

# Lesson 03 – Multimodal Prompting

*Prompt Engineering Mini-Series*

---

## 1. What & Why

Combine two or more input/output modalities (text, images, audio) to enrich context or produce new media.

---

Business gains	Examples
Faster insight	Screenshot dashboard “Summarize key trends.”
Creative assets	Text prompt generate banner ad images.
Accessibility	Auto-generate alt-text for uploaded images.

---

## 2. Supported Directions

---

Input → Output	Use-cases
<b>Image → Text</b>	Describe, tag, extract data, critique design
<b>Text → Image</b>	Generate diagrams, mock-ups, icons
<b>Image+Text → Text</b>	“Given chart + caption, write insights”
<b>Text → Image → Text</b>	Storyboard loop

---

## 3. Core Prompt Patterns

### A. Describe an Image

<attach image>

You are a UX researcher.

Describe layout, color palette, notable usability issues.

Output: 1) Description, 2) Issues list.

### B. Guided Image Generation

Generate an isometric icon set:

• cloud database • AI chatbot • vet clinic POS

Style: flat, pastel, minimal shadows.

Return only a DALL·E prompt.

### C. Image + Text Task

<attach product photo>

Specs: 10 cm × 8 cm × 2 cm, food-grade silicone.  
Write Amazon listing title ( 140 chars) + 5 bullet features.

---

## 4. Hands-On Templates

### Template A – Dashboard Insight Extractor

<attach dashboard screenshot>

Role: Data analyst.  
1 Summarize key trends ( 120 words).  
2 Rate each metric: Up / Down / Flat.  
3 Suggest action for lowest metric.

### Template B – Design Consistency Checker

<attach two mobile screens>

As senior UX reviewer, list up to 5 inconsistencies  
(Issue | Screen # | Fix 12 words).

### Template C – Diagram Generator & Explainer

Task: High-level architecture diagram for {{PROJECT\_DESCRIPTION}} ( 6 components).

- 1 Draft DALL·E prompt for SVG-style monochrome diagram.
  - 2 Provide 100-word explanation for non-technical stakeholder.
- 

## 5. Cheat-Sheet

Multimodal Prompting  
Pattern → (Img/PDF/Text)   Task   Output spec

Vision tips  
• Crop clutter • Ask for structured JSON if needed

Image-gen tips  
• Style + subject + composition + lighting + medium

---

## 6. Caveats

- Vision accuracy falls with tiny text / clutter – zoom or crop.
  - Upload limit ~4 MB per image (GPT-vision).
  - Redact sensitive info; generated images may require brand post-editing.
- 

## 7. 60-Minute Class Agenda

---

Min	Activity
0-5	Mini-lecture
5-20	Hands-on A
20-35	Hands-on B
35-50	Hands-on C
50-55	Q & A
55-60	Wrap-up & survey

---