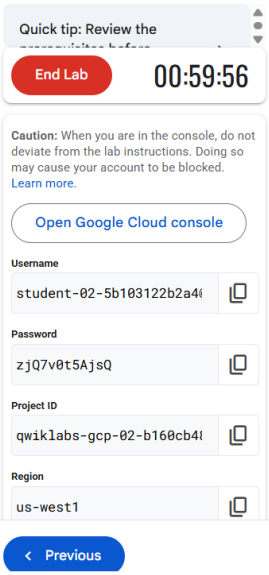


Multimodal Retrieval Augmented Generation (RAG) using the Gemini API in Vertex AI

Lab 1 hour No cost Intermediate

Rate Lab

This lab may incorporate AI tools to support your learning.

GSP1231

Google Cloud Self-Paced Labs

Overview

Gemini is a family of generative AI models developed by Google DeepMind that is designed for multimodal use cases.

LLMs to access external data and also as a mechanism for grounding to mitigate against hallucinations. RAG models are trained to retrieve relevant documents from a large corpus and then generate a response based on the retrieved documents. In this lab, you learn how to perform multimodal RAG where you perform Q&A over a financial document filled with both text and images.

Comparing text-based and multimodal RAG

Multimodal RAG offers several advantages over text-based RAG:

1. **Enhanced knowledge access**: Multimodal RAG can access and process both textual and visual information, providing a richer and more comprehensive knowledge base for the LLM.
2. **Improved reasoning capabilities**: By incorporating visual cues, multimodal RAG can make better informed inferences across different types of data modalities.

and multimodal embeddings, to build a document search engine.

Prerequisites

Before starting this lab, you should be familiar with:

- Basic Python programming.
- General API concepts.
- Running Python code in a Jupyter notebook on Vertex AI Workbench.

Objectives

In this lab, you learn how to:

- Extract and store metadata of documents containing both text and images, and generate embeddings the documents
- Search the metadata with text queries to find similar text or images
- Search the metadata with image queries to find similar images
- Using a text query as input, search for contextual answers using both text and images

Setup and requirements

Read these instructions. Labs are timed and you cannot pause them. The timer, which starts when you click **Start Lab**, shows how long Google Cloud resources are made available to you.

This hands-on lab lets you do the lab activities in a real cloud environment, not in a simulation or demo environment. It does so by giving you new, temporary credentials you use to sign in and access Google Cloud for the duration of the lab.

To complete this lab, you need:

- Access to a standard internet browser (Chrome browser recommended).

Note: Use an Incognito (recommended) or private browser window to run this lab. This prevents conflicts between your personal account and the student account, which may cause extra charges incurred to your personal account.

Note: Use only the student account for this lab. If you use a different Google Cloud account, you may incur charges to that account.

How to start your lab and sign in to the Google Cloud console

1. Click the **Start Lab** button. If you need to pay for the lab, a dialog opens for you to select your payment method. On the left is the Lab Details pane with the following:
    - The Open Google Cloud console button
    - Time remaining
    - Other information, if needed, to step through this lab
2. Click **Open Google Cloud console** (or right-click and select **Open Link in Incognito Window** if you are running the Chrome browser).

    The lab spins up resources, and then opens another tab that shows the Sign in page.

    Tip: Arrange the tabs in separate windows, side-by-side.

    Note: If you see the **Choose an account** dialog, click **Use Another Account**.

3. If necessary, copy the **Username** below and paste it into the **Sign in** dialog.

```
student-02-5b103122b2a4@qwiklabs.net
```

4. Click **Next**.

5. Copy the **Password** below and paste it into the **Welcome** dialog.

```
zjQ7v0t5AjsQ
```

You can also find the Password in the Lab Details pane.

6. Click **Next**.

Important: You must use the credentials the lab provides you. Do not use your Google Cloud account credentials.

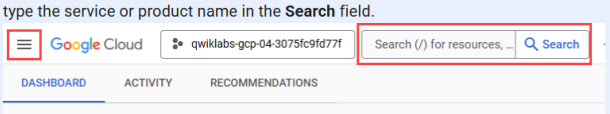
Note: Using your own Google Cloud account for this lab may incur extra

7. Click through the subsequent pages:

   - Accept the terms and conditions.
   - Do not add recovery options or two-factor authentication (because this is a temporary account).
   - Do not sign up for free trials.

After a few moments, the Google Cloud console opens in this tab.

Note: To access Google Cloud products and services, click the **Navigation menu** or type the service or product name in the **Search** field.



Task 1. Open the notebook in Vertex AI Workbench

1. In the Google Cloud console, on the **Navigation menu** (≡), click **Vertex AI > Workbench**.
2. Find the `vertex-ai-jupyterlab` instance and click on the **Open JupyterLab** button.

The JupyterLab interface for your Workbench instance opens in a new browser tab.

Task 2. Set up the notebook

1. Open the `intro_multimodal_rag` file.
2. In the **Select Kernel** dialog, choose **Python 3** from the list of available kernels.
3. Run through the **Getting Started** and the **Import libraries** sections of the notebook.
   - For **Project ID**, use `qwiklabs-gcp-02-b160cb4808d6`, and for **Location**, use `us-west1`.

Note: You can skip any notebook cells that are noted *Colab only*. If you experience a

Click **Check my progress** to verify the objective.

Install GenAI SDK for Python and import libraries

Check my progress

In the following sections, you run through the notebook cells to see how to use the Gemini API to build a multimodal RAG system.

Task 3. Download custom

The Gemini 2.0 Flash (`gemini-2.0-flash`) model is designed to handle natural language tasks, multiturn text and code chat, and code generation. In this section, you 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.

1. In this task, run through the notebook cells to load the model and download the helper functions and get the documents and images from Cloud Storage.

Click **Check my progress** to verify the objective.

Download image and documents from Cloud Storage

Check my progress

Task 4. Build metadata of documents containing text and images

The source data that you use in this lab is a modified version of Google-10K which provides a comprehensive overview of the company's financial performance, business operations, management, and risk factors. As the original document is rather large, you will be using a modified version with only 14 pages, split into two parts - Part 1 and Part 2 instead. Although it's truncated, the sample document still contains text along with images such as tables, charts, and graphs.

1. In this task, run through the notebook cells to extract and store metadata of text and images from a document.

Note: The cell to to extract and store metadata of text and images from a document

Click **Check my progress** to verify the objective.

Extract and store metadata of text and images from a document

Check my progress

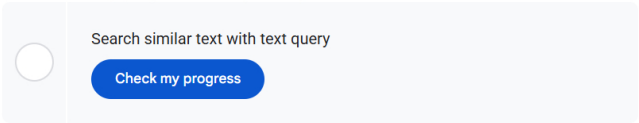
Task 5. Text Search

Let's start the search with a simple question and see if the simple text search using text

embeddings can answer it. The expected answer is to show the value of basic and

1. In this task, run through the notebook cells to search for similar text and images with a text query.

Click **Check my progress** to verify the objective.



Task 6. Image Search

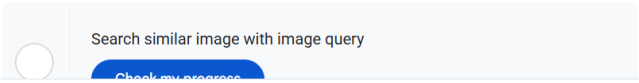
you want to find other images that look like it, from the same document or across multiple documents.

The ability to identify similar text and images based on user input, powered with Gemini and embeddings, forms a crucial foundation for the development of multimodal RAG systems, which explore in the next task.

1. In this task, run through the notebook cells to search for similar images with an image query.

Note: You may need to wait for a couple of minutes to get the score for this task.

Click **Check my progress** to verify the objective.

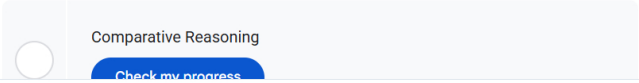


Comparative Reasoning

Imagine we have a graph showing how Class A Google shares did compared to other things like the S&P 500 or other tech companies. You want to know how Class C shares did compared to that graph. Instead of just finding another similar image, you can ask Gemini to compare the relevant images and tell you which stock might be better for you to invest in. Gemini would then explain why it thinks that way.

1. In this task, run through the notebook cells to compare two images and find the most similar image.

Click **Check my progress** to verify the objective.



Task 7. Multimodal retrieval augmented generation (RAG)

Let's bring everything together to implement multimodal RAG. You use all the elements that you've explored in previous sections to implement the multimodal RAG. These are the steps:

- **Step 1:** The user gives a query in text format where the expected information is available in the document and is embedded in images and text.
- **Step 2:** Find all text chunks from the pages in the documents using a method similar

`image_description` using a method identical to the one you explored in `Image Search`.

- **Step 4:** Combine all similar text and images found in steps 2 and 3 as `context_text` and `context_images`.
- **Step 5:** With the help of Gemini, we can pass the user query with text and image context found in steps 2 & 3. You can also add a specific instruction the model should remember while answering the user query.
- **Step 6:** Gemini produces the answer, and you can print the citations to check all relevant text and images used to address the query.

1. In this task, run through the notebook cells to perform multimodal RAG.

Note: You may need to wait for a couple of minutes to get the score for this task.

Click **Check my progress** to verify the objective.

Print the citations to check all relevant text and images

Congratulations!

In this lab, you've learned to build a robust document search engine using Multimodal Retrieval Augmented Generation (RAG). You learned how to extract and store metadata of documents containing both text and images, and generate embeddings for the documents. You also learned how to search the metadata with text and image queries to find similar text and images. Finally, you learned how to use a text query as input to search for contextual answers using both text and images.

Check out the following resources to learn more about Gemini:

- Gemini Overview
- Generative AI on Vertex AI Documentation
- Generative AI on YouTube
- Explore the Vertex AI Cookbook for a curated, searchable gallery of notebooks for Generative AI.
- Explore other notebooks and samples in the Google Cloud Generative AI repository.

Google Cloud training and certification

...helps you make the most of Google Cloud technologies. Our classes include technical skills and best practices to help you get up to speed quickly and continue your learning journey. We offer fundamental to advanced level training, with on-demand, live, and virtual options to suit your busy schedule. Certifications help you validate and prove your skill and expertise in Google Cloud technologies.

Manual Last Updated May 15, 2025

Lab Last Tested May 15, 2025

in Vertex AI