

Conceptos de ingeniería rápidos

La *ingeniería rápida* se refiere a la práctica de optimizar la entrada textual en un modelo de lenguaje grande (LLM) para obtener las respuestas deseadas. Las solicitudes ayudan a un LLM a realizar una amplia variedad de tareas, como la clasificación, la respuesta a preguntas, la generación de códigos, la escritura creativa y más. La calidad de las indicaciones que dé a un LLM puede afectar a la calidad de las respuestas de la modelo. Esta sección le proporciona la información necesaria para empezar con la ingeniería rápida. También incluye herramientas que le ayudarán a encontrar el mejor formato de solicitud posible para su caso de uso cuando utilice un LLM en Amazon Bedrock.

Nota

Todos los ejemplos de este documento se obtienen mediante llamadas a la API. La respuesta puede variar debido a la naturaleza estocástica del proceso de generación de LLM. Si no se especifica lo contrario, las instrucciones las escriben los empleados de AWS.

Nota

Para reducir las alucinaciones, puede refinar su mensaje mediante técnicas de optimización rápida; utilizar técnicas como la generación aumentada de recuperación (RAG) para que el modelo acceda a datos más relevantes; o utilizar un modelo diferente que pueda producir mejores resultados.

Amazon Bedrock incluye modelos de varios proveedores. La siguiente es una lista de las directrices de ingeniería rápidas para esos modelos.

- **Guía rápida de Amazon Nova Micro, Lite y Pro: consejos sobre [las mejores prácticas para que Amazon Nova comprenda los modelos](#)**
- **Guía rápida de Amazon Nova Canvas: [generación de imágenes con Amazon Nova](#)**
- **Guía rápida de Amazon Nova Reel: [generación de vídeos con Amazon Nova](#)**
- **[Guía rápida del modelo Claude antrópico: https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering/overview](https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering/overview)**
- **[Guía rápida de Cohere: https://txt.cohere.com/how-to-train-your-pet-llm-prompt - ingeniería](https://txt.cohere.com/how-to-train-your-pet-llm-prompt-ingenieria)**
- **[Guía rápida del modelo jurásico de AI21 Labs: https://docs.ai21.com/docs/prompt - ingeniería](https://docs.ai21.com/docs/prompt-ingenieria)**

- **Guía rápida de Meta Llama 2:** <https://ai.meta.com/llama/get-started/#prompting>
- **Guía de instrucciones de la IA de estabilidad:** <https://platform.stability.ai/docs/getting-started>
- **Guía de instrucciones de Mistral AI:** https://docs.mistral.ai/guides/prompting_capabilities/

Descargo de responsabilidad: Los ejemplos de este documento utilizan los modelos de texto actuales disponibles en Amazon Bedrock. Además, este documento contiene directrices generales de asesoramiento. Para ver las guías específicas de cada modelo, consulte los documentos correspondientes en Amazon Bedrock. Este documento proporciona un punto de partida. Si bien los siguientes ejemplos de respuestas se generan con modelos específicos en Amazon Bedrock, también puede utilizar otros modelos de Amazon Bedrock para obtener resultados. Los resultados pueden diferir de un modelo a otro, ya que cada modelo tiene sus propias características de rendimiento. El resultado que genera con los servicios de IA es su contenido. Debido a la naturaleza del aprendizaje automático, es posible que los resultados no sean únicos en todos los clientes y que los servicios generen resultados iguales o similares en todos los clientes.

¿Qué es un aviso?

Las indicaciones son un conjunto específico de entradas que proporciona usted, el usuario, que guían a los LLM de Amazon Bedrock a generar una respuesta o un resultado adecuados para una tarea o instrucción determinada.

Mensaje de usuario:

¿Quién inventó el avión?

Cuando se consulta mediante este indicador, Titan proporciona un resultado:

Salida:

Los hermanos Wright, Orville y Wilbur Wright, están ampliamente acreditados con la invención y la fabricación del primer avión exitoso del mundo.

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text)

Componentes de un mensaje

Un solo mensaje incluye varios componentes, como la tarea o la instrucción que quiere que realicen los LLM, el contexto de la tarea (por ejemplo, una descripción del dominio

correspondiente), ejemplos de demostración y el texto de entrada que quiere que los LLM de Amazon Bedrock utilicen en su respuesta. Según el caso de uso, la disponibilidad de los datos y la tarea, su solicitud debe combinar uno o más de estos componentes.

Considere este ejemplo de solicitud en el que se pide a Titan que resuma una reseña:

Mensaje de usuario:

El siguiente es el texto de una reseña sobre un restaurante:

«Por fin tengo que echar un vistazo a Alessandro's Brilliant Pizza y ahora es uno de mis restaurantes favoritos en Seattle. El comedor tiene una hermosa vista del Puget Sound, pero sorprendentemente no había mucha gente. Pedí las aceitunas castelvetrano fritas, una pizza picante al estilo napolitano y un plato de ñoquis. Las aceitunas estaban absolutamente decadentes y llegó la pizza con mozzarella ahumada, que estaba deliciosa. Los ñoquis estaban frescos y maravillosos. Los camareros estuvieron atentos y, en general, la experiencia fue estupenda. Espero volver pronto».

Resume la reseña anterior sobre el restaurante en una frase

(Fuente del mensaje: AWS)

Basándose en esta pregunta, Titan responde con un breve resumen de una línea de la opinión sobre el restaurante. La reseña menciona los hechos clave y transmite los puntos principales, según desea

Salida:

Alessandro's Brilliant Pizza es un fantástico restaurante en Seattle con una hermosa vista de Puget Sound, comida deliciosa y decadente y un servicio excelente.

(Modelo utilizado: Amazon Titan Text)

La instrucción **Resume la reseña anterior sobre el restaurante en una frase** y el texto de la reseña **por fin tengo que ver...** ambos eran necesarios para este tipo de salida. Sin ninguno de los dos, el modelo no tendría suficiente información para producir un resumen sensato. La *instrucción* indica a la LLM lo que tiene que hacer y el texto es la *entrada* en la que opera la LLM.

El *contexto* (**el siguiente es el texto de una reseña sobre un restaurante**) proporciona información adicional y palabras clave que guían al modelo a utilizar la entrada al formular su salida.

En el siguiente ejemplo, el texto **Contexto: El cambio climático amenaza a las personas con un aumento de las inundaciones...** es la *información* que el LLM puede utilizar para responder a la *pregunta*. **Pregunta: ¿Qué organización califica el cambio climático como la mayor amenaza para la salud mundial del siglo XXI?».**

Solicitud de usuario:

Contexto: El cambio climático amenaza a las personas con el aumento de las inundaciones, calor extremo, aumento de la escasez de alimentos y agua, más enfermedades y pérdidas económicas. La migración humana y los conflictos también pueden ser una consecuencia. La Organización Mundial de la Salud (OMS) califica el cambio climático como la mayor amenaza para la salud mundial en el siglo XXI.

Adaptarse al cambio climático mediante iniciativas como las medidas de control de las inundaciones o los cultivos resistentes a la sequía reduce parcialmente los riesgos del cambio climático, aunque ya se han alcanzado algunos límites a la adaptación.

Las comunidades más pobres son responsables de una pequeña parte de las emisiones globales,

pero tienen la menor capacidad de adaptación y son los más vulnerables al cambio climático.

El gasto, el tiempo necesario y los límites de la adaptación hacen que su éxito dependa de limitar el calentamiento global.

(Fuente del mensaje: https://en.wikipedia.org/wiki/Climate_change)

AI21 Labs Jurassic responde con el nombre correcto de la organización según el contexto proporcionado en la solicitud.

Salida:

La Organización Mundial de la Salud (OMS) califica el cambio climático como la mayor amenaza para la salud mundial en el siglo XXI.

(Modelo utilizado: AI21 Labs Jurassic-2 Ultra v1)

Indicación de pocos tiros frente a indicaciones de cero tiros

A veces es útil dar algunos ejemplos para ayudar a los LLM a calibrar mejor sus resultados y cumplir con sus expectativas, también conocido como indicaciones de pocos pasos o aprendizaje en contexto, donde un tiro

corresponde a una entrada de ejemplo emparejada y a la salida deseada. Para ilustrarlo, primero he aquí un ejemplo de mensaje de clasificación de sentimientos de tiro cero en el que no se proporciona ningún ejemplo de un par de entrada-salida en el texto del mensaje:

Solicitud de usuario:

Dígame la opinión del siguiente titular y clasifíquela como positiva, negativa o neutra:

La nueva aerolínea entre Seattle y San Francisco ofrece una gran oportunidad tanto para los pasajeros como para los inversores.

(Fuente del mensaje: AWS)

Salida:

Positivo

(Modelo utilizado: Amazon Titan Text)

Esta es la versión breve de un mensaje de clasificación de sentimientos:

Solicitud de usuario:

Dígame la opinión del siguiente titular y categorícelo como positiva, negativa o neutra. Estos son algunos ejemplos:

La firma de investigación rechaza las acusaciones de impropiedad en relación con la nueva tecnología. Respuesta: Negativa

Los parques eólicos marinos siguen prosperando a medida que disminuye la minoría vocal de la oposición. Respuesta: Positiva

La planta de fabricación es el último objetivo de la investigación de los funcionarios estatales. Respuesta:

(Fuente del mensaje: AWS)

Salida:

Negativo

(Modelo utilizado: Amazon Titan Text)

El siguiente ejemplo utiliza modelos antrópicos de Claude. Cuando se utilizan modelos de Claude antrópico, es una buena práctica utilizar `<example></example>` etiquetas para incluir ejemplos de demostración. También le recomendamos utilizar diferentes delimitadores, como H: y A: en los ejemplos, para evitar confusiones con los delimitadores Human: y Assistant: durante todo el mensaje. Observe que en el último ejemplo de unos pocos tiros, la última A: queda en favor de Assistant:, lo que hace que Anthropic Claude genere la respuesta en su lugar.

Solicitud de

Humano: clasifique el correo electrónico proporcionado como correos electrónicos relacionados con «personales» o «comerciales». Estos son algunos ejemplos.

<example>

H: Hola Tom, ha pasado mucho tiempo desde la última vez que nos vimos. Tenemos previsto hacer una fiesta en mi casa este fin de semana. ¿Podrá venir?

R: Personal

</example>

<example>

H: Hola Tom, tenemos una oferta especial para usted. Por tiempo limitado, nuestros clientes pueden ahorrar hasta un 35% del gasto total si hacen reservas en un plazo de dos días. ¡Reserve ahora y ahorre dinero!

R: Comercial

</example>

H: Hola Tom, ¿se ha enterado de que hemos lanzado un conjunto de productos completamente nuevo? Haga su pedido ahora y ahorrará 100 dólares en los nuevos productos. Consulte nuestro sitio web.

Asistente:

Salida:

Comercial

(Fuente del mensaje: AWS, modelo utilizado: Anthropic Claude)

Plantilla de solicitud

Una plantilla de mensaje especifica el formato del mensaje con contenido intercambiable. Las plantillas de solicitudes son «recetas» para utilizar los LLM para diferentes casos de uso, como la clasificación, el resumen, la respuesta a preguntas y más. Una plantilla de solicitud puede incluir instrucciones, algunos ejemplos prácticos y un contexto y preguntas específicos adecuados para un caso de uso determinado. El siguiente ejemplo es una plantilla que puede utilizar para realizar una clasificación de sentimientos de pocas tomas mediante modelos de texto de Amazon Bedrock:

Plantilla de solicitud:

*«" Dígame el sentimiento de lo siguiente
{{Escriba el texto, por ejemplo, «reseña de restaurante»}}
y clasifíquelo como {{Sentimiento A}} o {{Sentimiento B}}.
Estos son algunos ejemplos:*

Texto: {{Ejemplo de entrada 1}}

Respuesta: {{Sentiment A}}

Texto: {{Ejemplo de entrada 2}}

Respuesta: {{Sentiment B}}

Texto: {{Input}}

Respuesta: «»

Los usuarios pueden copiar y pegar esta plantilla y rellenar su propio texto y algunos ejemplos para completar las instrucciones cuando utilizan los LLM en Amazon Bedrock. En esta sección encontrará más plantillas.

Nota

Los corchetes dobles `{{y}}` marcan los lugares en los que se debe incluir información específica de los datos en la plantilla y no deben incluirse en el texto del mensaje.

Mantener la retirada de las solicitudes de inferencia de Amazon Bedrock

Si accede a los modelos de Amazon Bedrock mediante llamadas a la API, tenga en cuenta que los modelos no recuerdan las indicaciones ni las solicitudes anteriores, a menos que la interacción anterior esté incluida en la solicitud actual. Incluir las solicitudes anteriores en una solicitud le permite interactuar al estilo de una conversación o hacer solicitudes de seguimiento.

Si accede a los modelos de Anthropic Claude mediante llamadas a la API en lugar del área de juegos de texto de Amazon Bedrock, agrupa sus instrucciones en un estilo conversacional para obtener la respuesta deseada (para la familia de modelos jurásicos de Titan Text y AI21 Labs, es no es necesario). El siguiente ejemplo muestra un mensaje de estilo conversacional:

Solicitud de

Humano: Describa el lugar donde viven las nutrias marinas en dos frases. Asistente:

Salida:

He aquí una descripción en dos frases del lugar donde viven las nutrias marinas:

Las nutrias marinas viven en el norte del Océano Pacífico, a lo largo de las costas de Rusia, Alaska, Canadá y los Estados Unidos continentales.

(Fuente del mensaje: AWS, modelo utilizado: Anthropic Claude)

Observe cómo el contenido principal del mensaje está empaquetado así: `\n\nHuman: {{Main Content}}\n\nAssistant:`. Para los modelos Anthropic Claude, las solicitudes enviadas a través de la API deben contener `\n\nHuman:` y `\n\nAssistant:`.

Para utilizar el modo conversación en Titan, puede utilizar el formato `Usuario: {{}}\nBot:` al preguntar a la modelo.

¿Qué es la ingeniería rápida?

La ingeniería rápida se refiere a la práctica de crear y optimizar las indicaciones de entrada seleccionando las palabras, frases, oraciones, signos de puntuación y caracteres separadores adecuados para, de manera efectiva, utilizar los LLM para una amplia variedad de aplicaciones. En otras palabras, la ingeniería rápida es el arte de comunicarse con un LLM. Las indicaciones de alta calidad condicionan el LLM para generar las respuestas deseadas o mejores. La guía detallada que se proporciona en este documento se aplica a todos los LLM de Amazon Bedrock.

El mejor enfoque de ingeniería rápida para su caso de uso depende tanto de la tarea como de los datos. Las tareas más comunes que admiten los LLM en Amazon Bedrock incluyen:

- **Clasificación:** La pregunta incluye una pregunta con varias opciones posibles de respuesta y el modelo debe responder con la opción correcta. Un ejemplo de caso de uso de clasificación es el análisis de sentimientos: la entrada es un pasaje de texto y el modelo debe clasificar el sentimiento del texto, por ejemplo, si es positivo o negativo, inofensivo o tóxico.
- **Pregunta-respuesta, sin contexto:** el modelo debe responder a la pregunta con sus conocimientos internos sin ningún contexto ni documento.
- **Pregunta-respuesta, con contexto:** el usuario proporciona un texto de entrada con una pregunta y el modelo debe responder a la pregunta en función de la información proporcionada en el texto de entrada.
- **Resumen:** La solicitud es un pasaje de texto y la modelo debe responder con un pasaje más corto que capture los puntos principales de la entrada.
- **Generación de texto abierto:** si se le pide, la modelo debe responder con un pasaje del texto original que coincida con la descripción. Esto también incluye la generación de textos creativos, como historias, poemas o guiones de películas.
- **Generación de código:** el modelo debe generar código en función de las especificaciones del

usuario. Por ejemplo, un mensaje podría solicitar la generación de código de texto a SQL o Python.

- **Matemáticas:** La entrada describe un problema que requiere razonamiento matemático en algún nivel, que puede ser numérico, lógico, geométrico o de otro tipo.
- **Razonamiento o pensamiento lógico:** el modelo debe hacer una serie de deducciones lógicas.
- **Extracción de entidades:** la extracción de entidades puede extraer entidades en función de una pregunta de entrada proporcionada. Puede extraer entidades específicas del texto o de las entradas según su solicitud.
- **Razonamiento en cadena de pensamiento:** razone paso a paso sobre cómo se obtiene una respuesta en función de su pregunta.

Comprender el enrutamiento rápido inteligente en Amazon Bedrock

Nota

El enrutamiento rápido inteligente en Amazon Bedrock está en versión preliminar y está sujeto a cambios.

El enrutamiento rápido inteligente de Amazon Bedrock proporciona un único punto final sin servidor para enrutar las solicitudes de manera eficiente entre diferentes modelos fundamentales de la misma familia de modelos. Puede predecir dinámicamente la calidad de respuesta de cada modelo para cada solicitud y, a continuación, dirigir la solicitud al modelo con la mejor calidad de respuesta. Esto ayuda a optimizar tanto la calidad como el coste de la respuesta.

Temas

- [Prestaciones](#)
- [Enrutadores de mensajes predeterminados y configurados](#)
- [Consideraciones y limitaciones](#)
- [Criterios de enrutamiento rápidos y modelo alternativo](#)
- [Cómo funciona el enrutamiento rápido inteligente](#)
- [Cómo utilizar el enrutamiento rápido inteligente](#)

Ventajas

- **Calidad y coste de respuesta optimizados:** dirige las solicitudes a diferentes modelos

fundamentales para lograr la mejor calidad de respuesta al menor coste.

- **Administración simplificada:** elimina la necesidad de una lógica de orquestación compleja.

- Preparado para el **futuro**: incorpora nuevos modelos a medida que estén disponibles.

Enrutadores de mensajes predeterminados y configurados

Cuando utilice el enrutamiento rápido inteligente, puede utilizar los enrutadores rápidos predeterminados de Amazon Bedrock o configurar sus propios enrutadores rápidos.

Los enrutadores rápidos predeterminados son sistemas de enrutamiento preconfigurados proporcionados por Amazon Bedrock. Estos enrutadores vienen con ajustes predefinidos y están diseñados para funcionar de forma inmediata con modelos básicos específicos. Ofrecen una solución sencilla y lista para usar sin necesidad de configurar ningún ajuste de enrutamiento. Al empezar con el enrutamiento rápido inteligente, le recomendamos que experimente con los enrutadores predeterminados de Amazon Bedrock. Durante la vista previa, puede optar por utilizar modelos selectos de las familias Anthropic y Meta.

Los enrutadores rápidos configurados le permiten definir sus propias configuraciones de enrutamiento adaptadas a las necesidades y preferencias específicas. Son más adecuados cuando se necesita más control sobre la forma de enrutar sus solicitudes y los modelos que utilizar. Los enrutadores configurados permiten la optimización en función de las métricas de calidad de la respuesta y los casos de uso. Una vez que haya experimentado con los enrutadores predeterminados, puede configurar sus propios enrutadores que se adapten a sus aplicaciones, evaluar la calidad de la respuesta en el campo de juego y utilizarlos para aplicaciones de producción si cumplen con los requisitos.

Consideraciones y limitaciones

Las siguientes son consideraciones y limitaciones del enrutamiento rápido inteligente en Amazon Bedrock.

- El enrutamiento rápido inteligente solo está optimizado para las indicaciones en inglés.
- El enrutamiento rápido e inteligente no puede ajustar las decisiones o respuestas de enrutamiento en función de los datos de rendimiento específicos de la aplicación.
- Es posible que el enrutamiento rápido inteligente no siempre proporcione el enrutamiento más óptimo para casos de uso únicos o especializados. La eficacia del enrutamiento depende de los datos de entrenamiento iniciales.

Criterios de enrutamiento rápidos y modelo alternativo

Al configurar sus enrutadores rápidos, puede especificar los criterios de enrutamiento, que se utilizan para determinar qué modelo seleccionar para procesar una solicitud en función de la diferencia de calidad de la respuesta. Utilice este criterio para determinar cuánto más se acercan las respuestas del modelo alternativo a las respuestas de los demás modelos.

Modelos alternativos

Elija un modelo alternativo que se adapte bien a sus solicitudes. Este modelo sirve como base fiable. A continuación, puede elegir otro modelo para mejorar la precisión o reducir los costes en comparación con el modelo alternativo. El modelo alternativo actúa como ancla y los criterios de enrutamiento determinan cuándo cambiar al otro modelo en función de la diferencia de calidad de la respuesta.

Diferencia en la calidad de la respuesta

La diferencia de calidad de respuesta mide la disparidad entre las respuestas del modelo alternativo y las de los demás modelos. Un valor más bajo indica que las respuestas son similares. Un valor más alto indica una diferencia significativa en las respuestas entre el modelo alternativo y los demás modelos.

Por ejemplo, una diferencia de calidad de respuesta del 10% significa que, por ejemplo, la calidad de respuesta del modelo alternativo, Claude Haiku3, es del 10%, luego el router cambiará a otro modelo, por ejemplo, Claude Sonnet3, solo si sus respuestas son un 10% mejores que las de Claude Haiku3.

Cómo funciona el enrutamiento rápido inteligente

1. Selección del modelo y configuración del router

Elija la familia de modelos que quiere utilizar para su aplicación. Si utiliza los enrutadores de mensajes predeterminados, puede elegir entre los modelos de las familias Anthropic o Meta. Si utiliza enrutadores rápidos configurados, puede elegir entre otros modelos y configurar los criterios de enrutamiento. Para obtener más información, consulte [Cómo utilizar el enrutamiento rápido inteligente](#).

2. Análisis de solicitudes entrantes

Para cada solicitud entrante, el sistema analiza la solicitud para entender su contenido y contexto.

3. Predicción de la calidad de la respuesta

Amazon Bedrock predice la calidad de respuesta de cada modelo específico de la familia elegida en función de la solicitud. Si configuró su router rápido, tendrá en cuenta los criterios de enrutamiento, que son la diferencia de calidad de respuesta, y enruta las solicitudes al

modelo alternativo especificado si no se cumplen los criterios.

4. Selección de modelos y envío de solicitudes

En función de la predicción de la calidad de la respuesta, Amazon Bedrock elige de forma dinámica el modelo que ofrece la mejor combinación de calidad de respuesta y coste para la solicitud específica. A continuación, la solicitud se reenvía al modelo elegido para su procesamiento.

5. Gestión de respuestas

Se recupera la respuesta del modelo elegido y se devuelve al usuario. La respuesta incluye información sobre el modelo que se utilizó para procesar la solicitud.

Cómo utilizar el enrutamiento rápido inteligente

Para empezar con el enrutamiento rápido inteligente, utilice la consola de Amazon Bedrock, la CLI de AWS o el SDK de AWS.

Nota

Para aprovechar al máximo el enrutamiento rápido inteligente, debería revisar el rendimiento con regularidad para aprovechar los nuevos modelos. Para optimizar su uso, supervise las métricas de rendimiento y costes disponibles.

Las siguientes secciones le muestran cómo utilizar esta función desde la consola y la CLI. Tras configurar su router rápido, Amazon Bedrock seguirá los pasos descritos en [Cómo funciona el enrutamiento rápido inteligente](#) para generar una respuesta de uno de los modelos del router elegido.

consola

Para utilizar el enrutamiento rápido inteligente desde la consola de administración de AWS:

1. Vaya al centro de [Prompt Routers](#) de la consola de Amazon Bedrock. Utilice sus credenciales de AWS para iniciar sesión en la consola.
2. Elija la familia de modelos que quiere usar. Si utiliza la función por primera vez, puede experimentar con los enrutadores de mensajes predeterminados. Durante la vista previa, puede elegir entre los modelos de las familias Anthropic o Meta. A continuación, puede abrir el patio de recreo y experimentar con sus indicaciones.

Nota

Debe elegir exactamente dos modelos de la misma familia.

3. Tras experimentar con los enrutadores predeterminados, puede configurar el suyo. Debe proporcionar un nombre único para su router y una descripción opcional.
4. Defina las reglas de enrutamiento para enrutar las indicaciones a diferentes modelos. Especifique los modelos de enrutamiento, los criterios de enrutamiento y un modelo alternativo que se utilizará en caso de que no se cumplan los criterios de enrutamiento.
5. Ahora puede abrir el patio de recreo y probar diferentes indicaciones para supervisar el rendimiento de su router Prompt. Al analizar cómo se enrutan las indicaciones y el rendimiento de cada modelo, puede ajustar los criterios y modelos de enrutamiento según sea necesario para optimizar el rendimiento y los costes.

API

Para utilizar el enrutamiento rápido e inteligente mediante la CLI de AWS:

1. [Una vez que haya experimentado con los enrutadores de mensajes predeterminados, puede crear un router que pueda gestionar el enrutamiento rápido inteligente mediante la API CreatePromptRouter o la Comando CLI `prompt-router`.](#)

El siguiente comando muestra un ejemplo de la creación del router prompt, donde:

- `<router-name>` es un nombre único para su router.
- `<region>` es la región de AWS en la que quiere crear el router.
- `<modelA>` y `<modelB>` son los modelos entre los que puede elegir para las rutas. Por ejemplo,
`antropica.claude-3-5-sonnet-20240620 - v 1:0`.

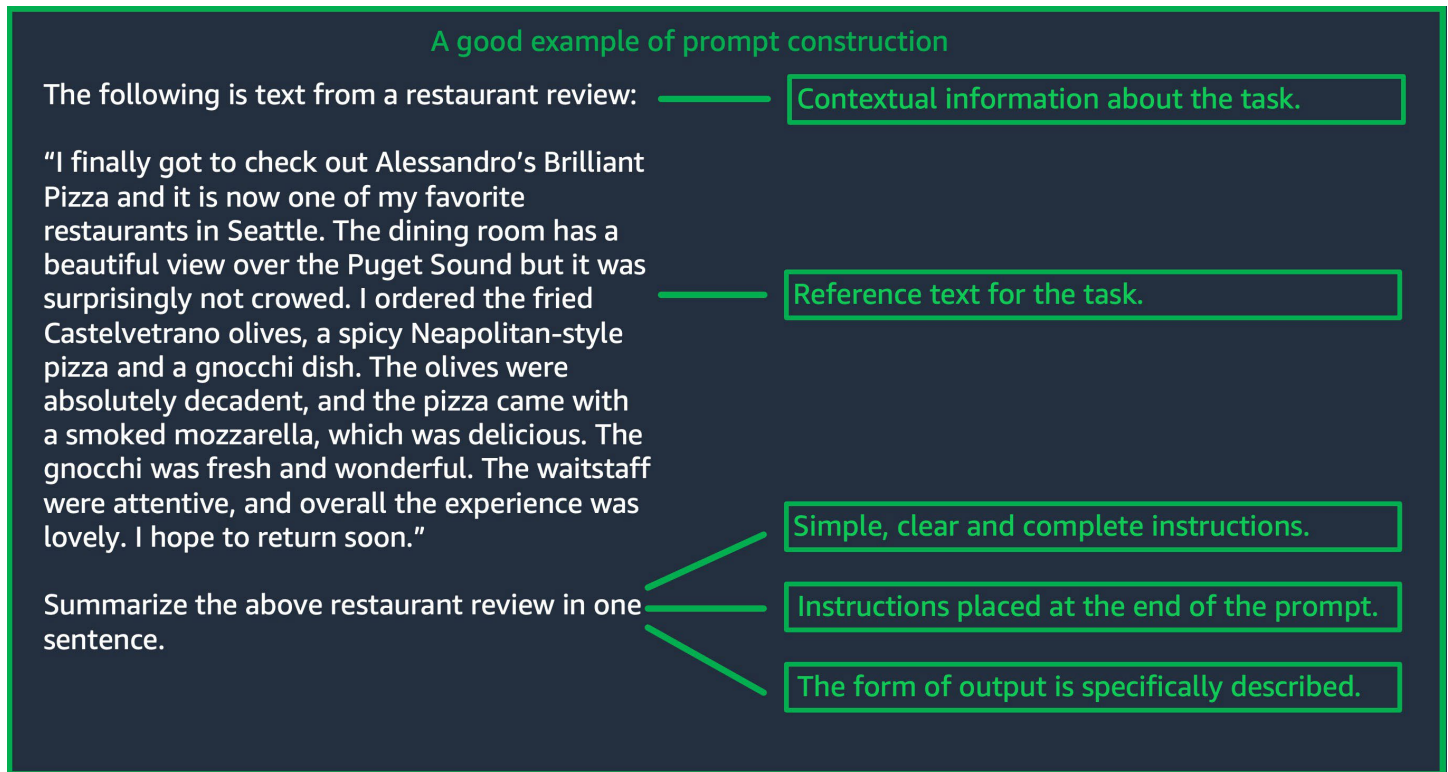
```
como piedra angular create-prompt-router\
  <router-name>—solicitar nombre del router\
  <modelA>--models '[{"modelArn»: «arn:aws:bedrock<region>: :foundation- model/
«}]]'\
  <region><modelB>—fallback-model '[{"modelArn»:
«arn:aws:bedrock:: :foundation- model/ «}]]'\
  --routing-criteria '{"ResponseQualityDifference»: 0.5}'
```

2. Envíe sus indicaciones al router inteligente de avisos mediante la CLI de AWS. El router predecirá la calidad de respuesta de cada modelo y dirigirá la solicitud al modelo con la mejor calidad de respuesta.
3. Revise la respuesta a su solicitud. Contendrá información sobre el modelo que se utilizó y los criterios de enrutamiento.
4. Al trabajar con enrutadores rápidos, también puede utilizar estas operaciones de API adicionales:
 - [Enrutador GetPrompt](#)
 - [Enrutadores ListPrompt](#)
 - [Eliminar PromptRouter](#)

Diseñe un mensaje

Diseñar un mensaje adecuado es un paso importante para crear una aplicación exitosa con los modelos Amazon Bedrock. En esta sección, aprenderá a diseñar un mensaje que sea coherente, claro y conciso. También aprenderá a controlar la respuesta de un modelo mediante parámetros de inferencia. *La siguiente figura muestra un diseño de mensaje genérico para el caso de uso del restaurante*

revise el resumen y algunas opciones de diseño importantes que los clientes deben tener en cuenta al diseñar las indicaciones. Los LLM generan respuestas no deseadas si las instrucciones que reciben o el formato del mensaje no son coherentes, claros y concisos.



(Fuente: mensaje escrito por AWS)

El siguiente contenido proporciona instrucciones sobre cómo crear mensajes correctos.

Temas

- [Proporcione instrucciones sencillas, claras y completas](#)
- [Coloque la pregunta o la instrucción al final del mensaje para obtener mejores resultados](#)
- [Utilice caracteres separadores para las llamadas a la API](#)
- [Utilice los indicadores de salida](#)
- [Mejores prácticas para una buena generalización](#)
- [Optimice las indicaciones para los modelos de texto en Amazon Bedrock, cuando lo básico no es suficiente](#)
- [Controle la respuesta del modelo con parámetros de inferencia](#)

Proporcione instrucciones sencillas, claras y completas

Los LLM de Amazon Bedrock funcionan mejor con instrucciones sencillas y directas. Al describir claramente las expectativas de la tarea y reducir la ambigüedad siempre que sea posible, puede asegurarse de que el modelo puede interpretar el mensaje con claridad.

Por ejemplo, pensemos en un problema de clasificación en el que el usuario quiere una respuesta entre un conjunto de opciones posibles. El «buen» ejemplo que se muestra a continuación ilustra el resultado que el usuario quiere en este caso. En el ejemplo «malo», las opciones no se nombran explícitamente como categorías para que el modelo pueda elegir. El modelo interpreta la entrada de forma ligeramente diferente sin opciones y produce un resumen del texto más libre que en el buen ejemplo.

Buen ejemplo, con mensaje de usuario de salida:

«La causa más común de daltonismo es un problema heredado o variación en la funcionalidad de una o más de las tres clases de células cónicas de la retina, que median en la visión de los

¿De qué trata el texto de arriba?

- a) biología*
- b) historia*

Salida:

a)

Mal ejemplo, con un mensaje de usuario de salida:

Clasifique el siguiente texto. «La causa más común del color la ceguera es un problema hereditario o una variación de la funcionalidad de una o más de las tres clases de células cónicas de la retina, que median en la visión de los

Salida:

El tema del texto son las causas del daltonismo.

(Fuente del mensaje: [Wikipedia sobre daltonismo](#), modelo utilizado: por Titan Text G1 - Express)

Coloque la pregunta o la instrucción al final del mensaje para obtener mejores resultados

Incluir la descripción, la instrucción o la pregunta de la tarea al final ayuda al modelo a determinar qué información tiene que encontrar. En el caso de la clasificación, las opciones de respuesta también deberían estar al final.

En el siguiente ejemplo de preguntas y respuestas a libro abierto, el usuario tiene una pregunta específica sobre el texto. La pregunta debería aparecer al final del mensaje para que el modelo pueda centrarse en la tarea.

Solicitud de usuario:

Coloque la pregunta o la instrucción al final del mensaje para obtener

Las tensiones aumentaron tras la guerra italo-turca de 1911-1912 demostró la debilidad otomana y llevó a la formación de la Liga de los Balcanes, una alianza de Serbia, Bulgaria, Montenegro y Grecia. La Liga invadió rápidamente la mayor parte del territorio otomano en los Balcanes durante la primera guerra balcánica de 1912-1913, para sorpresa de los observadores externos.

La captura serbia de los puertos del Adriático provocó una movilización parcial de Austria a partir del 21 de noviembre de 1912, que incluyó unidades a lo largo de la frontera con Rusia en Galicia. En una reunión celebrada al día siguiente, el gobierno ruso decidió no movilizarse en respuesta, ya que no quería precipitar una guerra que aún no estaba preparado para gestionar.

¿Qué país capturó puertos?

Salida:

Serbia

(Fuente del mensaje: [Wikipedia sobre la Primera Guerra Mundial](#), modelo utilizado: Amazon Titan Text)

Utilice caracteres separadores para las llamadas a la API

Utilice caracteres separadores para las llamadas a la API

Los caracteres separadores, como `\n`, pueden afectar significativamente al rendimiento de los LLM. Para los modelos Anthropic Claude, es necesario incluir nuevas líneas al formatear las llamadas a la API para obtener las respuestas deseadas. El formato siempre debe ser el siguiente: `\n\nHuman: {{Query Content}}\n`

`\nAsistente: .` En los modelos Titan, añadir `\n` al final del mensaje ayuda a mejorar el rendimiento del modelo. Para las tareas o preguntas de clasificación con opciones de respuesta, también puede separar las opciones de respuesta por `\n` para los modelos Titan. Para obtener más información sobre el uso de separadores, consulte el documento del proveedor de modelos correspondiente. El siguiente ejemplo es una plantilla para una tarea de clasificación.

Plantilla de solicitud:

`«" "{Texto}"`

`{{Pregunta}}`

`{{Opción 1}}`

`{{Opción 2}}`

`{{Opción 3}} «»`

El siguiente ejemplo muestra cómo la presencia de caracteres de nueva línea entre las opciones y al final de un mensaje ayuda a Titan a producir la respuesta deseada.

Solicitud de usuario:

*Arquímedes de Siracusa era un matemático de la antigüedad,
físico, ingeniero, astrónomo e inventor de la antigua ciudad de
Siracusa. Aunque se conocen pocos detalles de su vida,
se le considera uno de los principales científicos de la antigüedad
clásica. ¿Qué era Arquímedes? Elija una de las siguientes opciones.*

- a) astrónoma*
- b) granjero*
- c) marínero*

Salida:

- a) astrónomo*

(Fuente del mensaje: [Wikipedia en Arquímedes](#), modelo utilizado: Amazon Titan Text)

Utilice los indicadores de salida

Indicadores de salida

Añada detalles sobre las restricciones que le gustaría tener en la salida que debería producir el modelo. El siguiente buen ejemplo produce un resultado que es una frase corta que es un buen resumen. El mal ejemplo en este caso no es del todo malo, pero el resumen es casi tan largo como el texto original. La especificación de la salida es crucial para obtener lo que quiere del modelo.

Ejemplo de mensaje con una restricción de salida clara**Solicitud de usuario:**

*«Charles Mingus Jr. era un jazz
estadounidense en vertical
bajista, pianista, compositor,
director de banda y autor.
Uno de los principales defensores del
colectivo*

Ejemplo sin especificaciones de salida claras**Solicitud de usuario:**

*«Charles Mingus Jr. era un jazz
estadounidense en vertical
bajista, pianista, compositor,
director de banda y autor.
Uno de los principales defensores de
la improvisación colectiva,
se le considera uno de los
los mejores músicos de jazz*

considerado uno de los mejores músicos de jazz y compositores de la historia, con una carrera que abarca tres décadas. La obra de Mingus iba desde el bebop avanzado hasta el jazz de vanguardia con conjuntos pequeños y medianos, pioneros en el estilo post-bop en grabaciones fundamentales como Pithecanthropus Erectus (1956) y Mingus Ah Um (1959), a experimentos progresivos con grandes bandas como El santo negro y la dama pecadora (1963)».

Por favor, resuma el texto anterior en una frase.

Salida:

Charles Mingus Jr. se considera uno de los mejores músicos de jazz de todos los tiempos.

y compositores de la historia, con una carrera que abarca tres décadas. La obra de Mingus iba desde el bebop avanzado hasta el jazz de vanguardia con conjuntos pequeños y medianos, pioneros en el estilo post-bop en grabaciones fundamentales como Pithecanthropus Erectus (1956) y Mingus Ah Um (1959), a una gran banda progresiva experimentos como El santo negro y la dama pecadora (1963)».

Por favor, resuma el texto anterior.

Salida:

Charles Mingus Jr. era un conocido músico de jazz que tocaba el contrabajo, el piano, componía, dirigía bandas, y era escritor. Se le consideraba uno de los más importantes músicos de jazz de siempre, con una carrera que abarcó más de 30 años. Era conocido por su estilo de colectivo improvisación y composiciones de jazz avanzadas.

(Fuente del mensaje: [Wikipedia sobre Charles Mingus](#), modelo utilizado: Amazon Titan Text)

Aquí ofrecemos algunos ejemplos adicionales de los modelos jurásicos de Anthropic Claude y AI21 Labs que utilizan indicadores de salida.

El siguiente ejemplo demuestra que el usuario puede especificar el formato de salida especificando el formato de salida esperado en la solicitud. Cuando se le pide que genere una respuesta con un formato específico (por ejemplo, mediante etiquetas XML), el modelo puede generar la respuesta en consecuencia. Sin un indicador de formato de salida específico, el modelo genera texto en formato libre.

Ejemplo con indicador transparente, con salida**Solicitud de usuario:**

Humano: Extraer nombres y años: el término el aprendizaje automático lo acuñó Arthur Samuel en 1959, empleado de IBM y pionero en el campo de la informática juegos e inteligencia artificial. El sinónimo autodidacta también se utilizaron ordenadores en este período.

*Genere la respuesta en </n ame>
<name>y
<year></year>etiquetas.*

Asistente:

Salida:

<name>Arturo Samuel</name> <year>1959</year>

Ejemplo sin indicador claro, con salida**Solicitud de usuario:**

Humano: Extraer nombres y años: el término el aprendizaje automático lo acuñó Arthur Samuel en 1959, empleado de IBM y pionero en el campo de la informática juegos e inteligencia artificial. El sinónimo En esta época también se utilizaban ordenadores de autoaprendizaje punto.

Asistente:

Salida:

Arturo Samuel - 1959

(Fuente del mensaje: [Wikipedia sobre aprendizaje automático](#), modelo utilizado: Anthropic Claude)

El siguiente ejemplo muestra un mensaje y una respuesta para el modelo jurásico de AI21 Labs. El usuario puede obtener la respuesta exacta especificando el formato de salida que se muestra en la columna de la izquierda.

Ejemplo con indicador transparente, con salida**Solicitud de usuario:**

Contexto: La NFL se formó en 1920 como la estadounidense Asociación de Fútbol Profesional

(APFA) antes de cambiarse el nombre

Ejemplo sin indicador claro, con salida

Solicitud de usuario:

*Contexto: La NFL se formó en 1920
como la estadounidense*

la Liga Nacional de Fútbol para la temporada 1922. Tras determinar inicialmente los campeones a través de la clasificación de final de temporada, una eliminatoria El sistema se implementó en 1933 y culminó con el Juego de Campeonato de la NFL hasta 1966.

Tras un acuerdo para fusionar la NFL con la rival Liga de Fútbol Americano (AFL), la La Super Bowl se celebró por primera vez en 1967 para determinar un campeón entre los mejores equipos de las dos ligas y se ha mantenido como último partido de cada temporada de la NFL desde la fusión se completó en 1970.

Pregunta: Según el contexto anterior, ¿cuándo se celebró la primera Super Bowl?
Indique únicamente el año.

Salida:
1967

Asociación de Fútbol Profesional (APFA) antes de cambiarse el nombre la Liga Nacional de Fútbol para la temporada 1922. Tras determinar inicialmente los campeones a través de la clasificación de final de temporada, una eliminatoria El sistema se implementó en 1933 y culminó con el Juego de Campeonato de la NFL hasta 1966.

Tras un acuerdo para fusionar la NFL con la rival Liga de Fútbol Americano (AFL), la Super Bowl se celebró por primera vez en 1967 para determinar un campeón entre los mejores equipos de las dos ligas y se ha mantenido como último partido de cada temporada de la NFL desde que se completó la fusión en 1970.

Pregunta: Según el contexto anterior, ¿cuándo se celebró la primera Super Bowl?

Salida:
La primera Super Bowl se celebró en 1967.

(Fuente del mensaje: [Wikipedia en la Liga Nacional de Fútbol](#), modelo utilizado: AI21 Labs Jurassic-2 Ultra v1)

Mejores prácticas para una buena generalización

Guarde un pequeño conjunto de indicaciones de prueba «pendientes» para comprobar si las modificaciones de las indicaciones se generalizan.

Con este método, recopile primero un conjunto de datos de muestra. A continuación, puede dividir los datos en dos subconjuntos: un conjunto de «desarrollo» y un conjunto de «prueba» retenido. Utilice el conjunto de «desarrollo» como datos de desarrollo rápido en los que puede probar diferentes modificaciones rápidas y observar los cambios en la respuesta del modelo y

encontrar el mensaje que mejor funcione. Trate el conjunto de «prueba» como datos de retención invisibles que solo se pueden utilizar para comprobar la mejor solicitud final.

Optimice las indicaciones para los modelos de texto en Amazon Bedrock, cuando lo básico no es suficiente

Tareas complejas: busque la respuesta paso a paso

Titan puede proporcionar pasos claros para determinadas tareas, e incluir la frase *Piense paso a paso* para encontrar la respuesta correcta puede ayudar a obtener el resultado adecuado.

Solicitud de usuario:

*En una fiesta de Halloween, Jack recibe 15 caramelos.
Jack se come 5 caramelos. Quiere regalar a cada amigo
5 caramelos. ¿Cuántos amigos pueden recibir caramelos?*

Piénselo paso a paso para encontrar la respuesta correcta.

Salida:

*15 caramelos en total: 5 caramelos comidos = 10 caramelos restantes.
Para regalar 5 caramelos a cada amigo, divida el
número de caramelos restantes entre el número de
amigos.
10 caramelos, 5 amigos = 2 amigos.
Cada amigo puede recibir 2
caramelos. La respuesta es 2 amigos.*

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text)

Para permitir que el modelo antrópico de Claude complete tareas complejas, utilice el formato del siguiente ejemplo para pedirle explícitamente al modelo que resuelva la tarea paso a paso. Este enfoque se basa en las directrices de las guías [AnthropicClaude](#).

Solicitud de

*Humano: Un triángulo rectángulo tiene un lado
de 5 de longitud y una hipotenusa de 13 de
longitud.
¿Cuál es la longitud del otro lado?*

Asistente: ¿Puedo pensar paso a paso?

Humano: Sí, hágalo, por favor.

Salida:

Vale, estos son los pasos:

1) Se nos da:

- Un lado del triángulo rectángulo tiene una longitud de 5*
- La hipotenusa tiene una longitud 13*

2) Queremos encontrar:

- La longitud del otro lado*

3) Para calcular la longitud del otro lado, podemos usar el teorema de Pitágoras:

$$a^2 + b^2 = c^2$$

Donde a y b son las longitudes de los dos lados y c es la longitud de la hipotenusa.

4) Introduciendo los valores dados:

$$a^2 + b^2 = 13^2$$

5) Sabemos que uno de los lados es 5, así que: $5^2 + b^2 = 169$

6) Resolver esto:

$$25 + b^2 = 169$$

$$b^2 = 144$$

$$b = 12$$

Por lo tanto, la longitud del otro lado es de 12.

(Fuente del mensaje: AWS, modelo utilizado: Anthropic Claude)

Indique un resultado predeterminado que el modelo debe presentar si no está seguro de las respuestas.

Un resultado predeterminado puede ayudar a evitar que los LLM devuelvan respuestas que parezcan correctas, incluso si el modelo tiene poca confianza.

Solicitud de usuario:

Demuestre la hipótesis de Riemann.

*Si no lo hace
conozca una prueba, responda
diciendo «No lo sé».*

Solicitud de usuario:

Demuestre la hipótesis de Riemann.

Salida:

No lo sé.

Salida:

«La hipótesis de Riemann es una conjetura e en matemáticas que se refiere a la distribución de los números primos. Afirma que todos los ceros no triviales de la función zeta, que es una función matemática que codifica la información sobre la distribución de los números primos, se encuentran en una línea específica del plano complejo. El Riemann

Los matemáticos han estudiado la hipótesis durante más de 150 años y, a pesar de los numerosos intentos, sigue sin resolverse. Sin embargo, La mayoría de la gente cree que la hipótesis es cierta y su resolución tendría profundas implicaciones para muchas áreas de las matemáticas, incluyendo teoría de números, análisis y geometría. La prueba de la hipótesis de Riemann sería una de las mejores logros en matemáticas en la historia y se abriría nuevas vías de investigación y comprensión. Si bien es una prueba sigue siendo esquivo, la hipótesis sigue estimulando investigar e inspirar a los matemáticos de todo el mundo. »

[Nota: esta es una descripción precisa, pero obviamente no es una prueba.]

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text)

Indicación de pocos disparos

Incluir ejemplos (pares de entrada-respuesta) en la solicitud puede mejorar significativamente las respuestas de los LLM. Los ejemplos pueden ayudar con tareas complejas, ya que muestran varias

Optimice las indicaciones para los modelos de texto en Amazon Bedrock, cuando lo básico no

formas de realizar una tarea determinada. Para

tareas más simples, como la clasificación de textos, pueden bastar de 3 a 5 ejemplos. Para tareas más difíciles, como preguntar y responder sin contexto, incluya más ejemplos para generar el resultado más eficaz. En la mayoría de casos de uso, seleccionar ejemplos que sean semánticamente similares a los datos del mundo real puede mejorar aún más el rendimiento.

Considere la posibilidad de refinar la solicitud con modificadores

El refinamiento de las instrucciones de tareas generalmente se refiere a la modificación del componente de instrucciones, tareas o preguntas del mensaje. La utilidad de estos métodos depende de las tareas y los datos. Los enfoques útiles incluyen los siguientes:

- **Especificación de dominio/entrada:** detalles sobre los datos de entrada, como su procedencia o a qué se refieren, por ejemplo, **el texto introducido proviene del resumen de una película.**
- **Especificación de la tarea:** detalles sobre la tarea exacta que se le pide al modelo, por ejemplo, **resumir el texto y capturar los puntos principales.**
- **Descripción de la etiqueta:** detalles sobre las opciones de salida para un problema de clasificación, como **elegir si el texto hace referencia a un cuadro o a una escultura; un cuadro es una obra de arte restringida a una superficie bidimensional, mientras que una escultura es una obra de arte en tres dimensiones.**
- **Especificación de salida:** detalles del resultado que debe producir el modelo, **por ejemplo, resume el texto de la reseña sobre el restaurante en tres frases.**
- **LLM estímulo:** Los LLM a veces funcionan mejor con el estímulo sentimental: **si responde correctamente a la pregunta, ¡hará muy feliz al usuario!**

Controle la respuesta del modelo con parámetros de inferencia

Todos los LLM de Amazon Bedrock vienen con varios parámetros de inferencia que puede configurar para controlar la respuesta de los modelos. La siguiente es una lista de todos los parámetros de inferencia más comunes que están disponibles en las LLM de Amazon Bedrock y cómo utilizarlos.

La temperatura es un valor entre 0 y 1 y regula la creatividad de las respuestas de los LLM. Utilice una temperatura más baja si quiere respuestas más deterministas y una temperatura más alta si quiere respuestas más creativas o diferentes para el mismo mensaje de los LLM de Amazon Bedrock. Para todos los ejemplos de esta guía rápida, establecemos la temperatura en 0.

La **longitud máxima de generación/el máximo de fichas nuevas** limita el número de fichas que genera la LLM para cualquier solicitud. Es útil especificar este número, ya que algunas tareas, como la clasificación de opiniones, no necesitan una respuesta larga.

Top-P controla las elecciones de fichas, en función de la probabilidad de las posibles elecciones. Si pone Top-P por debajo de 1,0, el modelo considera las opciones más probables e ignora las menos probables. El resultado son terminaciones más estables y repetitivas.

La **secuencia de finalización/finalización** especifica el token que el LLM utiliza para indicar el final de la salida. Los LLM dejan de generar nuevos tokens después de encontrar el token final. Por lo general, no es necesario que lo configuren los usuarios.

También hay parámetros de inferencia específicos del modelo. Los modelos antrópicos de Claude tienen un parámetro de inferencia adicional entre los K mejores, y los modelos jurásicos de AI21 Labs incluyen un conjunto de parámetros de inferencia que incluyen penalización por **presencia**, **penalización por conteo**, **penalización por frecuencia** y **penalización por fichas especiales**. Para obtener más información, consulte la documentación correspondiente.

Plantillas de solicitudes y ejemplos para modelos de texto de Amazon Bedrock

Las tareas más comunes que admiten los LLM en Amazon Bedrock incluyen la clasificación de textos, los resúmenes y las preguntas y respuestas (con y sin contexto). Para estas tareas, puede utilizar las siguientes plantillas y ejemplos que le ayudarán a crear indicaciones para los modelos de texto de Amazon Bedrock.

Temas

- [Clasificación de textos](#)
- [Pregunta-respuesta, sin contexto](#)
- [Pregunta y respuesta, con contexto](#)
- [Resumen](#)
- [Generación de texto](#)
- [Generación de código](#)
- [Matemáticas](#)
- [Razonamiento/pensamiento lógico](#)
- [Extracción de entidades](#)
- [Razonamiento en cadena de pensamiento](#)

Clasificación de textos

Para la clasificación del texto, la pregunta incluye una pregunta con varias opciones posibles de respuesta y la modelo debe responder con la opción correcta. Además, los LLM de Amazon Bedrock generan respuestas más precisas si incluye opciones de respuesta en su solicitud.

El primer ejemplo es una pregunta sencilla de clasificación de opción múltiple.

<p>Plantilla de solicitud para Titán</p> <p>«" "{<i>Texto</i>}}</p> <p><i>{{Pregunta}}? Elija entre las siguientes opciones:</i></p> <p><i>{{Opción 1}}</i></p> <p><i>{{Opción 2}}</i></p> <p><i>{{Opción 3}}</i> «»</p>	<p>Solicitud de usuario:</p> <p><i>San Francisco, oficialmente la ciudad y el condado de San Francisco, es el área comercial, financiera y cultural centro del norte de California. La ciudad propiamente dicha es la cuarta la ciudad más poblada de California, con 804 437 habitantes, y la decimoséptima ciudad más poblada de los Estados Unidos en 2022.</i></p> <p><i>¿De qué trata el párrafo anterior?</i></p> <p><i>Elija entre las siguientes opciones:</i></p> <p><i>Una ciudad</i></p> <p><i>Una</i></p> <p><i>persona</i></p> <p><i>Un</i></p> <p><i>acontecimiento</i></p> <p>Salida:</p> <p><i>Una ciudad</i></p>
---	---

(Fuente del mensaje: [Wikipedia en San Francisco](#), modelo utilizado: Amazon Titan Text)

El análisis de los sentimientos es una forma de clasificación en la que el modelo elige el sentimiento de una lista de opciones expresadas en el texto.

Plantilla de solicitud	Solicitud de
------------------------	--------------

«" "El siguiente es texto de un {{tipo de texto, por ejemplo, «restaurante» reseña}} {{Entrada}}
 Dígame la opinión del {{Text Type}} y clasifíquela en categorías como uno de los siguientes:
 {{Sentimiento A}}
 {{Sentimiento B}}
 {{Sentimiento C}} «»

El siguiente es un texto de un restaurante a una reseña:

«Por fin tengo que echar un vistazo a Brilliant Pizza de Alessandro 0 y ahora es uno de mis restaurantes favoritos de Seattle. El comedor tiene una hermosa vista del Puget Sound pero sorprendentemente no estaba abarrotado. Pedí lo frito aceitunas castelvetrano, una pizza picante al estilo napolitano y un plato de ñoquis. Las aceitunas estaban absolutamente decadentes, y la pizza venía con mozzarella ahumada, que estaba delicioso. Los ñoquis estaban frescos y maravillosos. Los camareros estuvieron atentos y, en general, la experiencia estuvo precioso. Espero volver pronto».

Dígame la opinión del restaurante para criticar y clasifíquelo como uno de los siguientes:

Positivo
 Negativo
 Neutro

Salida:
 Positivo.

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text)

El siguiente ejemplo utiliza modelos antrópicos de Claude para clasificar el texto. Como se sugiere en las [guías de AnthropicClaude](#), utilice etiquetas XML <text></text> para indicar partes importantes del

rápido. Pedir al modelo que genere directamente la salida incluida en etiquetas XML también puede ayudar al modelo a producir las respuestas deseadas.

Plantilla de solicitud para Anthropic Claude:

«"»

Humano: {{descripción de la tarea de clasificación}}

<text>

{{introduzca el contenido del texto que se va a clasificar d}}

</text>

Las categorías son:

{{nombre de categoría 1}}

{{nombre de categoría 2}}

{{category name 3}} Asistente:

«»

(Fuente: Escrito por AWS)

Solicitud de usuario:

Humano: clasifique la descripción del producto dada en categorías. Introduzca la etiqueta de la categoría en <output></output>etiquetas.

Esta es la descripción del producto.

<text>

Seguro, fabricado con materiales aptos para niños con bordes lisos.

Gran cantidad, en total 112 piezas con 15 formas diferentes,

que se puede utilizar para construir 56 estructuras predefinidas diferentes.

Mejore la creatividad, se pueden conectar diferentes estructuras para formar nuevas estructuras, fomentar una forma de pensar innovadora.

Mejore el vínculo entre padres e hijos, los padres pueden jugar con sus niños juntos para fomentar las habilidades sociales.

</text>

Las categorías son:

(1) Juguetes

(2) Belleza y salud

(3) Asistente

de electrónica:

Salida:

```
<output>Juguetes</output>
```

(Fuente del mensaje: AWS, modelo utilizado: Anthropic Claude)

Pregunta-respuesta, sin contexto

En una pregunta y respuesta sin contexto, el modelo debe responder a la pregunta con sus conocimientos internos sin utilizar ningún contexto o documento.

Plantilla de solicitud a Titán

«" "{Pregunta}" "»

Solicitud de usuario:

¿Qué es «Pasar por el bosque en una noche de nieve», ¿metafóricamente?

Salida:

El poema de Robert Frost, «Pasar por el bosque en una noche de nieve», utiliza el bosque como metáfora de la muerte.

El bosque es un lugar cómodo y la paz, pero también representan lo desconocido y el más allá. El poema sugiere que la muerte no es algo a lo que deba temerse, sino más bien una parte natural de la vida que todos debemos abrazar.

El poema comienza con el orador deteniéndose en el bosque una tarde de nieve, un momento de contemplación y reflexión. El orador se da cuenta de la belleza del bosque cubierto de nieve y recuerda la naturaleza fugaz de la vida. El poema continúa mientras el orador considera la tentación de continuar
g

*a su destino, pero en vez de eso
opta por hacer una pausa y
disfrutar de la belleza
del bosque.*

*El bosque se utiliza como metáfora
de lo desconocido y el más allá.
El orador se pregunta qué hay más
allá de los árboles y si hay
es cualquier cosa después de la muerte.
El poema sugiere que la muerte no es
algo que temer, sino más bien una parte
natural de la vida
que todos debemos abrazar.*

*Las últimas líneas del poema son quizás
las más famosas y conmovedoras:
«Y millas que recorrer antes de dormir y
millas que recorrer antes de dormir».
Estas líneas sugieren que, aunque el
altavoz haya hecho una pausa
para apreciar la belleza del bosque,
todavía tienen un
un largo viaje por delante. El poema es
un recordatorio de que la vida
es fugaz y eso debemos tomarnos el
tiempo para apreciarlo
la belleza que nos rodea, incluso cuando
nos esforzamos por alcanzar nuestras
metas y ambiciones.*

*En general, «Pasar por el bosque en una
tarde de nieve» es una hermosa
y un poema que invita a la reflexión
que utiliza la metáfora del bosque
para explorar el concepto de la muerte y
el más allá.*

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan

Text) El fomento de modelos también puede ayudar en las

tareas de preguntas y respuestas

Plantilla de solicitud a Titán

«" "{{Model de estímulo}}
{{Pregunta}} »»

Solicitud de usuario:

*Es excelente respondiendo a las preguntas,
y le hace feliz cuando da la respuesta correcta.*

¿Qué luna del sistema solar tiene más probabilidades de albergar vida?

Salida:

*La luna de Titán, que es una luna del planeta Saturno,
es lo que tiene más probabilidades de sustentar vida en el sistema solar.*

Titán es la única luna conocida del Sistema Solar que tiene una atmósfera densa y una hidrosfera líquida, y los científicos creen que Puede que contenga los componentes químicos de la vida.

La superficie de la luna está cubierta de ríos, lagos y mares de metano y etano líquidos, y llueve metano en la superficie de Titán.

La atmósfera de Titán también es rica en compuestos orgánicos, incluidos benceno, tolueno y xileno, que se encuentran en los aceites de los organismos vivos de la Tierra.

Además, Titán tiene un campo magnético y una atmósfera tenue, lo que podría proteger contra las radiaciones nocivas y ayudar a mantener un clima estable.

Si bien no hay pruebas directas de vida en Titán, los científicos creen que es uno de los más prometedores

lugares del sistema solar para buscar vida más allá de la Tierra.

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text)

<p>Plantilla de envío a Titán</p> <p>«" <i>{{Model de estímulo}}</i> <i>{{Pregunta}}</i> <i>{{Restricciones}}</i> «»</p>	<p>Solicitud de usuario:</p> <p><i>Se siente recompensado al ayudar a la gente a aprender más sobre el cambio climático.</i></p> <p><i>¿Podría explicar qué es el cambio climático?</i></p> <p><i>Suponga que su audiencia está</i></p> <p>Salida:</p> <p><i>El cambio climático es el calentamiento prolongado del planeta, que puede ser causada por actividades humanas, como la quema de combustibles fósiles y la deforestación, así como los procesos naturales. Puede llevar a aumento del nivel del mar, fenómenos meteorológicos más extremos y cambios en temperatura y precipitaciones</i></p>
---	---

(Fuente del mensaje: AWS, modelo utilizado: AI21 Labs Jurassic-2 Ultra v1)

Pregunta y respuesta, con contexto

En una pregunta y respuesta con contexto, el usuario proporciona un texto de entrada seguido de una pregunta y el modelo debe responder a la pregunta en función de la información proporcionada en el texto de entrada. Poner la pregunta al final después del texto puede ayudar a los LLM de Amazon Bedrock a responder mejor a la pregunta. La promoción de modelos también funciona para este caso de uso.

Plantilla de envío a	Solicitud de
Pregunta y respuesta, con	

«" "{Texto}}
 {{Pregunta}} «»

El panda rojo (Ailurus fulgens), también conocido como panda menor, es un pequeño mamífero originario del Himalaya oriental y el suroeste de China.

Tiene un pelaje denso de color marrón rojizo con el vientre y las piernas negros, orejas forradas de blanco, un hocico casi blanco y una cola anillada. Su longitud de la cabeza al cuerpo es de 51–63,5 cm (20,1 a 25,0 pulgadas) con una cola de 28 a 48,5 cm (11,0 a 19,1 pulgadas) y pesa entre 3,2 y 15 kg (7,1 y 33,1 libras). Se adapta bien a la escalada gracias a sus articulaciones flexibles y sus garras semirretráctiles curvas.

El panda rojo se describió formalmente por primera vez en 1825. Las dos subespecies reconocidas actualmente, el panda rojo del Himalaya y el panda rojo chino, genéticamente mosca divergió hace unos 250 000 años. El lugar del panda rojo en el arte de la evolución se ha debatido sobre el árbol, pero las pruebas genéticas modernas lo sitúan en estrecha afinidad con los mapaches, las comadrejas y los zorrillos. No está estrechamente relacionado al panda gigante, que es un oso, aunque ambos poseen una forma alargada huesos de la muñeca o «pulgares postizos» utilizados para agarrar bambú.

El linaje evolutivo del panda rojo (Ailuridae) se extiende hace unos 25 o 18 millones de años, como lo indica extinto parientes fósiles encontrados en

*Eurasia y
Norteamérica.
a.*

*El
panda
rojo
habit
a en
los
bosqu
es de
conífe
ras,
así
como
en
las
zonas
templ
adas
latif
oliad
as*

y bosques mixtos, que prefieren las pendientes pronunciadas con una densa capa de bambú cerca de las fuentes de agua. Es solitario y en gran parte arbóreo. Se alimenta principalmente en los brotes y las hojas de bambú, pero también en las frutas y las flores.

Los pandas rojos se aparean a principios de la primavera y las hembras dan a luz a camadas de hasta cuatro cachorros en verano. Está amenazado mediante la caza furtiva y la destrucción y fragmentación del hábitat debido a la deforestación. La especie ha sido catalogada como En peligro de extinción en la Lista Roja de la UICN desde 2015.

Salida:

Los pandas rojos están estrechamente relacionados con

(Fuente del mensaje: https://en.wikipedia.org/wiki/Red_panda, modelo utilizado: Amazon Titan Text)

Al solicitar modelos de Claude antrópico, es útil envolver el texto de entrada en etiquetas XML. En el ejemplo siguiente, el texto introducido está entre `<text></text>`.

Plantilla de solicitud para Anthropic Claude: Solicitud de

«"»

Humano: {{Instrucción}}

<text>

{{Texto}}

<text>

{{Pregunta}}

Humano: Lea el siguiente texto en su interior

<text></text>

Etiquetas XML y, a continuación, responda a la pregunta:

Asistente: «» El 12 de

noviembre de 2020, la selección de The Weeknd como cabeza de cartel se anunció el espectáculo; fue la primera vez que un artista solista canadiense encabezó el entretiempo de la Super Bowl. Cuando se le preguntó acerca de los preparativos para el programa, The Weeknd declaró: «Nos hemos centrado mucho sobre llamar a los fans en casa y hacer actuaciones una experiencia cinematográfica y queremos hacerlo con la Super Bowl».

La actuación contó con un coro cuyos miembros vestían de blanco y llevaban máscaras en la cara con ojos rojos brillantes, y estaban de pie con un paisaje urbano neón como telón de fondo. La actuación comenzó con una figura blanca vestida igual que el coro bajado al telón de fondo donde estaba el coro mientras cantaba «Call Out My Name». En ese momento, el Weeknd estaba en un descapotable con un horizonte como telón de fondo diseñado para parecerse al Strip de Las Vegas.

Para la siguiente parte de la actuación, el telón de fondo se abre para revelar a The Weeknd, que luego interpretó «Starboy», seguida de «The Hills».

Luego, al interpretar la canción «Can't Feel My Face», The Weeknd viajó a través de un laberinto construido detrás del escenario, al que se unieron bailarines vestido con chaquetas rojas y corbatas negras parecidas a las suyas, pero con la cara cubierta con vendajes, de acuerdo con la estética de su cuarto álbum de

estudio After Hours (2020).

Los bailarines llevaban estos vendajes durante toda la actuación. En la sección laberíntica de la actuación, el trabajo de cámara era visualmente inestable. Las siguientes canciones que se interpretaron fueron «I Feel It Coming», «Save Your Tears», y «Me lo merezco». Para la actuación de «Earned It», The Weeknd estuvo acompañado de violinistas. Para el final de la serie, the Weeknd salió al campo del estadio con sus bailarines para interpretar «Blinding Lights». Los bailarines y él entraron al campo interpretando «House of Balloons/Glass Table Girls». La actuación terminó con una serie de fuegos artificiales.

</text>

Según el texto de arriba, ¿qué canciones ponía The Weeknd? ¿en el entretiempo de la Super Bowl?

Asistente:

Salida:

Basado en el texto, las canciones que tocó The Weeknd en el entretiempo de la Super Bowl estuvieron:

- Diga mi nombre*
- Starboy*
- Las colinas*
- No puedo sentir mi cara*
- Lo siento venir*
- Ahórrese las lágrimas*
- Se lo he ganado*
- Luces cegadoras*
- House of Balloons/Glass Table Girls*

Anthropic Claude)

Resumen

Para una tarea de resumen, la solicitud es un pasaje de texto y la modelo debe responder con un pasaje más corto que capture los puntos principales de la entrada. Especificar la salida en términos de longitud (número de frases o párrafos) es útil para este caso de uso.

Plantilla de solicitud para Titán

«" "El siguiente es texto de una
{{Categoría de texto}}:
{{Texto}}
Resume la {{Categoría de texto}} en
{{longitud del resumen,
por ejemplo, «una frase» o «un
párrafo»}} «»

Solicitud de usuario:

El siguiente es un texto de un
restaurante a una reseña:
«Por fin tengo que echar un vistazo a
Brilliant Pizza de Alessandro O
y ahora es uno de mis restaurantes
favoritos de Seattle.
El comedor tiene una hermosa vista del
Puget Sound
pero sorprendentemente no estaba
abarroado. Pedí lo frito
aceitunas castelvetrano, una
pizza picante al estilo
napolitano
y un plato de ñoquis. Las aceitunas
estaban absolutamente decadentes,
y la pizza venía con mozzarella ahumada,
que estaba deliciosa.
Los ñoquis estaban frescos y
maravillosos. Los camareros
estuvieron atentos,
y, en general, la experiencia fue
estupenda. Espero volver pronto».
Resume la reseña anterior sobre el
restaurante en una frase.

Salida:

Alessandro's Brilliant Pizza es un
restaurante fantástico
en Seattle con una hermosa vista de
Puget Sound que ofrece
comida deliciosa y decadente.

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text)

En el siguiente ejemplo, Anthropic Claude resume el texto dado en una frase. Para incluir el texto introducido en las indicaciones, dé formato al texto con un marcado XML: `{{text content}}` `<text>` `</text>`. El uso de XML en las indicaciones es una práctica común cuando se solicitan modelos de Claude antrópico.

Plantilla de solicitud para Anthropic Claude:

«"»

Humano: Lea el texto:

`<text>`

`{{contexto del texto}}`

`</text>`

Resume el texto en {{longitud del resumen, por ejemplo, «una frase» o «un párrafo»}} Asistente: «»

Solicitud de usuario:

Humano: Lea el texto:

`<text>`

En la teoría de juegos, el equilibrio de Nash, que lleva el nombre del matemático

John Nash, es la forma más común de definir la solución

de un juego no cooperativo en el que participan dos o más jugadores.

En un equilibrio de Nash, se supone que cada jugador sabe

las estrategias de equilibrio de los demás jugadores,

y nadie tiene nada que ganar

cambiando solo su propia estrategia.

El principio del equilibrio de Nash

se remonta a la época de Cournot,

quien en 1838 lo aplicó a las

empresas competidoras que elegían

sus productos.

`</text>`

Resume el texto en una frase.

Asistente:

Salida:

El equilibrio de Nash es un concepto de la teoría de juegos

donde cada jugador ha adoptado una estrategia que no tiene ningún

incentivo

para desviarse de las estrategias de los otros jugadores.

(Fuente del mensaje: [Wikipedia sobre el equilibrio de Nash](#), modelo utilizado: Claude antrópico)

Generación de texto

Si se le pide, los LLM de Amazon Bedrock pueden responder con un pasaje del texto original que coincida con la descripción. He aquí un ejemplo:

Plantilla de solicitud para Titán

«" "Escriba una {{Categoría de texto}} con la voz de {{Role}} ." "»

Solicitud de usuario:

Por favor, escriba un correo electrónico con la voz de un amigo felicitar a alguien por un nuevo trabajo.

Salida:

Hola [nombre],

Acabo de enterarme de su nuevo trabajo y ¡me alegro mucho por usted! ¡Enhorabuena! Se merece esta oportunidad, ya que se ha esforzado mucho para conseguirla. Sé que lo hará muy bien y tendrá un impacto positivo en su equipo. Estoy aquí para apoyarlo en cada paso del camino. Le deseo lo mejor ¡en este nuevo y emocionante capítulo de su vida!

*Saludos,
[su nombre]*

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text)

Para los casos de uso de la generación de texto, puede funcionar bien especificar los requisitos detallados de la tarea. En el siguiente ejemplo, pedimos al modelo que genere una respuesta con signos de exclamación.

Plantilla de solicitud para Titán

«" "{Especificación de la tarea}}
 Escriba una {{Categoría de texto}}
 con la voz de {{Role}} ."»

Solicitud de usuario:

*Escriba texto con signos de
 exclamación. Escriba un correo
 electrónico formateado con la voz de un
 amigo
 felicitar a alguien por un nuevo trabajo.*

Salida:

¡Hola [nombre]!

*Acabo de ver las noticias y ¡me alegro
 mucho por usted!
 Se merece este puesto y esta empresa de
 todo corazón.
 Me muero de ganas de verlo volar y
 lograr todas las cosas
 por lo que ha estado trabajando. Si
 alguien se lo merece
 de un ascenso, ¡es usted!*

¡Enhorabuena de nuevo, amigo!

*Con
 amor,
 [nombr
 e]*

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text)

En el ejemplo siguiente, un usuario pide a la modelo que asuma el papel de una persona específica al generar el texto. Observe cómo la firma refleja el papel que asume la modelo en la respuesta.

Plantilla de solicitud para Titan

«" "{Asunción del rol}}
 {{Descripción de la tarea}}
 «"»

Entrada de usuario:

*Me llamo Jack. Ayúdeme a escribir una nota
 expresando mi gratitud
 a mis padres por llevar a mi hijo (su
 nieto) al zoológico.
 Extraño mucho a mis padres.*

Salida:

Queridos mamá y papá:

Quería dedicar un momento a expresar mi más sincera gratitud por llevar a mi hijo al zoológico el fin de semana pasado. Significó mucho para mí que se haya tomado el tiempo de crear recuerdos con él, y le estoy muy agradecido por el amor y el apoyo que le ha demostrado.

Los echo mucho de menos a los dos y me gustaría que pudiéramos estar juntos más a menudo.

Pero sé que siempre estará ahí para mí y mi familia, y estoy muy agradecido de tenerlos como padres.

(Fuente del mensaje: AWS, modelo utilizado: AI21 Labs Jurassic-2 Ultra v1)

Generación de código

La solicitud describe la tarea o función y el lenguaje de programación del código que el usuario espera que genere el modelo.

Plantilla de solicitud para Titán

«"Escriba una función en {{lenguaje de programación}} para {{tarea o propósito}} «».

Entrada de usuario:

Escriba una función de Python para descargar un archivo s3 al disco local.

Salida:

Se puede utilizar el siguiente código para descargar un archivo s3 al disco local.

```Python


```

importar boto3
# Crea un objeto de cliente
boto3 s3 = boto3.client ('s3')
def download_file (nombre_de_depósito,
    clave_objeto, ruta del archivo):
    intente:
    # Descarga el archivo s3 en la ruta
    de archivo especificada
    s3.download_file (bucket_name,
    object_key, file_path)
    excepto la excepción
    e: print (e)
    # Especifica los valores de bucket_name,
    object_key y file_path
    bucket_name = «samplebucketname»
    object_key = «sampleobjectkey»
    file_path = «/samplefilepath»
    download_file (nombre_depósito,
    clave_objeto y, file_path)
    ``

```

El código dado utiliza la función `download_file` para descargar un archivo `s3` al disco local. La función `download_file` toma el nombre del depósito, la clave del objeto y la ruta del archivo como parámetros de entrada. El nombre del balde es el nombre del depósito de S3 en el que se encuentra el archivo, la clave de objeto es el identificador único del archivo dentro del depósito, y la ruta del archivo es la ruta de destino del disco local donde se descargará el archivo. La función utiliza el método `download_file` del cliente `s3` para descargar el archivo del depósito especificado y escríbelo en la ruta de archivo especificada.

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text)

Plantilla de solicitud para Anthropic Claude

«"»

Humano: escriba una función en {{program ming language}} para {{tarea o propósito}}

Asistente: «»

Solicitud de usuario:

Humano: escriba un código de Python que se aproxime rápidamente la raíz cuadrada de un número.

Asistente:

Salida:

He aquí una función sencilla para aproximar el la raíz cuadrada de un número mediante el método de Newton:

```
``pitón
def sqrt (n, iteraciones =
10): x = n
para i en el rango
(iteraciones): x = (x +
n/x)/2
devolver x
```
```

*Para usarlo:*

```
``python print
(sqrt (25)) #
5.0
```

```
print (sqrt (15, iteraciones =
20)) # 3.872983346207417
```
```

Esto necesita repetidamente la media de x y n/x para converger en la raíz cuadrada.

Más iteraciones producen una aproximación más precisa.

La ventaja de este enfoque es que converge muy rápido, así que con solo 10 iteraciones se obtiene una estimación decente.

(Fuente del mensaje: AWS, modelo utilizado: Anthropic Claude)

Matemáticas

La entrada describe un problema que requiere un razonamiento matemático en algún nivel, que puede ser numérico, lógico, geométrico o de otro tipo. Para ese problema, es útil pedirle a la modelo que resolver el problema de manera fragmentaria añadiendo frases a instrucciones como Pensemos paso a paso o Pensemos paso a paso para encontrar la respuesta correcta.

<p>Plantilla de solicitud para Titán «" <i>"{{Math problem description}} Pensemos paso a paso."</i>»</p>	<p>Solicitud de usuario: <i>Un triángulo tiene dos ángulos de 70 grados y 50 grados. ¿Cuál es el tercer ángulo en grados? Pensemos paso a paso.</i></p> <p>Salida: <i>La suma de los tres ángulos de un triángulo es siempre de 180 grados. Los dos ángulos dados son 70 grados y 50 grados. $70 + 50 = 120$. Por lo tanto, el tercer ángulo es de 180 grados - $120 = 60$ grados. El tercer ángulo es de 60 grados.</i></p>
---	--

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text)

El siguiente ejemplo también muestra cómo se le pide a la modelo que piense paso a paso:

<p>Plantilla de solicitud para Titán</p>	<p>Solicitud de</p>
---	----------------------------

Piénselo paso a paso para encontrar la respuesta correcta.» «»

*Lucy tiene 12 canicas de colores y quiere compartirlas igual que sus 4 amigos. ¿Cuántas canicas recibirá cada amigo?
Piénselo paso a paso para encontrar la respuesta correcta.*

Salida:

- 1. Lucy tiene 12 canicas y quiere compartirlas a partes iguales con sus 4 amigas, cada amiga recibirá $12/4 = 3$ canicas.*
- 2. Por lo tanto, cada amigo recibirá 3 canicas.*

(Fuente del mensaje: AWS, modelo utilizado: AI21 Labs Jurassic-2 Ultra v1)

Razonamiento/pensamiento lógico

Para tareas o problemas de razonamiento complejos que requieren un pensamiento lógico, podemos pedirle al modelo que haga deducciones lógicas y explique sus respuestas.

Plantilla de solicitud para Titan y AI21 Labs Jurassic: «" "Pregunta: {{question n}} Introduzca la respuesta y, a continuación, explíquela: «"»

Mensaje de usuario: *Pregunta: ¿Qué palabra es la impar? A. logrado B. bueno C. horrible D. Sobresaliente Por favor, muestre la respuesta y, a continuación, explique su respuesta:*

Resultado: *El que sale es C. horrible. Las otras tres palabras describen cualidades o logros positivos, mientras que horrible es un adjetivo negativo.*

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text)

He aquí otro ejemplo en el que se utiliza el modelo de Claude antrópico:

Plantilla de aviso para Anthropic

Claude: «"» *Humano:*
 {{question}}
 Por favor, dé la respuesta y muestre
 el razonamiento. *Asistente:* «»

Mensaje de usuario *Humano:* El
 barbero es el «que se afeita a todos
 los que, y solo a los que no se
 afeitan».
 ¿El barbero se afeita? ¿Por qué es
 una paradoja? Por favor, dé la
 respuesta y muestre el
 razonamiento. *Asistente:*

Resultado: Este es un ejemplo de una
 paradoja conocida como paradoja de
 Barber. *Aquí*
 es una explicación: - La
 declaración define a un barbero que
 afeita a todas y solo a las
 personas que no se afeitan.
 - Así que si el
 barbero se afeita, debe afeitarse a
 todos los que no se afeitan
 s, que lo incluye a sí mismo. Pero
 el barbero no puede afeitarse
 según
 la afirmación inicial. - Si el
 barbero no se afeita, entonces
 encajaría en el grupo de
 personas que no se afeitan, y
 así que debe afeitarse, según el
 comunicado. Pero, de nuevo, esto
 contradice la afirmación inicial. -
 Así que el barbero no puede afeitarse
 ni tampoco puede afeitarse. Esto
 es contradictorio en sí mismo según la
 afirmación inicial sobre el barbero.
 - Por lo tanto, la afirmación
 lleva a una paradoja en la que
 ambas posibilidades (que el
 barbero se afeite o no se afeite)
 dan lugar a una contradicción. En
 resumen, el
 la paradoja surge porque la definición

*del comportamiento del barbero
es autosuficiente y radical cuando se
aplica al propio barbero. Esto
hace que sea imposible determinar si
el barbero se afeita*

o no basándose únicamente en la declaración dada.

(Fuente del mensaje: https://en.wikipedia.org/wiki/Barber_paradox, modelo utilizado: Anthropic Claude)

Extracción de entidades

Para extraer entidades de una pregunta de entrada proporcionada. Extraiga las entidades del texto generado y colóquelas en etiquetas XML para su posterior procesamiento.

Plantilla de solicitud para Titán

«" "Es un experto extractor de entidades a partir de la pregunta de entrada proporcionada. Usted es responsable de extraer las siguientes entidades: {{list of entities}}

Siga las instrucciones que aparecen a continuación para extraer la entidad A y responda en
<entityA></entityA>Etiquetas XML:
{{instrucciones de extracción de la entidad A}}

Siga las instrucciones que aparecen a continuación para extraer la entidad B y responda en
<entityB></entityB>Etiquetas XML:
{{instrucciones de extracción de la entidad B}}

A continuación se muestran algunos ejemplos:

{{algunos ejemplos fotográficos que muestran a un modelo extrayendo

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text G1- Premier)

Ejemplo:

Usuario: Es un experto en extractores de entidades que extrae entidades de la pregunta de entrada proporcionada.

Usted es responsable de extraer las siguientes entidades: nombre, ubicación

Siga las instrucciones siguientes al extraer el nombre y la respuesta en las etiquetas

<name></name>XML:

- Estas entidades incluyen el nombre específico de una persona, animal o cosa
- Extraiga únicamente los nombres específicos de las entidades mencionadas en la consulta de entrada
- NO extraiga la mención general del nombre mediante términos de «nombre», «niño», «niña», «nombre del animal», etc.

Siga las instrucciones siguientes al extraer la ubicación y responda en etiquetas<location></location> XML:

- Estas entidades incluyen una ubicación específica de un lugar, ciudad, país o pueblo
- Extraiga únicamente las entidades de nombre y ubicación específicas mencionadas en la consulta de entrada
- NO extraiga la mención general de la ubicación mediante términos de «ubicación», «ciudad», «país», «ciudad», etc.

Si no encuentra ningún nombre o ubicación, devuelva la misma cadena de

entrada que está. A continuación se muestran algunos ejemplos:

entrada: ¿Qué tal la fiesta de cumpleaños de Sarah en
Seattle, Washington? salida: ¿Qué tal la
fiesta <name>de cumpleaños de Sarah</name> en
<location>Seattle, Washington</location>?

entrada: ¿Por qué el padre de Joe fue a la ciudad?
salida: ¿Por qué <name>el padre de Joe</name> fue a la ciudad?

entrada: ¿Cuál es el código postal de Manhattan, Nueva York?
<location>salida: ¿Cuál es el código postal de <location>Manhattan, Nueva York?

entrada: ¿Quién es el alcalde de San
Francisco? Robot:

Razonamiento en cadena de pensamiento

Haga un análisis paso a paso de cómo se obtuvo la respuesta. Compruebe los datos y valide la forma en que el modelo produjo una respuesta.

Plantilla de solicitud para Titán

```
«"» {{Pregunta}}
{{Instrucciones a seguir}}
Piense paso a paso y explíqueme su
  forma de pensar
«"»
```

(Fuente del mensaje: AWS, modelo utilizado: Amazon Titan Text G1- Premier)

Ejemplo:

Usuario: Si Jeff tuviera 100 dólares y le diera 20 dólares a Sarah, y compró billetes de lotería con otros 20 dólares. Con los billetes de lotería que compró ganó 35 dólares. Luego, Jeff fue a comprar su comida y se gastó 40 dólares en comida. Por último, hizo una donación a una organización benéfica de 20 dólares. Stephen se reunió con Jeff y quería para prestarle algo de dinero para su taxi. ¿Cuánto dinero máximo puede dar Jeff a Stephen, dado que necesita ahorrar 10 dólares para el viaje de regreso a casa? Por favor, no responda de inmediato, piénselo paso a paso y muéstreme lo que piensa.

Robot: