

Imágenes sintéticas impulsadas por Inteligencia Artificial: Su impacto en género.



1 ¿Qué tipo de datos existen?

» Datos de seres u objetos que existen o existieron¹

Son aquellos que se recopilan o se miden en el mundo real.

Pueden usarse para entrenar modelos de IA y estar basados en objetos, eventos o personas reales.

Datos sintéticos

Consisten en información anotada que las simulaciones por computadora o algoritmos generan como alternativa a los datos del mundo real.

Se crean en entornos digitales.

Reflejan datos del mundo real, matemática o estadísticamente. Sirven para entrenar modelos de IA².

Entornos simulados o Digital Twin

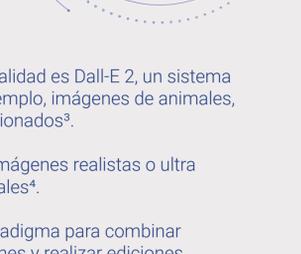
Sobre los datos de un objeto determinado y lo que se encuentra a su alrededor, es posible crear modelos computacionales que representan los comportamientos o estados del objeto físico.

Es la representación de un objeto, proceso o servicio.

Sirven para hacer representaciones digitales de objetos, procesos o servicios físicos y permiten realizar simulaciones para saber cómo funcionarán antes de crear e implementar los cambios en los objetos reales.

2 ¿Qué son las imágenes sintéticas impulsadas por la Inteligencia Artificial?

Son una especie del género dato sintético. Son imágenes artificiales generadas por algoritmos de inteligencia artificial.



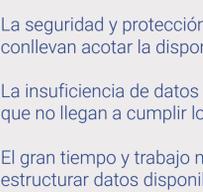
3 ¿Cómo se generan?

Generalmente, a partir de un comando de voz o por texto, es decir, por lenguaje natural, provisto por el usuario/a.

Uno de los sistemas más utilizados en la actualidad es Dall-E 2, un sistema que utiliza redes neuronales para crear, por ejemplo, imágenes de animales, objetos, o combinación de conceptos no relacionados³.

DALLE-2 se lanzó en 2022 y permite generar imágenes realistas o ultra realistas desde cero o a partir de imágenes reales⁴.

Este tipo de sistemas constituye un nuevo paradigma para combinar conceptos, atributos y estilos, expandir imágenes y realizar ediciones realistas en imágenes existentes⁵.



4 ¿Para qué sirven?

a.- Para entrenar sistemas inteligentes que requieren datos y solucionar problemas que se generan con datos no sintéticos, como:

- La seguridad y protección de los datos cuando las medidas a tomar conlleven acotar la disponibilidad necesaria para el entrenamiento.
- La insuficiencia de datos disponibles, cuando existen en cantidades que no llegan a cumplir los requerimientos.
- El gran tiempo y trabajo minucioso que en ciertos casos conlleva estructurar datos disponibles y suficientes en forma adecuada.

b.- Como una herramienta para inspirar y acelerar los procesos creativos de las personas⁶.

5 ¿Cómo lo hacen?

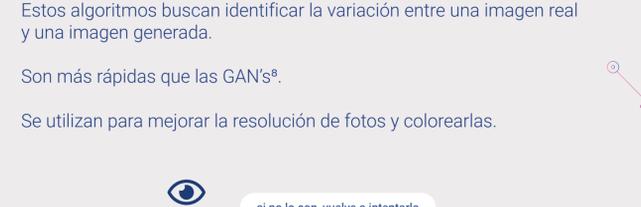
1. Lucha entre algoritmos. (Redes generativas adversarias)

Dos algoritmos participan de un juego de suma cero (**redes generativas adversarias - GAN**).

Un algoritmo reconoce imágenes de un tipo, el otro algoritmo genera imágenes inexistentes de un tipo específico que pasan a examen de su algoritmo contrincante (red adversaria).

El primer algoritmo decide si la imagen cumple con los requisitos del tipo de imagen que le exige el otro algoritmo. Como resultado, **tenemos imágenes cada vez más realistas**⁷.

Ilustración de una lucha entre algoritmos sobre una imagen sintética de un billete.



2. Buscadores algorítmicos de variaciones (Codificadores automáticos variacionales)

Estos algoritmos buscan identificar la variación entre una imagen real y una imagen generada.

Son más rápidas que las GAN's⁸.

Se utilizan para mejorar la resolución de fotos y colorearlas.



3. Algoritmos de IA que completan imágenes (modelos autorregresivos)

Con esta técnica se puede predecir un conjunto de datos que no existe, pero que el algoritmo permite completar a partir de la información existente⁹.

Ilustración de algoritmos que completan imágenes



6 Impacto de las imágenes sintéticas generadas por Inteligencia Artificial en cuestiones de género.

a. Todas estas técnicas algorítmicas tienen el potencial de reproducir estereotipos de género, prejuicios, creencias sexistas y valores socialmente arraigados sobre el rol de las mujeres y los varones en la sociedad.

Ello ocurre porque:

- Los modelos informáticos se nutren de datos reales que contienen estereotipos de género, creencias sexistas, valores y prejuicios.
- Falta de perspectiva de género y diversidad en las personas que toman decisiones en el ciclo de vida del modelo informático.

Por ejemplo, si coloco el término "secretaries working" en el sistema de IA Dall-E 2, el resultado arroja imágenes de personas del género femenino únicamente.

Del mismo modo, al ingresar la frase "doctor working during the pandemic", el sistema limita la búsqueda a personas de sexo masculino.



b. Además, si las imágenes sintéticas se combinan con imágenes de mujeres¹⁰ reales pueden dar lugar a:

- Generación de contenido grotesco, violento o sexual imaginario -deepfakes-.
- Incentivo al bullying y al acoso cibernético por las imágenes sintéticas de contenido sensible. Viralización de imágenes sintéticas no consentidas.

La disponibilidad abierta de estas herramientas no permite controlar de forma adecuada su utilización.

Equipo UBA IALAB

Dirección

Juan G. Corralán
Nieves Macchiavelli

Investigadoras

Mariana Sánchez Caparrós
Pilar Hidalgo León
Carina Papini

Coordinación

Giselle Heleg

Diseño Gráfico

Victoria Mafud
Abraham Lucena

Notas:

1. Se trata de datos que emergen de interacciones, procesos, vivencias, experiencias y que se corresponden a personas, objetos, lugares, procesos.

Es necesario tener en cuenta que los datos que se generan en el entorno virtual también pueden ser datos que existen o existieron y que surgieron de procesos. A su vez, en el mundo físico también pueden usarse datos artificiales.

Realizamos un esfuerzo por evitar el uso de la palabra real, debido a que el mundo real hoy en día incluye tanto el mundo físico como el entorno virtual. Ambos se combinan y retroalimentan de manera constante. Por lo tanto, el uso de la palabra real puede generar ambigüedad.

2. ¿Qué Son los Datos Sintéticos? | Blog de NVIDIA

3. DALL-E: Creating Images from Text (openai.com)

4. Los modelos son entrenados con imágenes reales y aprenden a reflejar las relaciones existentes entre las imágenes y el texto expresado en lenguaje natural por el usuario, así como la que hay entre diversos objetos y entre acciones y objetos. <https://openai.com/dall-e-2/>

5. DALL-E 2 (openai.com)

6. Open AI, "Dall-e Now available in beta", en <https://openai.com/blog/dall-e-now-available-in-beta/> [acceso el 20/9/2022].

7. Este análisis será el motor de la retropropagación de la red Generativa, que aprenderá cada vez a engañar mejor a su rival y por tanto a generar imágenes más reales cada vez. La perfección de las imágenes dependerá de lo bien entrenada que esté la red Adversaria, es decir, de lo perfeccionista que esta sea.

Ver más en Módulo Introducción a la Inteligencia Artificial Unidad 01 Cronología y límites de la Inteligencia Artificial, página 58. Bibliografía provista por la Maestría CEUPE.

Redes Neuronales Generativas Adversarias (GAN). Este tipo de redes neuronales presentadas en 2012 por primera vez, permiten generar datos que no existían previamente gracias a una competición entre una red que genera contenido y otra que valora su calidad. Gracias a este tipo de redes neuronales, que en ocasiones están formadas por redes neuronales de otros tipos en función del objetivo que se persiga, somos capaces de crear contenido totalmente nuevo que se parezca a otro contenido visto con anterioridad. Ver más en: Módulo Introducción a la Inteligencia Artificial Unidad 03 Métodos y tecnologías Fundacionales, página 14. Bibliografía provista por la Maestría CEUPE.

Las GAN se utilizan también para la restauración de imágenes y videos de otras épocas. Ver más en: Módulo Introducción a la Inteligencia Artificial Unidad 04 Aplicaciones e impacto en el mundo real, página 9. Bibliografía provista por la Maestría CEUPE. La primera, obtiene una estructura de datos o vector aleatorio de una muestra de imágenes, decodifica esa información y crea una nueva imagen.

La segunda, compara la imagen generada con la muestra de imágenes reales.

Si las imágenes coinciden en características que cumplen una etiqueta o clase, entonces la red que compara valida la imagen como correcta. Caso contrario, vuelve iterativamente hasta lograr el objetivo. Esto es, impulsar una imagen sintética muy parecida a las reales de la muestra de entrenamiento.

8. Hacen la comparación con imágenes de baja resolución (la distribución normal de las características de la imagen) y luego las maximizan comparando el error entre la imagen generada maximizada y la imagen de baja resolución original.

9. Pueden hacerlo utilizando la secuencia de píxeles de entrada y comparando su distribución normal con la distribución de los datos de entrenamiento de forma condicional, en un proceso iterativo, hasta completar la imagen.

10. Ver <https://medium.com/codex/editing-real-faces-is-back-in-openais-dall-e-2-1a4ec28dd2>. También <https://www.theguardian.com/technology/2022/sep/21/dall-e-2-users-allowed-upload-faces-for-first-times>