



Técnicas de prompting para IA Generativa: Prompting multimodal



01 Fundamentos del prompting multimodal

Qué es el prompting multimodal

El prompting multimodal consiste en dar instrucciones a modelos de IA capaces de procesar simultáneamente texto, imágenes, audio o vídeo para generar contenido más rico y preciso.

Modelos que combinan varios tipos de datos

En lugar de trabajar solo con texto, los modelos multimodales reciben instrucciones en distintos formatos: una descripción escrita, una imagen de referencia, quizás audio o fragmentos de vídeo. El modelo procesa estos datos de forma conjunta y genera resultados consistentes con todo ese contexto.

Modelos conversacionales como **ChatGPT** permiten subir una imagen, describir lo que se desea conseguir y obtener una respuesta visual ajustada a esa mezcla de información.

De un texto simple a una escena completa

Con prompting multimodal, en lugar de pedir solo “una foto de producto”, se puede añadir una **fotografía real del producto** como referencia y combinarla con una descripción detallada de la escena.

Por ejemplo, es posible solicitar un **shoot de producto** con fondo pastel o una **escena tipo influencer** usando el producto en su día a día. La IA analiza tanto el texto como la imagen y replica el producto con un grado de precisión muy alto.

- **Referencia visual:** fotografía real del producto u objeto principal.
- **Instrucciones textuales:** qué ambiente, estilo, iluminación o composición se desea.

- **Contexto narrativo:** cómo se usa el producto y qué se quiere comunicar con la escena.

Prompt solo texto frente a prompt multimodal

Ambos enfoques son útiles, pero el prompting multimodal aporta un control más fino sobre el producto y la escena generada.

Aspecto	Prompt solo texto	Prompt multimodal
Tipo de entrada	Descripción escrita de lo que se quiere ver.	Texto más imagen de referencia (producto, escenario, estilo), y potencialmente audio o vídeo.
Representación del producto	Interpretación aproximada según la descripción.	Representación casi exacta del producto real usado como referencia.
Control creativo de la escena	Limitado a lo que el modelo "imagina" a partir del texto.	Mayor control sobre composición, estilo, contexto y coherencia con el material aportado.
Precisión para marketing y publicidad	Puede ser suficiente para bocetos muy conceptuales.	Alta precisión, adecuada para piezas más cercanas a campañas reales.
Complejidad inicial	Muy sencillo de formular, solo requiere texto.	Requiere preparar imágenes y pensar mejor qué material se aporta.



La IA no realiza todo el trabajo creativo: potencia ideas claras y materiales bien preparados, pero la dirección conceptual sigue siendo humana.

Usos clave del prompting multimodal en contenidos visuales

El prompting multimodal abre nuevas formas de trabajar imagen y vídeo en áreas como marketing, publicidad, diseño y animación.

Fotografía de producto simulada

Permite generar imágenes de producto en diferentes fondos y ambientes (por ejemplo, fondos pastel) manteniendo el aspecto real del artículo.

Escenas tipo influencer

Facilita crear imágenes de personas usando el producto en situaciones cotidianas, mostrando su uso real sin organizar sesiones de fotos.

Apoyo a campañas de marketing

Ayuda a explorar múltiples enfoques visuales de un mismo mensaje antes de decidir qué concepto se desarrolla en profundidad.

Conceptos para diseño y animación

Genera escenas de referencia que sirven como base para diseño gráfico, motion graphics o animación, acelerando la fase de ideación.



02 Diseñar prompts con intención creativa

El papel de la intención en la generación visual

Aunque la IA es muy potente, el valor real surge cuando existe una idea clara y una intención creativa bien definida detrás de cada prompt.

Claridad de objetivo antes de generar imágenes

Antes de formular un prompt, resulta clave tener claro **qué se quiere conseguir**: mostrar el uso diario del producto, resaltar un beneficio concreto o simplemente presentar el objeto de forma atractiva.

Sin un objetivo definido, la IA solo produce imágenes “bonitas”, pero no necesariamente alineadas con las necesidades de marketing, diseño o comunicación.

Pensar como director creativo o de arte

El uso eficaz de la IA se parece al trabajo de un **director creativo, director de arte o diseñador de producto**. No se trata solo de estética, sino de controlar la escena para que tenga sentido con el producto y el mensaje.

- Decidir qué elementos deben aparecer y cuáles se deben evitar.
- Definir qué emoción, historia o contexto transmite la escena.
- Garantizar que el producto se vea de forma clara y relevante.

Pasos para preparar un prompt multimodal efectivo

Un buen resultado visual no depende solo del modelo de IA, sino de un proceso ordenado de preparación del prompt.

- 1 Definir el resultado deseado:** concretar qué se quiere mostrar (por ejemplo, el producto en uso diario, un bodegón minimalista o una escena aspiracional).
- 2 Seleccionar la referencia adecuada:** elegir una imagen del producto o material visual relevante que el modelo pueda analizar y replicar con precisión.
- 3 Describir la escena con detalle:** indicar fondo, estilo, tipo de entorno, iluminación y cualquier elemento adicional (por ejemplo, una persona tipo influencer interactuando con el producto).
- 4 Redactar el prompt en lenguaje natural:** formular la petición como si fuera una conversación profesional pero cercana, sin necesidad de tecnicismos complejos, siempre manteniendo la claridad.
- 5 Revisar, valorar e iterar:** analizar las imágenes generadas, detectar qué funciona y qué no, y ajustar el prompt para acercarse cada vez más al objetivo creativo.



El lenguaje del prompt puede ser sencillo y cotidiano; lo esencial es la precisión en la intención y en la descripción de la escena.

Equilibrio entre automatización y criterio humano

La colaboración entre IA generativa y criterio profesional redefine los procesos visuales, pero no elimina la necesidad de decisiones humanas informadas.

Fortalezas del modelo de IA

El modelo analiza **texto e imagen a la vez**, identifica el producto con gran exactitud y genera rápidamente múltiples variaciones visuales. Esto reduce tiempos de exploración y permite testar muchas ideas con poco esfuerzo operativo.

También ofrece consistencia: una vez entendida la referencia, es capaz de mantener el mismo producto en distintas escenas, fondos o estilos.

Responsabilidades creativas que siguen siendo humanas

Pese a su potencia, la IA no define la estrategia ni el concepto. La responsabilidad humana sigue siendo clave en aspectos como:

- **Definir el propósito** de cada imagen o vídeo dentro de una campaña o proyecto.
- **Elegir las ideas** que conectan mejor con la audiencia y con la propuesta de valor del producto.
- **Asegurar la coherencia** con la identidad visual y el posicionamiento de la marca.

El mayor impacto del prompting multimodal surge cuando la IA se utiliza como herramienta de amplificación de un trabajo previo de análisis, enfoque y diseño conceptual bien realizado.



Creado por Victoria, AI Founderz Fellow, y aprobado por el equipo de Founderz.



Última actualización 5 de diciembre de 2025



Este documento fue originalmente generado por la IA y revisado por nuestro equipo humano. En Founderz, utilizamos la IA de forma responsable y transparente.