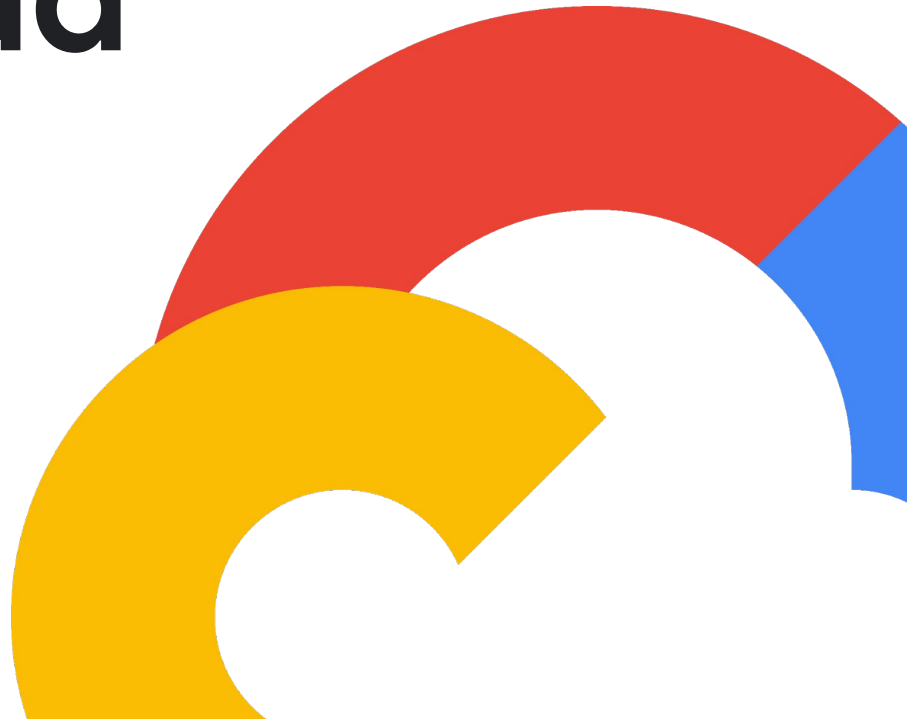


# GenAI on Google Cloud

Nov 23rd, 2023

Stefan Veis Pennerup

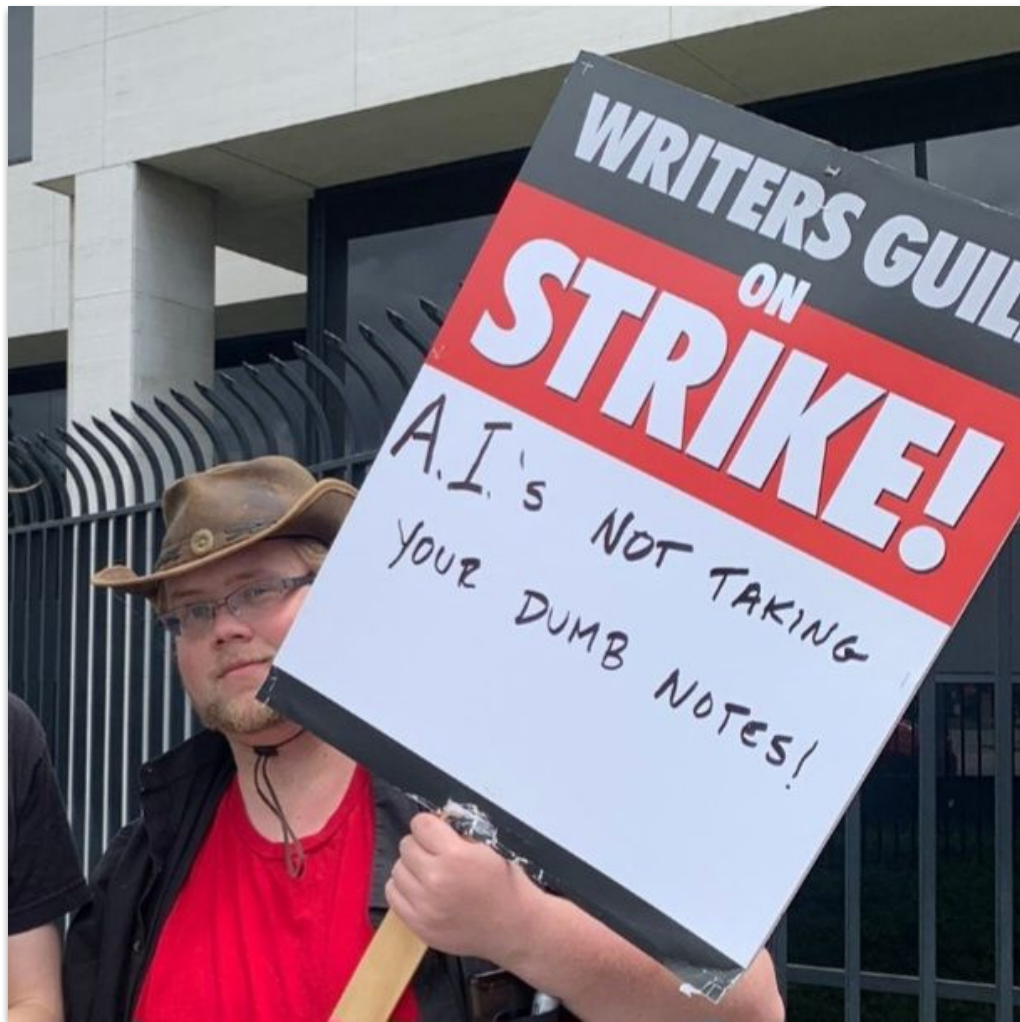
Google Cloud

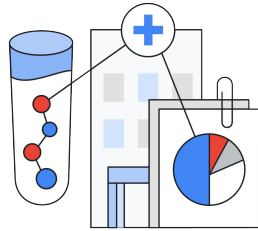












1784



Steam Power

1870



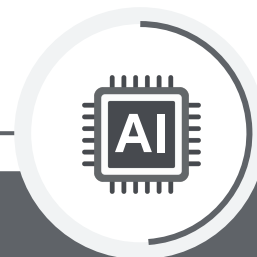
Electricity

1969



Information  
Technology

Today



Artificial Intelligence



## Large Language Models

LLMsLLMsLLMsLLMs

LLMsLLMsLLMsLLMs

LLMsLLMsLLMsLLMs

LLMsLLMsLLMsLLMs

LLMsLLMsLLMsLLMs

LLMsLLMsLLMsLLMs

LLMsLLMsLLMsLLMs

LLMsLLMsLLMsLLMs

LLMsLLMsLLMsLLMs

LLMsLLMsLLMsLLMs!



Go read this huuuuuuge pile of books.



So, you've learned about cats and millions of other concepts ... what's a cat?



A cat is a small, domesticated carnivorous mammal.



**Generative language models**

LaMDA, PaLM, GPT-3, etc.





**2017**  
Transformer



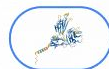
**2018**  
BERT



**2019**  
T5



**2020**  
LaMDA



**2021**  
AlphaFold



**2022**  
PaLM



**2023**  
Bard

# There is **no one LLM** to rule them all.

Different LLMs have different **strengths** and **weaknesses**

Different LLMs have very different processing **cost**.

To find **the right model for the right use-case** at the right cost will be a key to success.

LLMs come and go - **platforms remain**





## Consumers

Plan a 3 day trip to Patagonia

Create a valentine poem.

How to make GF pancakes?

A picture of a panda playing yahtzee

I want to write a novel. How do I get started?

Create a jazz song for a bday card

Bard + MakerSuite

Google



## Enterprises

How do we **control our data**

How do we deal with **fraud & security**

We need to be **accurate & explainable**

How will we **control costs**?

How do we integrate our existing **data & applications**

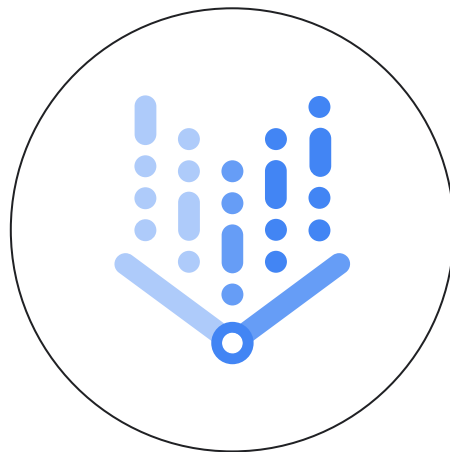
Vertex AI + Duet AI

Google Cloud



**Duet AI**

**For users**



**Vertex AI**

**For builders**



# Duet AI is everywhere, and more to come...

## Google Workspace

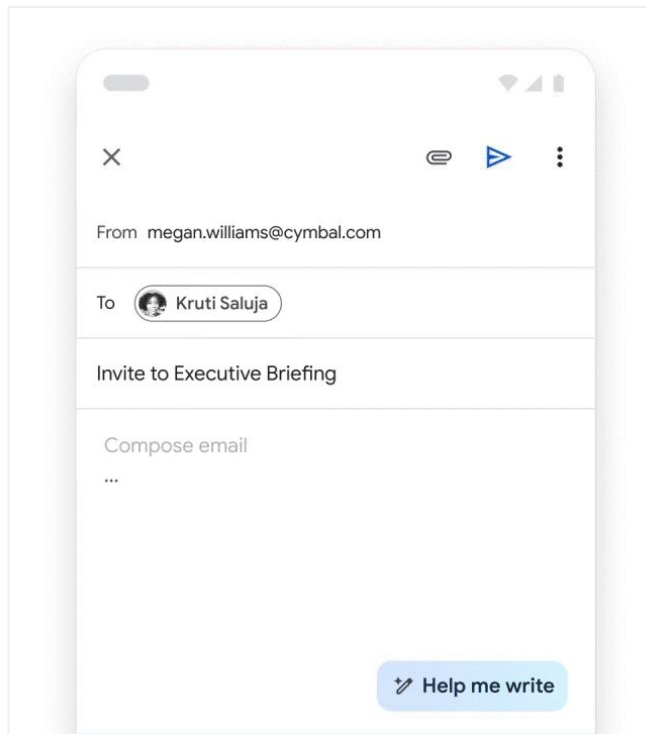
- Duet AI in Gmail
- Duet AI in Docs
- Duet AI in Sheets
- Duet AI in Slides
- Duet AI in Meet
- Duet AI in Chat
- Duet AI in Appsheet

## Google Cloud

- Duet AI in Apigee
- Duet AI in Cloud Logging
- Duet AI in Google Cloud Console
- Duet AI in Dataplex
- Duet AI in GKE
- Duet AI in Cloud Run
- Duet AI in Application Integration
- Duet AI in Colab Enterprise
- Duet AI in Looker
- Duet AI in BigQuery
- Duet AI in Cloud Spanner
- Duet AI in Database Migration Service
- Duet AI in Chronicle Security Operations
- Duet AI in Mandiant Threat Intelligence
- Duet AI in Security Command Center
- Duet AI in IDEs, fx VS Code

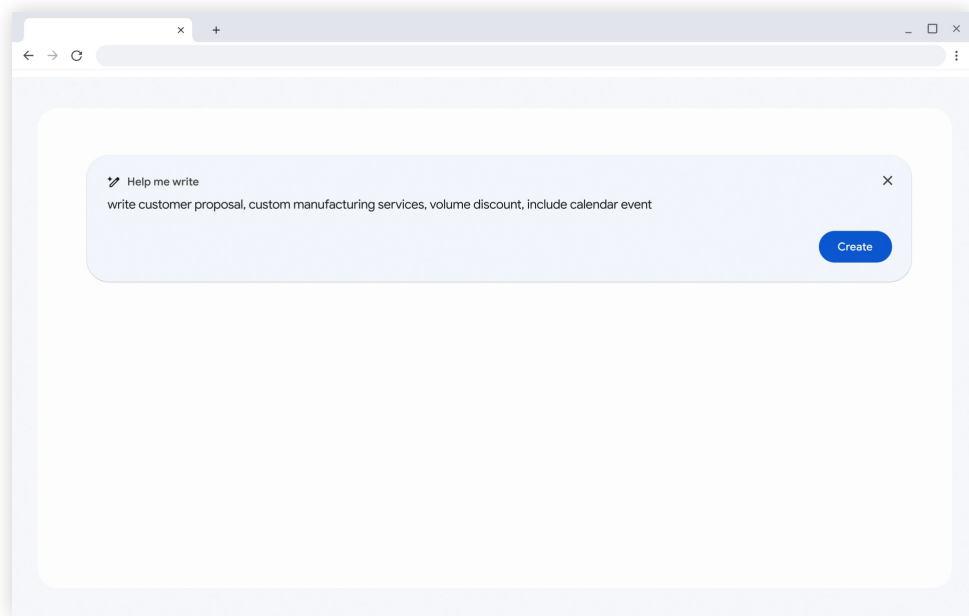
## With *help me write*, Sellers managing multiple accounts can use Duet AI in Gmail to write:

- Customized outreach
- Follow-up emails
- Event invites, and more



# With simple prompts, Duet AI helps Sellers:

- Create sales presentations
- Sales scripts
- Customer pitches, and more



# Duet AI in Google Cloud

## Assisted development

- Completes your code as you type
- Generates functions based on comments and definition
- Chat with AI pair programmer about code (e.g. Code explanation, Testing,...)
- Assist fix errors in code

## Assisted operations

- Retrieve “How-to” knowledge about GCP infrastructure, deployment and best practices
- Recommendations for optimizing cloud applications on cost, security, reliability and performance.

## Assisted data

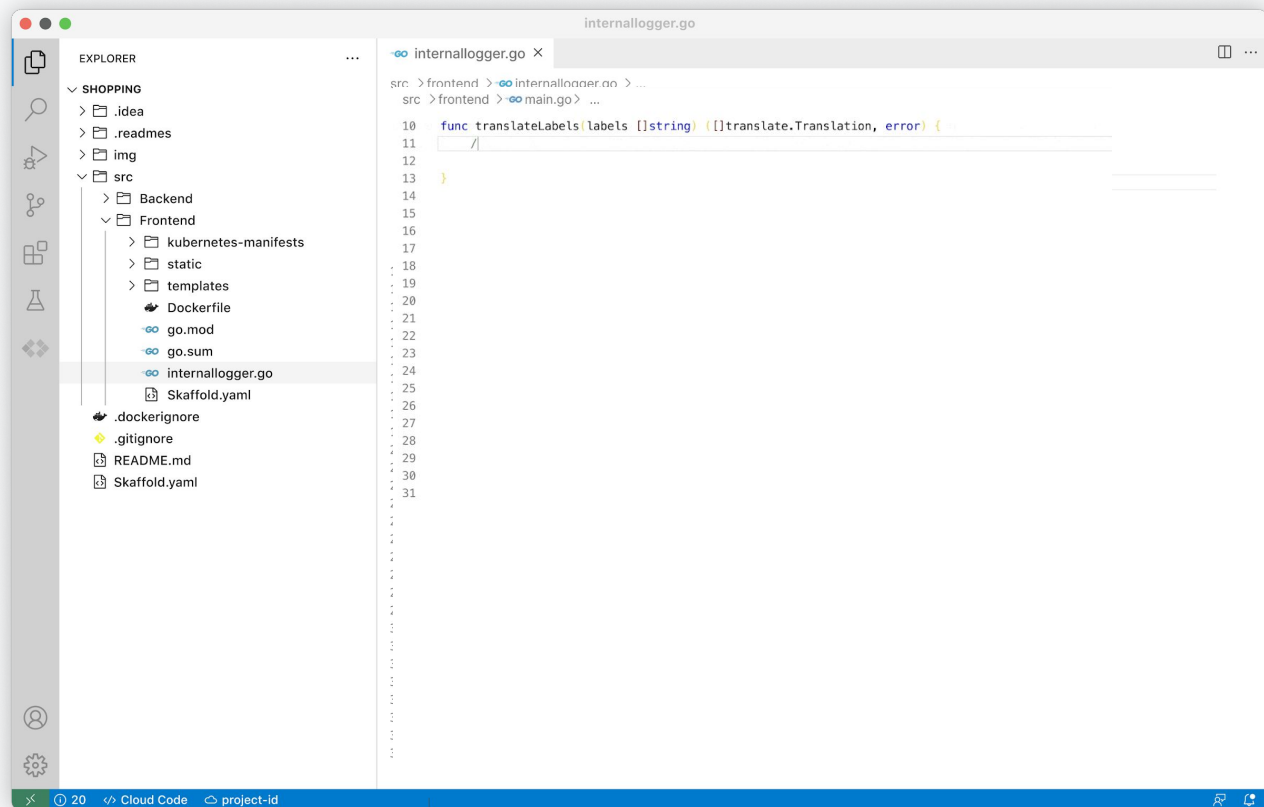
- Discover and Explore Data & Insights (queries, BI)
- Learn & Build: Assisted coding
- Translate code dialects
- Automated Data Enrichment and Wrangling for Analytics

## Assisted security

- Up-to-date security insights on major bad actors
- Automated alerts for breaches and lateral movements
- Guidance on threat remediation steps



# Assisted development



# Assisted data

The screenshot displays the Google Cloud BigQuery console interface. At the top, the Google Cloud logo is visible alongside a project name dropdown and a search bar. The main interface is divided into three sections: Explorer, Query Editor, and a bottom navigation bar.

**Explorer:** This section on the left lists resources under a 'Project name'. It includes 'External connections', 'Data set name 01' through 'Data set name 15', and 'Table set name 01'. A search bar at the top of the Explorer lists 'Type to search'. Below the list, a 'SHOW MORE' link is present.

**Query Editor:** The right section shows a query editor with a tab for '\*Unsaved Query 2'. The query text is as follows:

```
1 -- What are the most popular routes for NY Citi subscriber type?
2 |SELECT
  usertype,
  CONCAT(start_station_name, " to ", end_station_name) as route,
  COUNT(*) as num_trips,
  ROUND(AVG(cast(tripduration as int64) / 60),2) as duration
FROM `bigquery-public-data.new_york.citibike_trips`
GROUP BY
  start_station_name, end_station_name, usertype
ORDER BY
  num_trips DESC
LIMIT 10
```

Below the query text, a status message indicates 'The query will process in 10:00'. The bottom navigation bar includes tabs for 'SCHEMA', 'DETAILS', 'SAVED QUERIES', 'SNAPSHOTS', and 'TABLE EXPLORER'.

# Assisted data - Iteration

The screenshot displays the Google Cloud BigQuery console interface. On the left, the 'Explorer' pane shows a project hierarchy with various data sets and tables. The main editor area contains two queries, both using the BigQuery standard SQL dialect. The first query, labeled '\*Unsaved Query', is a basic SELECT statement that extracts the hour from the start\_time, calculates the average duration in minutes, and ranks subscribers by their average duration. The second query, labeled '\*Unsaved Query 2', is a more complex version that filters out null subscribers and rounds the average duration to single precision. The console also shows a status bar at the bottom with tabs for SCHEMA, DETAILS, SAVED QUERIES, SNAPSHOTS, and TABLE EXPLORER.

Google Cloud

Project Name

Search (/) for resources, docs, products, and more

Search

Explorer

+ ADD DATA

Type to search

Viewing all resources. [Show starred resources only.](#)

Project name

- External connections
- Data set name 01
- Data set name 02
  - Table set name 01
  - Table set name 01
  - Table set name 01
- Data set name 03
- Data set name 04
- Data set name 05
- Data set name 06
- Data set name 07
- Data set name 08
- Data set name 09
- Data set name 10
- Data set name 11
- Data set name 12
- Data set name 13
- Data set name 14
- Data set name 15

SHOW MORE

RUN SAVE SHARE SCHEDULE MORE

The query will process in 10:00

```
1 SELECT
2   EXTRACT(HOUR FROM start_time) AS hour,
3   subscriber_type,
4   AVG(duration_minutes) AS average_duration,
5   RANK() OVER (PARTITION BY subscriber_type ORDER BY AVG(duration_minutes) DESC) AS rank
6 FROM `bigquery-public-data.austin_bikeshare.bikeshare_trips`
7 GROUP BY 1,2
8 ORDER BY 1,2,3;
9
10 # using the query above filter out null subscribers and round all decimals to single precision
11 |SELECT
   EXTRACT(HOUR FROM start_time) AS hour,
   subscriber_type,
   ROUND(AVG(duration_minutes),1) AS average_duration,
   RANK() OVER (PARTITION BY subscriber_type ORDER BY AVG(duration_minutes) DESC) AS rank
 FROM `bigquery-public-data.austin_bikeshare.bikeshare_trips`
 WHERE subscriber_type IS NOT NULL
 GROUP BY 1,2
 ORDER BY 1,2,3;
```

SCHEMA DETAILS SAVED QUERIES SNAPSHOTS TABLE EXPLORER

# Security - Threat intelligence

The screenshot displays the Mandiant Advantage Threat Intelligence dashboard. The top navigation bar includes 'MANDIANT ADVANTAGE', 'What's New', and various tool categories like 'Reports & Analysis', 'Explore', 'Threat Monitoring', 'File Analysis', and 'Settings'. A search bar at the top right contains the query 'Ukraine'. Below the search bar, a 'Search' section shows 'Ukraine' as the query and a 'Show AI Summary' toggle. The main content area is divided into three sections: a 'Primary Filter' on the left, 'Search Results - 3,658' in the center, and a 'Generative AI Summary' on the right. The 'Primary Filter' section lists various categories with counts: All (3,658), Campaigns (5), Threat Actors (7), Malware & Tools (6), Vulnerabilities (3), Indicators (1,619), Reports & Analysis (2,618), and AI Summary References (5). The 'Search Results' section displays two results. The first result is a 'CAMPAIGN' titled 'Espionage Actor with Russia Nexus Conducts Email...' with a 'LATEST ACTIVITY' date of 'AUG 9, 2022'. It includes a 'Target Industries' filter set to 'Retail' and a brief description of the campaign. The second result is a 'REPORT' titled 'Hold the Line: Reviewing Mandiant Engagements in...' with a 'Published' date of 'DEC 22, 2022' and an 'Updated' date of 'AUG 2, 2022'. It also includes a 'Target Industries' filter and a brief description. The 'Generative AI Summary' section on the right provides a high-level overview of the cyber security landscape in Ukraine, mentioning the Russian invasion and the increase in attacks. It also lists 'Attacks Targeting Ukraine' and identifies several threat actors involved, including APT29, ROOTSAW, and Sandworm.

**MANDIANT ADVANTAGE** | What's New

THREAT INTELLIGENCE | Reports & Analysis | Explore | Threat Monitoring | File Analysis | Settings | Q Ukraine

**Search**

Q Ukraine | Show AI Summary

**Primary Filter**

- All (3,658)
- Campaigns (5)
- Threat Actors (7)
- Malware & Tools (6)
- Vulnerabilities (3)
- Indicators (1,619)
- Reports & Analysis (2,618)
- AI Summary References (5)

**Search Results - 3,658** | Sort By: Most Relevant

**CAMPAIGN** | LATEST ACTIVITY: AUG 9, 2022

**Espionage Actor with Russia Nexus Conducts Email...**

Target Industries: Retail (7 More)

Since at least October 1, 2022, UNC4322 has targeted Ukrainian organizations including government offices, entities in the media, energy industry, financial services, agriculture, and consumer goods. Phishing email include malicious Word document which

**REPORT** | PUBLISHED: DEC 22, 2022 | UPDATED: AUG 2, 2022

**Hold the Line: Reviewing Mandiant Engagements in...**

Target Industries: [Icons] | Motivations: [Icons]

Mandiant has continued to respond to dozens of Russian state-sponsored cyber threat incidents across Ukraine since the start of 2022. While we conducted engagements across nearly every sector of Ukrainian industry, our investigations overwhelmingly supported...

**REPORT** | PUBLISHED: DEC 22, 2022 | UPDATED: AUG 2, 2022

**War and Peace: Cyber Threat Activity Surrounding the...**

Target Industries: [Icons] | Motivations: [Icons]

In the first four months of 2022, Mandiant identified more Russian-sponsored disruptive cyberattacks against Ukraine than it has during the past eight years. Several of these

**Generative AI Summary**

This is a placeholder for the official, legal-approved, AI disclaimer appearing in Mandiant Advantage.

**Cyber security landscape in Ukraine**

The cyber security landscape in Ukraine has been transformed by the Russian invasion. Mandiant has observed a significant increase in the number of attacks targeting Ukraine.

The attacks targeting Ukraine are often sophisticated and well-coordinated. In some cases, they have been carried out by Russian government-backed hackers. In other cases, intrusions have been carried out by other countries that are interested in the situation in Ukraine.

**Attacks Targeting Ukraine**

Mandiant has identified several threat actors that have been involved in attacks targeting Ukraine, including:

- APT29 (also known as Cozy Bear or YTTIRUM): A Russian government-backed hacking group that has been active since at least 2010. APT29 has been linked to a variety of attacks, including the 2015 breach of the Democratic National Committee. Recently Mandiant has observed APT29's repeated use of the ROOTSAW and BEATDROP malware families to target various European MFAs likely in service of intelligence priorities on diplomatic responses to Russia's invasion of Ukraine.
- Sandworm (also known as Telebots or Voodoo Bear): A Russian military intelligence group that has been active since at least

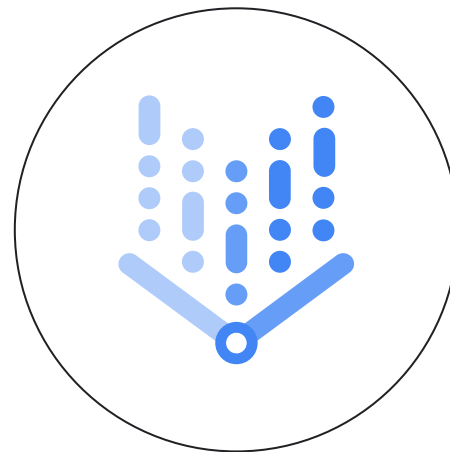
Copy





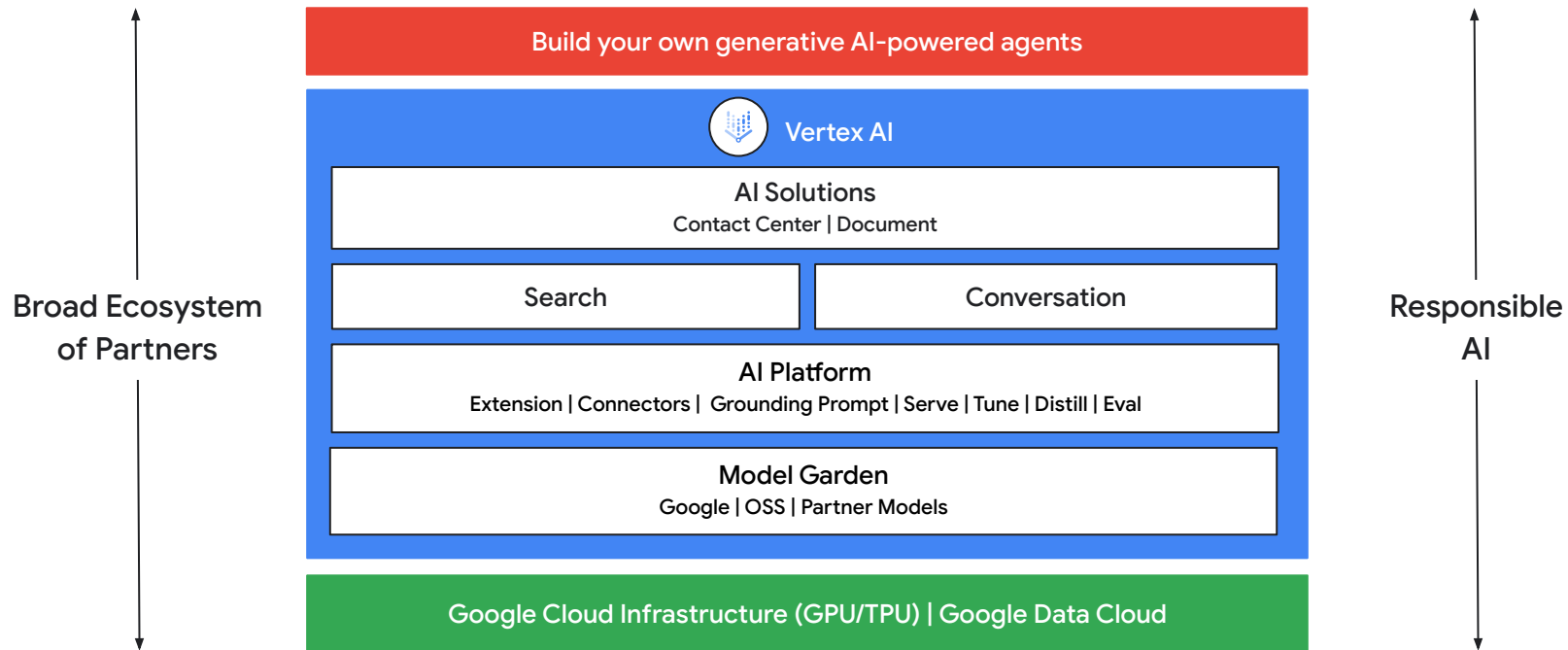
**Duet AI**

**For users**



**Vertex AI**

**For builders**



# You control your data, not Google

We ensure responsible use of data for Cloud AI/ML with these commitments to customers:



By default, we do not use customer data to train our models, in accordance with [GCP Terms](#) and [Cloud Data Processing Addendum](#). *Your customer data is only processed according to your instructions.*



You benefit from our privacy experience and commitments to transparency, compliance with regulations such as the GDPR, and privacy best-practices.



Our extensive data governance & security controls provide you with confidence in the protection of your data from other customers, users, attackers, and unauthorized access by Google employees.

# Live demo of Vertex AI Search.



# Foundation Models

Across a variety of model sizes to address use cases



Now in GA

## PaLM for Text

Custom language tasks



Now in GA

## PaLM for Chat

Multi-turn conversations with session context



Now in GA

## Imagen for Text to Image

Create and edit images from simple prompts



Now in GA

## Embeddings API for Text and Image

Extract semantic information from unstructured data



Now in GA

## Chirp for Speech to Text

Build voice enabled applications



Now in GA

## Codey for Code Generation

Improve coding and debugging



# **What is Prompt Design?**

**For people with  
photosensitive epilepsy,  
please look away now**





# Live demo of the Palm 2 API.

# **Good prompt design**

**01**

**Context**

**02**

**Clear goal**

**03**

**Breaking down tasks**

**04**

**Examples**

For the given review, return the sentiment.

Review: This is a shoe I will wear with black dress pants or jeans when I need comfort and a little style, but I am not impressed. This is a very flimsy shoe with little support at all. Unlike any other shoes I've purchased in the past. It looks nice, but it's not comfortable.

negative



For the given review, return the sentiment and explanation.

Review: This is a shoe I will wear with black dress pants or jeans when I need comfort and a little style, but I am not impressed. This is a very flimsy shoe with little support at all. Unlike any other shoes I've purchased in the past. It looks nice, but it's not comfortable.

negative

The review is negative because the reviewer is not impressed with the shoe. The reviewer says that the shoe is flimsy and has little support, and that it is not comfortable. The reviewer also says that the shoe looks nice, but it is not comfortable.

For the given review, return a JSON object that has the fields sentiment and explanation.

Review: This is a shoe I will wear with black dress pants or jeans when I need comfort and a little style, but I am not impressed. This is a very flimsy shoe with little support at all. Unlike any other shoes I've purchased in the past. It looks nice, but it's not comfortable.

```
{  
  "sentiment": "negative",  
  "explanation": "The shoe is not  
comfortable and has little support."  
}
```

For the given review, return a JSON object that has the fields sentiment and explanation. Acceptable values for sentiment are Positive or Negative. The explanation field contains text that explains the sentiment.

Review: This is a shoe I will wear with black dress pants or jeans when I need comfort and a little style, but I am not impressed. This is a very flimsy shoe with little support at all. Unlike any other shoes I've purchased in the past. It looks nice, but it's not comfortable.

```
{  
  "sentiment": "Negative",  
  "explanation": "The shoe is flimsy and  
has little support. It is not comfortable."  
}
```

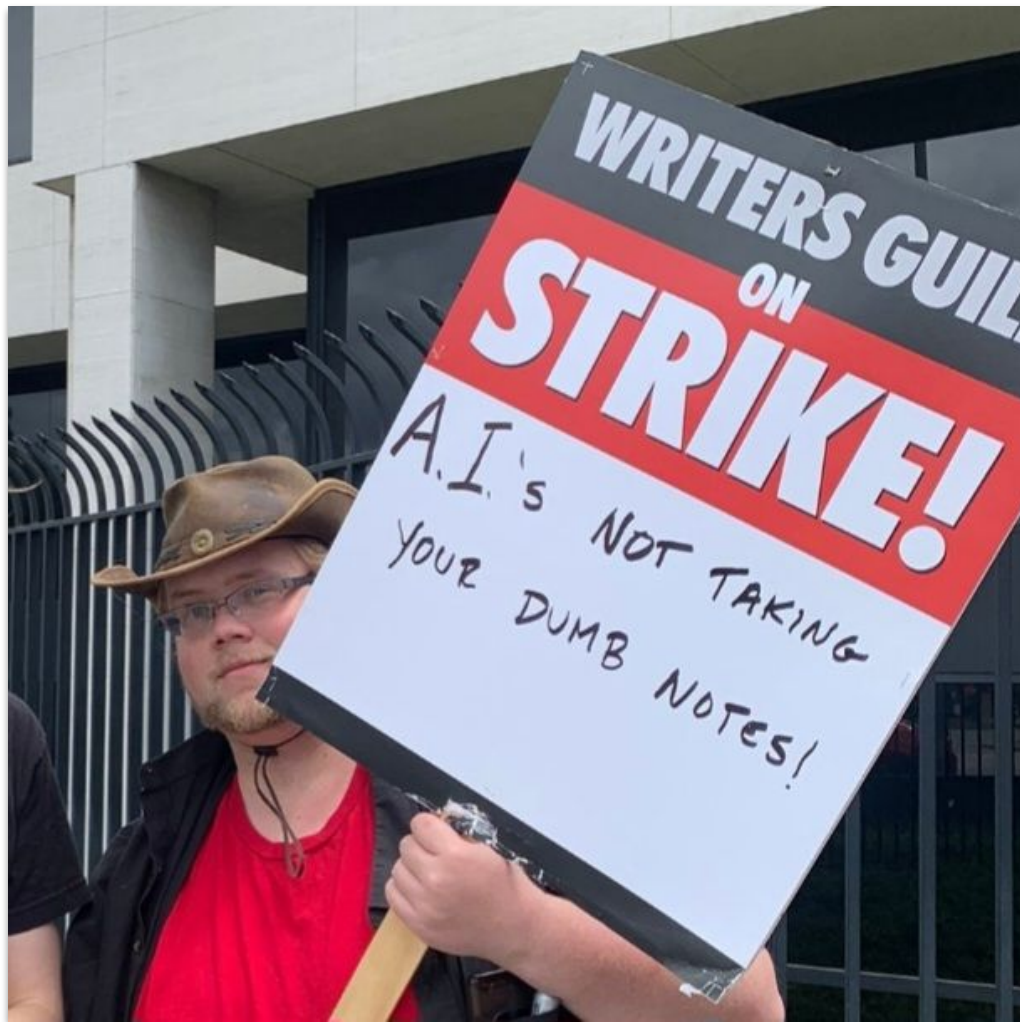


**You are a dog...**

**You are a junior data analyst for a financial company...**

**You are a marketer for a big retail company...**

**You are an developer for a system integrator...**



- 01** There is no one LLM to rule them all.
- 02** Start building a custom prompt library.
- 03** Let's go jointly go out and shape the future of AI.



**The Flip Flop Guy**  
flipflopGuy@google.com  
14:29 Today

Resolve

