



03

IA GENERATIVA

## UNIDAD 3: IA GENERATIVA

Les damos la bienvenida a Generación IA el primer curso de alfabetización digital sobre IA Generativa en Latinoamérica.

Generación IA es una formación desarrollada por Eidos Global y Microsoft que tiene como objetivo ampliar el horizonte de oportunidades laborales de las personas en Latinoamérica y Caribe ante la nueva economía mundial transformada por la inteligencia artificial.

Es una experiencia consolidada, un espacio seguro para aprender a hacer uso del potencial de esta tecnología que está transformando nuestras formas de trabajar. Generación IA es un espacio para aprender a utilizar esta tecnología de forma responsable y creativa. A través de módulos interactivos y recursos actualizados, explorarán los fundamentos de la IA y descubrirás cómo puede optimizar procesos, mejorar la toma de decisiones y beneficiar a la comunidad

### ¿Qué vamos a ver en esta unidad?

En esta unidad nos enfocaremos en la IA generativa, sus fundamentos y cómo diseñar prompts efectivos. Aprenderemos sobre elementos clave en los prompts y estrategias avanzadas para optimizar resultados. Además, investigaremos diversos formatos generativos como texto a audio y otros formatos, y examinamos aplicaciones prácticas. Finalmente, abordaremos el concepto de alucinaciones en IA para entender sus limitaciones y posibles errores.

# El arte de dar instrucciones

## ¿Qué es la IA generativa?

En unidades anteriores mencionamos que dentro del mundo de la IA podemos distinguir 2 tipos según su capacidad: descriptiva o generativa. Hablemos específicamente de la **IA generativa**.

La inteligencia artificial generativa es una disciplina dentro del Deep Learning, que tiene como objetivo generar contenidos en diferentes formatos como texto, imágenes, videos, modelos 3D, etc. Como aprendimos anteriormente, dentro del campo de la inteligencia artificial, existen otros subcampos, el más importante hoy en día es el Machine Learning. Mencionamos que el Machine Learning utiliza distintas técnicas para llevar a cabo diversas tareas (como por ejemplo, clasificar contenido), pero la técnica que saltó a la popularidad en los últimos años fueron las redes neuronales. La aparición de este tipo de algoritmo y el gran desarrollo que se viene gestando de esta técnica dieron paso al Deep Learning. Gracias a este gran avance es que la IA logró generar contenido de gran calidad, incluso a veces difícil de distinguir de algo producido por humanos.

Si bien podemos notar algunos rasgos característicos de contenido que fue creado con IA generativa, nada impide pensar que en un futuro cercano esta tecnología continúe avanzando y cada vez sea más difícil diferenciar material creado por humanos de aquello creado por la máquina. Pero no debemos inquietarnos; es importante entender que esta tecnología todavía sigue en desarrollo y, al sumergirnos en su estudio, tenemos la oportunidad de ser parte de una revolución que transformará nuestra manera de trabajar y aprender. Cada avance en IA abre nuevas puertas para la creatividad y la innovación, y al involucrarnos en su evolución, podemos contribuir a construir un futuro más fascinante y prometedor.

Pregunta para reflexionar: ¿Cómo crees que la IA generativa podría impactar tu campo profesional o tus hobbies?"

## ¿Qué es un prompt?

Hasta el momento, estuvimos trabajando en entender el interior de una IA. Lo que vamos a empezar a profundizar ahora es en cómo usarla, y para ello, vamos a comenzar trabajando con el concepto de **prompt**.

Un prompt es una instrucción o un conjunto de instrucciones que se le da a la IA para generar una respuesta o realizar una acción específica. Funciona como una instrucción o pregunta que guía al modelo en la tarea que se espera que realice, que puede variar desde responder preguntas, completar textos, generar ideas creativas, crear imágenes, hasta traducir idiomas o escribir código.

La naturaleza y estructura del prompt pueden influir significativamente en la calidad y relevancia de la respuesta del modelo. La calidad del prompt determina la calidad de la respuesta. Por eso, la elaboración de prompts efectivos es considerada una habilidad importante al trabajar con estos modelos. **Los prompts bien formulados son esenciales ya que son la manera que tenemos para comunicarnos con la IA generativa.**

## Elementos de un prompt

Un prompt bien estructurado es fundamental para obtener resultados precisos y relevantes de una IA. Para comenzar, hay que considerar los distintos elementos que idealmente puede tener un prompt. En este caso, vamos a profundizar en la estructura de un prompt para un modelo de texto a texto.

1. **Rol y contexto:** Es la parte inicial del prompt. Aquí le indicas a la IA cómo debe de actuar y cuál es la situación que se está abordando en este momento. Se debe ser tan específico como sea posible y sin caer en redundancias. Por ejemplo: *Actúa como un gestor de proyectos experimentado. La empresa te ha asignado la tarea de supervisar un proyecto de desarrollo de software para un nuevo producto.*
2. **Instrucción / pregunta:** Especificamos la acción que queremos que la IA realice o el objetivo que debe alcanzar. Debemos proporcionar claridad sobre qué se espera en la respuesta generada. Por ejemplo: *Elabora una lista de riesgos comunes que pueden afectar el proyecto y propón medidas para mitigarlos.*

3. **Formato de salida:** Aquí es donde le indicamos a la IA lo que esperamos obtener por respuesta. Es importante especificar el formato, sin embargo, si no necesitamos ninguno en particular, lo puedes omitir. También podemos incluir el tono que queremos que utilice (formal, técnico, poético, irónico, etc.) Por ejemplo: *La lista debe contener al menos 5 ítems en formato de viñetas. El tono de las descripciones debe ser formal.*

4. **Restricciones y aclaraciones:** Igual de importante es pedir qué hacer, como pedir qué no hacer. Sucede que a veces la IA ofrece información que no nos sirve para nada. Las restricciones no son indispensables, pero, acorde a lo que buscas obtener, pueden ser importantes para afinar la respuesta. Por ejemplo: *No repitas mi pregunta, no te disculpes, no expliques lo que estás haciendo ni por qué. Limitate a contestar única y exclusivamente lo que te pregunté.*

## Prompt Engineering

En el apartado anterior trabajamos sobre diversos actores que cumplen un rol fundamental en el mercado de la IA. Sin embargo, no podemos olvidarnos de un actor fundamental a la hora de utilizar modelos de IA: el usuario final. Como vimos anteriormente, quienes hagan uso de un modelo de IA deberá elaborar un prompt para indicarle al modelo qué precisa. Esta instrucción o conjunto de instrucciones que le damos al modelo, pueden alterar el resultado que obtengamos. En otras palabras, la naturaleza y estructura del prompt pueden influir significativamente en la calidad y relevancia de la respuesta del modelo. La calidad del prompt (entrada), determina la calidad de la respuesta (salida). Por eso, la elaboración de prompts efectivos es considerada una habilidad importante al trabajar con inteligencia artificial generativa.

¿Cómo podemos interactuar de manera efectiva con las máquinas que funcionan con IA? ¿Cuál es el lenguaje que manejan? ¿Cómo podemos guiarlas para que cumplan con nuestras necesidades y expectativas? Estas son algunas de las preguntas que debemos responder para convertir a la IA en una verdadera aliada.

El Prompt Engineering es la habilidad de diseñar indicaciones claras y precisas para guiar a la IA en la resolución de tareas. Al proporcionar instrucciones detalladas, aumentamos las posibilidades de que la IA genere resultados que se alineen con nuestras expectativas. Esta habilidad es fundamental para aprovechar al máximo el potencial de la IA, comparable a aprender un nuevo idioma que nos permite comunicarnos con estos agentes artificiales y orientarlos hacia los resultados deseados.

El **Prompt Engineering** se puede usar para una variedad de tareas y aplicaciones, dependiendo de nuestros objetivos. Algunos ejemplos incluyen:

- **Generación de contenido creativo:** Como escribir poemas, guiones, letras de canciones, o incluso código de programación.
- **Respuestas a preguntas:** Tanto preguntas abiertas (como "¿Cuáles son los beneficios de implementar IA en marketing?") como cerradas (como "¿Puede la IA mejorar las estrategias de marketing?").
- **Traducción de textos:** Indicar a la IA que traduzca un texto de un idioma a otro.
- **Resumen de información:** Pedirle a la IA que resuma textos largos en puntos clave.
- **Asistencia en la escritura:** Generar ideas, reescribir párrafos, o mejorar la claridad de un texto.
- **Automatización de tareas:** Optimizar procesos repetitivos o flujos de trabajo usando IA.
- **Resolución de problemas:** Guiar a la IA para resolver cuestiones complejas mediante prompts detallados.

El Prompt Engineering puede ser útil en cualquier escenario donde se quiera guiar a una IA para obtener un resultado específico.

## Técnicas avanzadas de Prompt Engineering

Como hemos visto hasta aquí, en el mundo de la inteligencia artificial (IA), específicamente cuando trabajamos con modelos de lenguaje como ChatGPT, es fundamental saber cómo formular indicaciones o prompts para obtener las respuestas más útiles y precisas. Esto se conoce como Prompt Engineering. A medida que profundizamos en este campo, encontramos técnicas avanzadas que nos permiten optimizar la interacción con estos modelos. A continuación, explicaremos algunas de estas técnicas de manera clara y sencilla.

- **Role Prompting (Indicación de Roles)** es una técnica donde se le asigna un rol específico al modelo de IA para que responda desde esa perspectiva. Por ejemplo, si estamos buscando una opinión experta sobre un tema de salud, podríamos indicarle al modelo que responda como si fuera un médico. Esto ayuda a que las respuestas sean más relevantes y apropiadas para el contexto deseado.

Ejemplo: "Como experto en derecho laboral, ¿qué recomendarías en un caso de despido injustificado?"

- Zero Shot Prompting (Indicación Sin Ejemplos) se refiere a pedirle al modelo que realice una tarea sin darle ningún ejemplo previo. Es como hacerle una pregunta directa sin darle ninguna guía o contexto adicional. Este método es útil cuando queremos ver cómo el modelo interpreta y responde a una solicitud con la mínima información posible.

Ejemplo: "Genera un informe sobre el impacto ambiental de la nueva legislación."

- Few Shot Prompting (Indicación con Pocos Ejemplos). A diferencia de Zero Shot, Few Shot Prompting implica darle al modelo uno o más ejemplos para guiar su respuesta. Estos ejemplos ayudan al modelo a entender mejor lo que se espera de él. Es especialmente útil cuando necesitamos que el modelo siga un formato o estilo específico.

Ejemplo: "Aquí tienes dos ejemplos de cómo deben presentarse los informes de auditoría. Ahora, por favor, genera uno similar basado en los datos proporcionados."

- Chain of Thoughts (Cadena de Pensamientos). Se utiliza para guiar al modelo paso a paso a través de un proceso de razonamiento. En lugar de darle una única instrucción, le damos una serie de instrucciones que lo llevan a desglosar el problema en pasos más pequeños y manejables, ayudándolo a llegar a una conclusión más precisa.

Ejemplo: "Primero, identifica los factores clave en la toma de decisiones. Luego, analiza cómo cada uno de esos factores influye en la política pública. Finalmente, concluye cuál debería ser la mejor opción."

- Chain of Verifications (Cadena de Verificaciones). Implica guiar al modelo a través de un proceso de verificación. Aquí, el enfoque está en validar cada paso o dato antes de llegar a una conclusión final. Esta técnica es útil cuando es crucial garantizar la exactitud de la información, especialmente en temas que requieren precisión, como cálculos o análisis detallados.

Ejemplo: "Primero, verifica los datos del informe financiero. Luego, confirma que los cálculos sean correctos. Finalmente, revisa las conclusiones basadas en esos datos."

- Prompt chaining sequence (secuencia de encadenamiento de indicaciones) es una técnica avanzada en la cual se utiliza una serie de prompts encadenados para guiar al modelo de lenguaje hacia una respuesta deseada. Este método permite desglosar

tareas complejas en pasos más manejables, asegurando una mayor precisión y coherencia en las respuestas generadas.

Ejemplo: Imaginemos que un empleado público necesita generar un informe sobre los efectos de una nueva política ambiental en una comunidad local. Esta tarea es compleja porque implica analizar múltiples factores. En lugar de pedirle al modelo que genere todo el informe de una sola vez, se pueden usar prompts encadenados para desglosar el proceso.

## **Paso 1: Analizar los efectos económicos**

Prompt: "Describe los efectos económicos más significativos que podría tener la nueva política ambiental en la comunidad local."

Respuesta del modelo: "La nueva política ambiental podría afectar a las empresas locales al aumentar los costos de producción debido a la implementación de tecnologías más limpias. Esto podría resultar en una reducción temporal de empleos mientras las empresas se adaptan."

## **Paso 2: Evaluar los efectos sociales**

Prompt siguiente: "Basándote en los efectos económicos identificados, analiza cómo estos cambios podrían impactar la cohesión social y la calidad de vida en la comunidad."

Respuesta del modelo: "La reducción de empleos podría llevar a un aumento de la tensión social, especialmente en sectores dependientes de las industrias afectadas. Sin embargo, la política también podría mejorar la salud pública a largo plazo al reducir la contaminación."

## **Paso 3: Resumir las conclusiones**

Prompt final: "A partir de los efectos económicos y sociales que has descrito, redacta una conclusión general sobre los impactos de la nueva política ambiental en la comunidad."

Respuesta del modelo: "En conclusión, aunque la nueva política ambiental podría generar desafíos económicos a corto plazo, como la pérdida de empleos, estos podrían ser compensados por mejoras en la salud pública y la calidad de vida a largo plazo. Es esencial que las autoridades acompañen esta política con medidas de apoyo para las empresas y los trabajadores afectados."

Este enfoque encadenado asegura que cada parte del análisis sea considerada y permite al modelo construir gradualmente una respuesta más coherente y detallada.



Estas técnicas avanzadas de *Prompt Engineering* no solo mejoran la calidad de las respuestas que obtenemos de los modelos de IA, sino que también permiten adaptar esas respuestas a las necesidades específicas de cada tarea o contexto.

## Aplicaciones y modelos de IA Generativa

### Tipos de aplicaciones de IA Generativa

Cómo hemos visto, la IA generativa se destaca por su finalidad de generar contenidos. Al mismo tiempo, este tipo de tecnología cuenta con la posibilidad de generar contenidos en múltiples formatos, como por ejemplo textos, códigos de programación en diversos lenguajes, imágenes, audios (música, voces, etc.), videos, modelos 3D, etc.

Ahora bien, los principales modelos de IA generativa que vamos a explorar en esta formación son aquellos que generan textos e imágenes.

### Modelos generativos de lenguaje

Los modelos generativos de lenguaje tienen la capacidad de aprender a reconocer patrones dentro de los lenguajes y así poder predecir, con mayor o menos precisión, qué palabra continua en una serie de palabras. Como hemos visto, gracias a la tecnología llamada “Deep Learning”, los modelos de IA generativa pueden entrenarse con una cantidad enorme de datos no estructurados. Por ejemplo, el modelo conocido como GPT, es un modelo generativo de lenguaje que se entrenó con enormes cantidades de textos tomados de internet. Gracias a este entrenamiento, el modelo generativo de lenguaje puede contar con millones de ejemplos para las diversas situaciones que se puede enfrentar el modelo a la hora de generar texto.

Por dar un ejemplo básico, un modelo de generación de lenguaje puede reconocer que siempre que se inicia una conversación, las personas dicen “Hola, ¿cómo estás?” y que la respuesta esperada suele ser “Bien, ¿y tú?”. Aunque estos modelos pueden generar texto que parece lógico y fluido, no entienden el contenido de la misma manera que lo haría un ser humano. Sus respuestas se basan en patrones y probabilidades, no en comprensión real.

## Modelos generativos de imagen

Un modelo generativo de imagen es un algoritmo de IA que puede crear nuevas imágenes a partir de datos de entrada, que pueden ser descripciones textuales (prompts) o imágenes iniciales. Su objetivo es producir imágenes que sean coherentes con el input proporcionado.

Estos modelos se entrenan con grandes cantidades de imágenes y sus descripciones asociadas para aprender las características visuales y estilos. Por ejemplo, pueden aprender a asociar descripciones como "una vaca sonriente" con imágenes reales de vacas.

Una vez entrenado, el modelo puede generar imágenes nuevas basadas en descripciones textuales o modificar imágenes existentes. Por ejemplo, si se le da un prompt que dice "diseña un logo moderno con una vaca sonriente", el modelo generará una imagen de un logo que cumpla con esta descripción.

Algunos modelos pueden tomar una imagen existente y realizar modificaciones en ella, como cambiar el estilo, eliminar elementos o agregar nuevos detalles.

## De texto a otros formatos

### Modelos text-to-text (texto-a-texto)

Los modelos text-to-text son aquellos que, como insumo, reciben texto (input) y como resultado de salida (output) generan texto. Algunos usos útiles de este tipo de modelos son para generar resúmenes de textos, traducciones, adaptaciones de estilos, extracción de datos relevantes, reescritura, clasificación, entre otros.

### Modelos text-to-image (texto-a-imagen)

Los modelos text-to-image son entrenados con una gran cantidad de imágenes, cada una acompañada con una breve descripción. Estos modelos son útiles para generar o editar imágenes, por ejemplo, eliminar fondos y hacer recortes de personas u objetos. También se utilizan para crear bocetos de logos o manuales de identidad de una marca, así como para generar fondos de pantalla, personajes, etc.

### **Modelos text-to-video (texto-a-video)**

Los modelos text-to-video transforman descripciones textuales en secuencias de video. Estos modelos pueden crear videos que representan visualmente las descripciones proporcionadas, como escenas específicas o acciones descritas en el texto. Se utilizan para generar contenido multimedia para marketing, entretenimiento, o formación, permitiendo la creación de videos a partir de simples instrucciones textuales. Por ejemplo, un modelo podría generar un video de "una escena en la playa al atardecer con personas jugando al voleibol" a partir de una descripción textual detallada.

### **Modelos text-to-audio (texto-a-audio)**

Los modelos text-to-audio convierten texto en contenido de audio. Esto incluye la generación de voces sintetizadas para leer texto en voz alta (text-to-speech), o la creación de efectos de sonido y música basados en descripciones textuales. Estos modelos son útiles para crear audiolibros, asistentes virtuales, y contenidos multimedia accesibles. Por ejemplo, un modelo puede tomar un texto escrito y producir una narración en audio de ese texto, o generar una pieza musical basada en una descripción de estilo y tema.

## **De imagen a otros formatos**

Los modelos de inteligencia artificial (IA) han revolucionado nuestra interacción con diversos tipos de datos, permitiendo transformaciones sorprendentes entre diferentes formatos. Entre las capacidades más innovadoras se encuentra la habilidad de convertir imágenes en una variedad de otros formatos, lo cual abre un abanico de posibilidades creativas y prácticas. Estos modelos no solo pueden describir visualmente una imagen, sino que también pueden transformar la información contenida en ella en texto descriptivo, generar imágenes de alta resolución o incluso animaciones dinámicas. Además, pueden ofrecer descripciones auditivas para personas con discapacidades visuales, aplicar estilos artísticos únicos y cambiar fondos, así como integrar imágenes en entornos de realidad aumentada para experiencias interactivas.

### **Image to text:**

- Descripción de Imágenes: Los modelos pueden analizar una imagen y generar una descripción detallada en texto. Esto es útil para la accesibilidad, como en el caso de crear descripciones para personas con discapacidad visual.
- Respuestas visuales a las preguntas: Al proporcionar una imagen, estos modelos pueden responder preguntas relacionadas con el contenido visual, facilitando la búsqueda de información específica dentro de una imagen.

- **Búsqueda de imágenes:** Los modelos pueden identificar elementos dentro de una imagen y ayudar en la búsqueda de imágenes similares o relacionadas en bases de datos.

## Image to image

- **Gran resolución:** Los modelos pueden aumentar la resolución de una imagen original, mejorando su calidad para aplicaciones que requieren detalles finos, como en la impresión de gran formato o en análisis detallados.
- **Completar imágenes:** Estos modelos pueden rellenar áreas faltantes en una imagen, una técnica conocida como "inpainting". Esto es útil para restaurar fotos antiguas o reparar imágenes dañadas.

## Image to audio:

Los modelos de IA también pueden convertir imágenes en contenido de audio, una tecnología especialmente valiosa para mejorar la accesibilidad y proporcionar una experiencia más inclusiva. Estos modelos pueden generar descripciones verbales detalladas de imágenes, facilitando a personas con discapacidades visuales el acceso a la información visual. La conversión de imágenes a audio puede incluir la creación de descripciones narrativas, así como la integración de detalles contextuales y emocionales que ayudan a entender mejor el contenido visual. Esta tecnología no solo amplía la accesibilidad, sino que también permite innovaciones en cómo interactuamos con el contenido multimedia.

## Image to video:

Los modelos pueden generar animaciones a partir de imágenes estáticas. Esto incluye la creación de movimientos o efectos visuales, como en la generación de animaciones para presentaciones o contenido multimedia.

Estas capacidades amplían enormemente las formas en que podemos utilizar y manipular imágenes, ofreciendo nuevas oportunidades en campos como la accesibilidad, la creación de contenido y la restauración digital. Los avances en estos modelos permiten transformar el contenido visual de maneras que antes eran inimaginables, abriendo nuevas posibilidades para la creatividad y la eficiencia en el procesamiento de imágenes.

## Modelos de lenguaje de gran tamaño (LLM)

Los [large language models](#) (LLM, modelos de lenguaje de gran tamaño) son una clase de modelo fundacional que existen antes de los nuevos surgimientos de los desarrollos de la IA Generativa. Los modelos LLM, se diseñaron por ejemplo para para mejorar las capacidades de comprensión del lenguaje natural (NLU) y de procesamiento del lenguaje natural (NLP). En pocas palabras, los LLM están diseñados para comprender y generar texto como un humano, además de otras formas de contenido, basándose en la enorme cantidad de datos utilizados para entrenarlos. Tienen la capacidad de inferir del contexto, generar respuestas coherentes y pertinentes para el contexto, traducir a idiomas distintos del inglés, resumir texto, responder preguntas (conversación general y preguntas frecuentes) e incluso ayudar en tareas de escritura creativa o de generación de código.

Los modelos fundacionales de tipo LLM pueden usar y adaptar muy rápidamente (en cuestión de minutos) a una gran diversidad de tareas como chat, redacción, traducciones, resúmenes, brainstorming (lluvias de ideas), generación de código, correcciones, poemas, formateo de emails, etc. Todo esto, sin tener que ser una persona experta en machine learning, porque lo único que se necesita saber para adaptar el modelo LLM a lo que uno busca generar, es poder escribirle utilizando prompts.

Por ejemplo, los modelos de generative pre-trained transformer (GPT, transformadores generativos preentrenados) de OpenAI son modelos fundacionales del tipo LLM. Otros LLM usados son Gemini (desarrollado por Google) o Llama 2 (desarrollado por Meta).

Lo que hace que los LLM sean especiales es su capacidad para realizar múltiples tareas. Pueden hacerlo porque contienen muchos parámetros que los hacen capaces de aprender conceptos avanzados.

Un LLM como el GPT-3 puede considerar miles de millones de parámetros y tiene la capacidad de generar contenido a partir de muy pocas entradas. A través de su exposición previa al entrenamiento a datos a escala de Internet en todas sus diversas formas y en una miríada de patrones, los LLM aprenden a aplicar sus conocimientos en una amplia gama de contextos.

## Alucinaciones en la IA

**¿Podemos confiar en lo que leemos y vemos?**



Las alucinaciones son un fenómeno que pueden experimentar varios seres vivos y no solo los seres humanos. Son percepciones que ocurren en ausencia de un estímulo externo que lo desencadene.

Ahora bien, en el caso de alucinaciones en la IA, nos referimos a errores o invenciones que los sistemas producen al generar contenido que parece plausible pero es inexacto o completamente ficticio. Por ejemplo, si la generación es text-to-text pueden aparecer oraciones sin sentido, palabras inventadas o con errores gramaticales. Lo mismo puede suceder con la generación de imágenes, donde quizás el resultado esté muy alejado de lo que se desea obtener.

# Ejemplos de alucinaciones

## Errores fácticos

Los errores fácticos son una de las formas más comunes de alucinaciones de IA, donde un modelo genera texto que parece verdadero pero no lo es. La esencia básica de la afirmación puede estar basada en la realidad y sonar plausible, pero los detalles son incorrectos.

Ejemplo: En febrero de 2023, el chatbot de Google Bard (ahora llamado Gemini) afirmó incorrectamente que el telescopio espacial James Webb tomó la primera imagen de un planeta fuera del sistema solar. Esto es incorrecto ya que las primeras imágenes de un exoplaneta se tomaron en 2004, según la NASA, y el telescopio espacial James Webb no se lanzó hasta 2021.

## Información inventada

Los generadores de texto y chatbots de IA han sido conocidos por proporcionar información que está completamente fabricada y no se basa en ningún hecho. Por ejemplo, generar personas que no existen, inventar páginas web o incluso hacer referencia a artículos de noticias inventados, libros y trabajos de investigación, lo cual puede ser perjudicial para alguien que esté utilizando la herramienta para investigar.

Ejemplo: En junio de 2023, se informó que un abogado de Nueva York utilizó una herramienta de IA para redactar una moción que resultó estar llena de opiniones judiciales y citas legales falsas. El abogado, quien luego fue sancionado y multado, afirmó que 'no comprendía que ChatGPT pudiera inventar casos.'

## Información sin sentido o irrelevante

Por naturaleza, los modelos de IA tienen el objetivo de generalizar y ser creativos con sus resultados. Esta creatividad a veces puede llevar a generar salidas inusuales, lo cual no es necesariamente un problema si la precisión no es el objetivo.

Ejemplo: Si le preguntas a la IA "¿cuál es la capital de Francia?" podríamos llegar a obtener esta respuesta: La capital de Francia es París. París es también el nombre de una cantante famosa.

## Errores gramaticales

Esto puede incluir errores como frases mal estructuradas, uso incorrecto de tiempos verbales, o combinaciones de palabras que no siguen las reglas del idioma.

Ejemplo: Si le preguntamos a un chatbot cuantas letras erre hay en la palabra carrera, seguramente nos diga que la respuesta es 2.

## Alucinaciones visuales

Ocurren cuando el sistema produce imágenes que contienen elementos o detalles que no existen en el mundo real o que son incoherentes. Esto puede suceder debido a una interpretación errónea de los datos de entrada, una mezcla de características aprendidas de manera incorrecta, o la generación de patrones que no siguen las leyes físicas o visuales. Como resultado, las imágenes pueden tener distorsiones, objetos irreales o combinaciones inusuales que no corresponden a ninguna referencia conocida.





## ¿Por qué se generan las alucinaciones?

- El modelo no fue entrenado con suficientes datos.
- El modelo fue entrenado con datos de baja calidad.
- El modelo no ha recibido suficiente contexto a la hora de recibir nuestra solicitud por medio de un prompt.
- El modelo no ha recibido suficientes restricciones a la hora de recibir el prompt.

Aunque los modelos de lenguaje están diseñados para producir texto fluido y coherente, no tienen comprensión de la realidad subyacente que están describiendo. Todo lo que hacen es predecir cuál será la siguiente palabra basándose en probabilidades, no en precisión. Tal como es el caso de las imágenes, al no tener comprensión de lo que se está haciendo, la IA genera una imagen que se aproxime lo mejor posible a la instrucción que le dimos, tenga o no sentido para nosotros.

**Aclaración:** Vale aclarar que, en el caso de la generación de imágenes, a veces las alucinaciones son consideradas parte del proceso creativo, buscando expandir nuestra imaginación. En ese caso lo podríamos considerar como algo más bien positivo.

## Prevenir las alucinaciones

La mejor forma de prevenir las alucinaciones es evitar que sucedan. Acá les mostramos algunos pasos para que el modelo funcione de la manera más óptima:

- Verifica la información con fuentes confiables: Es importante contrastar la información que obtenemos con otras fuentes para verificar que el resultado no contenga errores. Incluso también podríamos usar múltiples sistemas de IA y comparar resultados.
- Definir para qué vamos a utilizar el modelo de IA: Al generar un prompt, es crucial proporcionar la mayor cantidad de detalles que podamos y tener en claro qué tareas queremos realizar para que el modelo pueda comprendernos y ofrecer el mejor resultado posible.
- Limitar las respuestas: No conviene sobrecargar de preguntas e instrucciones un chatbot. Es mejor ir preguntando de a poco para ir probando y refinando el sistema continuamente.
- Confiar en el ojo humano: Debemos siempre revisar el material que obtenemos. Esto incluso puede significar ser escépticos ante afirmaciones sorprendentes o demasiado específicas. Somos nosotros quienes decidimos qué material es útil y qué no.

*Tanto en texto como en imágenes, las alucinaciones reflejan la capacidad limitada de la IA para comprender el contexto completo y la realidad, basándose únicamente en patrones aprendidos a partir de datos anteriores. Recuerda que la IA no tiene "conocimiento" real, solo patrones de datos. Las máquinas nos pueden ayudar de maneras increíbles pero al final somos los seres humanos quienes tenemos la última palabra y realizamos la verificación final.*

## Mitos de la IA

### “Todos construimos la IA”

La creencia de que "todos construimos la Inteligencia artificial" es un mito porque no todos tenemos acceso a las mismas herramientas, datos y recursos computacionales necesarios para desarrollar modelos de inteligencia artificial generativa. Las grandes corporaciones tecnológicas dominan la creación de estos modelos, concentrando el poder y los recursos.

La creencia de que "todos construimos la Inteligencia artificial" es un mito porque entrenar modelos de inteligencia artificial generativa exige una infraestructura computacional robusta y costosa, que solo está al alcance de unos pocos.

El mito de que "Todos construimos la IA" sugiere una participación universal en su desarrollo. Sin embargo, la realidad es más matizada. Es cierto que los proyectos de código abierto permiten a una amplia comunidad de desarrolladores, investigadores y entusiastas contribuir al avance de la IA generativa. Esta colaboración democratiza el acceso a la tecnología y fomenta la innovación al permitir que individuos de diversas partes del mundo y con distintos niveles de experiencia aporten su conocimiento y habilidades. Sin embargo, afirmar que todos construimos la IA puede ser engañoso. La participación en el desarrollo de IA, especialmente en niveles avanzados, requiere un conjunto específico de habilidades técnicas, acceso a recursos computacionales significativos y, a menudo, una comprensión profunda de conceptos matemáticos y algorítmicos complejos. Estas condiciones no están distribuidas equitativamente a nivel global, lo que resulta en una participación desigual en la construcción de estas tecnologías. Además, la contribución efectiva al desarrollo de la IAG a menudo está influenciada por la disponibilidad de datos, que puede estar limitada o sesgada según el contexto. Por lo tanto, aunque la colaboración abierta expande el alcance de quién

puede contribuir al desarrollo de la IAG, sigue habiendo barreras significativas que limitan la participación universal efectiva.