

Lab: Multimodal Prompts

Estimated Time: 30 mins

Learning Objectives

By the end of this lab, you'll be able to:

- Summarize multimodal content from a PDF document using generative AI
- Analyze and extract key data points from an image using multimodal prompt techniques
- Generate a creative narrative or description inspired by a visual image

Introduction

Multimodal prompts is one of the key advancements in the constantly evolving field of AI. It enables generative AI (GenAI) models to leverage content from different modalities (such as text and images) and provide a more robust, more structured response to the user queries.

In this lab, you will make use of publicly available GenAI tools like OpenAI's [ChatGPT](#) and Google's [Gemini](#) to explore how these tools interpret and respond to multimodal inputs. You will experiment with combining text and images in your prompts to observe how each tool handles contextual understanding, reasoning, and creativity across different modalities. By the end of the lab, you will gain practical experience in crafting effective multimodal prompts and develop a deeper understanding of the capabilities and limitations of current multimodal AI systems. For the tasks included in the lab, you can work with either of the two tools: ChatGPT or Gemini.

Important Note: All documents and images used in the examples that follow are synthetically generated for the purpose of this lab.

Task 1: Multimodal PDF content analysis

Scenario: You are part of a research team analyzing how GenAI is being adopted in everyday tasks. Your team has received a PDF survey report titled "Adoption of GenAI in Day-to-Day Tasks," which includes a mix of text, tables, and charts. In this task, you will use a multimodal GenAI model to extract insights from both the textual and visual elements of the document. Your goal is to understand how well the model interprets complex, multimodal content and to evaluate its ability to summarize findings, identify trends, and answer specific questions based on the PDF.

The required PDF document is available [here](#).

[Adoption of GenAI in Day-to-Day Tasks](#)

Exemplar prompts

1. Sample prompt to summarize the multimodal document.

Briefly Summarize the key findings of this PDF survey, including both the text and any visual data like charts or tables.

► [Click here for a sample response](#)

2. Sample prompt to extract statistic on most frequently performed daily tasks where users utilize GenAI, based on the information in the PDF.

What are the most common day-to-day tasks where users report adopting GenAI, based on the content of this document?

► [Click here for a sample response](#)

3. Sample prompt to generate recommendations to improve GenAI adoption

Based on the survey results in this PDF, what recommendations would you make to improve GenAI adoption in daily workflows?

► [Click here for a sample response](#)

Task 2: Image analysis

Scenario: Personal expense tracking

You are part of a consumer analytics team for a grocery chain. You've been given the receipt (as an image) from a customer's purchase at Fresh Foods Market. Your goal is to extract insights from the receipt to support various business and customer service objectives, such as inventory planning, price comparison, tax calculations, and trend analysis.

The image of the receipt is shared below. You can copy the image and paste it in the message box of a generative AI tool.

Fresh Foods Market			
789 Farmer's Road, Springfield, USA			
Phone: (555) 987-6543			
Date: 2024-07-26		Time: 10:22 AM	
Item	Qty	Price	Total
Organic Strawberries	1 lb	\$4.00/lb	\$4.00
Bananas	4	\$0.75/ea	\$3.00
Broccoli Crowns	1.2 lb	\$2.99/lb	\$3.59
Avocado	2	\$1.75/ea	\$3.50
Almond Milk (Half Gall)	1	\$3.80	\$3.80
Greek Yogurt	3	\$2.00/ea	\$6.00
Chicken Breast	1.5 lb	\$7.99/lb	\$11.99
Brown Rice Pasta	1	\$2.50	\$2.50
Fair Trade Coffee	1	\$12.00	\$12.00
Kalamata Olives	1	\$5.75	\$5.75
Organic Spinach	1 bag	\$3.29	\$3.29
Cheddar Cheese	8 oz	\$6.50	\$6.50
Whole Wheat Bread	1 loaf	\$4.20	\$4.20
Ground Beef	2 lb	\$8.50/lb	\$17.00
Large Eggs	1 dozen	\$3.75	\$3.75
Subtotal:			\$90.87
Sales Tax (6%):			\$5.45
TOTAL:			\$96.32
Paid with: Debit Card			
Card Number: XXXX-XXXX-XXXX-9876			
Transaction ID: 654321			
Customer Service: support@freshfoods.com			
Thank you for shopping at Fresh Foods Market!			

Exemplar prompts

1. Sample prompt to extract data from the image.

Extract the list of all items purchased, including quantities, unit prices, and total prices.

► Click here for the sample response

2. Sample prompt for conducting mathematical reasoning to calculate the average price per item category.

Calculate the average price per item category (produce, dairy, meat, pantry).

► Click here for the sample response

3. Sample prompt for decision-making and optimization: A prompt to analyze the purchased items and suggest which could be omitted or replaced with cheaper alternatives to stay within budget (of \$80).

Imagine this shopper had a strict budget of \$80 for this trip. Based on the items purchased, which items, if any, could they have potentially o

► [Click here for the sample response](#)

Task 3: Creative content generation

Scenario: You are a content designer for an immersive narrative app that generates personalized creative content based on photos users upload. Your task is to develop a written piece, inspired by a mysterious image sent by a user — a serene alpine landscape at sunset, with a small cabin, a glowing campfire, and four people gathered around it.

The relevant image is shared below. You can copy the image and paste it in the message box of a generative AI tool.



Exemplar prompts

1. Sample prompt to generate a short story.

Write a short story inspired by this image. Focus on the people around the campfire and what brings them to this remote location at sunset.

► [Click here for the sample response](#)

2. Sample prompt to generate a poetic response.

Generate a poem or a lyrical piece that captures the mood and beauty of this landscape at dusk, reflecting on themes of nature, connection, or solitude.

► [Click here for the sample response](#)

Conclusion

This lab demonstrated the power of multimodal artificial intelligence, showing how GenAI models can process and understand information not just from text, but also from images and complex documents like PDFs that combine both. By using multimodal prompts, you experienced firsthand how AI can interpret visual content, extract information from integrated text and images, and generate creative responses that draw upon both seeing and reading. This capability reflects GenAI's growing ability to interact with data in ways that are closer to human perception, opening up possibilities for richer analysis, more intuitive information retrieval, and innovative content creation from diverse sources.

Author(s)

[Abhishek Gagneja](#)



Skills Network