



**INO AI LAB**

[ AI EDUCATION • COURSE MATERIAL ]

# Gemini & Google AI Stack

*Gemini 2.5 Pro, AI Studio, Vertex*

LEVEL	DURATION	LESSONS
<b>Beginner</b>	<b>3 hours</b>	<b>7</b>



[ 00 ]

# Table of Contents

---

LESSON 01	<b>Gemini's Multimodal Edge</b>
LESSON 02	<b>Notebooks, Deep Research &amp; Workspace AI</b>
LESSON 03	<b>Function Calling &amp; Structured Output</b>
LESSON 04	<b>Veo, Imagen, and Generative Media on Google</b>



[ LESSON 01 ]

# Gemini's Multimodal Edge

Gemini 2.5 Pro's killer feature is native multimodality at scale — it accepts text, images, audio, and video in a single prompt with a 2M-token context window. Drop in a one-hour meeting recording and ask for action items; paste 50 product screenshots and ask for a UX audit. No other frontier model in 2026 matches this raw input bandwidth, and it changes which workflows are even possible.

Access it free via Google AI Studio for prototyping, then move to Vertex AI for production with proper logging, IAM, and enterprise SLAs. Studio is the fastest path from idea to working prompt; Vertex is the right home for anything customer-facing. Don't ship Studio API keys from a Studio account to production — quotas and stability differ meaningfully between the two surfaces.

## // KEY TAKEAWAYS

- › 2M context + native multimodal is unique.
- › AI Studio to prototype, Vertex to ship.
- › Don't ship Studio keys to production.

[ LESSON 02 ]

# Notebooks, Deep Research & Workspace AI

NotebookLM is the underrated star of the Gemini family. Upload 50 sources — PDFs, slides, websites, YouTube videos — and it builds a grounded knowledge base you can query, mind-map, or convert to an audio overview podcast. It's the best tool in 2026 for learning a new domain fast or briefing a team on a complex topic without writing the brief yourself.

Deep Research runs multi-step web investigations and returns a structured report with citations — typically 20–40 minutes for executive-grade output. In Workspace, Gemini is embedded across Gmail, Docs, Sheets, and Meet; the highest-leverage use is 'Help me write' for first drafts and meeting recap generation, which alone can return the subscription cost in a month.

## // KEY TAKEAWAYS

- › NotebookLM = grounded learning + audio overviews.
- › Deep Research replaces analyst briefs.
- › Workspace AI pays for itself in meeting recaps.

[ LESSON 03 ]

## Function Calling & Structured Output

Gemini's function-calling API is mature and reliable. Define your tools with JSON schemas, the model picks which to call and with what arguments, and you execute them in your code. This is how you turn a chat model into a backend assistant that books appointments, queries databases, or controls IoT devices. Always validate arguments before executing — models occasionally hallucinate fields under load.

For pure structured output without tools, use the `responseSchema` parameter to force JSON conforming to your schema. This is more reliable than prompt-only JSON requests, especially for nested objects and enum fields. Keep schemas as flat as possible; deeply nested schemas degrade both accuracy and latency. Version your schemas alongside your code so prompt and parser evolve together.

### // KEY TAKEAWAYS

- › Function calling turns chat into a backend tool.
- › Validate arguments before executing.
- › Flat schemas beat deeply nested ones.

[ LESSON 04 ]

## Veo, Imagen, and Generative Media on Google

Veo 3 generates short videos with synced audio from text or image prompts — quality is competitive with Sora 2 and noticeably better at physical realism. Imagen 4 handles stills with strong typography rendering, finally making it viable for posters, packaging mockups, and ad creative without a designer. Both ship through the Gemini API and the Vertex Media Studio.

Treat generative video like film direction: write shot lists, specify camera moves, name lenses, lock down lighting language. Vague prompts produce generic clips; specific prompts produce assets a marketing team can actually ship. Always run human review for brand safety and IP — generative tools will happily produce trademark-adjacent characters or unsafe combinations if you don't constrain them.

### // KEY TAKEAWAYS

- › Veo 3 leads on physical realism + audio sync.
- › Imagen 4 finally handles typography well.
- › Direct video like a film: shot, lens, lighting.



[ NEXT ]

# Keep Going

---

You've completed this course material. The real learning starts when you apply what you've read. Pick one idea from this PDF and run a small experiment this week. Document what worked and what didn't. Share your findings with the community.

Explore more free courses, daily AI tips, and curated tools at:

[innovationailab.com](https://innovationailab.com)

Have feedback or want to suggest a topic? We read every message.

[hello@innovationailab.com](mailto:hello@innovationailab.com)

— Innovation AI Lab Team —

// part of LumiLife Tech