



— HARAMBEE
BIBLIOTECA



GUÍA DE PROMPTING

Inteligencia Artificial
Generativa

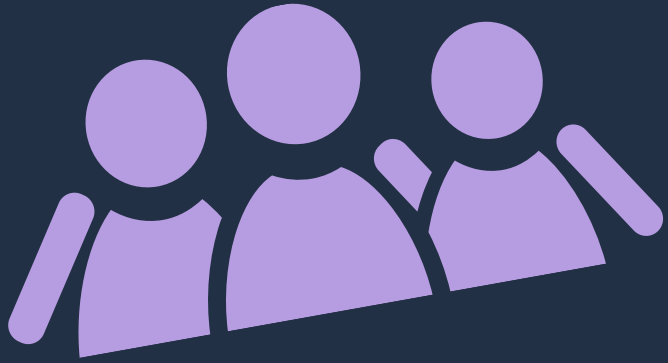


MANTENGAMOS EL PAÑO, APOYÁNDONOS ENTRE TODOS

- *Esta guía fue desarrollada por la comunidad Harambee.*
- *Está dirigida a usuarios de IA que buscan profundizar y mejorar.*
- *Se enfoca en soluciones de texto.*

COMUNIDAD DE PRÁCTICA EN IA

harambee-ia.cloud



GUÍA DESARROLLADA POR EL EQUIPO “DESAFÍO BIBLIOTECA”

Redacción:

- *Cristián Cortés*
- *Marcelo Taito*

Edición

- *Zinnia Leal*

COMUNIDAD DE PRÁCTICA EN IA

harambee-ia.cloud



¿POR QUÉ DEBERÍA INTERESARNOS?

Goldman Sachs (2023) proyectó que el uso de IA puede aumentar la productividad un 7 a 10% anual en los próximos 10 años.

De acuerdo a MIT, la productividad personal puede crecer hasta en un 40% (hacer casi lo mismo en la mitad del tiempo).

McKinsey estima que los sectores con mayor potencial son educación, consultoría, marketing y desarrollo de soluciones.

COMUNIDAD DE PRÁCTICA EN IA

harambee-ia.cloud



¿QUÉ ES PROMPTING ?

El **Prompt**... ¡el centro de todo! Para el uso de la IA Generativa, este concepto es crucial. Es el punto de inicio que **define la interacción entre humanos y máquinas**. El Prompt es la entrada de texto o instrucción que se le proporciona al modelo para generar una respuesta o realizar una tarea.

En esencia, **es el estímulo o la pregunta que se plantea al modelo**. Estos mensajes, conocidos como prompts, deben ser cuidadosamente elaborados para ser claros, concisos y efectivos.

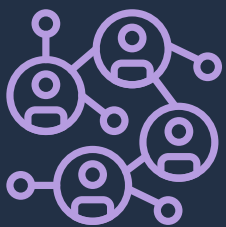
El objetivo principal del prompting es **lograr una interacción fluida y natural** entre el usuario y la máquina, de manera que el usuario pueda realizar sus tareas de manera eficiente y sin frustraciones.

HABLEMOS DE IA GENERATIVA

Los grandes modelos de lenguaje (LLM) están entrenados con billones de parámetros, por lo que logran generar conversaciones.

Los modelos más populares de IA generativa, como ChatGPT, Claude y Gemini [Bard], se basan en una estructura que se denomina “Transformer” y que, en muy simple, va calculando la probabilidad de la siguiente palabra en la conversación. De esta forma, se enfocan en conceptos centrales y luego van rellenando los espacios en la conversación. Igual como lo hacemos los humanos.

Algunos elementos a considerar



**Relevancia
contextual**



Longitud



**Formulación
clara y
específica**



**Pistas
adicionales**

FALLOS, “ALUCINACIONES” Y SESGOS

Independiente de su tremenda capacidad, hay tres potenciales problemas en el uso de la IA que se deben tener presentes.

1. La **IA** Generativa tiende a equivocarse cuando le solicitamos cálculos matemáticos o datos muy específicos
2. Dado que funciona buscando parámetros, puede encontrar patrones inexistentes y entregar respuestas incorrectas
3. La **IA** está cargada con enormes volúmenes de datos, que podrían tener sesgos.

No tengas “miedo” de corregirla cuando se equivoque. Siempre con respeto





VAMOS AL PROMPT

Imagina que vas a una cafetería y pides: “Algo de comer y de tomar”. Lo más seguro es que te llegue cualquier cosa y que sea muy distinto a lo que esperabas.

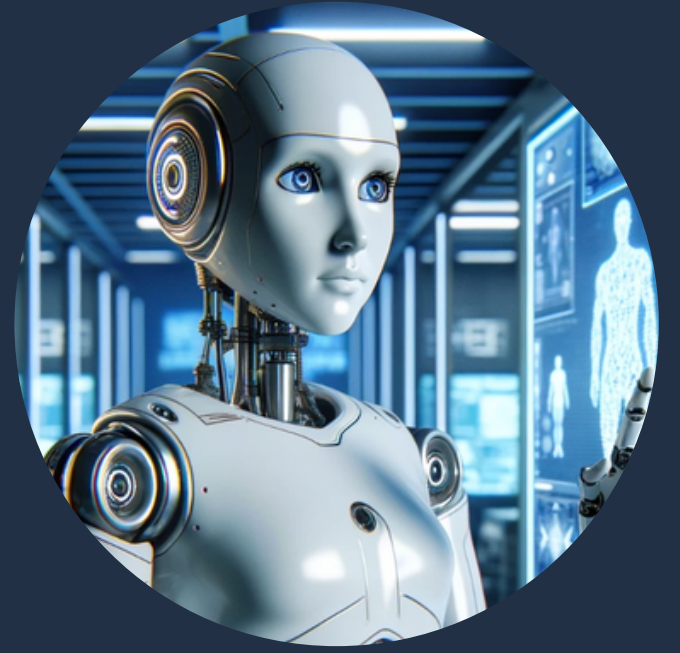
Para trabajar con IA es lo mismo. Las instrucciones deben ser claras y específicas.



ESTRATEGIAS DE PROMPT

*Usamos como base la información
contenida en “platform.openai.com”, que
también sirven para otras soluciones
como Claude.*

COMUNIDAD DE PRÁCTICA EN IA
harambee-ia.cloud



6 ESTRATEGIAS

1. Escribe instrucciones claras
2. Proporciona texto de referencia
3. Divide las tareas complejas en subtarefas más simples
4. Dale tiempo al modelo para pensar
5. Utiliza herramientas externas
6. Haz pruebas sistemáticas

*En las siguientes páginas,
revisemos punto por punto*

1. ESCRIBE INSTRUCCIONES CLARAS

Al igual que en el ejemplo de la cafetería, estos modelos no pueden leer tu mente. Debes solicitar claramente y con detalles lo que necesitas. Aquí algunas técnicas:

- Dale el **contexto** de tu solicitud: “Vamos a trabajar en un texto de marketing”.
- Píde al modelo que adopte una **personalidad**. “Serás un experto en Design Thinking”.
- Usa **delimitadores** para indicar claramente distintas partes de la entrada, por ejemplo ,cuando estás agregando texto que quieres que use como contexto (el más usado es la triple comilla “”).
- Especifica los **pasos** necesarios para completar una tarea. “Paso 1: te proporcionaré un texto entre “”; Paso 2: Traduce ese texto a inglés.
- Proporciona **ejemplos**.
- Especifica la **longitud** deseada de la salida: “en 50 palabras”, “Resume el texto en dos párrafos”.

1.1. EJEMPLOS: INSTRUCCIONES CLARAS

Veamos un ejemplo:

Trabajaremos escribiendo post para twitter. Actúa como un experto en textos de marketing y escribe tres post respecto a lo saludable y beneficioso para el medio ambiente de ir al trabajo en bicicleta. Agrega un emoji y dos # en cada post.

En este caso, entregamos contexto e indicamos el formato de salida (texto tipo twitter o X)

Cuando tengas tareas repetitivas, prueba con instrucciones que “programen” al chat para que actúe como un asistente.

**1er
Prompt**

Trabajaremos escribiendo post para twitter. Actúa como un experto en textos de marketing. Cada vez que te entregue un tema escribe tres post. Agrega un emoji y dos # en cada post.

**2do
Prompt**

ir en bicicleta al trabajo es saludable y beneficioso para el medio ambiente.

1.2. ALGUNAS TÉCNICAS ADICIONALES

- **Zero-Shot:** Pregunta sin ejemplos (sin entrenamiento previo) para obtener respuestas “novedosas”
- **Few Shot:** Entrega algunos ejemplos de lo que quieres como respuesta, para guiar la conversación
- **¿Entendiste?:** Cuando la tarea o instrucción (prompt) sea más compleja, asegúrate de que el modelo lo comprendió. Pregúntale si entendió y si necesita más información.



1.3 RGC PROMPTING

El RGC Prompting (Relevance, Goal, Context) [Relevancia, Objetivo, Contexto] presenta una estructura sólida que se puede aplicar universalmente a cualquier entrada, que lleva a un resultado esperado.

Su objetivo es establecer un formato estandarizado que optimice un marco de uso para ChatGPT y así mejorar su rendimiento en un amplio espectro de entradas.

Fórmula Express

Rol	Personaje de ChatGPT (Eres un experto en...)
Resultado	Resultado deseado (Cree x mensajes que...)
Objetivo	Propósito del resultado (El objetivo es crear en mis seguidores interés en...)
Contexto	Quién, qué, dónde y por qué (Los mensajes creados son para...)
Restricción	Limitaciones y pautas (Los mensajes deben ser amigables y tener menos de 60 palabras)

1.3 RGC PROMPTING

DIAGRAMA GENERAL



Voy a planificar un viaje a Roma!



1.3 RGC PROMPTING



Ejemplo de uso: **Planificar un viaje a Roma**

	Descripción
Rol	Usuario que está planeando un viaje.
Resultado	Obtener un itinerario detallado y sugerencias para el viaje.
Objetivo	Solicitar al modelo que genere un plan de viaje, incluyendo actividades, alojamiento y transporte.
Contexto	Especificar la duración del viaje, los intereses del viajero y cualquier restricción presupuestaria.
Restricción	Limitar el presupuesto total del viaje y considerar las preferencias alimenticias o restricciones dietéticas del viajero.

PROMPT FINAL:

Rol: Eres un usuario que está planeando un viaje. **Resultado Esperado:** Quiero obtener un itinerario detallado y sugerencias para el viaje a Roma. **Objetivo:** Genera un plan de viaje que incluya actividades, alojamiento y transporte. **Contexto:** El viaje tendrá una duración de 5 días. Mis intereses incluyen visitar lugares históricos y museos en Roma. El presupuesto total es de 2500 euros. Toma en cuenta las preferencias alimenticias o restricciones dietéticas, como la intolerancia a la lactosa y al gluten. **Restricciones:** Asegúrate de respetar el presupuesto total y considerar las preferencias alimenticias o restricciones dietéticas especificadas.

2. TEXTO DE REFERENCIA

Estos **modelos** son tan poderosos que, si no les das una guía, pueden terminar respondiendo lo que el modelo quiere. Es importante, sobre todo cuando trabajas con información muy específica.

Puedes proporcionar al modelo un **texto de referencia** entre “”” (triple comillas) y pedir que responda usando la información contenida en ese texto. También puedes proporcionar un documento para que responda teniéndolo de base.

O puedes usar **técnicas más avanzadas**. En el caso de OpenAI, “Fine Tuning” o “Embeddings”

Veamos estos 3 puntos en detalle

2.1. CAPACIDAD DE LOS MODELOS

Antes de seguir revisando la forma de entregar textos, hablemos de los dos principales “Competidores”, sus modelos y capacidades.

OpenAI - GPT. Los dos principales modelos de GPT de OpenAI son 3,5 y 4.

- GPT 4. La versión turbo soporta hasta 128 mil tokens (aprox 200 páginas de texto) y está actualizado hasta abril/2023. Se puede acceder desde el Chat Plus, Playground o la API.
- GPT 3.5. La versión turbo soporta hasta 16 mil tokens (aprox 25 páginas de texto) y está actualizado hasta sept/2021. Se puede acceder desde el Chat, Playground o la API.
- OpenAi tiene varios modelos interesantes de explorar para Imagen, Texto a Audio, Audio a Texto
- OpenAi no guarda tus “chats” si usas sus modelos en el Playground o la API

Anthropic - Claude.

- Las versión 2.1 (Pro-pagada y Normal) soporta hasta 200 mil tokens (aprox 350 páginas de texto) y está actualizada hasta dic/2022. Se puede acceder desde el Chat
- Anthropic no guarda tu información si usas la versión Pro.

2.2. ¿CÓMO ENTREGAR UN TEXTO DE REFERENCIA?

Puedes entregar un texto entre “” para darle un contexto desde el cual responder. Ahora que sabes las capacidades de los modelos, solo debes cuidar que sea breve.

Si tu texto es muy extenso, puedes subir documentos completos. En el caso de ChatGPT (chat.openai.com) solo es posible en la versión pagada. En Claude está disponible en la versión gratuita (claude.ai).

Algo **importante para considerar**. Si subes un documento de **texto** (Word, TXT, Pdf), el sistema lo leerá y te responderá respecto al documento, pero usando sus propias bases de conocimiento. Si subes un **documento estructurado** (Excel, CSV), analizará y responderá usando solo la información ahí contenida.

Avanzado: Puedes mejorar la precisión de la respuesta, subiendo un documento de texto estructurado (.JSON).

2.2. UN EJEMPLO DE TRIPLE COMILLAS “”””

Usar triple comillas puede ser útil para entregar un texto de referencia a modo de contexto y también para estructurar conversaciones, aumentando la claridad de lo solicitado en tus interacciones con, por ejemplo, ChatGPT.

Ten en cuenta que la calidad de las respuestas dependerá de la formulación del prompt y de las instrucciones que proporcionas.

Prueba el siguiente Prompt para **guiar la conversación:**
(*incluye las **triple comillas** al copiar el Prompt*)

*Haz la
prueba en
ChatGPT*

```
""" Usuario: Cuéntame sobre inteligencia artificial.  
Asistente: La inteligencia artificial (IA) es...  
Usuario: ¿Cuáles son algunos ejemplos de  
aplicaciones prácticas de la IA en la vida diaria?  
Asistente: Hay diversas aplicaciones de la IA en la  
vida diaria, tales como... """
```

¡Experimenta y ajusta según tus necesidades!

2.3. “BIBLIOGRAFÍA”

Ahora que sabemos las fechas de actualización y las capacidades de los modelos, también puedes solicitar que use como referencia “bibliografía”.

Prueba lo siguiente: Imagina que quieres trabajar en tu modelo de negocios. Sabemos que el libro **Business Model Generation** fue publicado antes de 2021, por lo que la IA debería conocerlo.

1er
Prompt

¿Conoces el libro “Business Model Generation” de A. Osterwalder y otros?

Resp.

Sí, conozco el libro Business Model Generation

Vamos a asegurarnos de que lo conoce bien

2do
Prompt

De acuerdo al libro, cuál es la definición del concepto de propuesta de valor?

Resp.

La propuesta de valor....

Una vez nos hayamos asegurado de que conoce el libro y que las preguntas las responde correctamente, podemos usar este libro como contexto.

2.4. TÉCNICAS AVANZADAS

Existen dos técnicas avanzadas para trabajar con los modelos que, aunque necesitan más conocimiento, logran resultados espectaculares. Estas técnicas están disponibles para OpenAI, en el caso de Claude, aún están en modo beta y cerrado.

La primera es “Fine Tuning”. Se puede usar en platform.openai.com/fine-tuning. Permite resultados de mayor calidad al personalizar o entrenar tus propios modelos a través de ejemplos. De esta forma, tu modelo entrenado, responderá de forma más apegada y consistente con lo que esperas.

La segunda técnica avanzada es “Embeddings”. Las Incrustaciones (platform.openai.com/docs/guides/embeddings) se usan para “guiar” al modelo, al convertir las cadenas de texto en números (vectores). Esta técnica generalmente se usa para agrupar información, para sistemas de recomendaciones, medición de diversidad y clasificación de texto, entre otras.

3. DIVIDIR TAREAS COMPLEJAS

Ayuda al modelo a enfocarse. Al igual como una persona se confunde cuando se le entregan tareas muy complejas y con diversas instrucciones o pierde el “hilo” en conversaciones muy largas, lo mismo le ocurre a los modelos de IA Generativa. Recordemos, además, que los modelos tienen limitaciones.

Veamos algunas técnicas:

- **Cadena de Pensamiento:** Desglosa explícitamente el proceso de razonamiento o los pasos a seguir, para que tu conversación se mantenga enfocada.
- **Resumen del chat anterior:** Cuando ya se ha alcanzado una longitud de una entrada o chat, muy amplia, podemos solicitar un resumen con los elementos principales y con eso iniciar un nuevo chat.
- **Avanzado:** *Podemos transformarlo en Embeddings, para mejorar los resultados.*
- **Resumir documentos largos:** Utiliza secuencias de consultas, respecto a partes de información, por ejemplo, cuando estamos trabajando con un documento. Las respuestas serán de mejor calidad que si solicitamos un “resumen detallado”.

3.1. EJEMPLO DE CADENA DE PENSAMIENTO

Prompt

Te comportarás como un experto en modelamiento de negocios. Revisaremos mi modelo de negocios. Te entregaré un documento y me ayudarás a revisar las distintas partes usando como referencia el libro “business model generation”....

En este caso, le estamos solicitando una tarea muy compleja en un solo paso. El modelo la logrará hacer sin problema, pero la información que nos entregue va a ser superficial.

Prompt

Te comportarás como un experto en modelamiento de negocios. Revisaremos mi modelo de negocios. Te entregaré un documento y me ayudarás a revisar parte por parte usando como referencia el libro “business model generation”. Comenzaremos con el perfil del cliente. Una vez que te responda que ya estoy conforme con la revisión, seguiremos con la propuesta de valor....

Solicitamos al modelo que vayamos revisando por partes, de manera que se enfoque en una tarea específica, mejorando los resultados

4. DAR TIEMPO AL MODELO PARA “PENSAR”

Si nos preguntan por el resultado de 135×1428 , lo más probable es que lo podamos resolver sin problema, solo necesitamos algo de tiempo para hacer el cálculo. Lo mismo pasa con los modelos de IA.

Algunas técnicas útiles:

- **Busca tu propia solución:** Siguiendo el ejemplo de la revisión del modelo de negocios. Podemos solicitar al modelo que, a partir de la información que entregamos, elabore una definición propia de “oferta de valor” y que luego la compare con la nuestra, de forma de tener una revisión “crítica” y asegurarnos, por ejemplo, que nuestra oferta está planteada correctamente.
- **Debate multi-agente:** Si le diste una personalidad, ¿por qué no darle algunas otras?, así lograrás que antes de contestar, piense y discuta la respuesta desde varias perspectivas... y tendrás a todo un equipo trabajando para ti.

5. HERRAMIENTAS EXTERNAS

Estas son prácticas avanzadas, pero que hacen llevar el uso de la IA generativa a otro nivel. Los modelos de IA Generativa, sobre todo el de OpenAi (dado que es el más “abierto” para uso por ahora), permiten infinidad de desarrollo de soluciones.

Por ejemplo:

- **Entrena tus propios modelos:** Abordamos este tema en el apartado 2.4. Puedes entrenar tus propios modelos entregando ejemplos (Fine Tuning) o guiándolo con cadenas de texto (Embeddings).
- **Conectar el modelo a tus procesos:** Puedes conectar el modelo a tus propios procesos, por ejemplo, para la revisión de una base de datos, contestar correos de forma automatizada o desarrollar un “chat bot” inteligente para Whatsapp. Las opciones de modelos de IA conectados con No-Code es infinita.
- **Conoce y utiliza otras soluciones:** Open AI tiene soluciones para transformar audio a texto, texto a audio, texto a imágenes, etc.

Las opciones y funciones, incluso sin programar, son infinitas.

6. PRUEBAS SISTEMÁTICAS

Importantes, sobre todo, cuando se está trabajando con asistentes virtuales o con prompts muy complejos y queremos evaluar si un cambio en la instrucción mejora el resultado.

Los “experimentos” deben considerar que usar solo algunos ejemplos para evaluar, no ayudarían tanto a distinguir entre mejoras reales y “suerte” aleatoria.

Un buen **diseño de evaluación** es:

1. Representativo y diverso.
2. Contiene muchos casos de prueba (10 es lo mínimo, de 30 hacia arriba es lo óptimo).
3. Fácil de sistematizar y repetir.

Una **táctica** utilizada es evaluar las respuesta del modelo, respecto a un resultado que nosotros conocemos y tenemos estandarizado.

ELEMENTOS FINALES

Diferencias de ChatGPT y Claude

Son muy pocas las diferencias y, en general, esta guía funciona en ambos modelos.

Claude reconoce etiquetas (apertura `<>` y cierre `</>`) que le permiten enfocarse en distintos temas dentro de un prompt y te permite “hacer llamadas” de temas específicos. Por ejemplo:

```
Utiliza este texto como <contexto> textoContexto  
</contexto>
```

```
Sigue estas reglas para responder a las preguntas de  
usuarios <reglas> Reglas </reglas>
```

Más sobre Ingeniería de prompts

- <https://platform.openai.com/docs/guides/prompt-engineering>
- <https://cookbook.openai.com/>
- <https://docs.anthropic.com/claude/docs/guide-to-anthropics-prompt-engineering-resources>



GRACIAS POR LLEGAR HASTA AQUÍ

*Únete a la comunidad y
colaboremos para que más personas
aumenten su productividad
utilizando herramientas de IA
Generativa*

¿TIENES COMENTARIOS PARA MEJORAR ESTA GUÍA?
contacto@harambee-ia.cloud