inspect_rich_documents_w_gemini_multimodality_and_multimodal_rag-v1.0.0

July 15, 2025

1 Inspect Rich Documents with Gemini Multimodality and Multimodal RAG Challenge Lab

Note: if you encounter an authentication error when running the cells in the notebook, go to **Vertex AI** > **Dashboard**, and click on **Enable All Recommended APIs**. Then, re-run the failed cell, and continue the lab.

1.1 Setup and requirements

1.1.1 Install Gen AI SDK for Python and other dependencies

Run the following four cells below before you get to Task 1. Be sure to add your current project ID to the cell titled **Define Google Cloud project information**.

[1]:
```
# "RUN THIS CELL AS IS"

# Install required python packages and other dependencies
%pip install --upgrade --quiet google-genai
%pip install --quiet pymupdf
```

```
Note: you may need to restart the kernel to use updated packages.
Note: you may need to restart the kernel to use updated packages.
```

1.1.2 Restart current runtime

You must restart the runtime in order to use the newly installed packages in this Jupyter runtime. You can do this by running the cell below, which will restart the current kernel.

[2]:
```
# "RUN THIS CELL AS IS"

import IPython

# Restart the kernel after libraries are loaded.

app = IPython.Application.instance()
app.kernel.do_shutdown(True)
```

[2]:
```
{'status': 'ok', 'restart': True}
```

The kernel is going to restart. Please wait until it is finished before continuing to the next step.

1.1.3 Define Google Cloud project information

[1]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

PROJECT_ID = "qwiklabs-gcp-01-2c3e5c2e25b1"  # @param {type:"string"}
LOCATION = "us-west1"  # @param {type:"string"}
```

1.1.4 Initialize Gen AI

Initialize the Gen AI SDK for Python for your project:

[2]:
```
# "RUN THIS CELL AS IS"

# Initialize Gen AI.
from google import genai
client = genai.Client(vertexai=True, project=PROJECT_ID, location=LOCATION)
```

1.2 Task 1. Generating Multimodal Insights with Gemini

Gemini is multimodal model that supports multimodal prompts. You can include text, image(s), and video in your prompt requests and get text or code responses.

To complete Task 1, follow the instructions at the top of each notebook cell: * Run the cells with the comment "RUN THIS CELL AS IS". * Complete and run the cells with the comment "COMPLETE THE MISSING PART AND RUN THIS CELL".

Note: Ensure you can see the weather related data in the response that is printed.

1.2.1 Setup and requirements for Task 1

Import libraries

[3]:
```
from vertexai.generative_models import GenerationConfig, GenerativeModel,␣
↪Image, Part
```

Load Gemini 2.0 Flash Model

[4]:
```
# "RUN THIS CELL AS IS"

multimodal_model = GenerativeModel("gemini-2.0-flash-001")
```

```
/opt/conda/lib/python3.10/site-
packages/vertexai/generative_models/_generative_models.py:433: UserWarning: This
feature is deprecated as of June 24, 2025 and will be removed on June 24, 2026.
For details, see https://cloud.google.com/vertex-ai/generative-
ai/docs/deprecations/genai-vertexai-sdk.
  warning_logs.show_deprecation_warning()
```

Define helper functions

[5]:
```
# "RUN THIS CELL AS IS"

import http.client
import typing
import urllib.request

import IPython.display
from PIL import Image as PIL_Image
from PIL import ImageOps as PIL_ImageOps


def display_images(
    images: typing.Iterable[Image],
    max_width: int = 600,
    max_height: int = 350,
) -> None:
    for image in images:
        pil_image = typing.cast(PIL_Image.Image, image._pil_image)
        if pil_image.mode != "RGB":
            # RGB is supported by all Jupyter environments (e.g. RGBA is not␣
 ↪yet)
            pil_image = pil_image.convert("RGB")
        image_width, image_height = pil_image.size
        if max_width < image_width or max_height < image_height:
            # Resize to display a smaller notebook image
            pil_image = PIL_ImageOps.contain(pil_image, (max_width, max_height))
        IPython.display.display(pil_image)


def get_image_bytes_from_url(image_url: str) -> bytes:
    with urllib.request.urlopen(image_url) as response:
        response = typing.cast(http.client.HTTPResponse, response)
        image_bytes = response.read()
    return image_bytes


def load_image_from_url(image_url: str) -> Image:
    image_bytes = get_image_bytes_from_url(image_url)
    return Image.from_bytes(image_bytes)


def display_content_as_image(content: str | Image | Part) -> bool:
    if not isinstance(content, Image):
        return False
    display_images([content])
    return True
```

```
def display_content_as_video(content: str | Image | Part) -> bool:
    if not isinstance(content, Part):
        return False
    part = typing.cast(Part, content)
    file_path = part.file_data.file_uri.removeprefix("gs://")
    video_url = f"https://storage.googleapis.com/{file_path}"
    IPython.display.display(IPython.display.Video(video_url, width=600))
    return True


def print_multimodal_prompt(contents: list[str | Image | Part]):
    """
    Given contents that would be sent to Gemini,
    output the full multimodal prompt for ease of readability.
    """
    for content in contents:
        if display_content_as_image(content):
            continue
        if display_content_as_video(content):
            continue
        print(content)
```

1.2.2 Task 1.1. Image understanding across multiple images

[6]:
```
# "RUN THIS CELL AS IS"

# You're going to work with provided variables in this task.
# First, review and describe the content/purpose of each variable below.

image_ask_first_1_url = "https://storage.googleapis.com/spls/gsp520/
 ↪Google_Branding/Ask_first_1.png"
image_dont_do_this_1_url = "https://storage.googleapis.com/spls/gsp520/
 ↪Google_Branding/Dont_do_this_1.png"
image_ask_first_1 = load_image_from_url(image_ask_first_1_url)
image_dont_do_this_1 = load_image_from_url(image_dont_do_this_1_url)

instructions = "Instructions: Consider the following image that contains text:"
prompt1 = "What is the title of this image"
prompt2 = """
Answer the question through these steps:
Step 1: Identify the title of each image by using the filename of each image.
Step 2: Describe the image.
Step 3: For each image, describe the actions that a user is expected to take.
Step 4: Extract the text from each image as a full sentence.
Step 5: Describe the sentiment for each image with an explanation.
```

```
Answer and describe the steps taken:
"""
```

Create an input for the multimodal model

[7]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Now, you're going to create an input for your multimodal model. Create your␣
 ↪contents list using the variables above. Ensure the structure matches the␣
 ↪format expected by the multimodal model.

contents = [instructions,
    prompt2,
    image_ask_first_1,
    image_dont_do_this_1]
```

Generate responses from the multimodal model

[8]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# In the next part of this task, you're going to generate responses from the␣
 ↪multimodal model. Capture the output of the model in the "responses"␣
 ↪variable by using your "contents" list.

responses =multimodal_model.generate_content(contents)
```

Display the prompt and responses

[9]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# In the last part of this task, you're going to print your contents and␣
 ↪responses with the prompt and responses title provided. Use descriptive␣
 ↪titles to help organize the output (e.g., "Prompts", "Model Responses") and␣
 ↪then display the prompt and responses by using the print() function.

# Hint: "\n" inserts a newline character for clearer separation between the␣
 ↪sections.

print("=== Prompts ===\n")
print_multimodal_prompt(contents)

print("\n=== Model Responses ===\n")
print(responses.text)
```

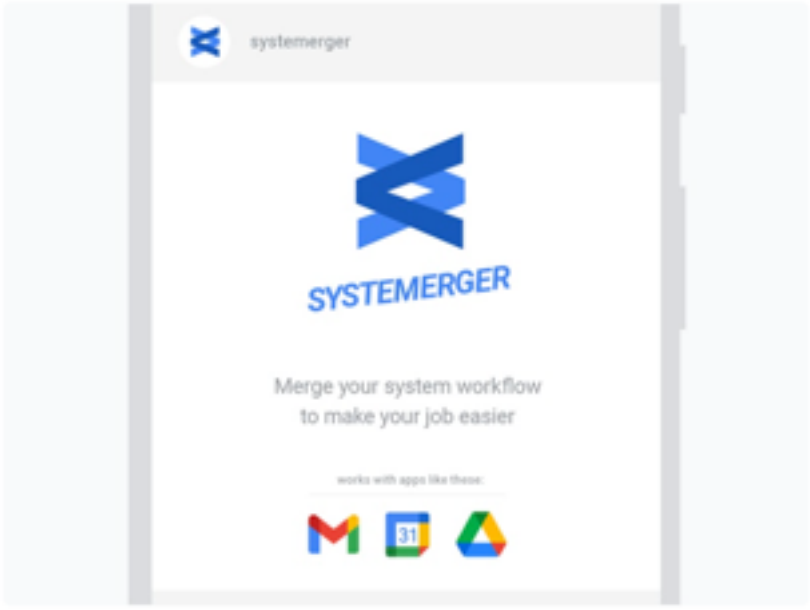
```
=== Prompts ===

Instructions: Consider the following image that contains text:

Answer the question through these steps:
```

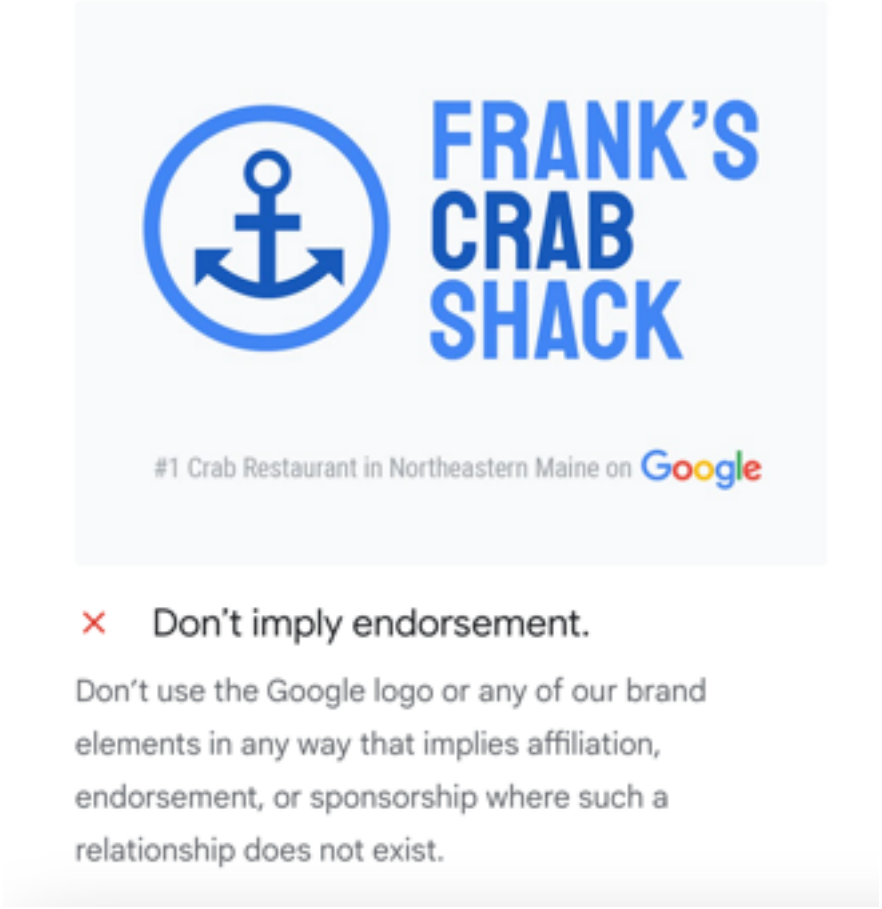
```
Step 1: Identify the title of each image by using the filename of each image.
Step 2: Describe the image.
Step 3: For each image, describe the actions that a user is expected to take.
Step 4: Extract the text from each image as a full sentence.
Step 5: Describe the sentiment for each image with an explanation.

Answer and describe the steps taken:
```



Product icons

Refer to our icon usage guidelines to see whether you can use certain product icons in association with your business.



=== Model Responses ===

Here's a breakdown of the image analysis, following the requested steps:

Image 1 Analysis:

*   **Step 1: Image Title:** Based on the context, there doesn't appear to be an explicit title given other than **Product icons** in the image.

*   **Step 2: Image Description:** The image is a screenshot of a logo design. It includes the logo for "Systemerger," which features an abstract blue icon resembling interlocking shapes. Below the logo is the tagline, "Merge your system workflow to make your job easier." Under the tagline are logos for Gmail, Google Calendar, and Google Drive, indicating app integrations. The overall color scheme is primarily white and blue. The text under the image is a guideline about product icons, urging users to refer to icon usage guidelines to determine if they can use certain product icons in association with their

business.

*   **Step 3: User Actions:** The user is expected to review the example logo. They are also expected to read and understand the guidelines regarding product icon usage. The primary action is to refer to a linked resource ("icon usage guidelines") to determine if they can use such icons.

*   **Step 4: Extracted Text:** "Merge your system workflow to make your job easier" and "works with apps like these:"

*   **Step 5: Sentiment:** The sentiment is neutral and informative. It presents a logo example and offers guidelines, aiming to educate the user on how to appropriately use product icons.

Image 2 Analysis:

*   **Step 1: Image Title:** Based on the context, there doesn't appear to be an explicit title given other than **Don't imply endorsement.** in the image.

*   **Step 2: Image Description:** The image displays a sample logo for "Frank's Crab Shack." The logo features a blue anchor inside a circle, followed by the restaurant's name in a bold, blue font. Below it reads, "#1 Crab Restaurant in Northeastern Maine on Google," with the Google logo. The text under the image is a guideline regarding implying endorsement, instructing users not to use the Google logo or brand elements in a way that suggests affiliation, endorsement, or sponsorship if such a relationship doesn't exist.

*   **Step 3: User Actions:** The user is expected to review the sample logo and understand the guidelines against implying endorsement or affiliation with Google when it doesn't exist.

*   **Step 4: Extracted Text:** "#1 Crab Restaurant in Northeastern Maine on Google"

*   **Step 5: Sentiment:** The sentiment is cautionary and informative. It warns against potentially misleading use of the Google logo or brand elements, ensuring that users do not falsely imply endorsement or sponsorship.

1.2.3 To verify your work for Task 1.1, click Check my progress in the lab instructions.

1.2.4 Task 1.2. Similarity/Differences between images

Explore the variables of the task

[10]:
```
# "RUN THIS CELL AS IS"

# You're going to work with provided variables in this task. First, review and␣
↪describe the content/purpose of each variable below.
```

```
image_ask_first_3_url = "https://storage.googleapis.com/spls/gsp520/
 ↪Google_Branding/Ask_first_3.png"
image_dont_do_this_3_url =  "https://storage.googleapis.com/spls/gsp520/
 ↪Google_Branding/Dont_do_this_3.png"
image_ask_first_3 = load_image_from_url(image_ask_first_3_url)
image_dont_do_this_3 = load_image_from_url(image_dont_do_this_3_url)

prompt1 = """
Consider the following two images:
Image 1:
"""
prompt2 = """
Image 2:
"""
prompt3 = """
1. What is shown in Image 1 and Image 2?
2. What is similar between the two images?
3. What is difference between Image 1 and Image 2 in terms of the text ?
"""
```

Create an input for the multimodal model

[11]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Now, you're going to create an input for your multimodal model. Create your␣
 ↪contents list using the variables above. Ensure the structure matches the␣
 ↪format expected by the multimodal model.

contents = [
    prompt1,
    image_ask_first_3,
    prompt2,
    image_dont_do_this_3,
    prompt3
]
```

Set configuration parameters

[12]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Now, you're going to set configuration parameters that will influence how the␣
 ↪multimodal model generates text. These settings control aspects like the␣
 ↪creativity and focus of the responses. Here's how:
# Temperature: Controls randomness. Lower values mean more predictable results,␣
 ↪higher values mean more surprising and creative output
```

```
# Top p / Top k: Affects how the model chooses words. Explore different values␣
 ↪to see how they change the results.
# Other parameters: Check the model's documentation for additional options you␣
 ↪might want to adjust.

# Store your configuration parameters in a generation_config variable. This␣
 ↪improves reusability, allowing you to easily apply the same settings across␣
 ↪tasks and make adjustments as needed.

generation_config = GenerationConfig(
    temperature=0.4,
    top_p=1.0,
    top_k=40,
    max_output_tokens=1024
)
```

Generate responses from the multimodal model

[13]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# In the next part of this task, you're going to generate responses from a␣
 ↪multimodal model. capture the output of the model in the "responses"␣
 ↪variable by using your "contents" list and the confiuguration settings.


responses = multimodal_model.generate_content(contents)
```

Display the prompt and responses

[14]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# In the last part of this task, you're going to print your contents and␣
 ↪responses with the prompt and responses title provided. Use descriptive␣
 ↪titles to help organize the output (e.g., "Prompts", "Model Responses") and␣
 ↪then display the prompt and responses by using the print() function.

# Hint: "\n" inserts a newline character for clearer separation between the␣
 ↪sections.

print("=== Prompts ===\n")
print_multimodal_prompt(contents)

print("\n=== Model Responses ===\n")
print(responses.text)
```

```
=== Prompts ===

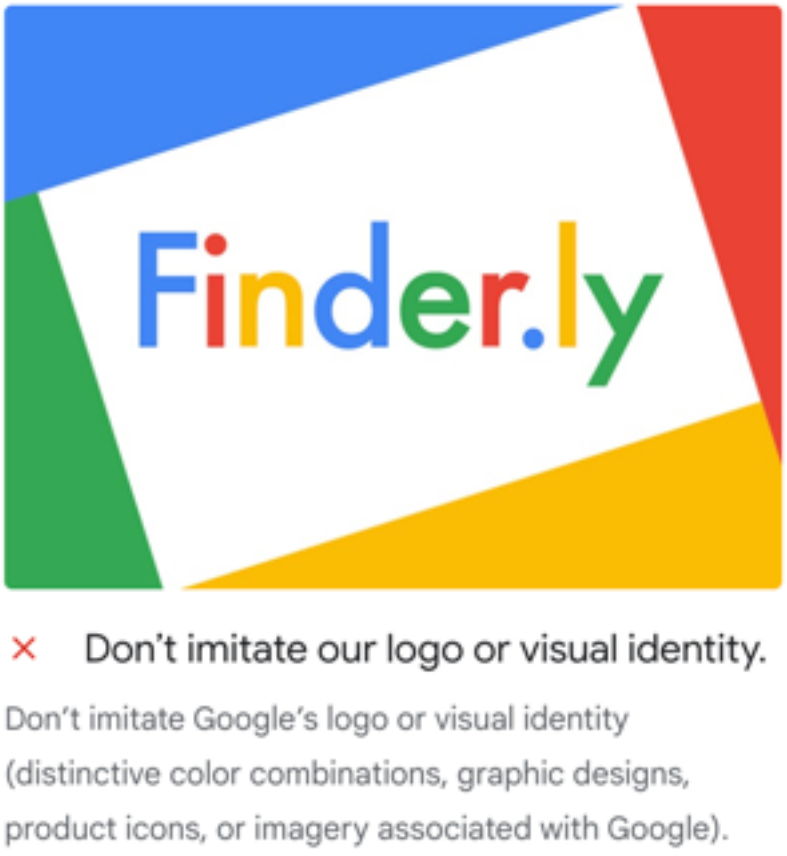

Consider the following two images:
```

Image 1:

Sponsorships and partnerships

If you have an existing sponsorship deal or business relationship with Google, reach out to your Google contact to submit your request internally for review.

Image 2:



1. What is shown in Image 1 and Image 2?
2. What is similar between the two images?
3. What is difference between Image 1 and Image 2 in terms of the text ?

=== Model Responses ===

Here's a breakdown of the images you provided:

1. What is shown in Image 1 and Image 2?

* **Image 1:** Shows an announcement from Wisahawkeen Public Library about a new "Grow with Google" partnership. It features the library's logo, the "Grow with Google" partner logo, and a snippet of text discussing sponsorships and partnerships. The text mentions contacting a Google contact to submit requests internally for review if you have an existing sponsorship deal or business relationship with Google.

```
*   **Image 2:** Displays the "Finder.ly" logo with a multicolored design. It
also has a warning message that reads "Don't imitate our logo or visual
identity" followed by a descriptive text that mentions not imitating Google's
logo or visual identity.

**2. What is similar between the two images?**

*   **Association with Google:** Both images are related to Google in some way.
Image 1 shows a partnership *with* Google, while Image 2 provides guidelines
*about* using Google's visual identity.
*   **Text and Image Combination:** Both images combine an image with
explanatory text.

**3. What is the difference between Image 1 and Image 2 in terms of the text?**

*   **Image 1:** The text provides information about existing sponsorship deals
and partnerships with Google, suggesting that a process exists for review and
approval.
*   **Image 2:** The text is a warning or guideline prohibiting the imitation of
Google's logo and visual identity.
*   **Tone:** The text in Image 1 has an informational tone, while the text in
Image 2 has a cautionary or advisory tone.
```

1.2.5 To verify your work for Task 1.2, click Check my progress in the lab instructions.

1.2.6 Task 1.3. Generate a video description

Explore the variables of the task

[15]:
```
# "RUN THIS CELL AS IS"

# You're going to work with provided variables in this task.
# First, review and describe the content/purpose of each variable below.


prompt = """
What is the product shown in this video?
What specific product features are highlighted in the video?
Where was this video filmed? Which cities around the world could potentially␣
 ↪serve as the background in the video?
What is the sentiment of the video?
"""
video = Part.from_uri(
    uri="gs://spls/gsp520/google-pixel-8-pro.mp4",
    mime_type="video/mp4",
)
```

Create an input for the multimodal model

[16]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Now, you're going to create an input for your multimodal model. Create your␣
 ↪contents list using the variables above. Ensure the structure matches the␣
 ↪format expected by the multimodal model.

contents = [prompt, video]
```

Generate responses from the multimodal model

[17]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# In the next part of this task, you're going to generate responses from a␣
 ↪multimodal model. Capture the output of the model in the "responses"␣
 ↪variable by using your "contents" list.


responses = multimodal_model.generate_content(contents)
```

Display the prompt and responses **Note:** If you encounter any authentication error below cell run, go to the **Navigation menu**, click **Vertex AI > Dashboard**, then click **"Enable all Recommended APIs"** Now, go back to cell 16, and run cells 16, 17 and below cell.

[18]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# In the last part of this task, you're going to print your contents and␣
 ↪responses with the prompt and responses title provided. Use descriptive␣
 ↪titles to help organize the output (e.g., "Prompts", "Model Responses") and␣
 ↪then display the prompt and responses by using the print() function.

# Hint: "\n" inserts a newline character for clearer separation between the␣
 ↪sections.


print("=== Prompts ===\n")
print(prompt)

print("\n=== Model Responses ===\n")
print(responses.text)
```

```
=== Prompts ===


What is the product shown in this video?
What specific product features are highlighted in the video?
Where was this video filmed? Which cities around the world could potentially
serve as the background in the video?
What is the sentiment of the video?
```

```
=== Model Responses ===

Here's a breakdown of the video:

*   **Product:** The video showcases the Google Pixel 8 Pro smartphone.
*   **Highlighted Features:**
    *   AI-enhanced photography features, specifically highlighting the ability
to:
        *   remove unwanted objects from photos using Magic Eraser.
        *   optimize sound in videos.
        *   manipulate faces within group photos to create the best possible
composite image.
*   **Filming Location:**
    *   The video was primarily filmed in London, as evidenced by the
architecture and streetscapes.
    *   Other cities around the world that could potentially serve as a similar
background include European cities like Paris, Rome, or Berlin.
*   **Sentiment:** The overall sentiment is positive and energetic. The video
aims to portray the Google Pixel 8 Pro as a fun, creative, and useful tool for
capturing and enhancing memories.
```

Proceed to Task 1.4 below (no progress check for Task 1.3 in lab instructions).

1.2.7 Task 1.4. Extract tags of objects throughout the video

Explore the variables of the task

[19]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# You're going to work with provided variables in this task. First, review and␣
↪describe the content/purpose of each variable below.

prompt = """
Answer the following questions using the video only:

Which particular sport is highlighted in the video?
What values or beliefs does the advertisement communicate about the brand?
What emotions or feelings does the advertisement evoke in the audience?
Which tags associated with objects featured throughout the video could be␣
↪extracted?
"""
video = Part.from_uri(
    uri="gs://spls/gsp520/google-pixel-8-pro.mp4",
    mime_type="video/mp4",
)
```

Create an input for the multimodal model

[20]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Now, you're going to create an input for your multimodal model. Create your␣
 ↪contents list using the variables above. Ensure the structure matches the␣
 ↪format expected by the multimodal model.

contents = contents = [prompt, video]
```

Generate responses from the multimodal model

[21]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# In the next part of this task, you're going to generate responses from a␣
 ↪multimodal model. capture the output of the model in the "responses"␣
 ↪variable by using your "contents" list and video input.


responses = multimodal_model.generate_content(contents)
```

Display the prompt and responses

[22]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# In the last part of this task, you're going to print your contents and␣
 ↪responses with the prompt and responses title provided. Use descriptive␣
 ↪titles to help organize the output (e.g., "Prompts", "Model Responses") and␣
 ↪then display the prompt and responses by using the print() function.

# Hint: "\n" inserts a newline character for clearer separation between the␣
 ↪sections.

print("=== Prompts ===\n")
print(prompt)

print("\n=== Model Responses ===\n")
print(responses.text)
```

```
=== Prompts ===


Answer the following questions using the video only:

Which particular sport is highlighted in the video?
What values or beliefs does the advertisement communicate about the brand?
What emotions or feelings does the advertisement evoke in the audience?
Which tags associated with objects featured throughout the video could be
extracted?
```

```
=== Model Responses ===

Okay, here are the answers to your questions based solely on the content of the
video:

*   **Which particular sport is highlighted in the video?**
    Basketball is highlighted in the video.

*   **What values or beliefs does the advertisement communicate about the
brand?**
    The advertisement communicates that the brand values creativity, ease of
use, and the ability to enhance and transform photos and videos. It emphasizes
features like AI-powered photo editing, the magic eraser, and the ability to
make photos and videos clearer and more engaging.

*   **What emotions or feelings does the advertisement evoke in the audience?**
    The advertisement evokes feelings of excitement, joy, wonder, and
playfulness. The fast-paced editing, upbeat music, and diverse range of people
create a sense of energy and inspire creativity.

*   **Which tags associated with objects featured throughout the video could be
extracted?**
    *   People: People taking selfies, baby, man shooting basketball.
    *   Technology: Smartphone, phone screen, photo editing app interface.
    *   Objects: Basketball, basketball hoop, buildings, furniture, beanie.

Hope this helps!
```

Proceed to Task 1.5 below (no progress check for Task 1.4 in lab instructions).

1.2.8 Task 1.5. Ask more questions about a video

Note: Although this video contains audio, Gemini does not currently support audio input and will only answer based on the video.

Explore the variables of the task

[24]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# You're going to work with provided variables in this task.
# First, review and describe the content/purpose of each variable below.


prompt = """
Answer the following questions using the video only:

How does the advertisement portray the use of technology, specifically AI, in␣
 ↪capturing and preserving memories?
```

```
What visual cues or storytelling elements contribute to the nostalgic␣
 ↪atmosphere of the advertisement?
How does the advertisement depict the role of friendship and social connections␣
 ↪in enhancing experiences and creating memories?
Are there any specific features or functionalities of the phone highlighted in␣
 ↪the advertisement, besides its AI capabilities?

Provide the answer JSON.
"""
video = Part.from_uri(
    uri="gs://spls/gsp520/google-pixel-8-pro.mp4",
    mime_type="video/mp4",
)
```

Create an input for the multimodal model

[25]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Now, you're going to create an input for your multimodal model. Create your␣
 ↪contents list using the variables above. Ensure the structure matches the␣
 ↪format expected by the multimodal model.

contents = [prompt, video]
```

Generate responses from the multimodal model

[26]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# In the next part of this task, you're going to generate responses from a␣
 ↪multimodal model. capture the output of the model in the "responses"␣
 ↪variable by using your "contents" list and video input.


responses =multimodal_model.generate_content(contents)
```

Display the prompt and responses

[27]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# In the last part of this task, you're going to print your contents and␣
 ↪responses with the prompt and responses title provided. Use descriptive␣
 ↪titles to help organize the output (e.g., "Prompts", "Model Responses") and␣
 ↪then display the prompt and responses by using the print() function.

# Hint: "\n" inserts a newline character for clearer separation between the␣
 ↪sections.
print("=== Prompts ===\n")
print(prompt)
```

```
print("\n=== Model Responses (JSON) ===\n")
print(responses.text)
```

```
=== Prompts ===


Answer the following questions using the video only:

How does the advertisement portray the use of technology, specifically AI, in
capturing and preserving memories?
What visual cues or storytelling elements contribute to the nostalgic atmosphere
of the advertisement?
How does the advertisement depict the role of friendship and social connections
in enhancing experiences and creating memories?
Are there any specific features or functionalities of the phone highlighted in
the advertisement, besides its AI capabilities?

Provide the answer JSON.


=== Model Responses (JSON) ===

Here's the JSON with the answers to your questions, based solely on the provided
video:

```json
{
  "AI_Memory_Preservation": "The advertisement portrays AI as a tool to enhance
and preserve memories by improving image quality (Magic Eraser to remove
unwanted objects), clarifying audio (removing noise from baby sounds), and
ensuring everyone looks their best in group photos.",
  "Nostalgic_Atmosphere": "There are no clear visual cues or storytelling
elements that specifically evoke nostalgia. The advertisement focuses more on
modern, vibrant, and youthful experiences.",
  "Friendship_Social_Connections": "The advertisement highlights friendship and
social connections by showing groups of friends taking selfies and sharing
experiences together, emphasizing how the phone helps capture those moments.
(Group pictures and recording basketball games)",
  "Phone_Features": "Besides AI capabilities, the advertisement highlights the
phone's Magic Eraser (object removal), the ability to clarify audio and the
ability to ensure everyone smiles and looks their best in group photos."
}
```
```

Proceed to Task 1.6 below (no progress check for Task 1.5 in lab instructions).

1.2.9 Task 1.6. Retrieve extra information beyond the video

Explore the variables of the task

[28]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# You're going to work with provided variables in this task.
# First, review and describe the content/purpose of each variable below.


prompt = """
Answer the following questions using the video only:

How does the advertisement appeal to its target audience through its messaging␣
 ↪and imagery?
What overall message or takeaway does the advertisement convey about the brand␣
 ↪and its products?
Are there any symbolic elements or motifs used throughout the advertisement to␣
 ↪reinforce its central themes?
What is the best hashtag for this video based on the description ?

"""
video = Part.from_uri(
    uri="gs://spls/gsp520/google-pixel-8-pro.mp4",
    mime_type="video/mp4",
)
```

Create an input for the multimodal model

[29]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Now, you're going to create an input for your multimodal model. Create your␣
 ↪contents list using the variables above. Ensure the structure matches the␣
 ↪format expected by the multimodal model.

contents = [prompt, video]
```

Generate responses from the multimodal model

[30]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# In the next part of this task, you're going to generate responses from a␣
 ↪multimodal model. capture the output of the model in the "responses"␣
 ↪variable by using your "contents" list and video input.

responses =  multimodal_model.generate_content(contents)
```

Display the prompt and responses

[31]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"
```

```python
# In the last part of this task, you're going to print your contents and␣
 ↪responses with the prompt and responses title provided. Use descriptive␣
 ↪titles to help organize the output (e.g., "Prompts", "Model Responses") and␣
 ↪then display the prompt and responses by using the print() function.

# Hint: "\n" inserts a newline character for clearer separation between the␣
 ↪sections.

print("=== Prompts ===\n")
print(prompt)

print("\n=== Model Responses ===\n")
print(responses.text)
```

```
=== Prompts ===


Answer the following questions using the video only:

How does the advertisement appeal to its target audience through its messaging
and imagery?
What overall message or takeaway does the advertisement convey about the brand
and its products?
Are there any symbolic elements or motifs used throughout the advertisement to
reinforce its central themes?
What is the best hashtag for this video based on the description ?



=== Model Responses ===

Okay, here are the answers to your questions based on the provided video:

*   **How does the advertisement appeal to its target audience through its
messaging and imagery?**

    *   The ad appeals to its target audience by showcasing the Google Pixel 8
Pro's ability to enhance everyday moments through photo editing using AI. It
does this by showing how the phone can "make it pop" with color, "make it clear"
by removing unwanted noise and "change everything" by removing unwanted objects
in pictures. The fast-paced editing process appeals to consumers who want to
enhance their images as fast as possible.

*   **What overall message or takeaway does the advertisement convey about the
brand and its products?**

    *   The overall message is that the Google Pixel 8 Pro, thanks to AI can
```

enhance photos by editing them quickly and effeciently. The phrase "AI, only in Pixel" emphasizes this message by showcasing AI enhancing the images in a snap.

*   **Are there any symbolic elements or motifs used throughout the advertisement to reinforce its central themes?**

    *   The use of AI in the editing and enhancement of photos. The background beat also adds to the enhance photos with AI due to the catchy beat and upbeat tempo.

*   **What is the best hashtag for this video based on the description?**

    *   Based on the video, the best hashtag would be #OnlyInPixel.

1.2.10 To verify your work for Task 1.6, click Check my progress in the lab instructions.

1.3 Task 2. Retrieving and integrating knowledge with multimodal retrieval augmented generation (RAG)

To complete Task 2, follow the instructions at the top of each notebook cell: * Run the cells with the comment "RUN THIS CELL AS IS". * Complete and run the cells with the comment "COMPLETE THE MISSING PART AND RUN THIS CELL".

For additional information about the available data and helper functions for Task 2, review the section named **Available data and helper functions for Task 2** in the lab instructions.

1.3.1 Setup and requirements for Task 2

Import libraries

[32]:
```
# "RUN THIS CELL AS IS"

# Import necessary packages and libraries.

from IPython.display import Markdown, display
from vertexai.generative_models import (
    Content,
    GenerationConfig,
    GenerationResponse,
    GenerativeModel,
    HarmCategory,
    HarmBlockThreshold,
    Image,
    Part,
)
```

Load the Gemini 2.0 Flash model

[33]:
```
# "RUN THIS CELL AS IS"
```

```
# Load the Gemini 2.0 Flash model.

multimodal_model = GenerativeModel("gemini-2.0-flash-001")
```

Download custom Python modules and utilities The cell below will download some helper functions needed for this notebook, to improve readability. You can also view the code (`intro_multimodal_rag_utils.py`) directly on GitHub.

[34]:
```
# "RUN THIS CELL AS IS"

# Import necessary packages and libraries.
import os
import urllib.request
import sys

if not os.path.exists("utils"):
    os.makedirs("utils")


# Download the helper scripts from utils folder.
url_prefix = "https://raw.githubusercontent.com/GoogleCloudPlatform/
 ↪generative-ai/main/gemini/use-cases/retrieval-augmented-generation/utils/"
files = ["intro_multimodal_rag_utils.py"]

for fname in files:
    urllib.request.urlretrieve(f"{url_prefix}/{fname}", filename=f"utils/
 ↪{fname}")
```

Get documents and images from Cloud Storage

[35]:
```
# "RUN THIS CELL AS IS"

# Download documents and images used in this notebook.

!gsutil -m rsync -r gs://spls/gsp520 .
print("Download completed")
```

```
Building synchronization state…
Starting synchronization…
Copying gs://spls/gsp520/Google_Branding/Dont_do_this_3.png…
Copying gs://spls/gsp520/Google_Branding/Ask_first_2.png…
Copying gs://spls/gsp520/Google_Branding/.DS_Store…
Copying gs://spls/gsp520/Google_Branding/Ask_first_3.png…
Copying gs://spls/gsp520/Google_Branding/Dont_do_this_1.png…
Copying gs://spls/gsp520/Google_Branding/Go_for_it_1.png…
Copying gs://spls/gsp520/Google_Branding/Ask_first_1.png…
Copying gs://spls/gsp520/Google_Branding/Dont_do_this_5.png…
Copying gs://spls/gsp520/Google_Branding/Dont_do_this_4.png…
```

```
Copying gs://spls/gsp520/Google_Branding/Go_for_it_2.png…
Copying gs://spls/gsp520/Google_Branding/Dont_do_this_2.png…
Copying gs://spls/gsp520/Google_Branding/Dont_do_this_6.png…
Copying gs://spls/gsp520/Google_Branding/Go_for_it_3.png…
Copying gs://spls/gsp520/Google_Branding/Go_for_it_4.png…
Copying gs://spls/gsp520/Google_Branding/Go_for_it_5.png…
Copying gs://spls/gsp520/Google_Branding/Go_for_it_6.png…
Copying gs://spls/gsp520/google-pixel-8-pro.mp4…
Copying gs://spls/gsp520/Google_Branding/Google_terms_of_service_en_us.pdf…
\ [18/18 files][ 10.8 MiB/ 10.8 MiB] 100% Done
Operation completed over 18 objects/10.8 MiB.
Download completed
```

1.3.2 Task 2.1. Build metadata of documents containing text and images

Note: These steps may take a few minutes to complete.

Import helper functions to build metadata

[36]:
```
# "RUN THIS CELL AS IS"

# Import helper functions from utils.
from utils.intro_multimodal_rag_utils import get_document_metadata
```

```
/opt/conda/lib/python3.10/site-
packages/vertexai/_model_garden/_model_garden_models.py:278: UserWarning: This
feature is deprecated as of June 24, 2025 and will be removed on June 24, 2026.
For details, see https://cloud.google.com/vertex-ai/generative-
ai/docs/deprecations/genai-vertexai-sdk.
  warning_logs.show_deprecation_warning()
```

Explore the variables of the task

[37]:
```
# "RUN THIS CELL AS IS"

# You're going to work with provided variables in this task.
# First, review and describe the content/purpose of each variable below.


# Specify the "PDF folder path" with single PDF and "PDF folder" with multiple␣
↪PDF.

pdf_folder_path = "Google_Branding/"  # if running in Vertex AI Workbench.

# Specify the image description prompt. Change it
image_description_prompt = """Explain what is going on in the image.
If it's a table, extract all elements of the table.
If it's a graph, explain the findings in the graph.
Do not include any numbers that are not mentioned in the image.
"""
```

Extract and store metadata of text and images from a document

[38]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Call the "get_document_metadata" function from the utils file to extract text␣
 ↪and image metadata from the PDF document. Store the results in two different␣
 ↪DataFrames: "text_metadata_df" and "image_metadata_df".
# text_metadata_df: This will contain extracted text snippets, their␣
 ↪corresponding page numbers, and potentially other relevant information.
# image_metadata_df: This will contain descriptions of the images found in the␣
 ↪PDF (if any), along with their location within the document.

text_metadata_df, image_metadata_df = get_document_metadata(
    multimodal_model,
    pdf_folder_path,
    image_save_dir="images",
    image_description_prompt=image_description_prompt,
    embedding_size=1408,
)

print("\n\n --- Completed processing. ---")
```

```
 Processing the file: ----------------------------------
Google_Branding/Google_terms_of_service_en_us.pdf


Processing page: 1
Processing page: 2
Processing page: 3
Processing page: 4
Processing page: 5
Processing page: 6
Processing page: 7
Processing page: 8
Processing page: 9
Processing page: 10
Processing page: 11
Processing page: 12
Processing page: 13
Processing page: 14
Processing page: 15
Processing page: 16


 --- Completed processing. ---
```

Inspect the processed text metadata

[39]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Explore the text_metadata_df dataframe by displaying the first few rows of␣
 ↪the dataframe.

text_metadata_df.head()
```

[39]:

	file_name	page_num	text	text_embedding_page	chunk_number	chunk_text	text_embedding_chunk
0	Google_terms_of_service_en_us.pdf	1	GOOGLE TERMS OF SERVICE\nEffective January 5, …	[-0.012991457246243954, 0.00233408878557384, 0…	1	GOOGLE TERMS OF SERVICE\nEffective January 5, …	[-0.005565959494560957, 0.0008999718702398241,…
1	Google_terms_of_service_en_us.pdf	1	GOOGLE TERMS OF SERVICE\nEffective January 5, …	[-0.012991457246243954, 0.00233408878557384, 0…	2	ongs to you, Google, or others\nIn case of pro…	[-0.02242434397339821, -0.02287789061665535, -…
2	Google_terms_of_service_en_us.pdf	2	Google services are provided by, and youre con…	[-0.027467481791973114, -0.03269978240132332, …	1	Google services are provided by, and youre con…	[-0.030542202293872833, -0.03709811717271805, …
3	Google_terms_of_service_en_us.pdf	2	Google services are provided by, and youre con…	[-0.027467481791973114, -0.03269978240132332, …	2	\nto use our services if you agree to follow t…	[-0.010060024447739124, -0.006401533726602793,…
4	Google_terms_of_service_en_us.pdf	3	apps and sites (like Search and Maps)\nplatfor…	[-0.024841992184519768, -0.016507208347320557,…	1	apps and sites (like Search and Maps)\nplatfor…	[-0.022818822413682938, -0.026401124894618988,…

Import the helper functions to implement RAG

[41]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Import helper functions from utils.
from utils.intro_multimodal_rag_utils import (
```

```
    get_similar_text_from_query,
    print_text_to_text_citation,
    get_similar_image_from_query,
    print_text_to_image_citation,
    get_gemini_response,
    display_images,
)
```

Proceed to Task 2.2 below (no progress check for Task 2.1 in lab instructions).

1.3.3 Task 2.2. Create a user query

Explore the variables of the task

[42]:
```
# "RUN THIS CELL AS IS"

# You're going to work with provided variables in this task.
# First, review and describe the content/purpose of each variable below.

query = """Questions:
 - What are the key expectations that users can have from Google regarding the␣
 ↪provision and development of its services?
- What specific rules and guidelines are established for users when using␣
 ↪Google services?
- How does Google handle intellectual property rights related to the content␣
 ↪found within its services, including content owned by users, Google, and␣
 ↪third parties?
- What legal rights and remedies are available to users in case of problems or␣
 ↪disagreements with Google?
- How do the service-specific additional terms interact with these general␣
 ↪Terms of Service, and which terms take precedence in case of any conflicts?
 """
```

Proceed to Task 2.3 below (no progress check for Task 2.2 in lab instructions).

1.3.4 Task 2.3. Get all relevant text chunks

Retrieve relevant chunks of text based on the query

[46]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Call the "get_similar_text_from_query" function from the utils file to␣
 ↪retrieve relevant chunks of text based on the query. Store the results in a␣
 ↪dictionary called "matching_results_chunks_data".
# matching_results_chunks_data: This dictionary will contain file_name,␣
 ↪page_num, cosine_score, chunk_number and chunk_score. The dictionary␣
 ↪represents a search result for a query related to the text_metadata_df.

matching_results_chunks_data = get_similar_text_from_query(
    query,
```

```
    text_metadata_df,
    column_name="text_embedding_chunk",
    top_n=10,
    chunk_text=True,
)
```

Display the first item of the text chunk dictionary

[47]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Explore the first item in your matching_results_chunks_data dictionary by␣
 ↪displaying the first item.

print_text_to_text_citation(
    matching_results_chunks_data,
    print_top=False,
    chunk_text=True,
)
```

```
Citation 1: Matched text:
Score: 0.87
File name: Google_terms_of_service_en_us.pdf
Page: 1
Page number: 1
Chunk number: 1
Chunk text: GOOGLE TERMS OF SERVICE
Effective January 5, 2022
Archived versions
Whats covered in these terms
We know its tempting to skip these Terms of Service, but
its important to establish what you can expect from us as
you use Google services, and what we expect from you.
These Terms of Service reflect the way Googles business works, the laws that
apply to our company, and
certain things weve always believed to be true. As a result, these Terms of
Service help define Googles
relationship with you as you interact with our services. For example, these
terms include the following topic
headings:
What you can expect from us, which describes how we provide and develop our
services
What we expect from you, which establishes certain rules for using our services
Content in Google services, which describes the intellectual property rights to
the content you find in our
services  whether that content belongs to you, Google, or others
In case of problems or disagreements, which describes other legal rig
Citation 2: Matched text:
Score: 0.83
```

File name: Google_terms_of_service_en_us.pdf
Page: 13
Page number: 13
Chunk number: 1
Chunk text: About these terms
By law, you have certain rights that cant be limited by a contract like these terms of service. These terms are
in no way intended to restrict those rights.
These terms describe the relationship between you and Google. They dont create any legal rights for other
people or organizations, even if others benefit from that relationship under these terms.
We want to make these terms easy to understand, so weve used examples from our services. But not all
services mentioned may be available in your country.
If these terms conflict with the service-specific additional terms, the additional terms will govern for that
service.
If it turns out that a particular term is not valid or enforceable, this will not affect any other terms.
If you dont follow these terms or the service-specific additional terms, and we dont take action right away,
that doesnt mean were giving up any rights that we may have, such as taking action in the future.
We may update these terms and service-speci
Citation 3: Matched text:
Score: 0.8
File name: Google_terms_of_service_en_us.pdf
Page: 2
Page number: 2
Chunk number: 2
Chunk text:
to use our services if you agree to follow these terms, which reflect how Googles business works and how we
earn money. When we speak of Google, we, us, and our, we mean Google LLC and its affiliates.
What you can expect from us
Provide a broad range of useful services
We provide a broad range of services that are subject to these terms, including:

Citation 4: Matched text:
Score: 0.76
File name: Google_terms_of_service_en_us.pdf
Page: 11
Page number: 11
Chunk number: 1
Chunk text: To the extent allowed by applicable law:
Google is liable only for its breaches of these terms or applicable service-

specific additional terms
Google isnt liable for:
loss of profits, revenues, business opportunities, goodwill, or anticipated savings
indirect or consequential losses
punitive damages
Googles total liability arising out of or relating to these terms is limited to the greater of (1) $200 or (2)
the fees paid to use the relevant services in the 12 months before the dispute
For business users and organizations only
If youre a business user or organization:
To the extent allowed by applicable law, youll indemnify Google and its directors, officers, employees,
and contractors for any third-party legal proceedings (including actions by government authorities)
arising out of or relating to your unlawful use of the services or violation of these terms or service-
specific additional terms. This indemnity covers any liability or expense arising from claims, losses,
damages, judgm
Citation 5: Matched text:
Score: 0.75
File name: Google_terms_of_service_en_us.pdf
Page: 13
Page number: 13
Chunk number: 2
Chunk text: s that we may have, such as taking action in the future.
We may update these terms and service-specific additional terms (1) to reflect changes in our services or how
we do business for example, when we add new services, features, technologies, pricing, or benefits (or
remove old ones), (2) for legal, regulatory, or security reasons, or (3) to prevent abuse or harm.
If we materially change these terms or service-specific additional terms, well provide you with reasonable
advance notice and the opportunity to review the changes, except (1) when we launch a new service or
feature, or (2) in urgent situations, such as preventing ongoing abuse or responding to legal requirements. If
you dont agree to the new terms, you should remove your content and stop using the services. You can also
end your relationship with us at any time by closing your Google Account.
DEFINITIONS
affiliate
An entity that belongs to the Google group of companies, which means Google LLC and its subsidiaries,
includi
Citation 6: Matched text:

Score: 0.74
File name: Google_terms_of_service_en_us.pdf
Page: 3
Page number: 3
Chunk number: 2
Chunk text: to our services, and start offering new services or stop
offering old ones. When a service requires or includes downloadable software,
that software sometimes
updates automatically on your device once a new version or feature is available.
Some services let you adjust
your automatic update settings.
If we make material changes that negatively impact your use of our services or
if we stop offering a service,
well provide you with reasonable advance notice, except in urgent situations
such as preventing abuse,
responding to legal requirements, or addressing security and operability issues.
Well also provide you with an
opportunity to export your content from your Google Account using Google
Takeout, subject to applicable
law and policies.
What we expect from you
Follow these terms and service-specific additional terms
The permission we give you to use our services continues as long as you comply
with:
these terms

Citation 7: Matched text:
Score: 0.74
File name: Google_terms_of_service_en_us.pdf
Page: 4
Page number: 4
Chunk number: 1
Chunk text: service-specific additional terms, which could, for
example, include things like additional age
requirements
You also agree that our Privacy Policy applies to your use of our services.
Additionally, we provide resources
like the Copyright Help Center, Safety Center, and descriptions of our
technologies from our policies site to
answer common questions and to set expectations about using our services.
Although we give you permission to use our services, we retain any intellectual
property rights we have in the
services.
Respect others
We want to maintain a respectful environment for everyone, which means you must
follow these basic rules
of conduct:
comply with applicable laws, including export control, sanctions, and human

trafficking laws
respect the rights of others, including privacy and intellectual property rights
dont abuse or harm others or yourself (or threaten or encourage such abuse or harm) for example, by
misleading, defrauding, illegally impersonating, defaming, bullying,
Citation 8: Matched text:
Score: 0.74
File name: Google_terms_of_service_en_us.pdf
Page: 5
Page number: 5
Chunk number: 1
Chunk text: to provide. If you choose to upload or share content,
please make sure you have the necessary rights to do so
and that the content is lawful.
License
Your content remains yours, which means that you retain any intellectual property rights that you
have in your content. For example, you have intellectual property rights in the creative content you
make, such as reviews you write. Or you may have the right to share someone elses creative content
if theyve given you their permission.
We need your permission if your intellectual property rights restrict our use of your content. You
provide Google with that permission through this license.
Whats covered
This license covers your content if that content is protected by intellectual property rights.
Whats not covered
This license doesnt affect your privacy rights its only about your intellectual property rights
This license doesnt cover these types of content:
publicly-available factual information that you provide, such as corrections to the
Citation 9: Matched text:
Score: 0.74
File name: Google_terms_of_service_en_us.pdf
Page: 7
Page number: 7
Chunk number: 1
Chunk text: using content youve shared publicly to promote the
services. For example, to promote a
Google app, we might quote a review you wrote. Or to promote Google Play, we might show a
screenshot of the app you offer in the Play Store.
developing new technologies and services for Google consistent with these terms
Duration
This license lasts for as long as your content is protected by intellectual

property rights.
If you remove from our services any content thats covered by this license, then our systems will stop
making that content publicly available in a reasonable amount of time. There are two exceptions:
If you already shared your content with others before removing it. For example, if you shared a
photo with a friend who then made a copy of it, or shared it again, then that photo may
continue to appear in your friends Google Account even after you remove it from your Google
Account.
If you make your content available through other companies services, its possible that search
engines, inc
Citation 10: Matched text:
Score: 0.73
File name: Google_terms_of_service_en_us.pdf
Page: 8
Page number: 8
Chunk number: 1
Chunk text: Using Google services on behalf of an organization or
business
Many organizations, such as businesses, non-profits, and schools, take advantage of our services. To use
our services on behalf of an organization:
an authorized representative of that organization must agree to these terms
your organizations administrator may assign a Google Account to you. That administrator might
require you to follow additional rules and may be able to access or disable your Google Account.
Service-related communications
To provide you with our services, we sometimes send you service announcements and other information. To
learn more about how we communicate with you, see Googles Privacy Policy.
If you choose to give us feedback, such as suggestions to improve our services, we may act on your feedback
without obligation to you.
Content in Google services
Your content
Some of our services give you the opportunity to make your content publicly available for example, you
might post a product or restaurant

Proceed to Task 2.4 below (no progress check for Task 2.3 in lab instructions).

1.3.5 Task 2.4. Create context text

Create a list to store the combined chunks of text

[48]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"


# Create an empty list named "context_text". This list will be used to store␣
 ↪the combined chunks of text.
context_text = list()
```

Iterate through each item in the text chunks dictionary

[49]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Create a for loop to iterate through each item in the␣
 ↪matching_results_chunks_data dictionary in order to combine all the selected␣
 ↪relevant text chunks

for key, value in matching_results_chunks_data.items():
    context_text.append(value["chunk_text"])
```

Join all the text chunks and store in a list

[50]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Take all of the individual text chunks stored in the context_text list and␣
 ↪join them together into a single string named final_context_text. Use "\n"␣
 ↪part inserts a newline character between each chunk, effectively creating␣
 ↪separate lines or paragraphs.

final_context_text = "\n".join(context_text)
```

Proceed to Task 2.5 below (no progress check for Task 2.4 in lab instructions).

1.3.6 Task 2.5. Pass context to Gemini

Explore the variables of the task

[51]:
```
# "RUN THIS CELL AS IS"

# You're going to work with provided variables in this task. First, review and␣
 ↪describe the content/purpose of each variable below.

prompt = f""" Instructions: Compare the images and the text provided as Context:
 ↪ to answer multiple Question:
Make sure to think thoroughly before answering the question and put the␣
 ↪necessary steps to arrive at the answer in bullet points for easy␣
 ↪explainability.
If unsure, respond, "Not enough context to answer".

Context:
 - Text Context:
 {final_context_text}
```

```
{query}

Answer:
"""
```

Generate Gemini response with streaming output

[52]:
```
# "COMPLETE THE MISSING PART AND RUN THIS CELL"

# Call "get_gemini_response" function from utils module in order to generate␣
 ↪Gemini response with streaming output. This function uses a multimodal␣
 ↪Gemini model, a text prompt, and configuration parameters and instructs the␣
 ↪Gemini model to generate a response using the provided prompt. As Gemini␣
 ↪model enables streaming, you will receive chunks of the response as they␣
 ↪were produced.
# Format the streamed output using Markdown syntax for easy readability and␣
 ↪conversion to HTML.

Markdown(
    get_gemini_response(
        multimodal_model,
        model_input=[prompt],
        stream=True,
        generation_config=GenerationConfig(temperature=0.2,␣
 ↪max_output_tokens=2048),
    )
)
```

[52]:
Here's a breakdown of the answers to your questions, based on the provided Google Terms of Service excerpt:

- **What are the key expectations that users can have from Google regarding the provision and development of its services?**
  - Google provides a broad range of useful services.
  - Google will provide reasonable advance notice if they make material changes that negatively impact the use of services or if they stop offering a service, except in urgent situations.
  - Google will provide an opportunity to export content from a Google Account using Google Takeout, subject to applicable law and policies, if they make material changes or stop offering a service.
  - Google may update its software automatically when a new version or feature is available.
- **What specific rules and guidelines are established for users when using Google services?**
  - Users must follow the Terms of Service and any service-specific additional terms.

  - Users must comply with applicable laws, including export control, sanctions, and human trafficking laws.
  - Users must respect the rights of others, including privacy and intellectual property rights.
  - Users must not abuse or harm others or themselves.
  - If users upload or share content, they must have the necessary rights to do so and ensure the content is lawful.

- **How does Google handle intellectual property rights related to the content found within its services, including content owned by users, Google, and third parties?**
  - Users retain intellectual property rights in their content.
  - Google requires a license from users to use their content if intellectual property rights restrict Google's use.
  - This license covers content protected by intellectual property rights but doesn't affect privacy rights.
  - If users remove content, Google's systems will stop making it publicly available in a reasonable time, with exceptions for content already shared or available through other services.
  - Google retains any intellectual property rights it has in the services.

- **What legal rights and remedies are available to users in case of problems or disagreements with Google?**
  - Users have certain legal rights that cannot be limited by the Terms of Service.
  - Google is liable only for its breaches of the terms or applicable service-specific additional terms.
  - Google isn't liable for loss of profits, revenues, business opportunities, goodwill, or anticipated savings; indirect or consequential losses; or punitive damages.
  - Google's total liability is limited to the greater of $200 or the fees paid for the relevant services in the 12 months before the dispute.
  - Business users and organizations indemnify Google for third-party legal proceedings arising from their unlawful use of the services or violation of the terms.

- **How do the service-specific additional terms interact with these general Terms of Service, and which terms take precedence in case of any conflicts?**
  - Users must follow both the general Terms of Service and any service-specific additional terms.
  - If the general Terms of Service conflict with the service-specific additional terms, the additional terms will govern for that service.

1.3.7 To verify your work for Task 2.5, click Check my progress in the lab instructions.