

UNIVERSIDAD DEL VALLE DE GUATEMALA

Security Data Science

Sección 10

Jorge Yass



Laboratorio 2

Curso intensivo de Gen AI

Abner Iván García Alegría 21285

Requisitos Antes de iniciar

Siga los siguientes pasos para prepararse antes de comenzar con las tareas diarias:

- ☐1. Regístrate para obtener una [cuenta de Kaggle](#) y aprende cómo funcionan los cuadernos . Asegúrate de [verificar tu cuenta](#) por teléfono, ya que es necesario para los laboratorios de código del curso.
- ☐2. Regístrate para obtener una cuenta de [AI Studio](#) y asegúrese de poder generar una clave API .
- ☐3. Regístrate en una cuenta de Discord y únete a nosotros en el [servidor de Discord de Kaggle](#) . Visita [#5dgai-general-chat](#) para encontrar anuncios oficiales de cursos y grabaciones de transmisiones en vivo.

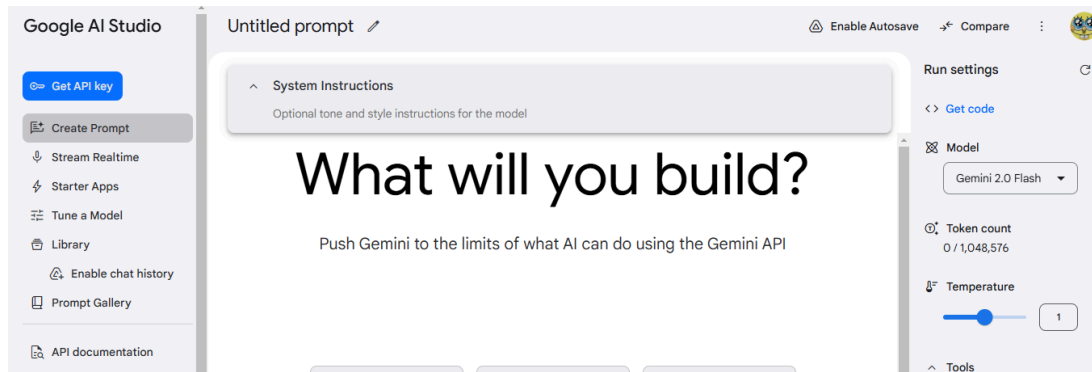
Cuenta en kaggle

- **Día 3 : Agentes de IA generativos: aprenda a crear agentes de IA sofisticados comprendiendo sus componentes principales y el proceso de desarrollo iterativo.**
- **Día 4 : LLM de dominio específico: profundice en la creación y aplicación de LLM especializados como SecLM y Med-PaLM, con conocimientos de los investigadores que los crearon.**
- **Día 5 : MLOps para IA generativa: descubra cómo adaptar las prácticas de MLOps para la IA generativa y aprovechar las herramientas de Vertex AI para modelos básicos y aplicaciones de IA generativa.**

¡Mucha suerte!

- ofrecido por Anant Nawalgaria, Mark McDonald, Paige Bailey y muchos otros colaboradores de Google.

Cuenta en AI Studio



Día 1 (Modelos lingüísticos básicos y diseño de propuestas)

Modelos lingüísticos básicos y generación de textos

Autores: Mohammadamin Barektain, Anant Nawalgaria, Daniel J. Mankowitz, Majd Al Merey, Yaniv Leviathan, Massimo Mascaro, Matan Kalman, Elena Buchatskaya, Aliaksei Severyn y Antonio Gulli

Introducción

La aparición de los modelos de lenguaje grandes (LLM, por sus siglas en inglés) representa un cambio radical en el mundo de la inteligencia artificial. Su capacidad para procesar, generar y comprender la intención del usuario está cambiando fundamentalmente la forma en que interactuamos con la información y la tecnología.

Un LLM es un sistema avanzado de inteligencia artificial que se especializa en procesar, comprender y generar texto similar al humano. Estos sistemas se implementan normalmente como una red neuronal profunda y se entrenan con cantidades masivas de datos de texto. Esto les permite aprender los patrones intrincados del lenguaje, lo que les da la capacidad de realizar una variedad de tareas, como traducción automática, generación creativa de texto, respuesta a preguntas, resumen de texto y muchas más.

Ingeniería rápida

Autor: Lee Boonstra

Introducción

Al pensar en la entrada y salida de un modelo de lenguaje grande, un mensaje de texto (a veces acompañado de otras modalidades, como mensajes de imagen) es la entrada que utiliza el modelo para predecir una salida específica. No es necesario ser un científico de datos o un ingeniero de aprendizaje automático: todos pueden escribir un mensaje. Sin embargo, elaborar el mensaje más eficaz puede ser complicado. Muchos aspectos de su mensaje afectan su eficacia: el modelo que utiliza, los datos de entrenamiento del modelo, las configuraciones del modelo, su elección de palabras, el estilo y el tono, la estructura y el contexto: todo es importante. Por lo tanto, la ingeniería de mensajes es un proceso iterativo. Los mensajes inadecuados pueden generar respuestas ambiguas e inexactas, y pueden obstaculizar la capacidad del modelo para proporcionar una salida significativa. No es necesario ser un científico de datos o un ingeniero de aprendizaje automático: todos pueden escribir un mensaje.

Cuando chateas con el chatbot Gemini, básicamente escribes indicaciones; sin embargo, este informe técnico se centra en escribir indicaciones para el modelo Gemini dentro de Vertex AI o mediante la API, porque al indicar indicaciones al modelo directamente, tendrás acceso a la configuración, como la temperatura, etc.



```

chat = flash.start_chat(history=[])
response = chat.send_message('Hello! My name is Abner.')
print(response.text)

```

Hello Abner! It's nice to meet you. How can I help you today?

Día 2 (Incorporaciones y almacenes de vectores/bases de datos)

Incrustaciones y tiendas de vectores


Autores: Anant Nawalgaria y Xiaoqi Ren

Introducción

El aprendizaje automático moderno se nutre de datos diversos: imágenes, texto, audio y más. Este informe técnico explora el poder de las incrustaciones, que transforman estos datos heterogéneos en una representación vectorial unificada para un uso sin inconvenientes en diversas aplicaciones. Lo guiaremos a través de:


- **Comprender las incrustaciones:** por qué son esenciales para manejar datos multimodales y sus diversas aplicaciones.
- **Técnicas de incrustación:** métodos para mapear diferentes tipos de datos en un espacio vectorial común.
- **Gestión eficiente:** técnicas para almacenar, recuperar y buscar grandes colecciones de incrustaciones.
- **Bases de datos vectoriales:** sistemas especializados para gestionar y consultar incrustaciones, incluidas consideraciones prácticas para la implementación en producción.
- **Aplicaciones en el mundo real:** ejemplos concretos de cómo las incrustaciones y las bases de datos vectoriales se combinan con modelos de lenguaje grandes (LLM) para resolver problemas del mundo real.

Día 4




+


 Crear




 Hogar




 Competiciones




 Conjuntos de datos




 Modelos



 Código




 Discusiones




 Aprender

▼

 Más



 Tu trabajo



 Ver eventos activos

Buscar

Solución de problemas específicos del dominio mediante LLM

Autores: Christopher Semturs, Shekoofeh Azizi, Scott Coull, Umesh Shankar y Wieland Holfelder

Introducción

Los modelos de lenguaje grandes (LLM) han surgido como herramientas poderosas para abordar desafíos complejos en numerosos dominios. Si bien las primeras iteraciones se centraron en tareas de propósito general, los desarrollos recientes han resaltado el potencial de ajustar los LLM para abordar problemas específicos dentro de campos especializados. Este informe técnico explora estos conceptos en dos dominios distintos: ciberseguridad y medicina. Cada uno muestra la capacidad única de los LLM para mejorar los flujos de trabajo existentes y desbloquear nuevas posibilidades.

La ciberseguridad presenta una serie de desafíos únicos para los LLM, incluida la escasez de datos disponibles públicamente, una amplia diversidad de conceptos altamente técnicos e información sobre amenazas que cambia a diario. Además, los casos de uso sensibles, como el análisis de malware, requieren consideraciones específicas para el desarrollo de modelos. Abordamos estos desafíos centrándonos en contenido y tareas específicas de ciberseguridad, combinando modelos de lenguaje centrados en la

Google AI Studio

Untitled prompt

Enable Autosave

Compare

Get API key

Create Prompt

Stream Realtime

Starter Apps

Tune a Model

Library

Enable chat history

Prompt Gallery

System Instructions

Optional tone and style instructions for the model

What will you build?

Push Gemini to the limits of what AI can do using the Gemini API

Run settings

Get code

Model

Gemini 2.0 Flash

Token count

0 / 1,048,576

Temperature

Google AI Studio

Top Halloween Costumes 2023

Enable Autosave

Compare

Get API key

Create Prompt

Stream Realtime

Starter Apps

Tune a Model

Library

Enable chat history

Prompt Gallery

API documentation

Developer forum

Changelog NEW

Gemini makes mistakes, so double-check it

System Instructions

Optional tone and style instructions for the model

characters are seeing a resurgence.

- Wednesday Addams:** Riding the wave of the Netflix show's popularity, this is a reliably popular choice.
- Marvel/DC Superheroes:** Classic and always in demand, especially with new movie and show releases.
- AI / Robot:** Leaning into the tech-heavy current events, this could be a relevant and creative costume.

Important Considerations:

- Timing:** What's trendy in early October might change by Halloween.
- Age Group:** Kids' costumes differ significantly from adults'.
- Regional Variations:** Popularity can differ based on location and local events.

To get a better picture of what's currently trending in your specific area, I recommend checking out local costume shops, party supply stores, and online retailers in your region. They often have insights into what's selling well locally.

👍

🗨

Run settings

Structured output

Edit

Code execution

Function calling

Edit

Grounding with Google Search

☒

Source: Google Search

Advanced settings

When and where is Taylor Swift's next concert?

Taylor Swift's Eras Tour concluded in Vancouver on December 8, 2024.[1] While there are no scheduled tour dates for 2025, there are reports that she may be planning a new tour for 2026.[2][3] However, this is still unconfirmed, and she is expected to take some time to finalize her plans.[1] Keep an eye out for announcements regarding her future plans, likely for 2026 at the earliest.[2]

Search Sources ?

- 1. hindustantimes.com
- 2. reddit.com
- 3. imdb.com

Code execution

Function calling

Grounding with Google Search

Source: Google Search

```
chunks = rc.grounding_metadata.grounding_chunks
for chunk in chunks:
    print(chunk)
```

```
GroundingChunk(begin_offset=12, end_offset=41, content='Taylor Swift\'s next concert',
data_sources=[DataSource(title='Taylor Swift - The Eras Tour - Wikipedia',
uri='https://en.wikipedia.org/wiki/Taylor_Swift_-_The_Eras_Tour')],
confidence_score=0.9324609)
GroundingChunk(begin_offset=45, end_offset=77, content='February 16, 2025, in Melbourne,
Australia', data_sources=[DataSource(title='Taylor Swift - The Eras Tour - Wikipedia',
uri='https://en.wikipedia.org/wiki/Taylor_Swift_-_The_Eras_Tour')],
confidence_score=0.9324609)
GroundingChunk(begin_offset=81, end_offset=113, content='Melbourne Cricket Ground as
part of The Eras Tour', data_sources=[DataSource(title='Taylor Swift - The Eras Tour -
```

Code execution

Function calling

Grounding with Google Search

Source: Google Search

Advanced settings

HTML(rc.grounding_metadata.search_entry_point.rendered_content)

```
<div class="v063rb">The <span class=" TLWKJc rQEFKc">Eras Tour</span> is the ongoing
sixth headlining concert tour by American singer-songwriter <span class=" TLWKJc
rQEFKc">Taylor Swift</span>. Described as her most ambitious tour yet, the tour is a
retrospective of her career as a tribute to all of her "musical eras". The tour began on
March 17, 2023, in Glendale, Arizona, and is currently set to conclude on December 8,
2024, in Vancouver, Canada. The tour is Swift's most commercially and critically
successful to date. It is estimated to gross over $2 billion, making it the highest-
grossing music tour of all time.
<div class="wHYlTd">Background</div>
<div class="NFQFxe ViZG9d">After the release of her seventh and eighth studio albums,
<span class=" TLWKJc rQEFKc">Lover</span> (2019) and <span class=" TLWKJc
rQEFKc">Folklore</span> (2020), Swift had planned to embark on her sixth headlining
```

Code execution

Function calling

Grounding with Google Search

Source: Google Search

Advanced settings

```
supports = rc.grounding_metadata.grounding_supports
for support in supports:
    print(support)
```

```
GroundingSupport(text_segment='Taylor Swift - The Eras Tour - Wikipedia',
data_source=DataSource(title='Taylor Swift - The Eras Tour - Wikipedia',
uri='https://en.wikipedia.org/wiki/Taylor_Swift_-_The_Eras_Tour'))
```

Code execution

Function calling

Grounding with Google Search

```
Markdown(markdown_buffer.getvalue())
```



Supported text:

* Taylor Swift - The Eras Tour - Wikipedia

Citations:

* 1: [Taylor Swift's next concert]

* 2: [February 16, 2025, in Melbourne, Australia]

* 3: [Melbourne Cricket Ground as part of The Eras Tour]

```
Markdown(r.text)
```

Australia won the 2023 Cricket World Cup, defeating India in the final.



```
HTML(r.candidates[0].grounding_metadata.search_entry_point.rendered_content)
```

```
<div class="v063rb"><span class=" TLWKJc rQEFKc">Australia</span> beat <span class="
TLWKJc rQEFKc">India</span> to win their sixth <span class=" TLWKJc rQEFKc">World
Cup</span> title. <span class=" TLWKJc rQEFKc">Australia</span> defeated <span class="
TLWKJc rQEFKc">India</span> by 6 wickets in the final to win the 2023 <span class="
TLWKJc rQEFKc">Cricket World Cup</span> at the Narendra Modi Stadium in Ahmedabad on
Sunday.
</div>
```

First turn: score=0.0

Second turn: score=0.9999851

Australia won the 2023 Cricket World Cup, defeating India in the final.



Use code with caution.

HTML(rc.grounding_metadata.search_entry_point.rendered_content)

```
<div class="v063rb"><span class=" TLWKJc rQEFKc">Australia</span> beat <span class="
TLWKJc rQEFKc">India</span> to win their sixth <span class=" TLWKJc rQEFKc">World
Cup</span> title. <span class=" TLWKJc rQEFKc">Australia</span> defeated <span class="
TLWKJc rQEFKc">India</span> by 6 wickets in the final to win the 2023 <span class="
TLWKJc rQEFKc">Cricket World Cup</span> at the Narendra Modi Stadium in Ahmedabad on
Sunday.
</div>
```

Día 5

Puesta en práctica de la IA generativa en Vertex AI mediante MLOps

Autores: Anant Nawalgaria, Gabriela Hernández Larios, Elia Secchi, Mike Styer, Christos Aniftos y Onofrio Petragallo

Introducción

La aparición de los modelos de base y la IA generativa (IA gen) ha inaugurado una nueva era para la creación de sistemas de IA. Seleccionar el modelo adecuado entre una amplia gama de arquitecturas y tamaños, conservar los datos, diseñar indicaciones óptimas, ajustar los modelos para tareas específicas, fundamentar los resultados de los modelos en datos del mundo real y optimizar el hardware son solo algunos de los nuevos desafíos que presentan los modelos de gran tamaño.



Este informe técnico profundiza en los principios fundamentales de MLOps y las adaptaciones necesarias para el dominio de la IA gen y los modelos de base. También examinamos la amplia gama de productos Vertex AI, diseñados específicamente para abordar las demandas únicas de los modelos de base y las aplicaciones basadas en IA gen. A través de esta exploración, descubrimos cómo Vertex AI, con sus sólidas bases de infraestructura de IA y herramientas MLOps, amplía sus capacidades para proporcionar

The screenshot shows a GitHub repository interface. The left sidebar displays the file tree with folders like .github, audio, conversation, embeddings, gemini, agents, batch-prediction, chat-completions, code-execution, and context-caching. The main area shows the commit history for the 'e2e-gen-ai-app-starter-pack' directory. A commit by 'eliasecchig' is highlighted, with a message 'fix: starter pack live api async tool calls (#1679)'. Below this is a table of commit history.

Name	Last commit message	Last commit date
..		
app	fix: starter pack live api async tool calls (#1679)	2 weeks ago
deployment	feat: Add Agent Builder Search pattern with Data Ingestion (#1380)	3 months ago
notebooks	fix: getting_started notebook starter pack eval prompt template (#1611)	last month
streamlit	fix: streamlit event handler, comments langgraph (#1268)	4 months ago
tests	feat: Add Agent Builder Search pattern with Data Ingestion (#1380)	3 months ago
.gitignore	feat: add e2e gen ai app starter pack (#1225)	4 months ago
CONTRIBUTING.md	feat: add 2e gemini starter pack (#1225)	4 months ago

🚀 End-to-End Gen AI App Starter Pack 🚀

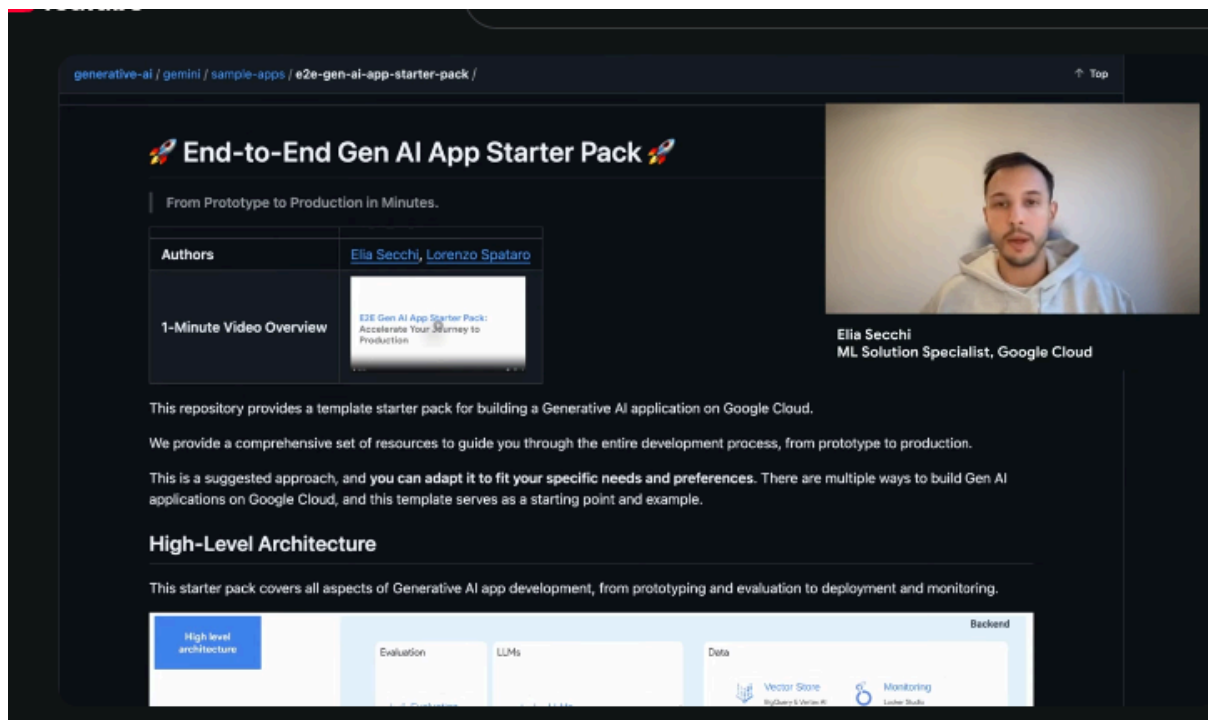
From Prototype to Production in Minutes.

Contact	e2e-gen-ai-app-starter-pack@google.com
1-Minute Video Overview	
20-Minute Video Walkthrough	

This repository provides a template starter pack for building a Generative AI application on Google Cloud.

We provide a comprehensive set of resources to guide you through the entire development process, from prototype to production.


This is a suggested approach, and **you can adapt it to fit your specific needs and preferences**. There are multiple ways to build Gen AI applications on Google Cloud, and this template serves as a starting point and example.



generative-ai / gemini / sample-apps / e2e-gen-ai-app-starter-pack /

End-to-End Gen AI App Starter Pack

From Prototype to Production in Minutes.

Authors	Elia Secchi, Lorenzo Spataro
1-Minute Video Overview	

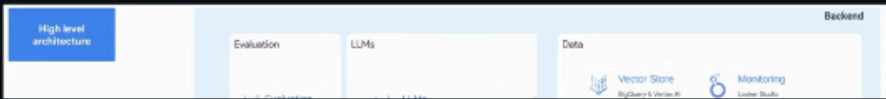
This repository provides a template starter pack for building a Generative AI application on Google Cloud.

We provide a comprehensive set of resources to guide you through the entire development process, from prototype to production.

This is a suggested approach, and **you can adapt it to fit your specific needs and preferences**. There are multiple ways to build Gen AI applications on Google Cloud, and this template serves as a starting point and example.

High-Level Architecture

This starter pack covers all aspects of Generative AI app development, from prototyping and evaluation to deployment and monitoring.



Elia Secchi
ML Solution Specialist, Google Cloud

Respondiendo las preguntas

1. ¿Hubo alguna aplicación o caso de uso de los LLMs que le llamó más la atención? ¿Por qué?

R// De los casos presentados, me resultó particularmente interesante el uso de LLMs en el sector financiero para automatizar flujos de trabajo de asesoría. La capacidad de un LLM para analizar grandes cantidades de información, como historiales de clientes, tendencias del mercado y regulaciones financieras, y luego generar recomendaciones personalizadas, representa un avance significativo. Me atrae la idea de cómo la IA puede democratizar el acceso a asesoramiento financiero de calidad, haciéndolo más rápido y accesible para un público más amplio. Además, este caso

demuestra el potencial de los LLMs para aumentar la productividad en industrias altamente reguladas, donde la precisión y la personalización son clave.

2. Proponga un caso de ciberseguridad que considere se puede solucionar mediante un LLM y describa de forma general cómo lo resolvería.

R// En el ámbito de la ciberseguridad, un caso que se podría abordar con un LLM es la detección y respuesta a ataques de phishing. Los correos electrónicos de phishing son cada vez más sofisticados y personalizados, lo que los hace difíciles de detectar para los filtros tradicionales. Un LLM podría ser entrenado para analizar el contenido, la estructura y el contexto de los correos electrónicos, identificando patrones y anomalías que sugieran un intento de phishing. Además, el LLM podría aprender continuamente de nuevos ataques y adaptarse a las tácticas cambiantes de los ciberdelincuentes.

Para implementar esta solución, se alimentará al LLM con un conjunto masivo de datos de correos electrónicos legítimos y de phishing, permitiéndole aprender a distinguir entre ambos. Luego, el LLM se integraría en un sistema de seguridad de correo electrónico, donde analizaría los mensajes entrantes en tiempo real. Si el LLM identifica un correo electrónico sospechoso, podría marcarlo como tal, enviarlo a una cuarentena o alertar al usuario. Además, el LLM podría proporcionar información sobre por qué considera que un correo electrónico es sospechoso, educando a los usuarios y fortaleciendo su capacidad para detectar futuros ataques de phishing.