Objectives

This tutorial walks you through using the Gemini API in Vertex AI via the Vertex AI SDK for Python. You will:

- Install the Vertex AI SDK for Python
- Use Gemini API to interact with the Gemini 2.0 Flash model
- Generate text from text prompts
- Explore configuration options
- Generate text from image(s) + text
- Generate text from video + text

---

1. Installation of Vertex AI SDK for Python

To install or upgrade to the latest version:

%pip install --upgrade --user google-cloud-aiplatform

> ✅ *Note: Your environment shows that version 1.91.0 was successfully installed.*

---

2. What You'll Learn (Essential for GenAI Students)

a. Working with Gemini Models in Vertex AI

- Using `gemini-2.0-flash` for fast and cost-effective generation.
- Best suited for multi-modal inputs (text, image, video).

b. Text Generation from Prompts

- How to call the model with a simple text input.
- Modify temperature, max tokens, top-k, top-p for tuning generation.

c. Multi-Modal Input

- Learn how to:
    - Combine text and image as input.
    - Use video snippets as part of the prompt for enriched responses.

d. Real-World Use Cases

- Resume summarization from uploaded image
- Product description generation from videos
- Image captioning
- Education bots that can read diagrams or explain screenshots

---

3. Recommended Skills to Develop in GenAI

- **Python Programming**
- **Google Cloud Platform (GCP)**
  - IAM, Billing, Vertex AI setup
- **Prompt Engineering**
  - Experimenting with prompt design for best model responses
- **Handling Multimedia Inputs**
  - Convert images/videos to base64
  - Use `PIL`, `OpenCV`, etc.
- **Model Evaluation & Fine-Tuning**
  - Understand response quality and refine prompts
- **Ethics and Safety in AI**
  - Avoid misuse of generated content
  - Bias detection

---

4. Learning Resources

- Google Cloud documentation on Vertex AI
- Kaggle notebooks and competitions
- GenAI specialization courses (Coursera, Udacity)
- Open-source LLMs (Gemma, Mistral, LLaMA) for local testing

✨ Gemini Flash, Movie Recommender & Multimodal Prompting – Student Guide

---

🔹 1. Getting Started with Gemini 1.5 Flash via Vertex AI SDK

Key Topics:

- Using Gemini Flash with Vertex AI Python SDK
- Prompting basics
- Chat & streaming modes
- Safety settings
- Prompt tuning
- Bonus: Movie recommender use case

Notebook Summary:

- **Install Vertex AI SDK**
- **Initialize Vertex AI**
- **Send simple prompts** like: `"Why is the sky blue?"`
- **Use chat mode:** Model remembers past messages
- **Enable streaming:** Real-time responses
- **Safety filters:** Filter out harmful/inappropriate content
- **Prompt tuning:** Adapt output to a specific tone (e.g., always reply like a pirate 🏴‍☠️)
- **Fun demo:** Recommend a movie based on user favorites

📎 *Link to Colab Notebook: (Insert your link here)*

🔹 2. Movie Recommender Use Case: Example Interaction

User Input:
"My name is Ned. You are my personal assistant. My favorite movies are *The Lord of the Rings* and *The Hobbit*."

Model Suggests: *Willow (1988)*

🔍 Why this recommendation?

- **High Fantasy World:** Quest, magic, evil force – a lighthearted version of LOTR
- **Magical Creatures & World-building:** Unique races and mystical lands
- **Hero's Journey:** A reluctant hero, Willow, must protect a child destined to defeat an evil queen
- **Old-school Effects:** Uses practical effects and puppetry like early LOTR films
- **Directed by Ron Howard, story by George Lucas (Star Wars)**

📘 Follow-up Prompt:

"Are my favorite movies based on a book series?"

Response:

> Yes, *The Lord of the Rings* and *The Hobbit* are both based on books by **J.R.R. Tolkien**.
>
> - LOTR = Trilogy: *The Fellowship of the Ring*, *The Two Towers*, *The Return of the King*
> - The Hobbit = Prequel to LOTR

🔹 3. Image + Text Prompting (Multimodal)

Gemini supports combining **text** + **images** as prompts. You can load images via:

- Local files
- Cloud Storage URI (`gs://`)
- Direct image links (URL)

Example Prompt:

Prompt: *“Describe this image?”*
Image: Tabby cat in snow

Response:

- Cat is mid-step with a raised paw
- Yellow-green eyes, tabby stripes
- Background: snowy with tire tracks

🛠️ Helper Functions to Know

- `display_images()` – display images in notebooks
- `get_image_bytes_from_url()` – load image from URL
- `load_image_from_url()` – convert image to Gemini-compatible format
- `get_url_from_gcs()` – convert GCS URI to HTTP URL

🔹 4. Few-Shot Prompting with Multiple Images

You can do:

- **Few-shot learning** with labeled examples

- **Multi-image input** in a single prompt

Example:

json
CopyEdit

```
{"city": "London", "Landmark": "Big Ben"}
{"city": "Paris", "Landmark": "Eiffel Tower"}
```

Prompt Gemini to recognize a third image using these as examples.

🔚 Conclusion

This guide walks you through:
✅ Gemini Flash setup
✅ Movie recommender logic
✅ Book vs movie context
✅ Multimodal prompting
✅ Writing structured prompts for both image & text tasks

Takeaways for Students

- You can query multimedia files (images, videos) using structured prompts and get structured outputs like JSON.
- Google Gemini supports direct file ingestion via `gs://` or `https://` and can generate bounding boxes and labels for objects.
- With proper schemas, you can *control* and *validate* the output format using `response_schema` and `generation_config`.
- This capability is especially useful in applications like:
    - Resume parsing and career suggestion
    - Media analysis and summarization
    - Visual QA systems
    - Surveillance and event detection