



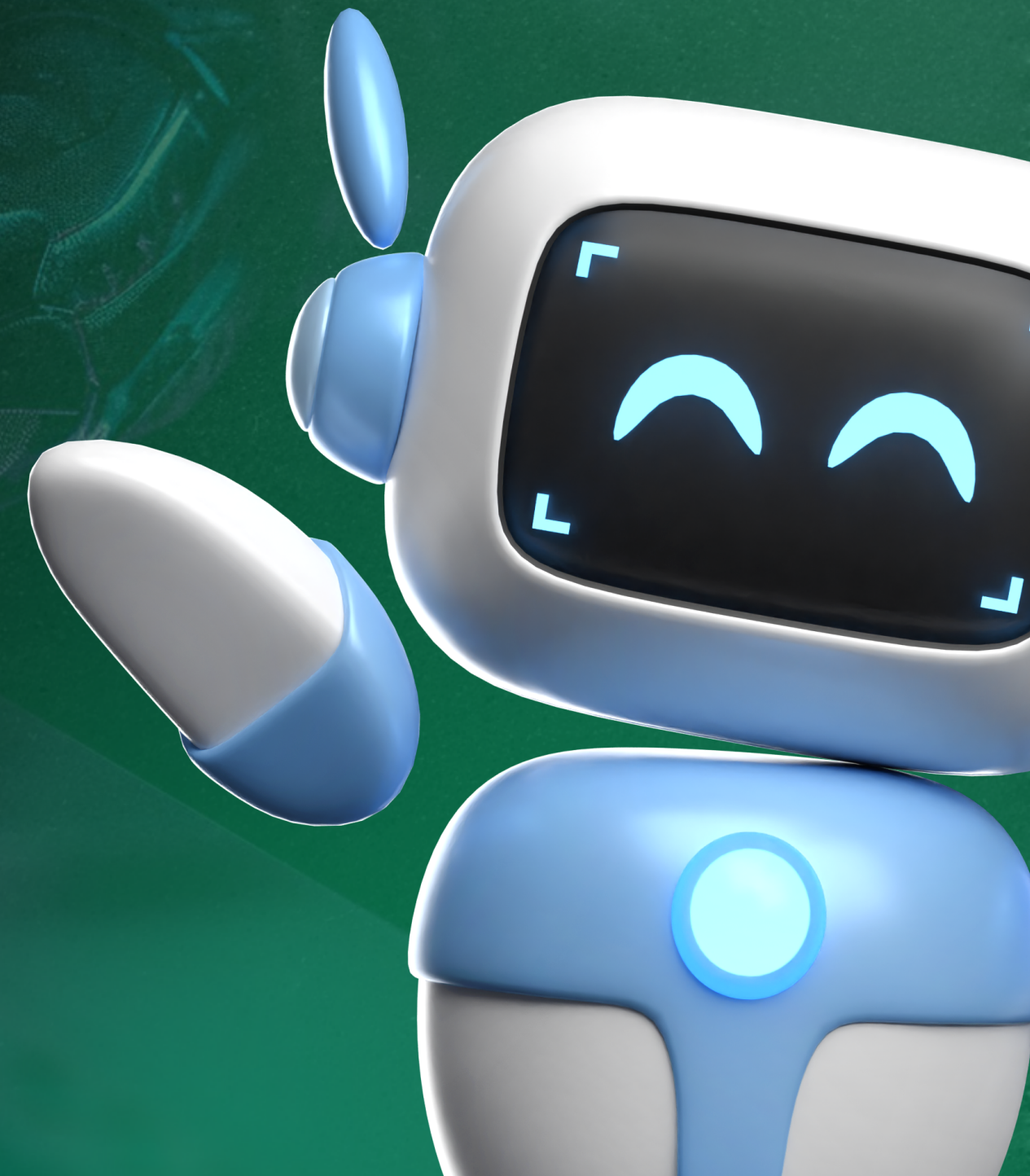
IA.HEROES
ACADEMY

¿QUÉ VEREMOS HOY?



IA.HEROES
ACADEMY

- ¿Dónde se está aplicando la Generación de Imagen mediante IA?
- Conociendo los participantes
 - a) Dall-E 3 b) Midjourney c) Stable Diffusion
- Cómo funcionan los modelos generación de Imagen por detrás?
 - a) GAN (Redes Adversarias Generativas)
 - b) VAE (Redes Generativas Autoencoder)
 - c) Modelos de Difusión
- Qué Modelos utilizan los 3 participantes?
- Cosas que la IA no puede hacer (aún)



1. ¿DÓNDE SE ESTÁ APLICANDO LA GENERACIÓN DE IMAGEN?

1. COMERCIO ELECTRÓNICO (E-COMMERCE) Y VENTA MINORISTA

- **Personalización de Productos:** Los minoristas en línea utilizan IA para generar imágenes realistas de productos personalizables, permitiendo a los clientes visualizar diferentes configuraciones antes de realizar una compra.
- **Pruebas Virtuales:** Las herramientas impulsadas por IA permiten a los clientes probarse virtualmente ropa, accesorios o maquillaje, superponiendo los artículos en la imagen del cliente.



2. BIENES RAÍCES

- **Escenificación Virtual:** Las empresas de bienes raíces utilizan IA para escenificar propiedades virtualmente generando imágenes de habitaciones amuebladas con varios artículos, ahorrando costos en comparación con la escenificación física.
- **Visualización de Propiedades:** Las herramientas de IA pueden generar imágenes fotorrealistas de propiedades en desarrollo, permitiendo a los posibles compradores visualizar la propiedad terminada.



3. PUBLICIDAD Y MARKETING

- **Contenido Personalizado:** La IA puede generar imágenes adaptadas a segmentos específicos de audiencia, mejorando la efectividad de las campañas de marketing.
- **Creación Dinámica de Anuncios:** Los sistemas de IA pueden generar innumerables variaciones de un anuncio, probándolos y optimizándolos en tiempo real para un mejor rendimiento.

4. ENTRETENIMIENTO Y MEDIA

- **Cine y Animación:** La IA se puede utilizar para crear personajes CGI realistas, entornos y efectos especiales de manera más eficiente que los métodos tradicionales.
- **Desarrollo de Videojuegos:** Los desarrolladores de juegos utilizan IA para generar texturas, paisajes y modelos de personajes de forma procedural, acelerando el proceso de desarrollo.



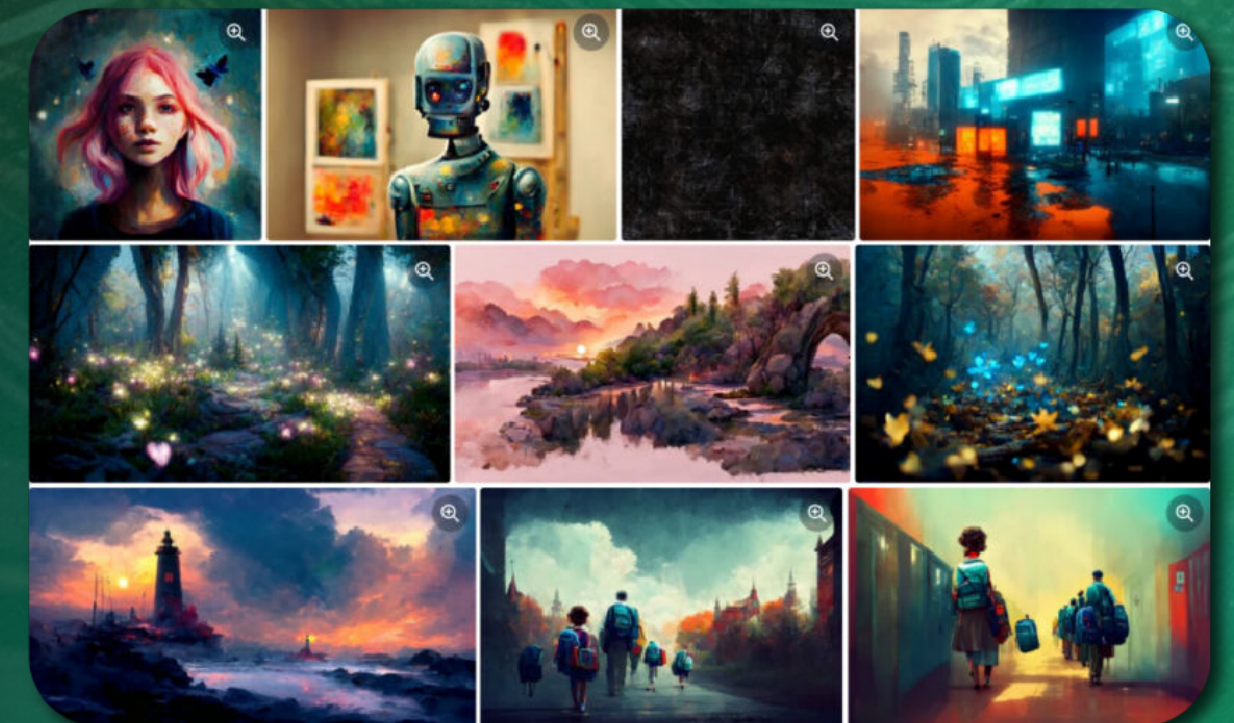
5. DISEÑO Y ARTE

- **Herramientas de Diseño Automatizado:** La IA puede producir elementos de diseño, plantillas y gráficos, ahorrando tiempo para diseñadores y artistas.
- **Creación de Arte con IA:** Los artistas utilizan IA para crear obras de arte únicas, y algunas de estas obras de arte generadas por IA se han vendido por cantidades sustanciales de dinero.



6. FOTOGRAFÍA DE STOCK

- **Generación de Imágenes de Stock:** La IA puede crear una amplia gama de fotos de stock que se pueden vender en plataformas como Shutterstock o Adobe Stock.





7. EDUCACIÓN Y CAPACITACIÓN

- **Material Educativo:** La IA puede generar imágenes y diagramas para materiales educativos, mejorando la experiencia de aprendizaje de los estudiantes.



8. CUIDADO DE LA SALUD

- **Imágenes Médicas:** La IA puede mejorar las imágenes médicas, hacer predicciones o crear representaciones visuales que ayuden en el diagnóstico y cuidado del paciente.



2. CONOCIENDO LOS PARTICIPANTES



IA.HEROES
ACADEMY



Midjourney



DALL·E

DALL·E 3

stability.ai

Stable Diffusion

2.1 DALL-E 3



- **Desarrollador:** OpenAI, una empresa que comenzó con la misión de desarrollar IA de manera que beneficie a la humanidad en su conjunto.
- **Resumen:** DALL-E 3 puede generar imágenes bastante buenas desde el principio, sin necesidad de indicaciones complejas, pero la desventaja es que lo que puede hacer con él es bastante limitado. No puede generar imágenes de personas famosas o imágenes que contengan ciertas palabras. En general, es excelente para comenzar y tener una idea de lo que la IA puede hacer, pero no es tan bueno si está tratando de obtener lo mejor de lo mejor de los modelos de difusión.
- Capaz de generar texto
- **Enlace:** <https://openai.com/dalle-e-3>



2.1 DALL-E 3

COMPLEJIDAD DE USO: ● ○ ○ ○ ○

LIBERTAD DE USO: ● ○ ○ ○ ○

COSTE: 20\$/MES ** INCLUIDO CON CHATGPT PLUS



2.2. MIDJOURNEY



IA.HEROES
ACADEMY

- **Desarrollador:** Midjourney es un laboratorio de investigación independiente que produce una IA patentada con el mismo nombre.
- **Resumen:** A la gente le encanta Midjourney porque genera imágenes con un estilo muy artístico que, por lo general, se ven muy bien de inmediato. Tiene un buen equilibrio entre lo que te permite hacer y la complejidad de usarlo, pero las imágenes generadas tienden a tener un aspecto más artístico que cualquier otra cosa (piensa en arte digital, pinturas, ilustraciones, diseño de personajes, etc.). curva de aprendizaje, pero una vez que aprendes a hacerlo, los resultados son en promedio muy buenos.
- **Enlace:** <https://www.midjourney.com/home/>

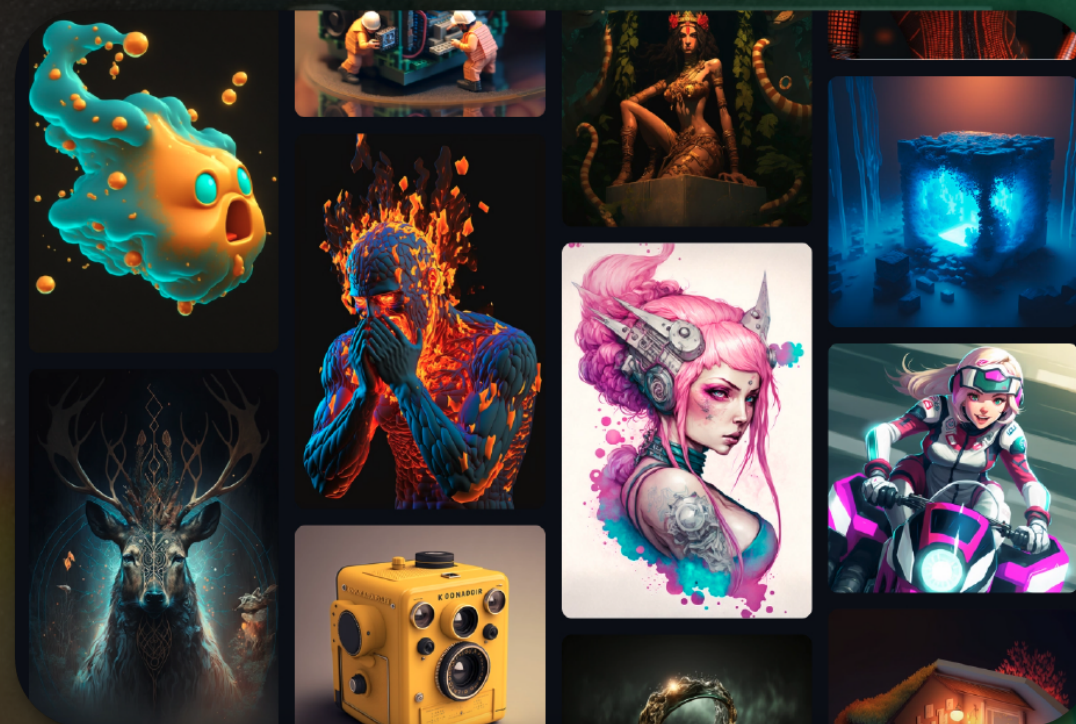


2.2. MIDJOURNEY

COMPLEJIDAD DE USO: ●●○○○

LIBERTAD DE USO: ●●○○○

COSTE: \$10/M (\$0,05 POR IMAGEN) O \$30/M IMÁGENES ILIMITADAS



2.3. STABLE DIFFUSION

- **Desarrollador:** Stability AI, una empresa privada que crea herramientas de IA de código abierto. Lanzaron Stable Diffusion como un modelo de difusión de código abierto en agosto de 2022.
- **Resumen:** Puede ejecutarse en computadoras relativamente comunes como la Macbook M1, eliminando la necesidad de tener que comprar y configurar supercomputadoras enormes con tarjetas gráficas muy costosas. Es de uso gratuito si lo ejecuta en su propia computadora, y puede hacer casi cualquier cosa, pero es más difícil obtener buenos resultados. Tiene una curva de aprendizaje empinada, pero una vez que dominas todos los trucos puedes crear cualquier cosa.
- **Enlace:** <https://stablediffusionweb.com/#demo>



2.3. STABLE DIFFUSION

COMPLEJIDAD DE USO: ● ● ● ● ○

LIBERTAD DE USO: ● ● ● ● ●

COSTE: GRATIS





	VIRTUD	COMPLEJIDAD	LIBERTAD	PRECIO	DESARROLLADOR
DALL-E 3	Fácil de Usar	● ○ ○ ○ ○	● ○ ○ ○ ○	20\$ / MES	Open AI
MIDJOURNEY	Resultados de alta calidad	● ● ○ ○ ○	● ● ○ ○ ○	\$10/MES (\$0,05 por imagen) o \$30/MES imágenes ilimitadas	Midjourney
STABLE DIFFUSION	Personalización y Control	● ● ● ● ○	● ● ● ● ●	Gratis	Stability AI

3. ¿CÓMO FUNCIONAN LOS MODELOS GENERACIÓN DE IMAGEN POR DETRÁS?



IA.HEROES
ACADEMY

- GAN (Redes Adversarias Generativas)
- VAEs (Redes Generativas Autoencoder)
- Modelos de Difusión

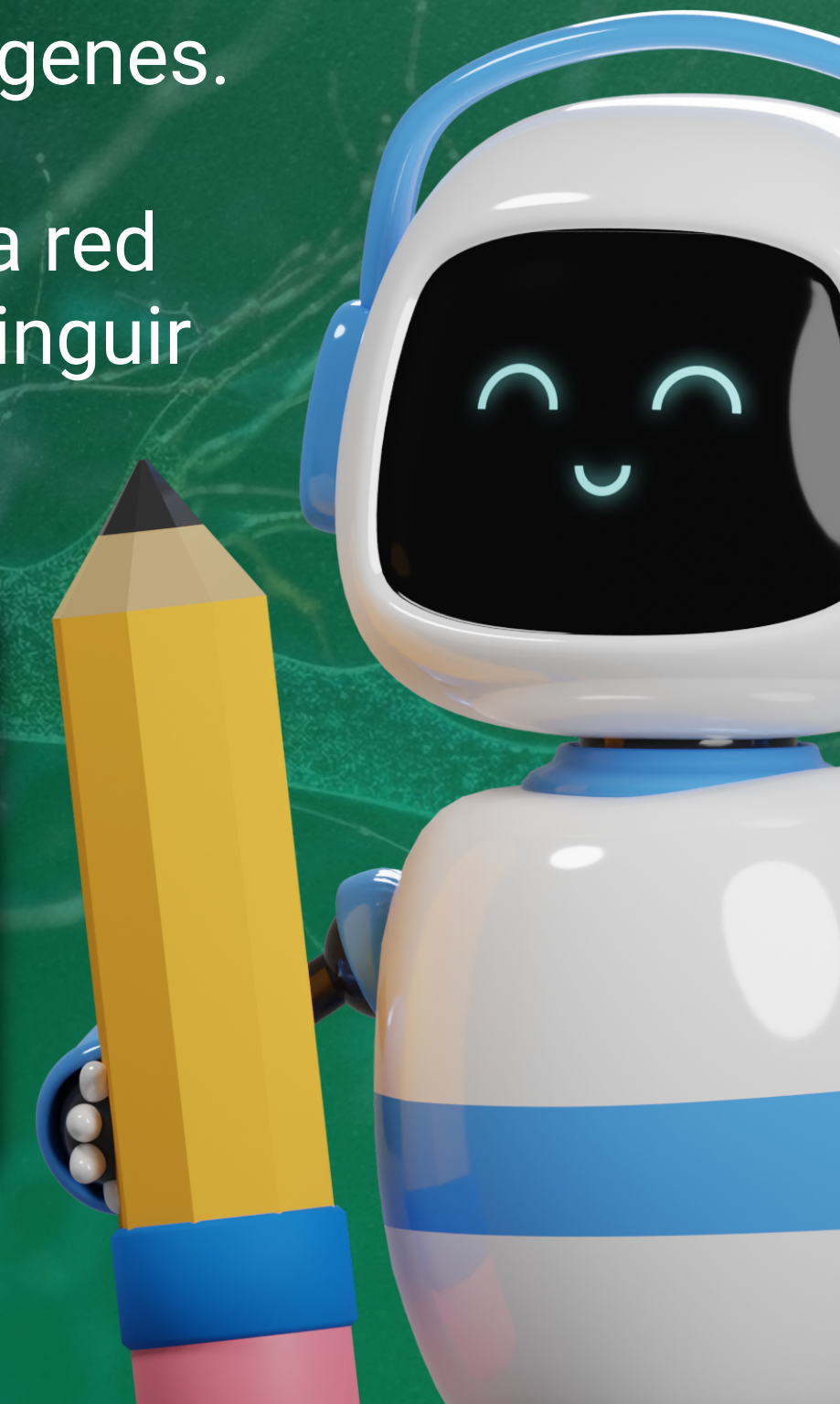
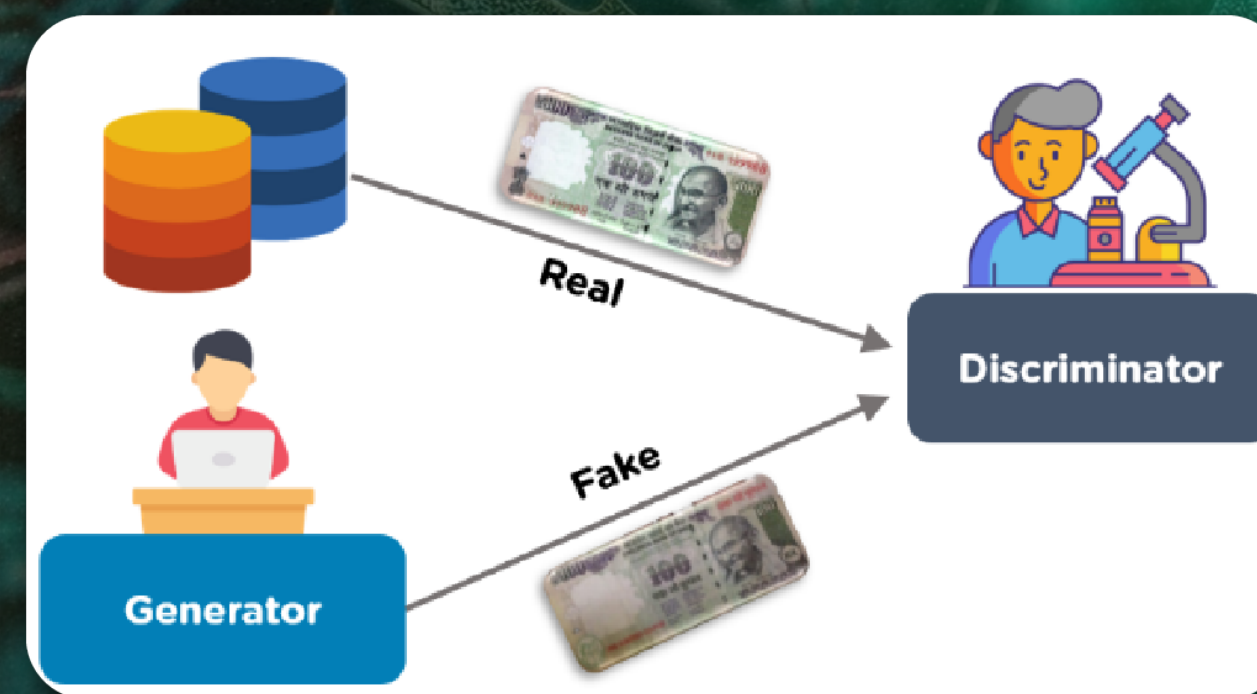
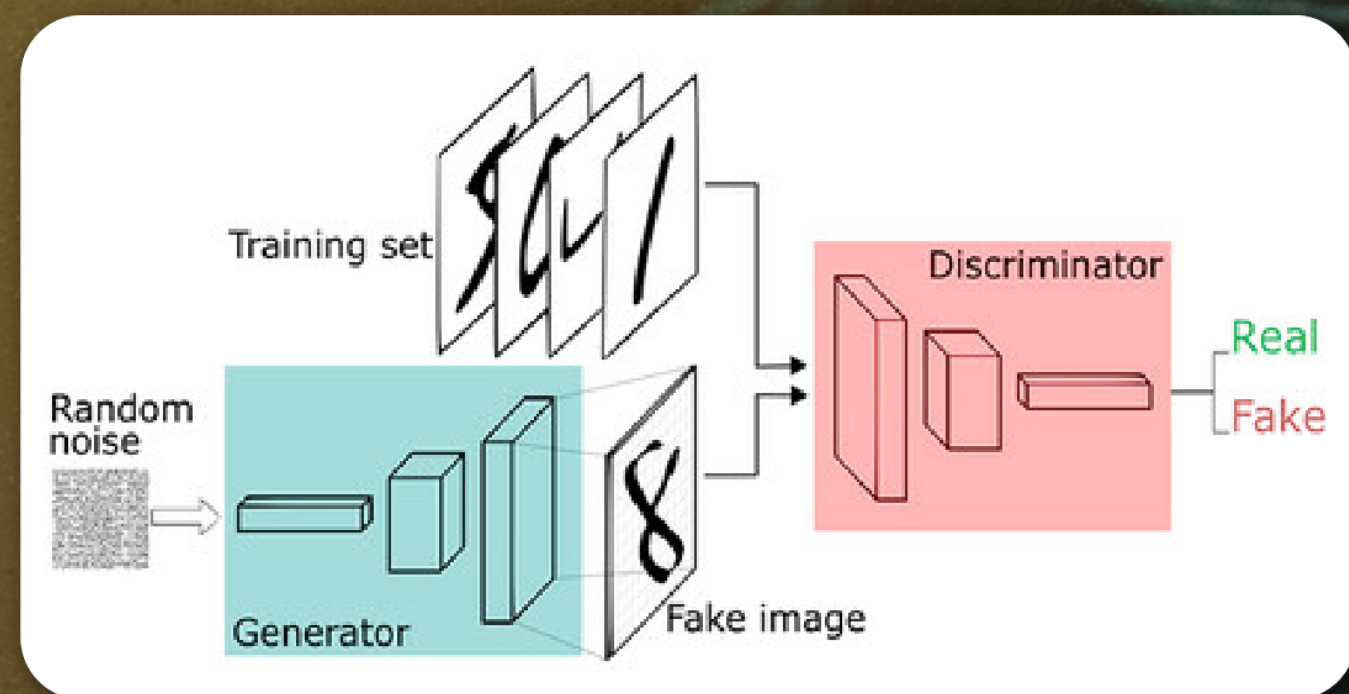
Cualquiera de estos modelos no "copia imágenes existentes" ni "hace collages" de diferentes imágenes. Las genera / crea!



3.1. GAN (REDES ADVERSARIAS GENERATIVAS)

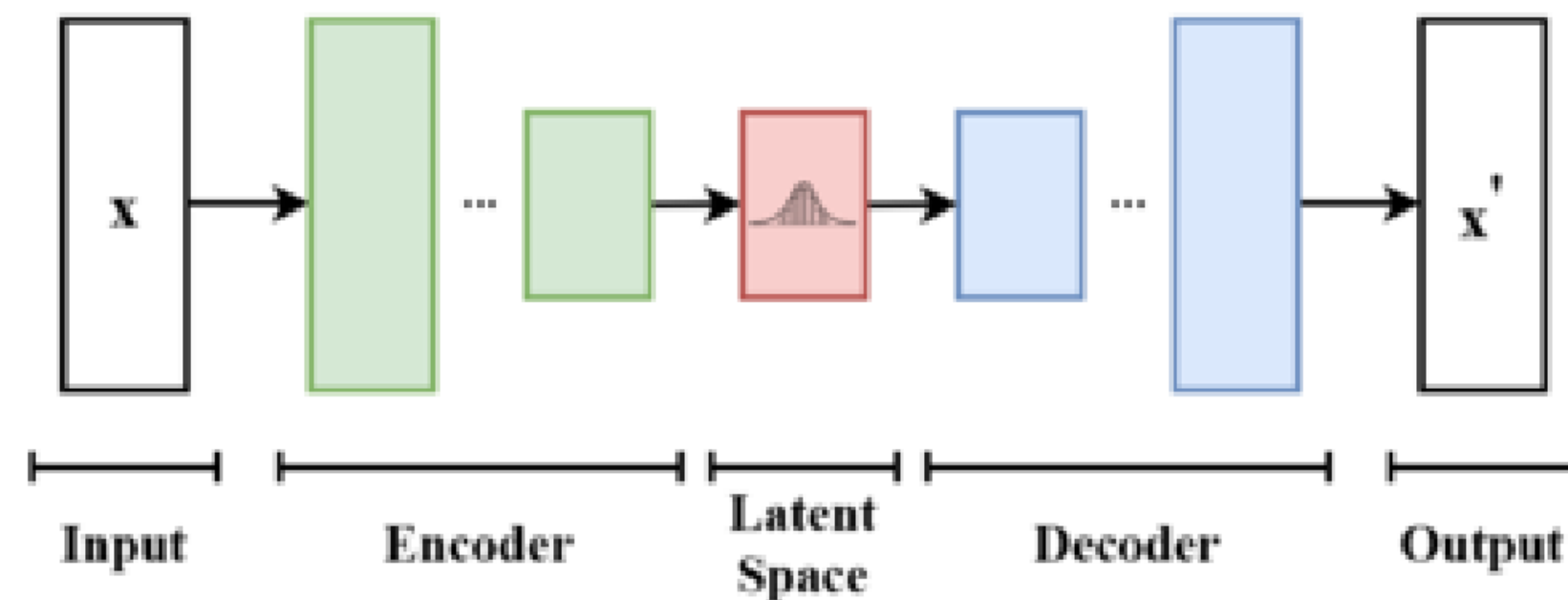
Este tipo de modelo ha sido el más utilizado para la generación de imágenes.

Las GANs consisten en dos redes neuronales que trabajan a la vez: una red generativa que crea imágenes y una red discriminativa que intenta distinguir entre las imágenes reales y las generadas por la red generativa.



3.2. VAES (REDES GENERATIVAS AUTOENCODER)

Las VAEs son un tipo de modelo de generación de imágenes que utiliza una red neuronal para comprimir una imagen en una representación de menor dimensión, y luego otra red para reconstruir la imagen a partir de esta representación comprimida. Las VAEs a menudo se utilizan en aplicaciones donde se requiere una comprensión más detallada de los datos, como la generación de imágenes médicas.



3.3. MODELOS DE DIFUSIÓN



IA.HEROES
ACADEMY

Los modelos de difusión son un tipo relativamente nuevo de generadores de imágenes en inteligencia artificial. En lugar de generar una imagen desde cero o transformar una imagen de entrada como hacen los GANs o VAEs, los modelos de difusión trabajan modificando iterativamente una imagen inicial (a menudo un ruido aleatorio) hasta que se convierte en una muestra del modelo de datos objetivo. Esto se logra a través de un proceso llamado cadena de Markov, donde cada iteración o "paso" del modelo se basa en el estado anterior.

Proceso de Difusión

A través de un proceso llamado difusión, se agrega ruido a una imagen hasta que se vuelve irreconocible.



Proceso de Difusión



IA.HEROES
ACADEMY

Primero, gradualmente agrega ruido a la imagen a través de una serie de pasos hasta que es 100% ruido



Para que entonces, la IA pueda pasar del ruido a la imagen deseada en función de un texto de entrada (Text PROMPT).



¿Que raro no? Y por que se realiza así este proceso? 🤔

Es más fácil "eliminar el ruido" de una imagen (hacerla un poco menos borrosa) que generar una nueva imagen desde cero



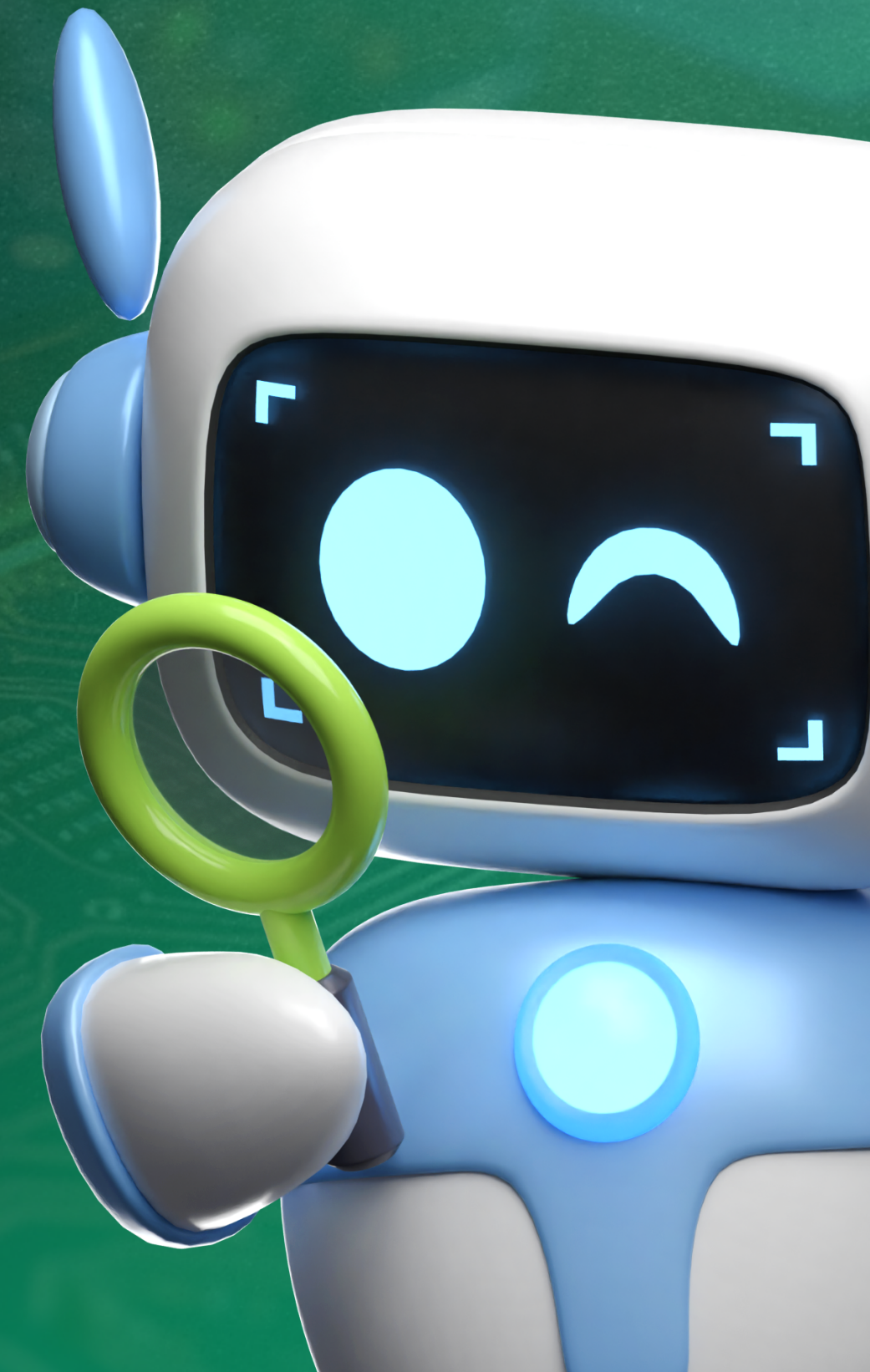
Sé que suena raro, pero, es simple. Así es como la IA aprende a generar imágenes:

- Coge la imagen original y agrega un poco de ruido.
- Luego agrega un poco más de ruido nuevamente.
- Sigue haciéndolo hasta que la imagen es todo ruido.



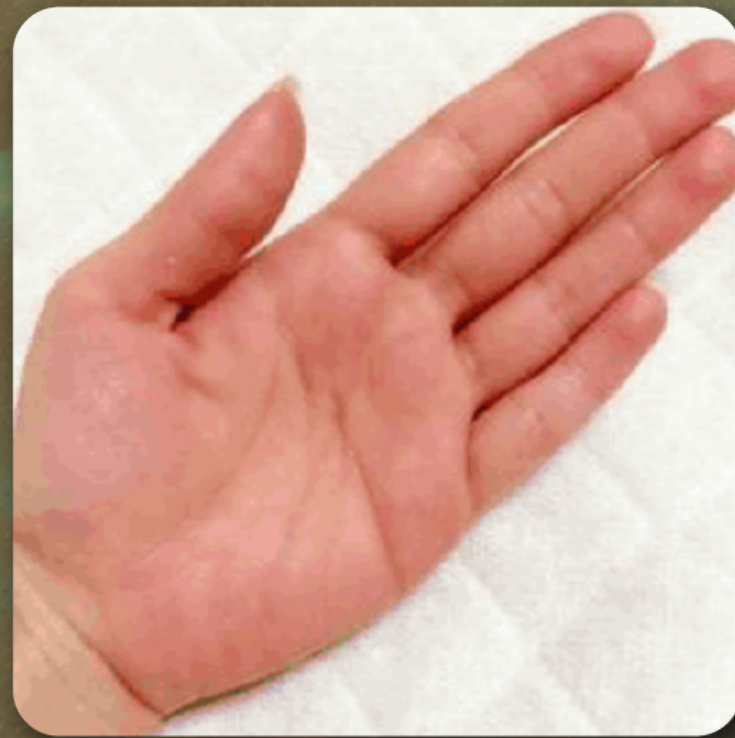


- Ahora, para cada imagen con ruido, intenta aprender a eliminar el ruido hasta que obtiene la imagen anterior con menos ruido.
- Sigue haciendo esto hasta que aprende a eliminar todo el ruido y genera la imagen original a partir de ruido puro. Todo el tiempo, la IA sabe lo que debe representar la imagen original ("un gato tirado en el suelo"), por lo que aprende a eliminar el ruido de una manera que se acerca más a un gato en el suelo, y no solo eliminando el ruido al azar.
- Una vez que la IA ha repetido este proceso suficientes veces con millones de imágenes, aprende cómo pasar de un ruido aleatorio a una imagen clara, según una descripción de cómo debería verse la imagen. Aprende eficazmente a representar un gato, a dibujar una flor o a pintar al estilo de Van Gogh.



Inconvenientes de los modelos de Difusión

Ahora que sabemos cómo la IA aprende a crear imágenes, podemos tener una idea intuitiva de por qué ciertos tipos de imágenes son difíciles para los modelos de difusión. El peligro, como siempre, está en los **detalles**.



Cuando agregamos ruido a las imágenes, perdemos los detalles finos. Las cosas que originalmente parecían claras ahora se ven borrosas, y es difícil ver qué es qué, cuándo termina algo y cuándo comienza algo.



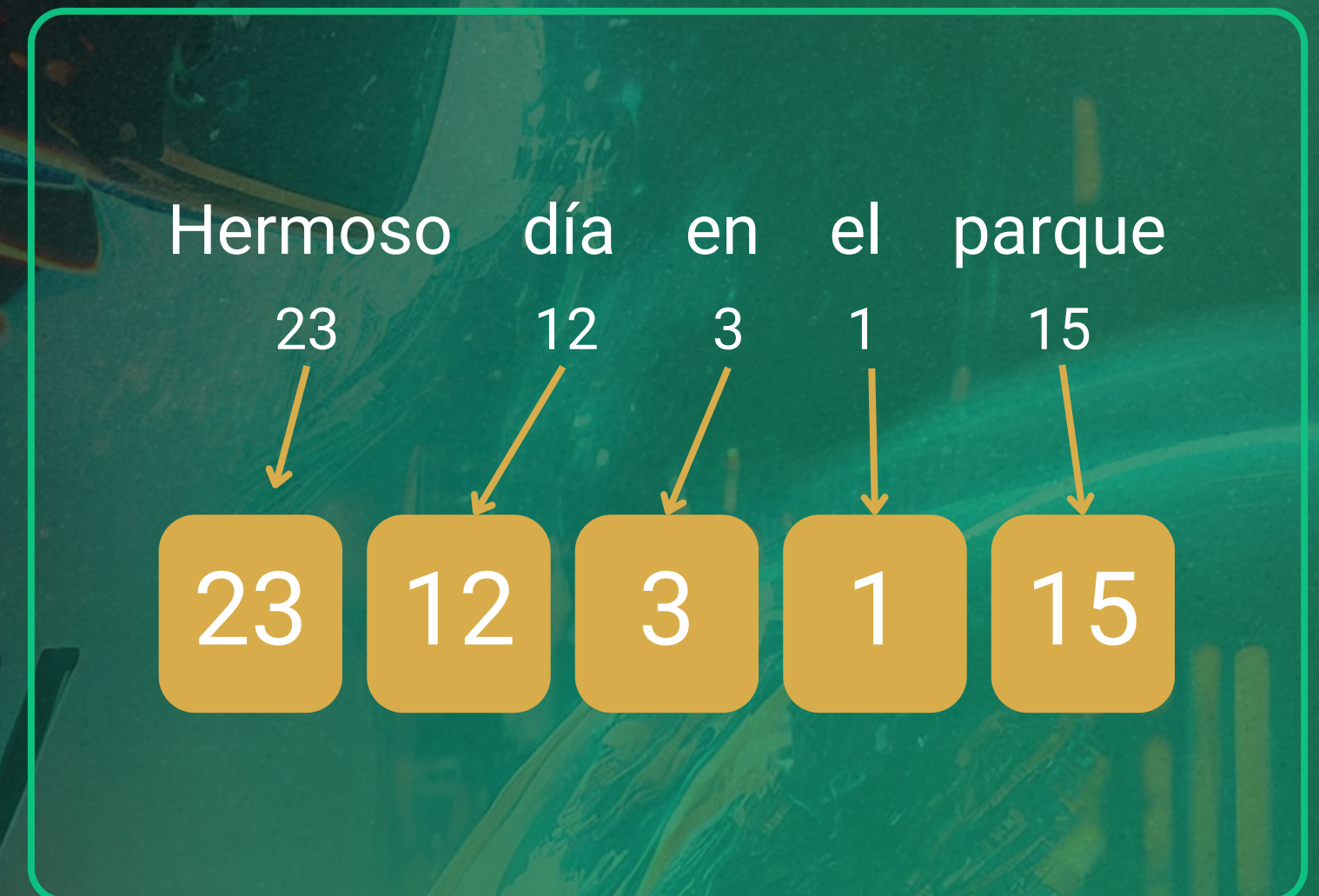
4. ¿QUÉ MODELOS UTILIZAN LOS 3 PARTICIPANTES?

DALL·E 3 consiste principalmente de 2 partes: una para convertir la entrada del usuario (input) en la representación de una imagen (llamada Prior) y otra para convertir esta representación en una foto real (llamada Decoder).



Las incrustaciones (embeddings) de texto e imágenes utilizadas provienen de otra red llamada **CLIP** (Contrastive Language-Image Pre-training), creada por OpenAI. **CLIP es una red neuronal que devuelve el mejor prompt para una imagen de entrada.** Hace lo contrario de lo que hace DALL·E 3: conversión de texto a imagen. El objetivo de CLIP es aprender la conexión entre la representación visual y textual de un objeto.

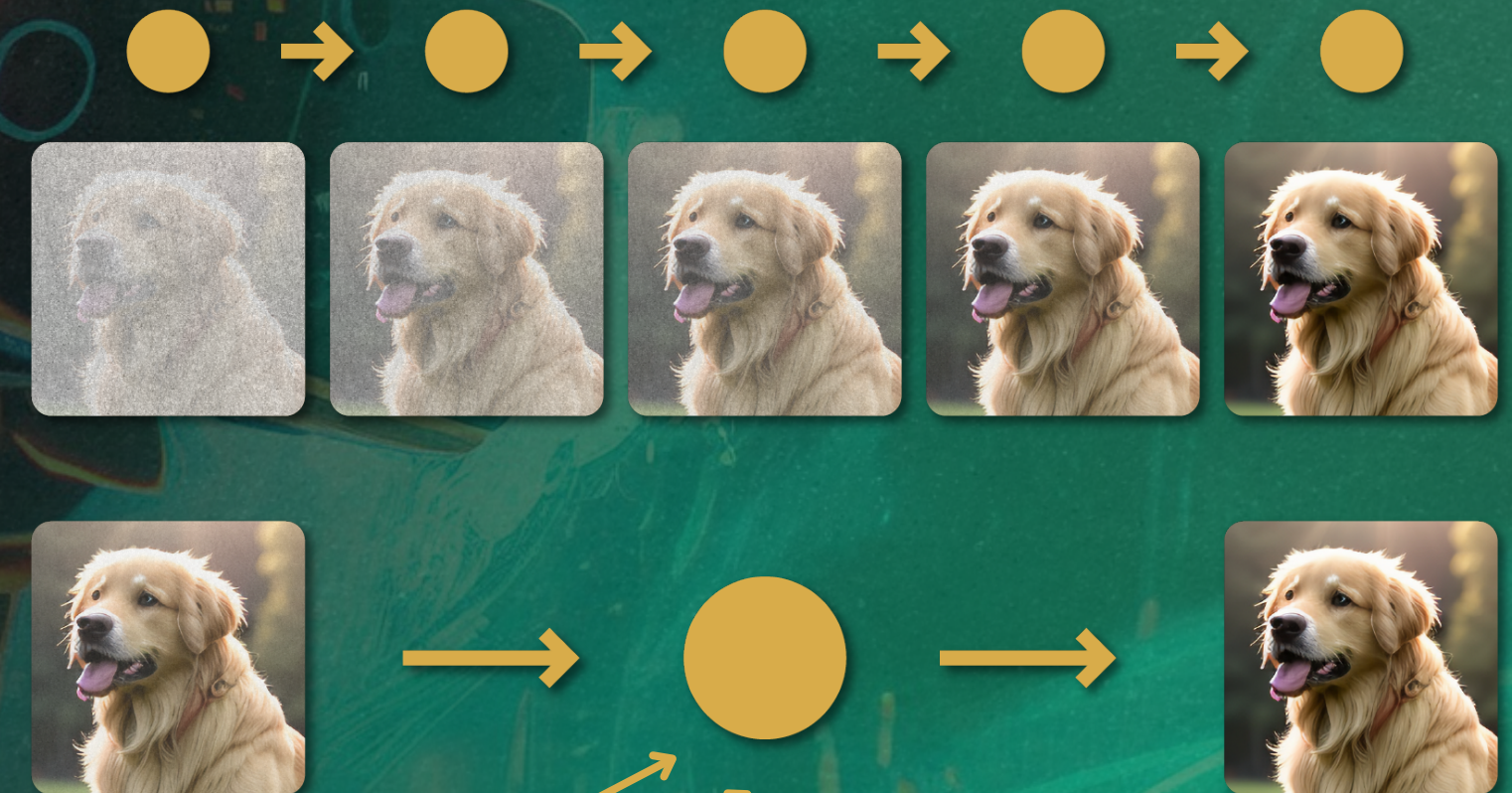
4. ¿QUÉ MODELOS UTILIZAN LOS 3 PARTICIPANTES?



4. ¿QUÉ MODELOS UTILIZAN LOS 3 PARTICIPANTES?



IA.HEROES
ACADEMY



- 1 El prompt se transforma en un embedding de texto CLIP mediante una red neuronal.
- 2 Prior reduce la dimensionalidad del embedding de texto mediante el análisis de componentes principales o PCA.
- 3 El embedding de la imagen se crea utilizando el embedding de texto.
- 4 En el paso del decodificador, se utiliza un modelo de difusión para transformar el embedding de la imagen a la imagen final.

4. ¿QUÉ MODELOS UTILIZAN LOS 3 PARTICIPANTES?



IA.HEROES
ACADEMY

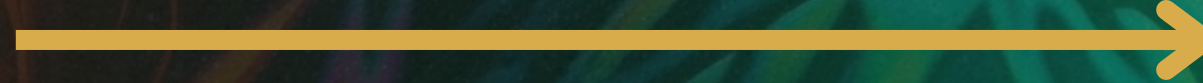
No se sabe a ciencia cierta debido a que es una empresa privada

MIDJOURNEY



Modelo de
Difusión

STABLE DIFFUSION



Modelo de
Difusión

5. COSAS QUE LA IA NO PUEDE HACER (AÚN)



IA.HEROES
ACADEMY

HAY ALGUNAS TAREAS QUE LA IA GENERATIVA NO PUEDE HACER... TODAVÍA.

En IAHeroes, hemos descubierto que los modelos de difusión son mejores para generar cosas que son:

- **Arte** (pinturas, fondos, cosas abstractas, imágenes puramente artísticas)
- **Inspiración** (edificios, personajes, vibraciones generales y estética)
- **Diversión** (la moda de los avatares de IA es un buen ejemplo)

Es excelente en cuanto a la etapa creativa, pero **la tecnología aún no está lista para generar imágenes donde los detalles son importantes.**

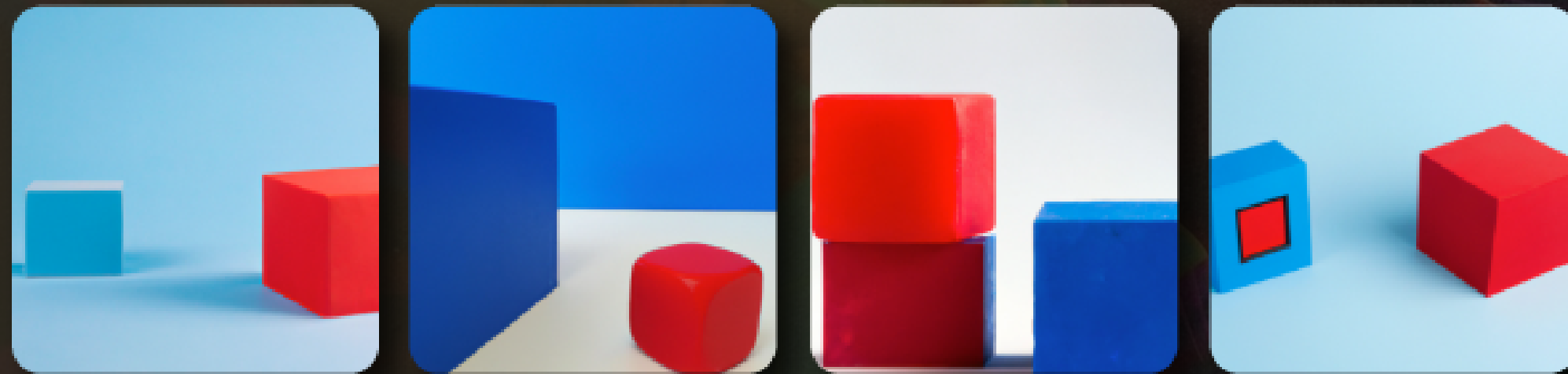




TAREAS CONCRETAS QUE LA IA NO PUEDE HACER

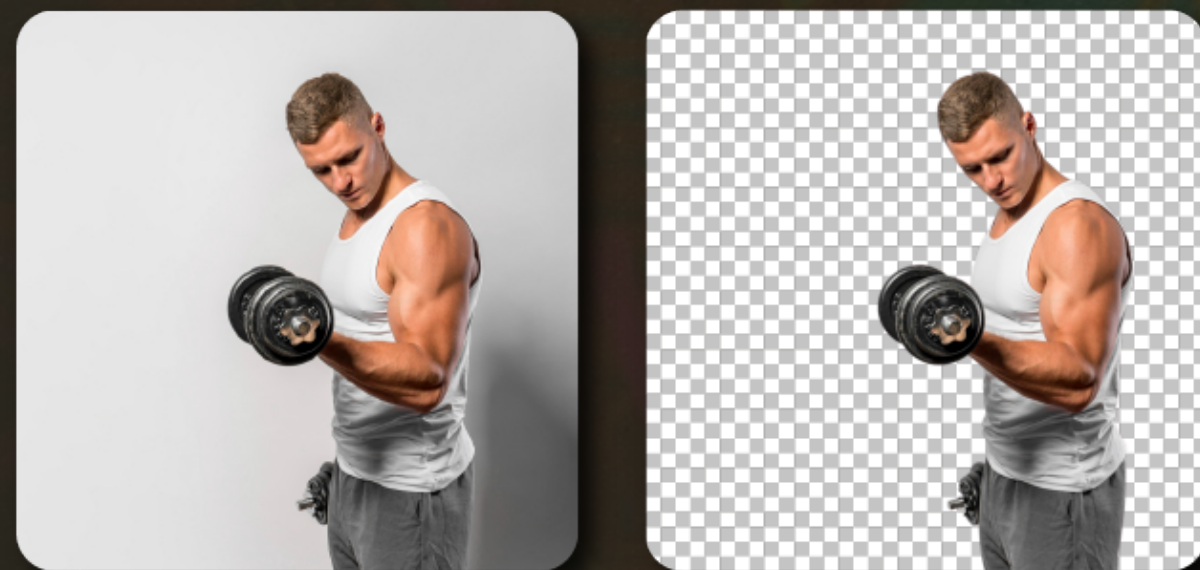
Noción del espacio: los modelos de difusión a menudo carecen de visión espacial.

Cubo rojo encima de un cubo azul



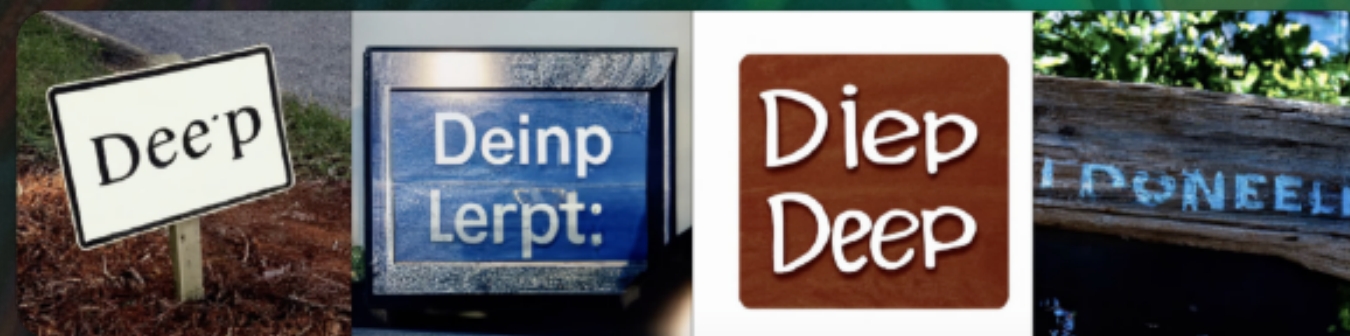
“vemos como el cubo rojo no se sitúa encima del cubo azul”

Transparencia



Texto

Imagen de un letrero con la frase “Deep Learning” en él... ¿O sí?





IA.HEROES
ACADEMY