



De Cero a Agente Conversacional: **Cómo Construir un Chatbot Inteligente**

International Women's Day 2025



Laura Orozco

Data Scientist,
Nowports



Table of contents

- 01 Que es un agente
- 02 Casos de uso
- 03 Gemini
- 04 Parametros
- 05 Vertex AI
- 06 Prompt

Table of contents

07	Demo
08	Datos Taxi
09	Bases de datos para agentes
10	Disponibilización
11	Recursos



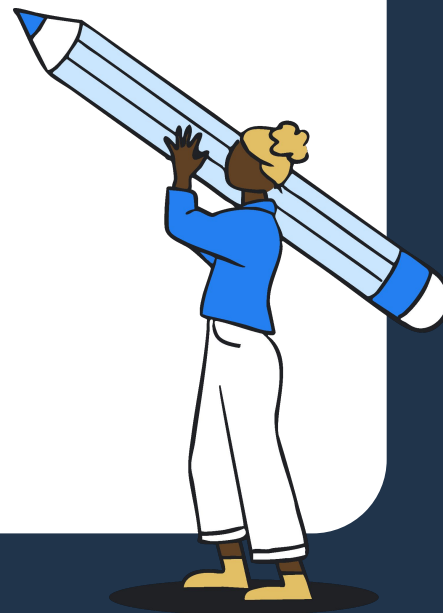
Los **agentes** conversacionales no son sólo máquinas que responden preguntas, sino **sistemas inteligentes** que transforman la manera en que **interactuamos** con el conocimiento y la información.”

Ray Kurzweil

Director de Ingeniería, Google

[W] Women Techmakers

01 Que es un agente?



Agentes

Definición: Un agente de IA es un sistema que interactúa con usuarios en lenguaje natural, ejecuta acciones y toma decisiones basadas en datos.

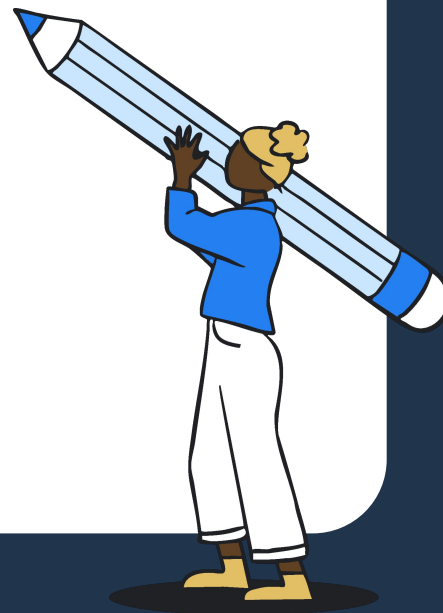
Características:

- Comprensión del lenguaje natural
- Integración con bases de datos y APIs
- Capacidad de aprendizaje y mejora





01 Casos de uso



Casos de uso

Programación y Desarrollo de Software

Los modelos de Gemini pueden asistir en tareas de codificación y depuración de código.

Ejemplos:

- Generación de código en múltiples lenguajes
- Explicación de código fuente
- Optimización de queries SQL

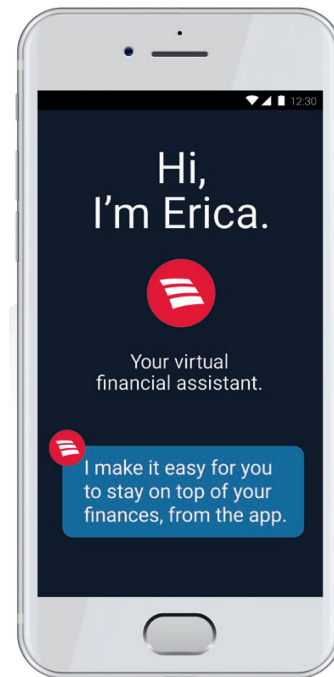


Casos de uso

Los chatbots y agentes de voz permiten ofrecer asistencia automatizada, reduciendo tiempos de espera y costos operativos.

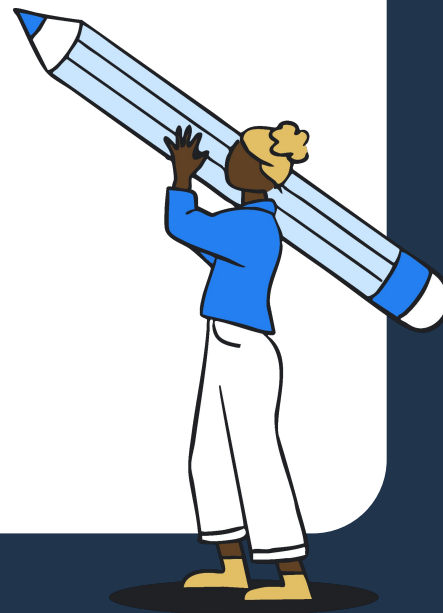
Ejemplos:

- Respuestas a preguntas frecuentes (FAQs)
- Soporte técnico automatizado con troubleshooting
- Seguimiento de pedidos y envíos
- Registro y actualización de tickets en CRM
- Ejemplo real: El bot de atención de Bank of America ("Erica"), que ayuda a los clientes con pagos, transacciones y consultas financieras.



[W] Women Techmakers

01 Gemini



Que es Gemini?

Descripción:

- Gemini es la familia de modelos de IA de Google que puede entender texto, imágenes, código y más.
- Gemini 1.5 Pro es ideal para agentes porque tiene memoria de contexto de largo plazo y razonamiento avanzado.



Modelos de razonamiento

Gemini ofrece múltiples variantes optimizadas para diferentes necesidades, desde generación rápida hasta razonamiento avanzado y multimodalidad. Estos modelos pueden procesar texto, imágenes, audio y video, con capacidades avanzadas de pensamiento y generación.

Modelo	Salida	Optimizado para
Gemini 2.5 Pro Experimental (gemini-2.5-pro-exp-03-25)	Texto	🧠 Pensamiento y razonamiento mejorados, comprensión multimodal y programación avanzada
Gemini 2.0 Flash (gemini-2.0-flash)	Texto, imágenes (exp.), audio (próx.)	⚡ Velocidad, transmisión en tiempo real y generación multimodal
Gemini 2.0 Flash-Lite (gemini-2.0-flash-lite)	Texto	💰 Rentabilidad y baja latencia

[Documentacion Google AI Developers/Modelos/](#)

Otros modelos



OpenAI
GPT-4o

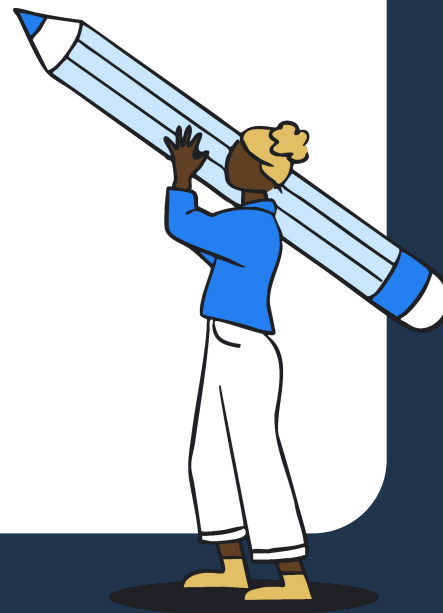
ANTHROPIC
Claude 3.5 Sonnet



 **LLaMA** 

[W] Women Techmakers

01 Parametros



Parametros

Los parámetros afectan **precisión**, **creatividad** y **fluidez** de las respuestas de Gemini. Ajustarlos correctamente según el caso de uso puede mejorar significativamente la calidad de las respuestas generadas.

Parámetro	Descripción
temperature	Controla la creatividad (0 = más preciso, 1 = más aleatorio)
top_p	Filtra palabras menos probables para mejorar coherencia
max_output_tokens	Define la longitud máxima de la respuesta
stop_sequences	Define frases donde el modelo debe detenerse

Parametros

Temperature (0.0 - 2.0)

Qué hace: Controla la aleatoriedad de la respuesta.

Valores bajos (0.1-0.3): Respuestas más determinísticas, precisas y repetibles.

Valores altos (1.5-2.0): Mayor creatividad y variabilidad en la respuesta.

Uso recomendado:

- **Baja temperatura:** Para tareas donde la precisión es clave (código, matemáticas, respuestas factuales).
- **Alta temperatura:** Para generación de ideas, escritura creativa o brainstorming.

Parametros

Max Tokens (Cantidad de palabras generadas)

Qué hace: Define la cantidad máxima de tokens que el modelo generará en su respuesta.

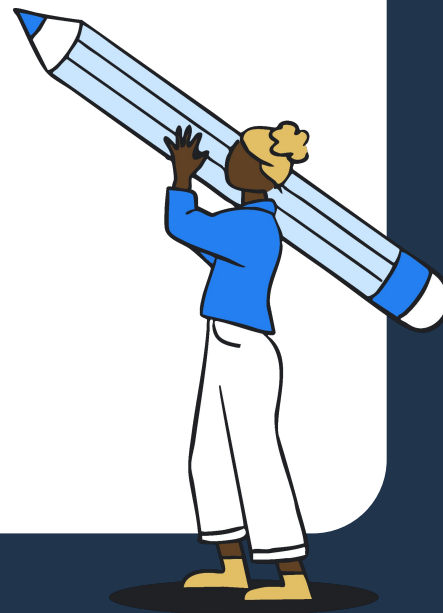
Ejemplo: Un valor bajo (50) da respuestas cortas, mientras que un valor alto (500) permite explicaciones más detalladas.

Uso recomendado:

- Valores bajos para respuestas rápidas y concisas.
- Valores altos para explicaciones más completas y detalladas.

[W] Women Techmakers

01 Prompt



Introducción a Prompt Engineering

Prompt Engineering optimiza las instrucciones dadas a modelos de IA para obtener mejores respuestas. Un buen prompt mejora la precisión, relevancia y claridad de las respuestas generadas por IA.

Estrategias Clave en Prompt Engineering

- **Uso de comillas:** Ayuda a delimitar términos clave y evitar confusiones.
- **Agregar contexto:** Especifica el entorno para respuestas más precisas.
- **Definir un rol:** Personaliza el estilo y el nivel de detalle de la respuesta.
- **Formato de salida:** Permite estructurar mejor la información generada.
- **Instrucciones claras:** Minimiza ambigüedades y mejora la coherencia.
- **Ejemplos en el prompt:** Guían al modelo para respuestas más alineadas con lo esperado.

Introducción a Prompt Engineering

Métodos Avanzados

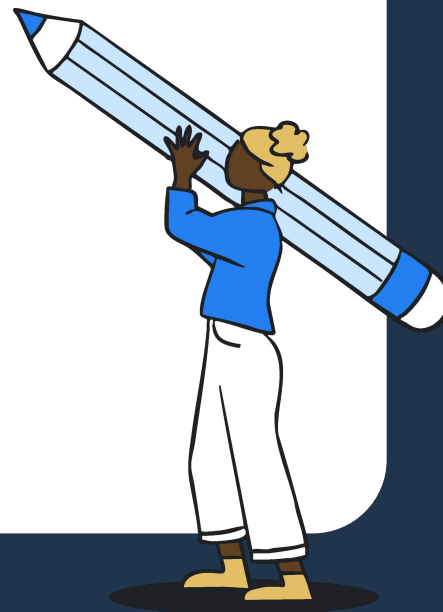
- Zero-shot prompting: Solicita una tarea sin ejemplos previos, útil para generalización.
- Few-shot prompting: Incluye ejemplos en el prompt para mejorar la precisión.
- Chain-of-thought (CoT): Divide problemas complejos en pasos intermedios para mayor claridad.

Frameworks para Crear Prompts Efectivos

- SALT (Simplicidad, Adaptación, Lógica y Tono): Optimiza la claridad y relevancia del prompt.
- FLART (Formato, Longitud, Audiencia, Relevancia, Tareas): Guía la estructura y el contenido del prompt según el objetivo.

[W] Women Techmakers

01 VertexAI



Gemini con Vertex AI SDK

```
# Usar el modelo  
model = GenerativeModel("gemini-pro")  
response = model.generate_content("que dia es hoy?")  
print(response.text)
```

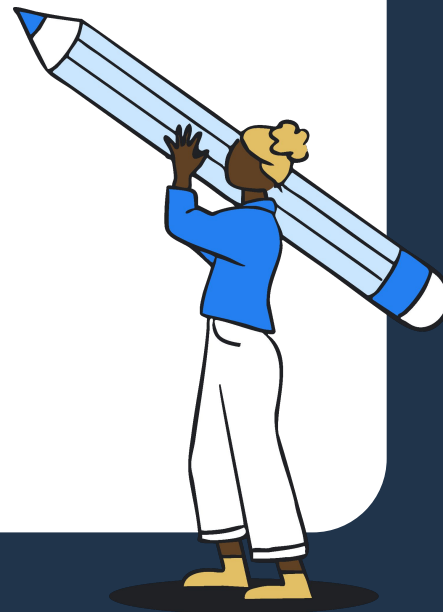
La fecha de hoy es 2023-10-26.

¿Te gustaría saber qué día de la semana es, o alguna otra información relacionada con la fecha?

[Repositorio/Notebooks](#)

[W] Women Techmakers

01 Demo



Otros modelos

Deploy

Guía de Preguntas

Ejemplos de preguntas que puedes hacer:

💰 Tarifas y Pagos:

- ¿Cuál es la tarifa promedio por viaje?
- ¿Cuánto se gana en propinas en hora pico?
- ¿Cuál es el pago total promedio por viaje?

🕒 Tiempo y Patrones:

- ¿Cuáles son las horas más ocupadas?
- ¿Qué día de la semana hay más viajes?
- ¿Cuánto duran los viajes en promedio?

Análisis de Taxis NY

¿Qué deseas saber sobre los taxis de NY?

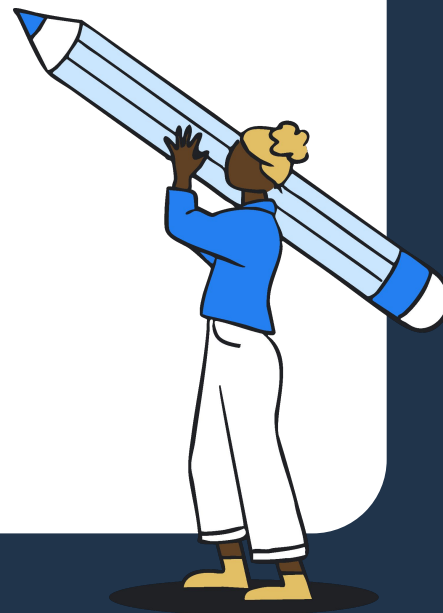
>

Controles

 Women Techmakers

[W] Women Techmakers

01 Datos Taxi



Taxis Nueva York

La base de datos pública New York Taxi Trips en Google Cloud Platform (GCP) contiene registros de viajes en taxi en la ciudad de Nueva York. Esta base de datos es mantenida por el NYC Taxi & Limousine Commission (TLC) y proporciona información detallada sobre cada viaje, incluyendo:

Estructura de los Datos

Los datos están organizados en tablas dentro del conjunto de datos `bigquery-public-data.new_york_taxi_trips`. Algunas tablas clave incluyen:

- [`yellow_taxi_trips`](#): Datos de los taxis amarillos (Yellow Taxis)
- `green_taxi_trips`: Datos de los taxis verdes (Green Taxis)
- `fhv_taxi_trips`: Datos de los For-Hire Vehicles (FHV).

Taxis Nueva York

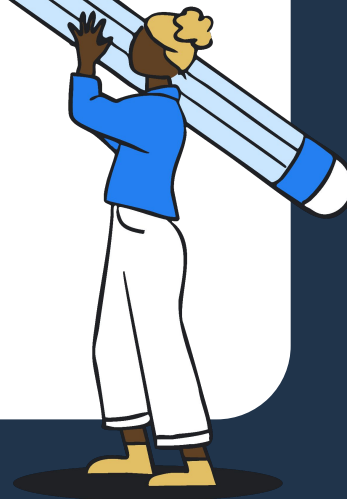
Campos Principales

Algunos de los campos más relevantes incluyen:

- *pickup_datetime, dropoff_datetime*: Tiempos de recogida y llegada.
- *pickup_location_id, dropoff_location_id*: Identificadores de las zonas de recogida y destino.
- *passenger_count*: Número de pasajeros.
- *trip_distance*: Distancia recorrida en millas.
- *fare_amount*: Monto de la tarifa base en USD.
- *tip_amount*: Monto de la propina en USD.
- *total_amount*: Monto total pagado (incluye peajes y cargos adicionales).
- *payment_type*: Método de pago (efectivo, tarjeta, etc.).



01 Bases de datos para agentes



Bases de datos vectoriales

¿Qué son las Bases de Datos Vectoriales?

Las bases de datos vectoriales almacenan y recuperan información en forma de vectores de alta dimensión. Estos vectores representan datos como texto, imágenes o audio en un espacio numérico, permitiendo búsquedas eficientes basadas en similitud.

¿Por qué se usan en Agentes IA?

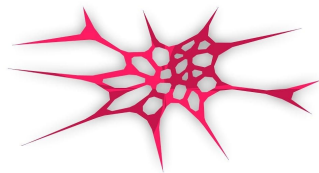
- Los agentes de IA utilizan bases de datos vectoriales para:
- Búsqueda semántica: Encuentran información relevante basada en significado, no solo en palabras clave.
- Memoria a largo plazo: Permiten que los agentes recuerden y relacionen información de manera eficiente.
- Optimización del rendimiento: Recuperan respuestas rápidas y precisas, mejorando la interacción con usuarios.

Bases de datos vectoriales



Chroma

FAISS
Scalable Search With Facebook AI



Weaviate

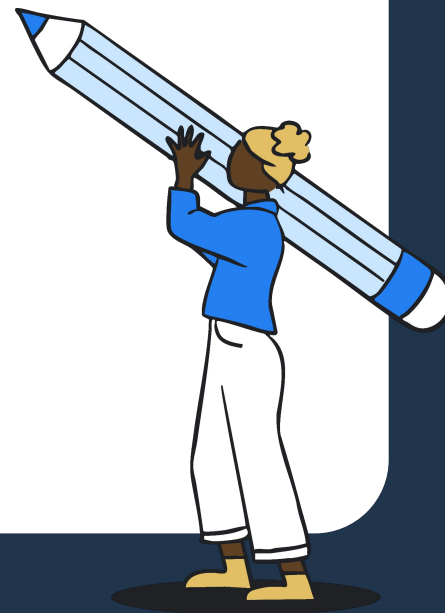


Pinecone

[Datacamp/the-top-5-vector-databases](https://www.datacamp.com/publications/the-top-5-vector-databases)



01 Disponibiliza ción

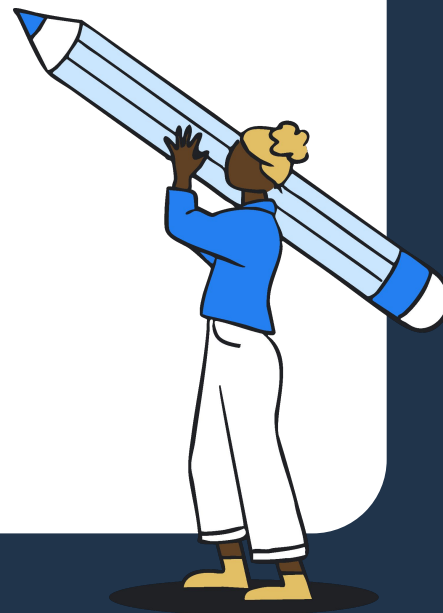


Disponibilización





01 Conversational Analytics



Conversational Analytics

¿Qué es Conversational Analytics?

Es una función que permite a los usuarios interactuar con los datos mediante lenguaje natural, sin necesidad de conocimientos técnicos en SQL.

¿Cómo crear un chat con los datos?

- Conectar Looker Studio a BigQuery para acceder a las fuentes de datos.
- Habilitar la funcionalidad de preguntas en lenguaje natural para generar análisis automáticos.
- Utilizar filtros y segmentaciones para personalizar las respuestas

Conversational Analytics

¿Cómo crear un agente más específico?

- Definir el nombre y la descripción del agente.
- Conectar el agente a las fuentes de datos en BigQuery u otras bases compatibles.
- Especificar instrucciones para guiar las respuestas del agente.
- Publicar y comenzar a interactuar con los datos mediante preguntas en lenguaje natural.

Limitaciones de Looker Studio Pro

- La funcionalidad de agentes conversacionales solo está disponible en la versión Pro.
- Requiere configuración específica y permisos adecuados.
- Puede haber restricciones en el acceso a ciertos conjuntos de datos dependiendo de los permisos.

Conversational Analytics

Chat with your data

Select data to begin

Agents

Data sources

🔍 Search

🔗 Users - The Look

📁 Word analysis

📄 Initial audit: product labels

Don't see what you're looking for? [Connect to data](#)

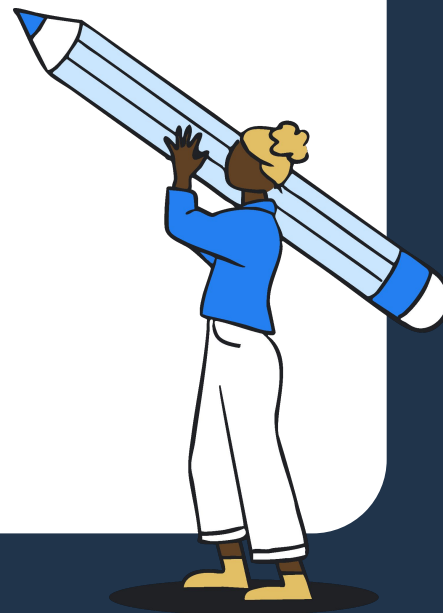
🔗 Looker

📄 Google Sheets

📁 File Upload

[W] Women Techmakers

01 Recursos



Recursos

- [Repositorio Chat Conversacional con VertexAI](#)
- [Develop GenAI Apps with Gemini and Streamlit](#)
- [Create Generative AI Apps on Google Cloud](#)



Thank you!

