

# Module 3. Ứng dụng GenAI trên Google Cloud

## 1. Hướng dẫn sử dụng Qwiklabs trong Google Cloud

Nền tảng Qwiklabs, một phần của Google Cloud, cung cấp trải nghiệm thực hành trên các dịch vụ của Google Cloud. Qwiklabs cấp cho người dùng thông tin đăng nhập tạm thời để truy cập vào Cloud Console mà không yêu cầu chi phí.

Để bắt đầu, hãy đăng nhập vào Coursera trong một cửa sổ ẩn danh (hay còn gọi là duyệt web riêng tư hoặc InPrivate, tùy thuộc vào trình duyệt). Việc đăng nhập bằng cửa sổ ẩn danh giúp đảm bảo rằng bạn không vô tình sử dụng tài khoản Google cá nhân, tránh các hóa đơn phát sinh không mong muốn. Bạn có thể tham khảo các bài viết hướng dẫn về trình duyệt trong phần tải xuống của video.

Sau khi đăng nhập Coursera thành công bằng cửa sổ ẩn danh, hãy quay lại khóa học và truy cập trang bài thực hành. Nếu được yêu cầu, hãy chấp nhận Quy tắc Danh dự và nhập tên của bạn. Tiếp theo, nhấp vào "Mở Công cụ" để mở phòng lab trong một tab mới. Trong tab này, nhấp vào "Bắt đầu Phòng thí nghiệm" và đợi thông tin kết nối hiển thị ở phía bên trái. Mỗi phòng lab có bộ đếm thời gian, và phòng lab sẽ tự động kết thúc khi hết thời gian.

Để truy cập Cloud Console, hãy nhấp vào "Mở Google Console" và đăng nhập bằng tên người dùng và mật khẩu được cung cấp trong trang chi tiết kết nối. Sao chép và dán thông tin đăng nhập này vào trang đăng nhập. Lưu ý rằng thông tin này sẽ khác nhau cho mỗi phòng lab. Qwiklabs tạo một tài khoản mới cho mỗi phòng lab, do đó bạn cần nhấp qua các bước thiết lập tài khoản ban đầu. Tiếp theo, bạn cần chấp nhận các điều khoản và dịch vụ của Google Cloud khi vào Cloud Console.

Sau khi vào dự án, hãy kiểm tra ID dự án hiển thị ở góc trên. Đảm bảo ID này trùng khớp với ID trong Qwiklabs. Trong một số phòng lab, có thể có nhiều ID dự án hoặc người dùng khác nhau. Kiểm tra kỹ các thông tin này để tránh nhầm lẫn.

Bạn có thể kiểm tra tên người dùng hiển thị trong Cloud Console. Đây không phải là email cá nhân của bạn mà là tài khoản do Qwiklabs tạo. Đảm bảo thông tin này khớp với thông tin đăng nhập trong Qwiklabs.

Một số phòng lab có tính năng theo dõi tiến độ, hiển thị điểm số ở góc trên bên phải của cửa sổ Qwiklabs. Điểm số sẽ tăng khi bạn hoàn thành các mục tiêu. Bạn có thể nhấp vào điểm số để xem chi tiết các bước được chấm điểm.

Sau khi hoàn thành phòng lab, điểm số của bạn sẽ được cập nhật. Bạn có thể nhấp vào "Kết thúc Phòng thí nghiệm", xác nhận và cung cấp phản hồi. Sau khi kết thúc phòng lab, dự án do Qwiklabs cung cấp và mọi tài nguyên trong dự án đó sẽ bị xóa. Cuối cùng, đóng trang phòng lab Qwiklabs.

Quay lại trang bài thực hành, bạn sẽ thấy thông báo bạn đã hoàn thành thành công phòng lab. Nếu bạn kiểm tra điểm số của khóa học, bạn sẽ thấy điểm số đầy đủ cho phòng lab này.

Lưu ý quan trọng, hãy luôn khởi chạy Coursera trong cửa sổ ẩn danh và sử dụng thông tin đăng nhập do Qwiklabs cung cấp để đăng nhập vào Cloud Console.

## Tổng quan

Vertex AI là một nền tảng phát triển máy học toàn diện cung cấp cả khả năng AI dự đoán và AI tạo sinh. Nó cho phép bạn đào tạo, đánh giá và triển khai các mô hình máy học dự đoán cho mục đích dự báo. Ngoài ra, bạn có thể sử dụng nền tảng này để khám phá, tinh chỉnh và cung cấp các mô hình AI tạo sinh để tạo ra nội dung.

Vertex AI Studio cho phép bạn nhanh chóng thử nghiệm và tùy chỉnh các mô hình AI tạo sinh để bạn có thể tận dụng khả năng của chúng trong các ứng dụng của mình. Nó cung cấp nhiều công cụ và tài nguyên, bao gồm cả giao diện người dùng (UI) và các ví dụ về mã, giúp bạn dễ dàng bắt đầu với AI tạo sinh, ngay cả khi bạn không có nền tảng về máy học.

Bài thực hành này sẽ hướng dẫn bạn khám phá Vertex AI Studio, nơi bạn sẽ mở khóa tiềm năng của các mô hình AI tạo sinh tiên tiến. Bạn sẽ khám phá Gemini và sử dụng nó để phân tích hình ảnh, thiết kế lời nhắc và tạo các cuộc hội thoại trực tiếp trên bảng điều khiển Google Cloud. Không cần API hoặc SDK Python - tất cả đều có thể truy cập thông qua giao diện người dùng trực quan.

## Mục tiêu

Trong bài thực hành này, bạn sẽ học cách:

- Phân tích hình ảnh bằng Gemini
- Khám phá chế độ Freeform của Vertex AI Studio
- Thiết kế lời nhắc văn bản cho các phương pháp prompting zero-shot, one-shot và few-shot
- Tạo các cuộc hội thoại bằng lời nhắc trò chuyện

## Thiết lập và yêu cầu

### Trước khi bạn nhấp vào nút Start Lab

Hãy đọc những hướng dẫn sau. Các bài thực hành được tính thời gian và bạn không thể tạm dừng chúng. Bộ đếm thời gian, bắt đầu khi bạn nhấp vào Start Lab, hiển thị khoảng thời gian tài nguyên Google Cloud sẽ được cung cấp cho bạn.

Bài thực hành thực hành này cho phép bạn tự thực hiện các hoạt động trong một môi trường đám mây thực, không phải trong môi trường mô phỏng hoặc demo. Điều này được thực hiện bằng cách cung cấp cho bạn thông tin xác thực tạm thời mới mà bạn sử dụng để đăng nhập và truy cập Google Cloud trong thời gian diễn ra bài thực hành.

Để hoàn thành bài thực hành này, bạn cần:

- Truy cập trình duyệt internet tiêu chuẩn (khuyến nghị trình duyệt Chrome).
- **Lưu ý:** Sử dụng cửa sổ trình duyệt ẩn danh hoặc riêng tư để chạy bài thực hành này. Điều này ngăn chặn mọi xung đột giữa tài khoản cá nhân của bạn và tài khoản Sinh viên, có thể gây ra các khoản phí phát sinh cho tài khoản cá nhân của bạn.
- Thời gian để hoàn thành bài thực hành---hãy nhớ rằng, một khi bạn bắt đầu, bạn không thể tạm dừng bài thực hành.
- **Lưu ý:** Nếu bạn đã có tài khoản hoặc dự án Google Cloud cá nhân của riêng mình, đừng sử dụng nó cho bài thực hành này để tránh các khoản phí phát sinh cho tài khoản của bạn.

## Cách bắt đầu bài thực hành của bạn và đăng nhập vào bảng điều khiển Google Cloud

1. Nhấp vào nút **Start Lab**. Nếu bạn cần thanh toán cho bài thực hành, một cửa sổ bật lên sẽ mở ra để bạn chọn phương thức thanh toán. Ở bên trái là bảng **Lab Details** với nội dung sau:
  - Nút **Open Google Cloud console**
  - Thời gian còn lại
  - Thông tin đăng nhập tạm thời mà bạn phải sử dụng cho bài thực hành này
  - Các thông tin khác, nếu cần, để thực hiện bài thực hành này
2. Nhấp vào **Open Google Cloud console** (hoặc nhấp chuột phải và chọn **Open Link in Incognito Window** nếu bạn đang chạy trình duyệt Chrome).

Bài thực hành sẽ khởi tạo tài nguyên và sau đó mở một tab khác hiển thị trang **Sign in**.

**Mẹo:** Sắp xếp các tab trong các cửa sổ riêng biệt, cạnh nhau.

**Lưu ý:** Nếu bạn thấy hộp thoại **Choose an account**, hãy nhấp vào **Use Another Account**. Nếu cần, hãy sao chép **Username** bên dưới và dán vào hộp thoại **Sign in**.

"Username"

Bạn cũng có thể tìm thấy **Username** trong bảng **Lab Details**.

Nhấp vào **Next**.

Sao chép **Password** bên dưới và dán vào hộp thoại **Welcome**.

"Password"

Bạn cũng có thể tìm thấy **Password** trong bảng **Lab Details**.

Nhấp vào **Next**.

**Quan trọng:** Bạn phải sử dụng thông tin đăng nhập mà bài thực hành cung cấp cho bạn. Không sử dụng thông tin đăng nhập tài khoản Google Cloud của bạn.

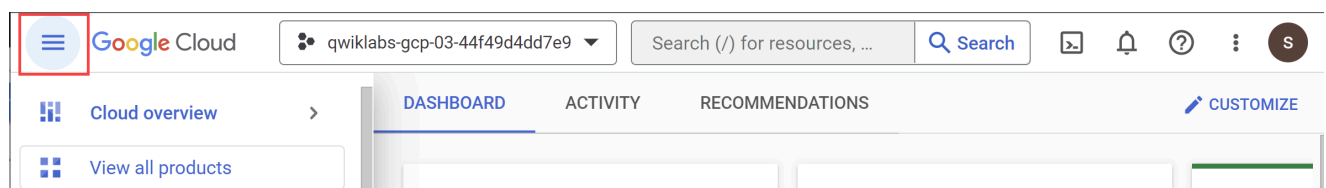
**Lưu ý:** Việc sử dụng tài khoản Google Cloud của riêng bạn cho bài thực hành này có thể phát sinh

Hãy nhấp qua các trang tiếp theo:

- Chấp nhận các điều khoản và điều kiện.
- Không thêm các tùy chọn khôi phục hoặc xác thực hai yếu tố (vì đây là tài khoản tạm thời).
- Không đăng ký các gói dùng thử miễn phí.

Sau vài giây, bảng điều khiển Google Cloud sẽ mở ra trong tab này.

**Lưu ý:** Để xem menu danh sách các sản phẩm và dịch vụ của Google Cloud, hãy nhấp vào menu Điều hướng ở phía trên bên trái.



## Bài 1: Phân tích hình ảnh bằng Gemini ở chế độ Freeform

Trong phần này, bạn sẽ sử dụng Gemini để phân tích một hình ảnh và trích xuất thông tin từ nó. Ở chế độ Freeform, bạn có thể thiết kế các prompt (lời nhắc) cho nhiều tác vụ khác nhau như phân loại, trích xuất và tạo nội dung. Không có lịch sử hội thoại ở chế độ Freeform, vì vậy mỗi prompt là một yêu cầu hoàn toàn mới đối với mô hình.

### Bật API Vertex AI

1. Trong Google Cloud console, nhập **Vertex AI API** vào thanh tìm kiếm trên cùng.
2. Nhấp vào kết quả **Vertex AI API** trong mục **Marketplace & APIs**.
3. Nhấp vào **Enable** (Bật).



## Vertex AI API

[Google Enterprise API](#)

Train high-quality custom machine learning models with minimal machine learning expertise and...

ENABLE

TRY THIS API [↗](#)

## Phân tích hình ảnh bằng Gemini

1. Trong Google Cloud console, từ menu Navigation (biểu tượng menu), chọn **Vertex AI > Vertex AI Studio > Overview**.

## Vertex AI Studio

Test, tune and deploy enterprise-ready generative AI

[Try a tutorial](#)

[Documentation](#)

[API Reference](#)

Generate with Gemini

Open Freeform

Open Chat

Translate text →

Generate images →

Convert or synthesize speech →



Prompt management

Saved prompts and tools to make them better.



Tuning

Customize a model to your task



Evaluation

Understand how a model performs on your data

2. Dưới phần **Generate with Gemini**, nhấp vào **Open Freeform**.

Giao diện người dùng bao gồm ba phần chính:

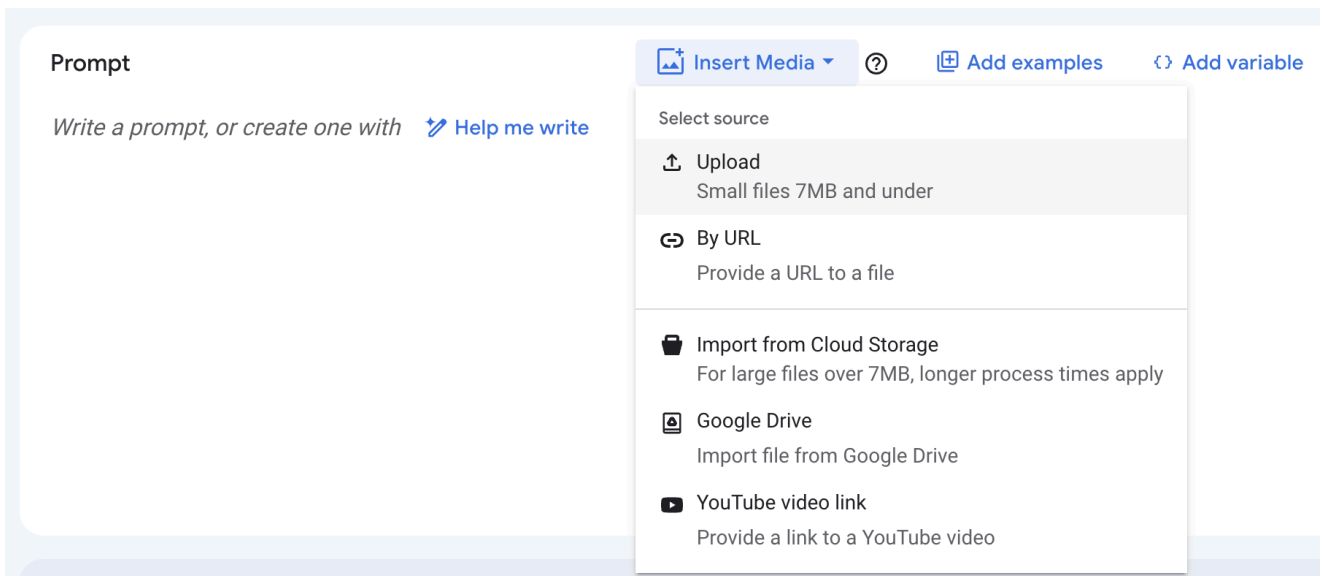
- **Prompt** (nằm ở giữa): Tại đây, bạn có thể tạo một prompt sử dụng các khả năng đa phương thức.
- **Configuration** (nằm ở bên phải): Phần này cho phép bạn chọn mô hình, cấu hình các tham số và lấy mã tương ứng.
- **Response** (nằm ở dưới cùng): Phần này hiển thị kết quả của prompt của bạn.

The screenshot shows the Google AI Studio interface. On the left, the 'Prompt' section is highlighted with a red box. It contains a 'System instructions' field with the text 'Give the model context to understand the task and provide tailored responses'. Below this is a large text area for the prompt, with a 'Help me write' button. On the right, the 'Configuration' section is also highlighted with a red box. It includes settings for the model (gemini-1.5-flash-002), region (us-central1 (Iowa)), temperature (1), output token limit (8192), and grounding (Source: Google Search). There are also options for 'Add stop sequence', 'Output format' (Plain text), and 'Safety Filter Settings'. At the bottom right, there are links for 'Share feedback', 'Report inappropriate responses', and 'View service disclaimers'.

- Ở phía trên bên trái, nhấp vào **Untitled Prompt** và đổi tên prompt của bạn thành **Image Analysis**.
- Trong phần **Configuration** ở phía trên bên phải, nhấp vào menu thả xuống **Model** sau đó chọn mô hình **gemini-1.5-pro-002**.  
*Lưu ý: Tên và phiên bản mô hình có thể thay đổi khi có các mô hình mới được phát hành.*
- Tải xuống ảnh mẫu. Nhấp chuột phải vào hình ảnh bảng thời gian và sau đó lưu nó vào máy tính của bạn.



6. Ở phía trên bên phải của phần **Prompt**, nhấp vào **Insert media > Upload**. Tải lên hình ảnh bằng thời gian mà bạn đã tải xuống. Media có thể ở dạng hình ảnh, video, văn bản hoặc tệp âm thanh.



Hình ảnh sẽ được hiển thị bên trong phần **Prompt**. Sao chép văn bản sau và dán nó bên dưới hình ảnh và nhấp vào nút **Submit** ở phía dưới bên phải của phần **Prompt**.

Title the image.

Hoặc cụ thể hơn:

Title the image in 3 words.

Tiêu đề có đáp ứng mong đợi của bạn không? Hãy thử sửa đổi prompt để xem bạn có nhận được các kết quả khác nhau không.

7. Mô tả hình ảnh. Thay thế prompt trước bằng prompt sau và nhấp vào nút **Submit**.

Describe the image in detail.

8. Điều chỉnh tham số. Trong phần **Configuration**, điều chỉnh nhiệt độ bằng cách cuộn từ trái (0) sang phải (2). Gửi lại prompt để quan sát bất kỳ thay đổi nào trong kết quả so với kết quả trước đó.

*Lưu ý: Nhiệt độ kiểm soát mức độ ngẫu nhiên trong việc lựa chọn token. Nhiệt độ thấp phù hợp với các prompt mong đợi câu trả lời đúng, trong khi nhiệt độ cao hơn có thể dẫn đến kết quả đa dạng hơn, bất ngờ hoặc có khả năng bị sai lệch. Với nhiệt độ là 0, token có xác suất cao nhất luôn được chọn.*

9. Trích xuất văn bản từ hình ảnh. Thay thế prompt trước bằng prompt sau:

Read the text in the image.

10. Tiếp theo, nếu bạn muốn định dạng đầu ra thành danh sách, hãy thay thế prompt trước bằng prompt sau:

Parse the time and city in this image into a list with two columns: time and city.

11. Đến lượt bạn - hãy thử một số prompt khác nhau! Kết quả này khác với trước như thế nào?
12. Phân tích thông tin trên hình ảnh. Thay thế prompt trước bằng prompt sau:

Calculate the percentage of the flights to different continents.

Kết quả có đáp ứng mong đợi của bạn không? Bạn nên thử các prompt khác nhau cho các tác vụ khác nhau. Bạn cũng nên thử nghiệm với các cài đặt nhiệt độ khác nhau để quan sát những thay đổi trong kết quả.

13. Sau khi bạn hoàn thành thiết kế prompt, hãy lưu prompt bằng cách nhấp vào **Save** ở phía trên bên phải của phần **Configuration**. Đối với khu vực, hãy chọn **Region** từ menu thả xuống và sau đó xác nhận bằng cách nhấp vào **Save**.
14. Để tìm các prompt đã lưu của bạn, trên menu điều hướng bên trái, hãy điều hướng đến **Prompt Management**.

*Lưu ý: Sau khi chọn **Save**, hãy cho các prompt vài giây để lưu đúng cách và sau đó tiếp tục với bài thực hành. Nhấp vào "Try again" nếu được nhắc "failed to update history."*

15. Nhấp vào **Check my progress** để xác minh các mục tiêu.

## Bài 2: Khám phá các khả năng đa phương thức

Ngoài hình ảnh, văn bản và âm thanh, Gemini có khả năng chấp nhận video làm đầu vào và tạo ra văn bản làm đầu ra.

- Điều hướng đến **Cloud Storage > Buckets** và sao chép tên của bucket Cloud Storage của bạn và lưu lại để sử dụng trong bước tiếp theo.
- Nhấp vào **Activate Cloud Shell** (biểu tượng Activate Cloud Shell) ở đầu Google Cloud console.
- Trong terminal Cloud Shell của bạn, hãy chạy lệnh bên dưới để sao chép video mẫu `gs://spl/spls/gsp154/video/train.mp4` (xem trước) vào bucket Cloud Storage của bạn. Thay thế `<Your-Cloud-Storage-Bucket>` bằng tên bucket bạn đã sao chép trước đó.



```
gcloud storage cp gs://splis/gsp154/video/train.mp4 gs://<Your-Cloud-Storage-Bucket>
```

*Lưu ý: Đảm bảo thay thế <Your-Cloud-Storage-Bucket> bằng tên bucket của bạn.*

4. Từ menu Navigation (biểu tượng menu), chọn **Vertex AI > Vertex AI Studio > Overview**.

5. Dưới phần **Generate with Gemini**, nhấp vào **Open Freeform**.

6. Ở phía trên bên phải, dưới phần **Model**, chọn mô hình `gemini-1.5-pro-002`.

*Lưu ý: Tên và phiên bản mô hình có thể thay đổi khi có các mô hình mới được phát hành.*

7. Nhấp vào **Insert Media > Import from Cloud Storage**.
8. Nhấp vào tên bucket của bạn, sau đó nhấp vào video mẫu, ví dụ: `train.mp4` và nhấp vào **Select**.
9. Tạo thông tin về video bằng cách chèn prompt sau và nhấp vào nút **Submit**.

Title the video.

Chế độ Freeform cung cấp nhiều khả năng như viết truyện từ hình ảnh, phân tích video và tạo quảng cáo đa phương tiện. Khám phá thêm các trường hợp sử dụng freeform bằng cách nhấp vào **Prompt gallery**.

### Tóm tắt các bước thực hiện:

1. **Lấy tên bucket Cloud Storage:** Điều hướng đến Cloud Storage và sao chép tên bucket của bạn.
2. **Mở Cloud Shell:** Kích hoạt Cloud Shell từ Google Cloud console.
3. **Sao chép video mẫu:** Sử dụng lệnh `gcloud storage cp` để sao chép video `train.mp4` vào bucket của bạn. Đảm bảo thay thế `<Your-Cloud-Storage-Bucket>` bằng tên bucket thực tế của bạn.
4. **Mở Vertex AI Studio Freeform:** Điều hướng đến Vertex AI Studio và mở chế độ Freeform.
5. **Chọn mô hình:** Chọn mô hình `gemini-1.5-pro-002` (hoặc mô hình tương ứng).
6. **Nhập video từ Cloud Storage:** Chèn video `train.mp4` từ bucket Cloud Storage của bạn.
7. **Tạo prompt:** Sử dụng prompt "Title the video." để tạo tiêu đề cho video.
8. **Khám phá thêm:** Tìm hiểu thêm về các khả năng khác của chế độ Freeform thông qua Prompt gallery và tài liệu hướng dẫn.

### Lưu ý:

- Hãy đảm bảo bạn đã bật Vertex AI API trước khi thực hiện các bước này.

- Thay thế `<Your-Cloud-Storage-Bucket>` bằng tên bucket Cloud Storage thực tế của bạn.
- Các tên và phiên bản mô hình có thể thay đổi, hãy sử dụng các thông tin tương ứng nếu có sự thay đổi.

## Bài 3: Thiết kế prompt văn bản

Trong phần này, bạn sẽ khám phá cách thiết kế prompt văn bản trong Vertex AI Studio. Bạn sẽ tìm hiểu về các phương pháp prompting zero-shot, one-shot và few-shot.

### Thiết kế prompt

Bạn có thể cung cấp văn bản đầu vào mong muốn, ví dụ như một câu hỏi, cho mô hình. Sau đó, mô hình sẽ đưa ra phản hồi dựa trên cách bạn cấu trúc prompt của mình. Quá trình tìm hiểu và thiết kế văn bản đầu vào (prompt) tốt nhất để nhận được phản hồi mong muốn từ mô hình được gọi là Thiết kế Prompt (Prompt Design).

#### Các phương pháp thiết kế prompt

Có ba phương pháp chính để thiết kế prompt:

- **Zero-shot prompting:** Đây là phương pháp mà LLM (Mô hình ngôn ngữ lớn) chỉ được cung cấp một prompt mô tả nhiệm vụ và không có dữ liệu bổ sung. Ví dụ: nếu bạn muốn LLM trả lời một câu hỏi, bạn chỉ cần prompt "thiết kế prompt là gì?".
- **One-shot prompting:** Đây là phương pháp mà LLM được cung cấp một ví dụ duy nhất về nhiệm vụ mà nó được yêu cầu thực hiện. Ví dụ: nếu bạn muốn LLM viết một bài thơ, bạn có thể cung cấp cho nó một ví dụ về bài thơ.
- **Few-shot prompting:** Đây là phương pháp mà LLM được cung cấp một số ví dụ nhỏ về nhiệm vụ mà nó được yêu cầu thực hiện. Ví dụ: nếu bạn muốn LLM viết một bài báo, bạn có thể cung cấp cho nó một vài bài báo để đọc.

#### Các tham số

- **Temperature (Nhiệt độ)** và **Token limit (Giới hạn token)** là hai tham số quan trọng mà bạn có thể điều chỉnh để ảnh hưởng đến phản hồi của mô hình.
  - **Temperature** kiểm soát tính ngẫu nhiên trong việc lựa chọn token. Nhiệt độ thấp phù hợp khi bạn mong đợi một câu trả lời đúng. Nhiệt độ 0 có nghĩa là token có xác suất cao nhất luôn được chọn. Nhiệt độ cao hơn có thể dẫn đến kết quả đa dạng, bất ngờ hoặc có khả năng bị sai lệch. Mô hình `gemini-1.5-pro` có phạm vi nhiệt độ từ 0 - 2 và giá trị mặc định là 1.
  - **Output token limit** xác định lượng văn bản tối đa đầu ra từ một prompt. Một token xấp xỉ bốn ký tự.

### Zero-shot prompting

Hãy thử zero-shot prompting trong chế độ Freeform.

1. Điều hướng quay lại trang **Vertex AI Studio > Overview** và nhấp vào **Open Freeform**.
2. Ở phía trên bên phải, dưới phần **Model**, chọn mô hình `gemini-1.5-pro-002`.  
*Lưu ý: Tên và phiên bản mô hình có thể thay đổi khi có các mô hình mới được phát hành.*
3. Sao chép nội dung sau vào trường nhập prompt:

What is a prompt gallery?

4. Nhấp vào nút **Submit**.  
Mô hình sẽ phản hồi bằng một định nghĩa toàn diện về thuật ngữ "prompt gallery".
5. Dưới đây là một số bài tập khám phá để bạn tìm hiểu thêm:
  - Điều chỉnh tham số **Output Token limit** thành 1024 và nhấp vào nút **SUBMIT**.
  - Điều chỉnh tham số **Temperature** thành 0.5 và nhấp vào nút **SUBMIT**.
  - Điều chỉnh tham số **Temperature** thành 2.0 và nhấp vào nút **SUBMIT**.
  - Kiểm tra xem các phản hồi thay đổi như thế nào khi bạn thay đổi các tham số.

## One-shot prompting

Bạn có thể thiết kế prompt theo cách có tổ chức hơn. Bạn có thể cung cấp **Context** (Bối cảnh) và **Examples** (Ví dụ) trong các trường nhập liệu tương ứng. One-shot prompting là phương pháp mà mô hình được cung cấp một ví dụ duy nhất về nhiệm vụ mà nó được yêu cầu thực hiện. Trong phần này, bạn sẽ yêu cầu mô hình hoàn thành một câu.

1. Bắt đầu bằng cách xóa bất kỳ văn bản nào khỏi hộp **Prompt**.
2. Bên trong hộp **Prompt**, nhấp vào **Add examples**. Thao tác này sẽ mở một cửa sổ mới nơi bạn có thể thêm ví dụ cho prompt.

### Examples

Examples help the model understand what an appropriate model response looks like. You can write your own example input and output or use the Test section to save a real response as an example. You can also add a prefix which will be appended to every example (for instance, "question" and "answer").

Import

Add input column

INPUT	OUTPUT
Add an input prefix	Add an output prefix
<div>Write an input to add a new row</div>	<div>Write an output to add a new row</div>

0 tokens

Cancel

Add examples

3. Thêm nội dung này vào trường **INPUT**:

The color of the grass is

4. Thêm nội dung này vào trường **OUTPUT**:

The color of the grass is green

5. Nhấp vào nút **Add examples**.

6. Trong trường **Test**, sao chép nội dung sau vào trường **Input**.

The color of the sky is

7. Nhấp vào nút **Submit**. Bạn sẽ nhận được phản hồi từ mô hình tương tự như sau:

The color of the sky is blue

Thay vì hoàn thành câu, mô hình đã đưa ra một câu đầy đủ làm phản hồi vì bạn đã cung cấp một ví dụ để mô hình dựa vào đó để tạo ra kết quả đầu ra. Để thay đổi phản hồi chỉ đơn giản là hoàn thành câu, bạn có thể điều chỉnh ví dụ được cung cấp trong trường **OUTPUT**.

8. Nhấp vào nút **Examples** trong hộp **Prompt** và thay đổi trường **OUTPUT** thành:

Green

9. Nhấp vào nút **Add examples**.

10. Trong trường **Test**, sao chép nội dung sau vào trường **Input**.

The color of the sky is

11. Nhấp vào nút **Submit**. Bạn sẽ nhận được phản hồi từ mô hình tương tự như sau:

blue

Bạn có thể thấy rằng mô hình giờ đây đã hoàn thành câu dựa trên ví dụ bạn đã cung cấp. Bạn đã thành công trong việc tác động đến cách mô hình tạo ra phản hồi.

## Few-shot prompting

Đối với bài thực hành tiếp theo, bạn sẽ sử dụng mô hình để thực hiện phân tích cảm xúc trên một câu, chẳng hạn như xác định xem đánh giá phim là tích cực hay tiêu cực bằng cách sử dụng few-shot prompting.

1. Trong trường **Prompt**, xóa các ví dụ của bạn từ phần trước. Để xóa các ví dụ của bạn, hãy di chuột qua **Examples** và nhấp vào nút **X (Remove File)**.

Prompt

Insert Media ▾

Examples | 3 added

X

Add variable

Clear Prompt

Remove file

2. Nhấp vào nút **Add examples** để thêm nhiều ví dụ hơn.
3. Thêm các ví dụ sau:

INPUT	OUTPUT
A well-made and entertaining film	positive
I fell asleep after 10 minutes	negative
The movie was ok	neutral

4. Sau khi bạn đã thêm các ví dụ, hãy nhấp vào nút **Add examples**.

Examples

Examples help the model understand what an appropriate model response looks like. You can write your own example input and output or use the Test section to save a real response as an example. You can also add a prefix which will be appended to every example (for instance, "question" and "answer").

Import

Add input column

INPUT	Add an input prefix	OUTPUT	Add an output prefix
A well-made and entertaining film		positive	X
I fell asleep after 10 minutes		negative	X
The movie was ok		neutral	X
<div>Write an input to add a new row</div>		<div>Write an output to add a new row</div>	

40 tokens

Cancel

Add examples

5. Trong trường **Test**, sao chép nội dung sau vào trường **Input**.

It was a time well spent!

6. Nhấp vào nút **Submit**.

Test ?

Batch Testing

Add to examples

53 tokens

INPUT

Add an input prefix

OUTPUT

Add an output prefix

Markdown

It was a time well spent!

positive

Mô hình hiện cung cấp một cảm xúc cho văn bản đầu vào. Đối với văn bản `It was a time well spent!`, cảm xúc được gán nhãn là `positive`.

7. Sau khi bạn hoàn thành thiết kế prompt, hãy đặt tên cho prompt là `Sentiment Analysis`.
8. Nhấp vào **Check my progress** để xác minh các mục tiêu.

### Tóm tắt các bước thực hiện:

#### 1. Zero-shot:

- Mở Freeform, chọn mô hình.
- Nhập câu hỏi đơn giản.
- Thay đổi các tham số (Output Token Limit, Temperature) và quan sát kết quả.

#### 2. One-shot:

- Thêm ví dụ về INPUT và OUTPUT.
- Test với input mới.
- Điều chỉnh ví dụ OUTPUT để thay đổi cách phản hồi.

#### 3. Few-shot:

- Xóa ví dụ cũ.
- Thêm nhiều ví dụ INPUT và OUTPUT cho bài toán phân tích cảm xúc.
- Test với input mới.
- Đặt tên cho prompt và lưu.

## Bài 4: Tạo cuộc trò chuyện

Chế độ Chat là một chế độ trò chuyện cho phép bạn có một cuộc trò chuyện tự do với mô hình. Mô hình sử dụng lịch sử hội thoại làm ngữ cảnh cho các phản hồi trong tương lai. Trong phần này, bạn sẽ tạo một prompt trò chuyện và có một cuộc trò chuyện với mô hình.

1. Từ menu bên trái, điều hướng đến **Chat** để tạo một prompt trò chuyện mới.
2. Ở phía trên bên phải, dưới phần **Model**, chọn mô hình `gemini-1.5-flash-002`.

*Lưu ý: Tên và phiên bản mô hình có thể thay đổi khi có các mô hình mới được phát hành.*

3. Trong phần này, bạn sẽ thêm ngữ cảnh vào cuộc trò chuyện và để mô hình phản hồi dựa trên ngữ cảnh đã cung cấp.
4. Nhấp vào **System instructions** và thêm ngữ cảnh sau:

```
Your name is Roy.
You are a support technician of an IT department.
You only respond with "Have you tried turning it off and on again?" to any queries.
```

5. Chèn prompt sau:

My computer is so slow! What should I do?

6. Nhấp vào nút **Submit**.

Mô hình sẽ trả lời như sau:

Have you tried turning it off and on again?

7. Trong **System instructions**, nhấp vào **Clear value**.

8. Nhấp vào **System instructions** và thêm ngữ cảnh sau:

Your name is Roy.  
You are a support technician of an IT department.  
You are here to support the users with their queries.

9. Chèn prompt sau:

My computer is so slow! What should I do?

10. Nhấp vào nút **Submit**.

Mô hình bây giờ sẽ hữu ích hơn trong phản hồi của mình đối với truy vấn của người dùng.

11. Hãy thoải mái thử nghiệm với các prompt và ngữ cảnh khác nhau để xem mô hình phản hồi như thế nào. Bạn cũng có thể thêm nhiều ngữ cảnh hơn vào prompt trò chuyện để xem mô hình phản hồi dựa trên ngữ cảnh đã cung cấp như thế nào.
12. Sau khi bạn hoàn thành thiết kế prompt, hãy đặt tên cho prompt là **Support Technician Helper**.
13. Nhấp vào **Check my progress** để xác minh các mục tiêu.

**Tóm tắt các bước thực hiện:**

1. **Mở chế độ Chat:** Điều hướng đến mục Chat trong Vertex AI Studio.
2. **Chọn mô hình:** Chọn mô hình **gemini-1.5-flash-002**.
3. **Thiết lập ngữ cảnh (lần 1):** Thêm ngữ cảnh vào phần System instructions:
  - Tên là Roy
  - Kỹ thuật viên hỗ trợ IT
  - Chỉ trả lời: "Have you tried turning it off and on again?"
4. **Gửi prompt (lần 1):** Nhập prompt: "My computer is so slow! What should I do?" và kiểm tra phản hồi.
5. **Xóa ngữ cảnh:** Xóa ngữ cảnh ở phần System instructions.

6. **Thiết lập ngữ cảnh (lần 2):** Thêm ngữ cảnh mới:

- Tên là Roy
- Kỹ thuật viên hỗ trợ IT
- Hỗ trợ người dùng

7. **Gửi prompt (lần 2):** Nhập lại prompt "My computer is so slow! What should I do?" và kiểm tra phản hồi.

8. **Thử nghiệm:** Thử với các prompt và ngữ cảnh khác nhau.

9. **Đặt tên và lưu prompt:** Đặt tên prompt là `Support Technician Helper` và lưu lại.