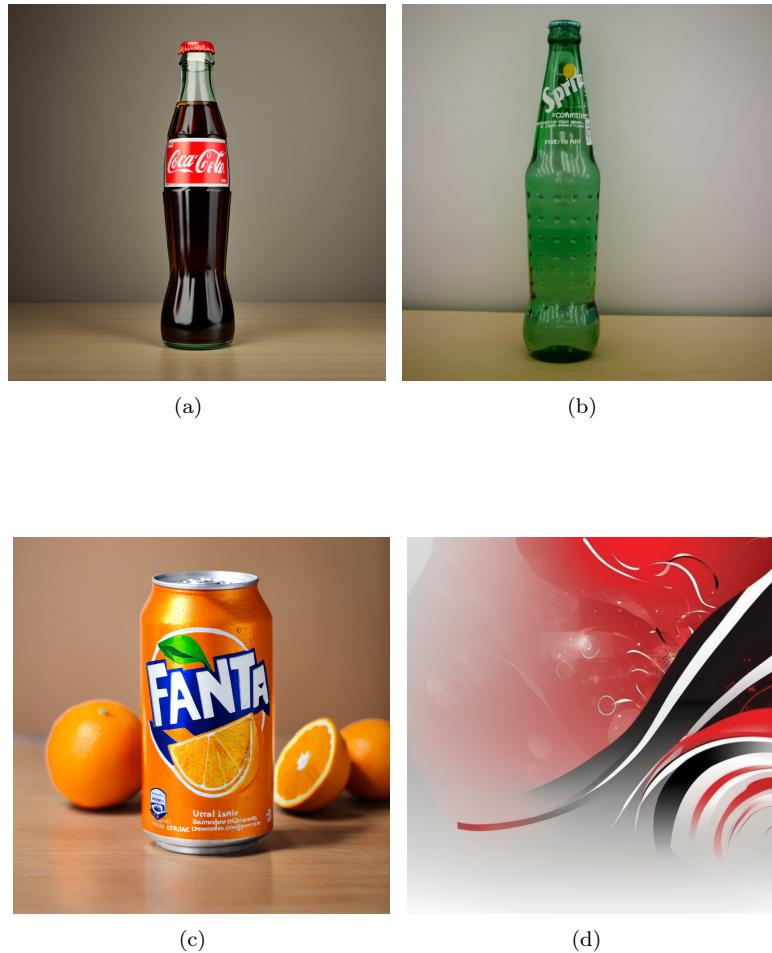


Figura 8: Imágenes con modelos LoRA

## 10. Automatic 1111 como UI

AUTOMATIC1111 es una interfaz gráfica de usuario (GUI) para el modelo de difusión de texto a imagen Stable Diffusion. Está desarrollada por el equipo de Stability AI y está disponible de forma gratuita (Open-source)[1].

En la figura "Fig.9", observamos que se solicitó la generación de una imagen mediante el prompt *A photo of a sks bottle*. Esta elección se debe a que el modelo de Coca-Cola fue entrenado específicamente con la palabra clave *sks* como desencadenante (trigger word). La trigger word se vincula al embedding y a la modificación realizada por el modelo LoRA. Este enfoque posibilita la creación de imágenes mejoradas sin comprometer la integridad del modelo original.

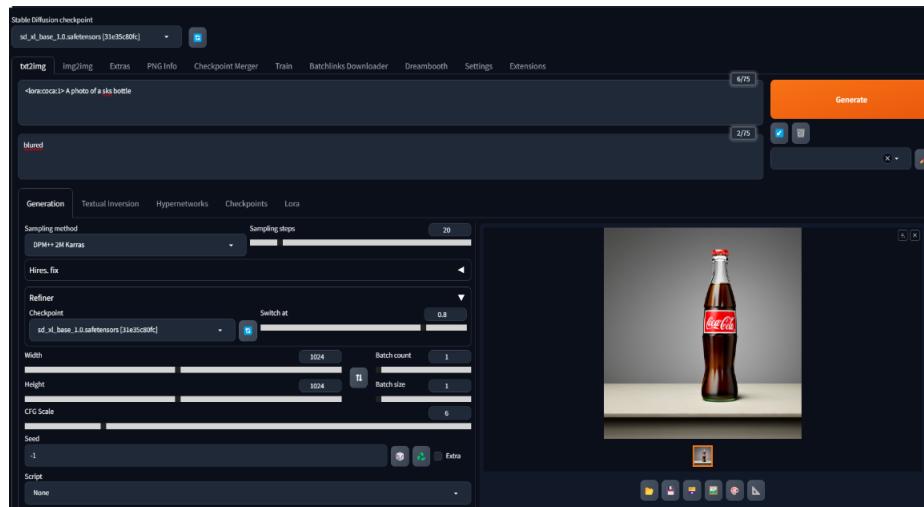


Figura 9: Interfaz AUTOMATIC1111

La elección del modelo LoRA es un proceso sencillo. Simplemente se accede a la ventana de LoRA, donde previamente se han almacenado los modelos ya entrenados "Fig.10". Después, se selecciona el modelo deseado utilizando la sintaxis <lora:coca (nombre del modelo)>. Posteriormente, se puede introducir el prompt de manera convencional. Es importante destacar que la carga es rápida debido al peso ligero de los modelos LoRA. Como recomendación, se aconseja evitar la inclusión de varios LoRAs en el mismo prompt, ya que en nuestras pruebas hemos observado que esto puede llevar a confusiones por parte del modelo, resultando en la generación de imágenes sin sentido o con contenido de ruido.

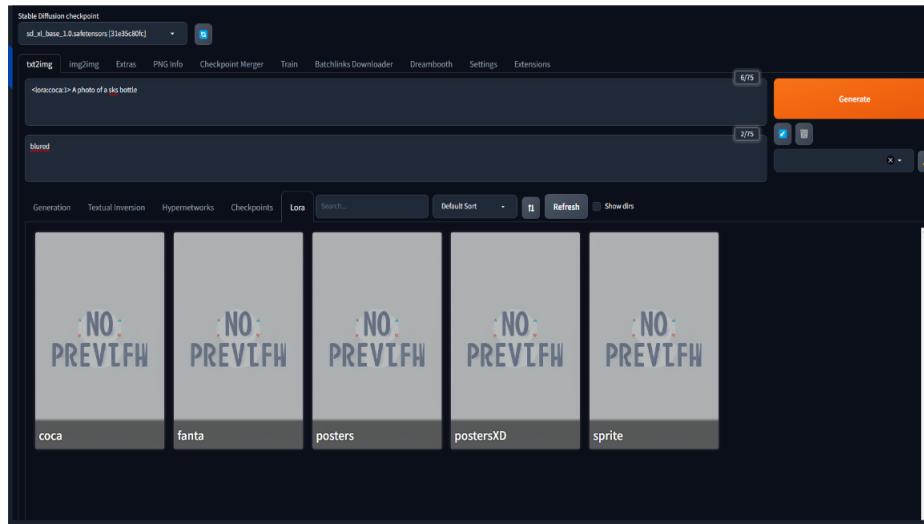


Figura 10: LoRAs AUTOMATIC1111

## 10.1. Beneficios

**AUTOMATIC1111** se destaca por ofrecer varias ventajas en su integración con Stable Diffusion, proporcionando una experiencia eficiente y adaptable a diversas necesidades:

En primer lugar, su interfaz de usuario es amigable y de fácil comprensión, brinda una notable facilidad de uso. Incluso para aquellos usuarios con poca experiencia en aprendizaje automático, la graphical user interface(GUI) resulta intuitiva.

La eficiencia es otro punto fuerte de **AUTOMATIC1111**, ya que acelera significativamente el proceso de generación de imágenes. Esta característica lo convierte en una herramienta ideal tanto para proyectos creativos como para investigaciones, donde la rapidez en la obtención de resultados es esencial.

La personalización es una ventaja clave que **AUTOMATIC1111** aporta al proceso de generación de imágenes. Los usuarios tienen la capacidad de personalizar cada aspecto, lo que contribuye a la obtención de resultados más creativos y expresivos, adaptándose a las necesidades específicas de cada proyecto.

En el contexto de su aplicación con Stable Diffusion, **AUTOMATIC1111** ofrece características específicas que potencian aún más su utilidad:

- **Compatibilidad con modelos Stable Diffusion XL:** **AUTOMATIC1111** se integra sin problemas con el modelo Stable Diffusion XL, proporcionando una calidad de imagen significativamente mejorada en comparación con versiones anteriores.
- **Herramientas de edición:** El conjunto de herramientas de edición incorporadas en **AUTOMATIC1111** permite a los usuarios ajustar con precisión los detalles de las imágenes generadas, ofreciendo un control más refinado sobre el resultado final.

**Compatibilidad con extensiones:** **AUTOMATIC1111** es compatible con diversas extensiones que agregan nuevas funciones y capacidades, ampliando así su versatilidad y potencial.

## 10.2. Herramientas mas importantes

La elección de presentar **AUTOMATIC1111** como la interfaz gráfica de usuario principal se fundamenta principalmente en la activa y colaborativa comunidad que contribuye a su mejora continua y a la creación de herramientas poderosas para el perfeccionamiento de la generación de imágenes. Dentro de las diversas funcionalidades que ofrece, destacamos las siguientes herramientas clave:

- **Text-2-image:** Este componente, común en todas las GUI de inteligencia artificial generativa, permite la creación de imágenes a partir de instrucciones específicas o "prompts" proporcionados por el usuario. Facilita la materialización visual de conceptos expresados a través del texto.
- **Image-2-Image:** La función *Image-2-Image* se centra en la transformación y generación de imágenes basada en otras imágenes. Proporciona herramientas para modificar y adaptar visualmente una imagen de entrada, permitiendo una amplia gama de posibilidades creativas.

- **Negative prompts:** es una característica valiosa que permite explorar variaciones opuestas a las instrucciones positivas, ampliando la versatilidad creativa y posibilitando la generación de imágenes con enfoques contrastantes.
- **png-info:** La inclusión de *png-info* como herramienta provee información detallada sobre archivos en formato PNG. Esto resulta beneficioso para comprender las características específicas de las imágenes utilizadas en el proceso creativo, mejorando así la manipulación y comprensión de los datos visuales.
- **Inpainting:** La función *Inpainting* se destaca por su capacidad para restaurar y completar áreas faltantes en una imagen, permitiendo la corrección visual de imperfecciones o la integración de elementos ausentes de manera coherente con el conjunto, contribuyendo así a la mejora estética de las creaciones visuales.

Si se examinan detenidamente los ejemplos en la "Fig.8", se aprecia que tanto la botella de Sprite como la de Fanta presentan caracteres que parecerían ser texto en algún idioma extraño. Sin embargo, esta aparente escritura se debe a que los modelos de inteligencia artificial generativa para imágenes no reconocen ni comprenden lo que constituye texto. Por lo tanto, las imágenes de entrenamiento, especialmente en el caso de bebidas, a menudo contienen elementos escritos que el modelo interpreta como patrones a replicar, generando contenido sin sentido.

Para abordar este inconveniente, utilizamos la función de *inpainting* de **AUTOMATIC1111**. Simplemente colocamos la imagen y superponemos una capa sobre la sección que deseamos eliminar. La solución óptima que hemos encontrado es indicar al modelo, a través de un prompt, el color con el cual deseamos sustituir el "texto". En el caso de la "Fig.11", se especificó *orange* como el color deseado, resolviendo así el problema sin inconvenientes. Cabe mencionar que en ocasiones puede ser necesario realizar varios intentos para corregir este tipo de errores, ya que el modelo podría llenar la región de manera incorrecta o con un tono ligeramente diferente en algunos casos.

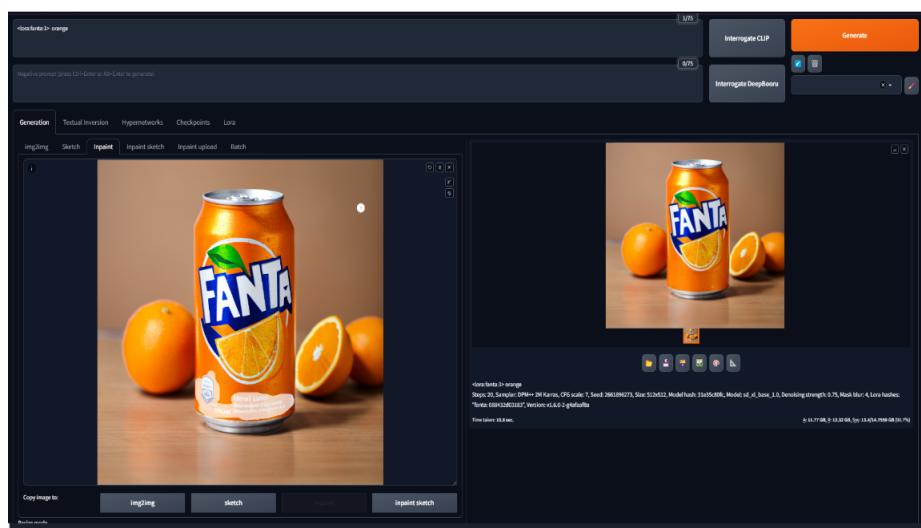


Figura 11: Inpainting Fanta

## 11. Microsoft Designer

La integración de Microsoft Designer en nuestro proceso de generación y modificación de imágenes ha aportado diversos beneficios sustanciales, enriqueciendo la experiencia de edición y personalización de las creaciones visuales. Al hacer uso de esta herramienta, se han experimentado resultados notables que contribuyen significativamente al desarrollo de nuestro proyecto.

### 11.1. Beneficios

La herramienta ofrece una diversidad de estilos de texto que van desde tipografías elegantes hasta opciones más informales, permitiendo a los usuarios adaptar el contenido textual de las imágenes según el tono y propósito de su proyecto. Las herramientas de recorte avanzadas de Microsoft Designer son esenciales para perfeccionar la composición visual. La capacidad de manipular y ajustar la imagen generada garantiza resultados finales estéticamente atractivos. Además de la edición, Microsoft Designer posibilita la creación de nuevas imágenes, permitiendo a los usuarios explorar diferentes enfoques creativos dentro del mismo entorno de edición. La interfaz intuitiva y accesible de Microsoft Designer facilita su uso, asegurando que incluso aquellos sin experiencia previa en diseño gráfico puedan aprovechar al máximo sus funciones, promoviendo así una experiencia de usuario positiva. En la “Fig.12” se puede mostrar lo que se diseño con el apoyo de esta herramienta.

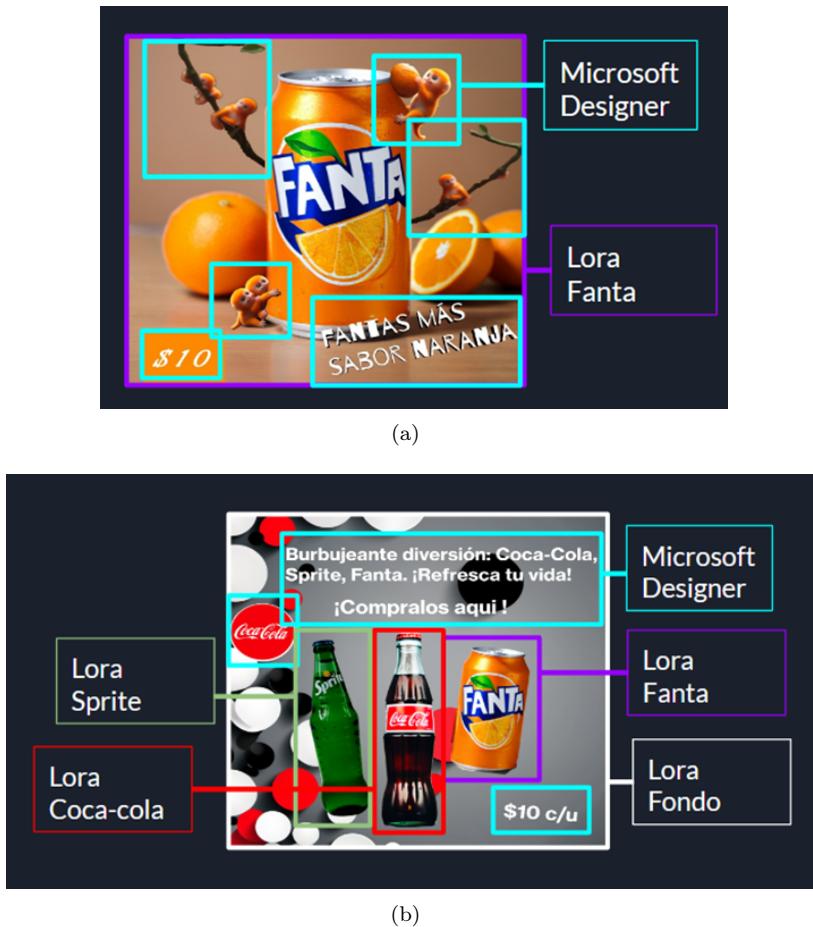


Figura 12: Muestreo de integración de imágenes generadas y Microsoft Designer

## 11.2. Resultados

La combinación de las capacidades de generación de imágenes y la posterior edición en Microsoft Designer ha permitido una personalización detallada de las creaciones visuales. Desde ajustes mínimos hasta transformaciones más significativas, los usuarios tienen el control total sobre el aspecto final de sus creaciones. La diversidad de estilos de texto disponibles en Microsoft Designer contribuye a la creación de contenido textual impactante, mejorando la calidad estética general de las imágenes generadas. Gracias a las herramientas de recorte avanzadas, se ha logrado la optimización de imágenes para diversos fines. Microsoft Designer facilita la adaptación de las imágenes a diferentes contextos y plataformas.

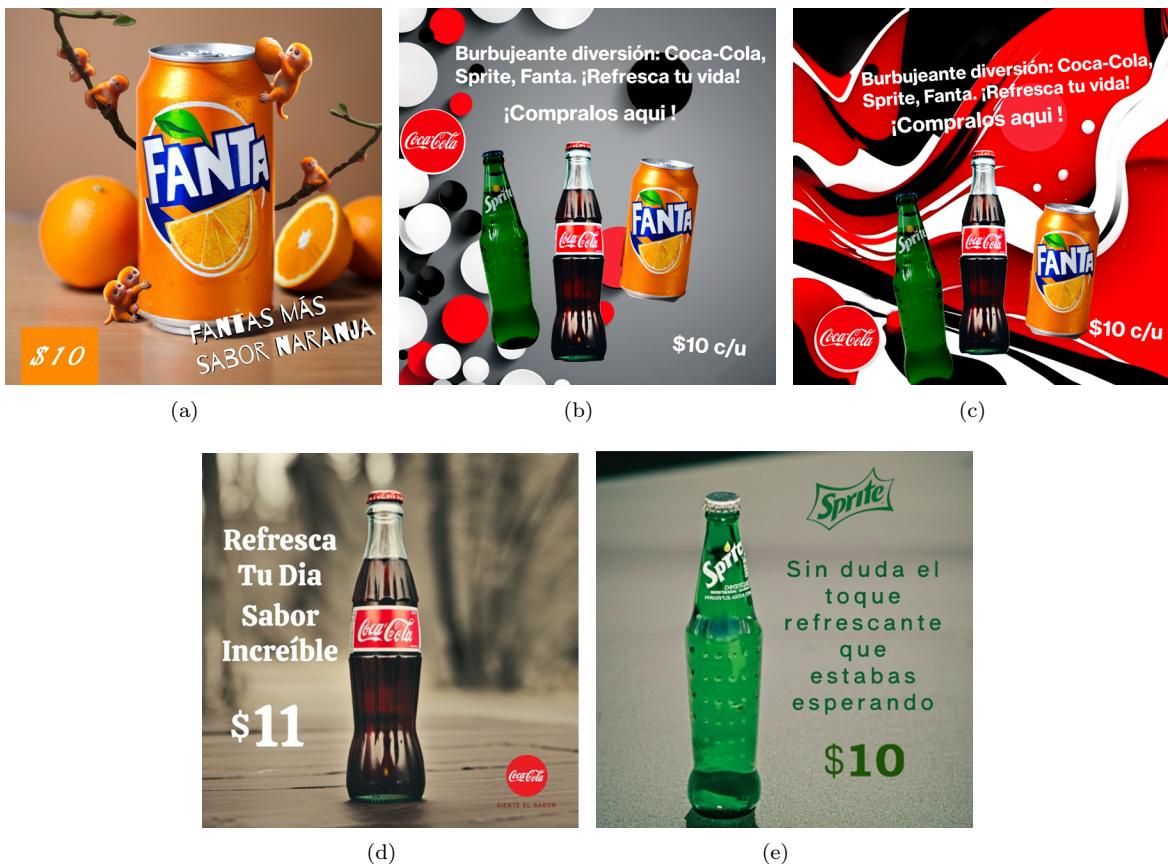


Figura 13: Resultado final integración de Microsoft Designer

## 12. STEFANN

STEFANN(Scene Text Editor using Font Adaptive Neural Network) es un software de editor de escenarios de texto por medio de redes neuronales adaptativas de fuente como es indicado en sus siglas, este proporciona una gran ventaja al ser un editor de texto de libre uso comercial que funciona siendo una herramienta muy intuitiva para el cambio de texto, el cual no es muy común.

### 12.1. Objetivo

Hoy en día es un problema notorio que los generadores de imágenes en base a texto no son capaces de crear texto legible en la mayoría de los casos, todavía más difícil algo que mantenga raciocinio, para casos como estos las opciones suelen tratarse de la utilización de impainting para quitar el texto y ponerle nuevo, tal como es el caso anterior de la utilización de Automatic 1111 y Microsoft Designer, pero para muchos casos donde el texto es legible, puede ser un difícil imponer el texto con la fuente que la imagen tenía originalmente, por lo que corregir los textos conservando la fuente es la mejor opción para dichos casos.

## 12.2. Uso

Como se menciona anteriormente Stefann es un programa que ofrece una solución deseable para correcciones de texto en muchos casos de manera intuitiva y rápida para cualquier persona una vez instalado, esto al simplemente seleccionar la foto en los archivos, realizar clic en 4 puntos de la imagen seleccionada por medio del ratón para señalar el texto a cambiar y utilizar el teclado para teclear las letras que se desea que cambien en el resultado final.

## 12.3. Resultados

Stefann a mostrado ser una herramienta de mucho valor para corrección de textos para imágenes generadas, e incluso para imágenes normales, donde se buscan otros objetivos como cambiar el idioma de las palabras dentro de la imagen. Si bien sus modelos pueden ser reentrenados para objetivos más específicos como lo que es la empresa de Arca, esto requiere más recursos y tiempo para terminar con un resultado deseable.

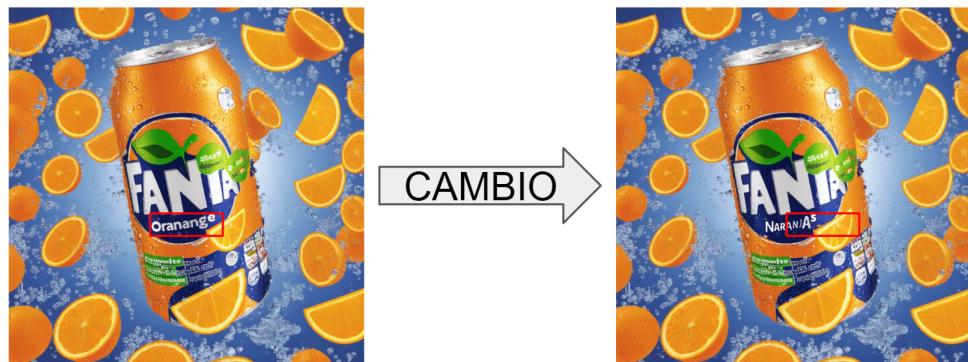


Figura 14: Fanta Corrección con STEFANN

## 13. Conclusión

Podemos concluir este proyecto con resultados positivos en cuanto a llegar a un punto donde tenemos herramientas de simple uso para la generación de anuncios comerciales para productos de la familia Coca-Cola con ciertos grados de éxito, y si bien hay apartados que se pueden mejorar, en general se realizó un buen entregable en correspondencia al tiempo y los recursos otorgados dado para trabajar estas herramientas, destacando el desarrollo del modelo de Stable Diffusion XL al entrenarlo con un método combinado Dreambooth-LoRA, generando excelentes imágenes de los productos deseados.

Gracias a este proyecto fuimos capaces de comprobar que si bien es complicado realizar modelos generativos que sean competitivos con aquellos que están actualmente activos en el mercado, contamos con la habilidad y conocimiento suficiente para utilizar modelos y herramientas que ya estén disponibles e incluso reentrenarlos para situaciones específicas como lo fue en esta ocasión los productos de la familia CocaCola.

## 14. Anexos

En esta sección se encuentran los anexos relacionados con el proyecto. La carpeta denominada *Fotos con ChatGPT* contiene más imágenes y los prompts utilizados. El proceso de generación de prompts se detalla en la Sección 5. Enlace:

<https://drive.google.com/drive/folders/1mrgg2eTcBZTEy4Hw4PucSpte0kupkmb9?usp=sharing>

Las imágenes de entrenamiento se encuentran disponibles en el siguiente enlace del Google Drive:

<https://drive.google.com/drive/folders/1DhIvqF9DM5YsWJ38Fe6c27C6BIOJQcRQ?usp=sharing>

**Nota:** Dentro de la carpeta de Fondos, se encuentra una subcarpeta llamada *Posters*. A estas imágenes se les eliminó cualquier elemento que no representara un fondo, y así se realizaron los procesos de entrenamiento.

Para visualizar más imágenes generadas utilizando los modelos entrenados LoRA, se puede acceder al siguiente enlace:

[https://drive.google.com/drive/folders/1I1W3RUQhQNFGK3\\_nBCcvxT6zNzB5V9bH?usp=sharing](https://drive.google.com/drive/folders/1I1W3RUQhQNFGK3_nBCcvxT6zNzB5V9bH?usp=sharing)

Para acceder a los modelos LoRA se puede usar el siguiente enlace a la carpeta de drive:

<https://drive.google.com/drive/folders/1p1wbo9VMACaA3D45YZlpB0eHzow0V44n?usp=sharing>

## Referencias

- [1] S. AI. "AUTOMATIC1111." (2023), dirección: <https://github.com/AUTOMATIC1111/stable-diffusion-webui>.
- [2] O. AI). "Business terms." (2023), dirección: <https://openai.com/policies/business-terms>.
- [3] S. AI). "LICENSE." (2023), dirección:  
<https://github.com/AUTOMATIC1111/stable-diffusion-webui/blob/master/LICENSE.txt>.
- [4] S. AI). "LICENSE-SDXL1.0." (2023), dirección: [https://github.com/Stability-AI/generative-models/blob/main/model\\_licenses/LICENSE-SDXL1.0](https://github.com/Stability-AI/generative-models/blob/main/model_licenses/LICENSE-SDXL1.0).
- [5] S. AI). "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis." (2023), dirección: <https://huggingface.co/papers/2307.01952>.
- [6] H. Face). "AutoTrain." (2023), dirección: <https://huggingface.co/docs/autotrain/index>.
- [7] koiboi. "LoRA vs Dreambooth vs Textual Inversion vs Hypernetworks." (2023), dirección: [https://www.youtube.com/watch?v=dVjMiJsuR5o&list=PLFSYi0xdWnuVL-RovOW-wvoorY6TPXKz\\_&index=13&t=435s](https://www.youtube.com/watch?v=dVjMiJsuR5o&list=PLFSYi0xdWnuVL-RovOW-wvoorY6TPXKz_&index=13&t=435s).
- [8] Microsoft). "Designer for Web Image Generator and Brand Kit Terms Preview." (2023), dirección: <https://designer.microsoft.com/termsOfUse.pdf>.
- [9] O. Mishra). "Textual Inversion: A method to finetune Stable Diffusion Model." (2023), dirección: <https://medium.com/@onkarmishra/how-textual-inversion-works-and-its-applications-5e3fda4aa0bc>.
- [10] PEFT). "DreamBooth fine-tuning with LoRA." (2023), dirección: <https://github.com/huggingface/peft/tree/main>.
- [11] S. Raschka). "Parameter-Efficient LLM Finetuning With Low-Rank Adaptation (LoRA)." (2023), dirección: <https://lightning.ai/pages/community/article/lora-llm/>.
- [12] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein y K. Aberman, *DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation*, 2022. DOI: 10.48550/ARXIV.2208.12242. dirección: <https://arxiv.org/abs/2208.12242>.