

Multimodal AI: When AI Has Eyes and Ears

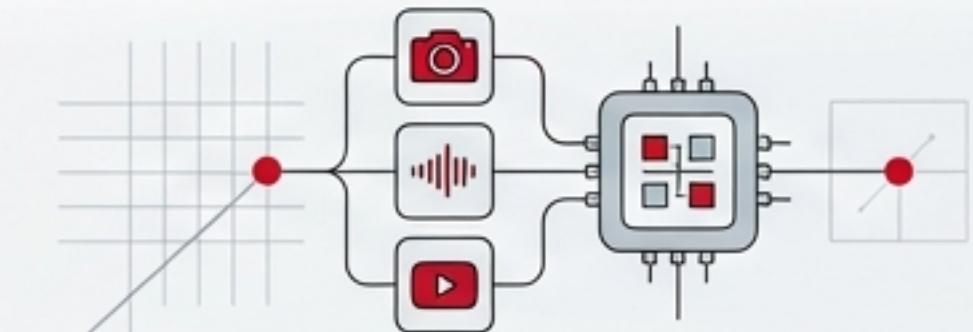
Analyzing and Creating with
Images, Audio, and Video

UBUS 670 | AI for Business Leaders | Day 4



1

Explain how multimodal AI processes images, audio, and video—not just text.



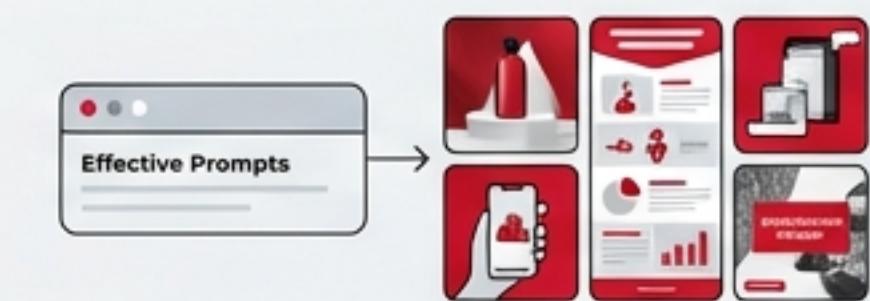
2

Analyze business content across multiple modalities using Gemini.



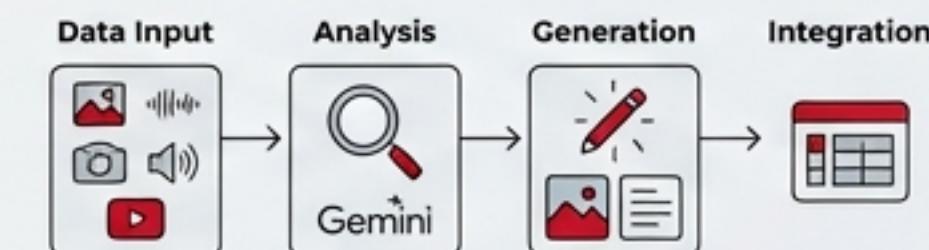
3

Generate marketing visuals using AI image generation with effective prompts.



4

Design a structured multimodal workflow that combines analysis and generation.



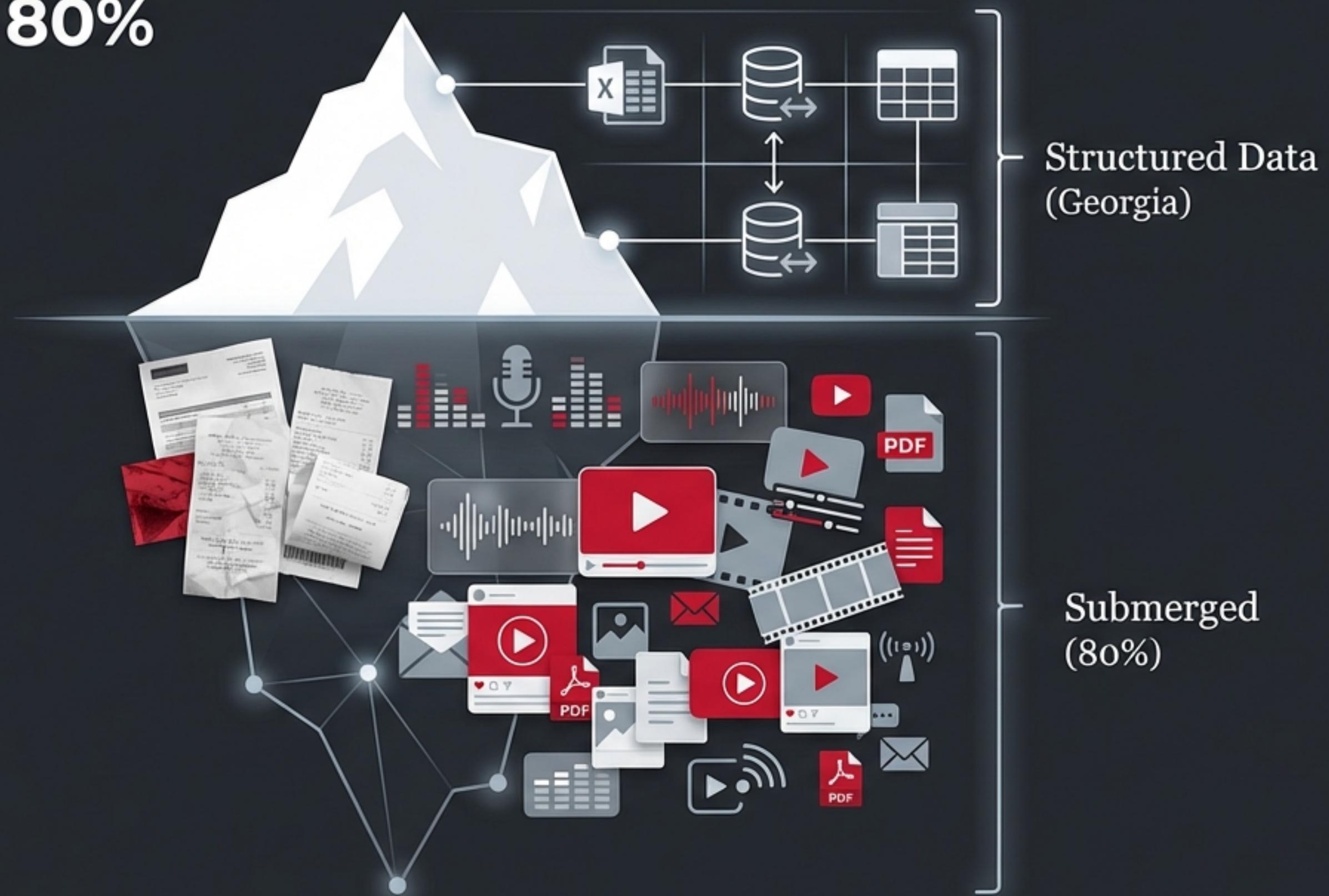
Key Skill: Multimodal AI: From Text → Images, Audio, Video

THE INVISIBLE 80%

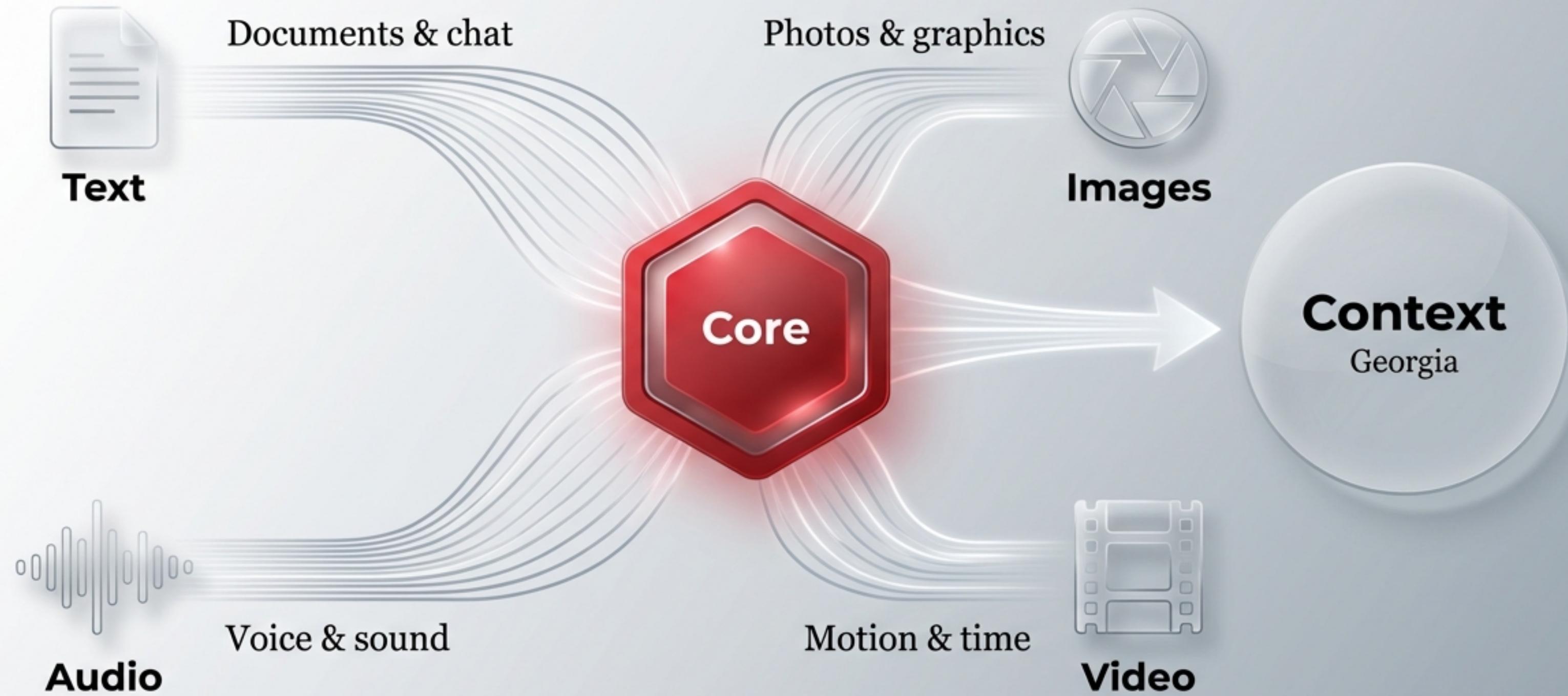
80%

of business data
is unstructured.

→ Photos, Audio, Video

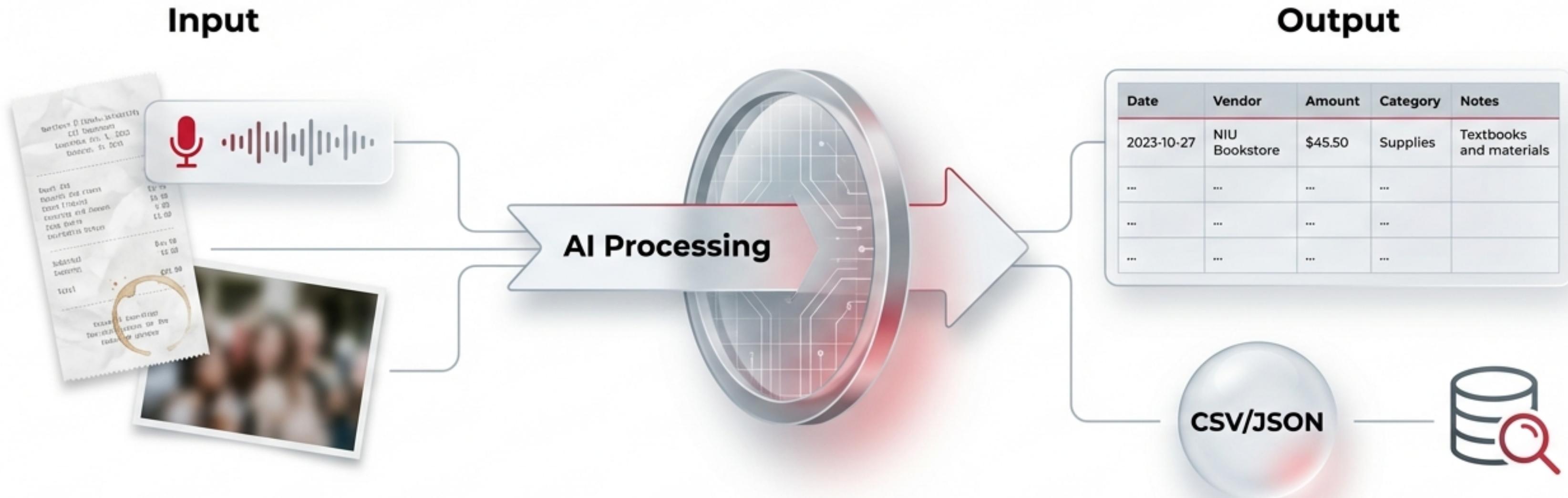


The Multimodal Hub



Multimodal AI processes and understands all inputs simultaneously.

Structured Outputs from Unstructured Media



“The modality changes; the workflow stays the same.”

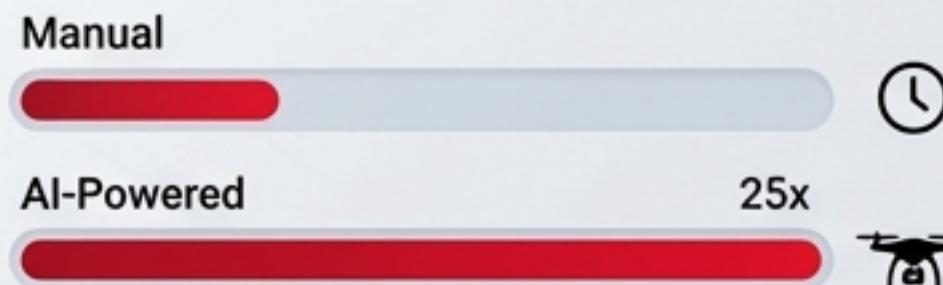
Real-World Multimodal Intelligence

Corporate editorial and high-tech consultancy

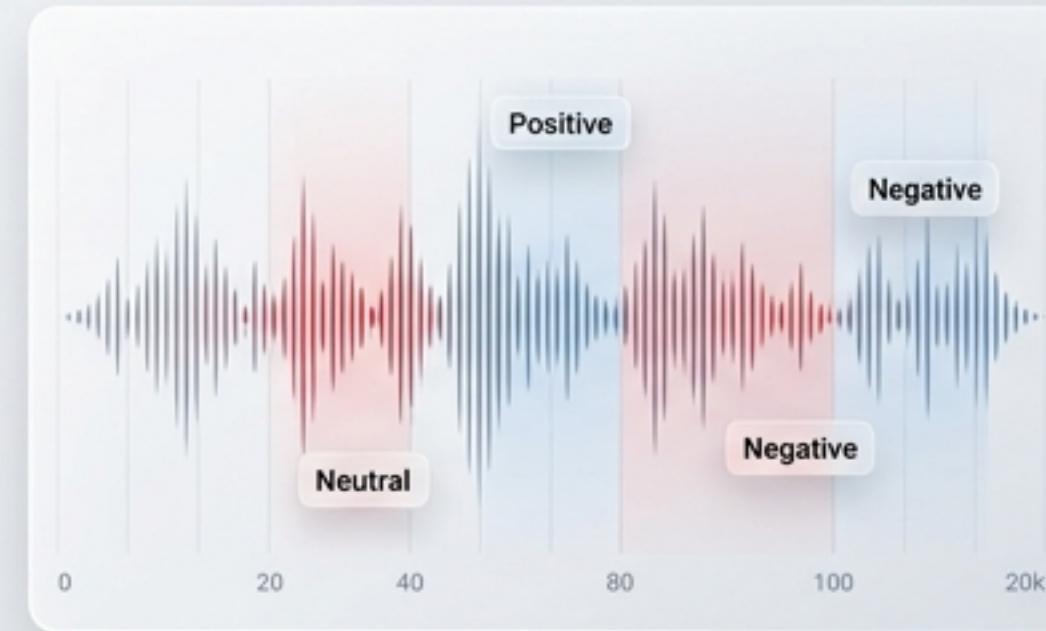
Vision (Warehouse)



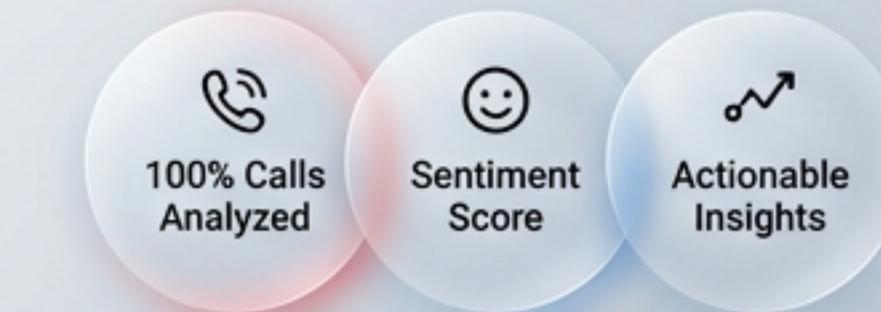
Gather AI: 25x faster
inventory counting.



Audio (Call Center)



Observe.AI: 100% of calls
analyzed for sentiment.



Video (Retail)



RetailNext: Heatmaps to
optimize layout.



Image Understanding: What AI Sees

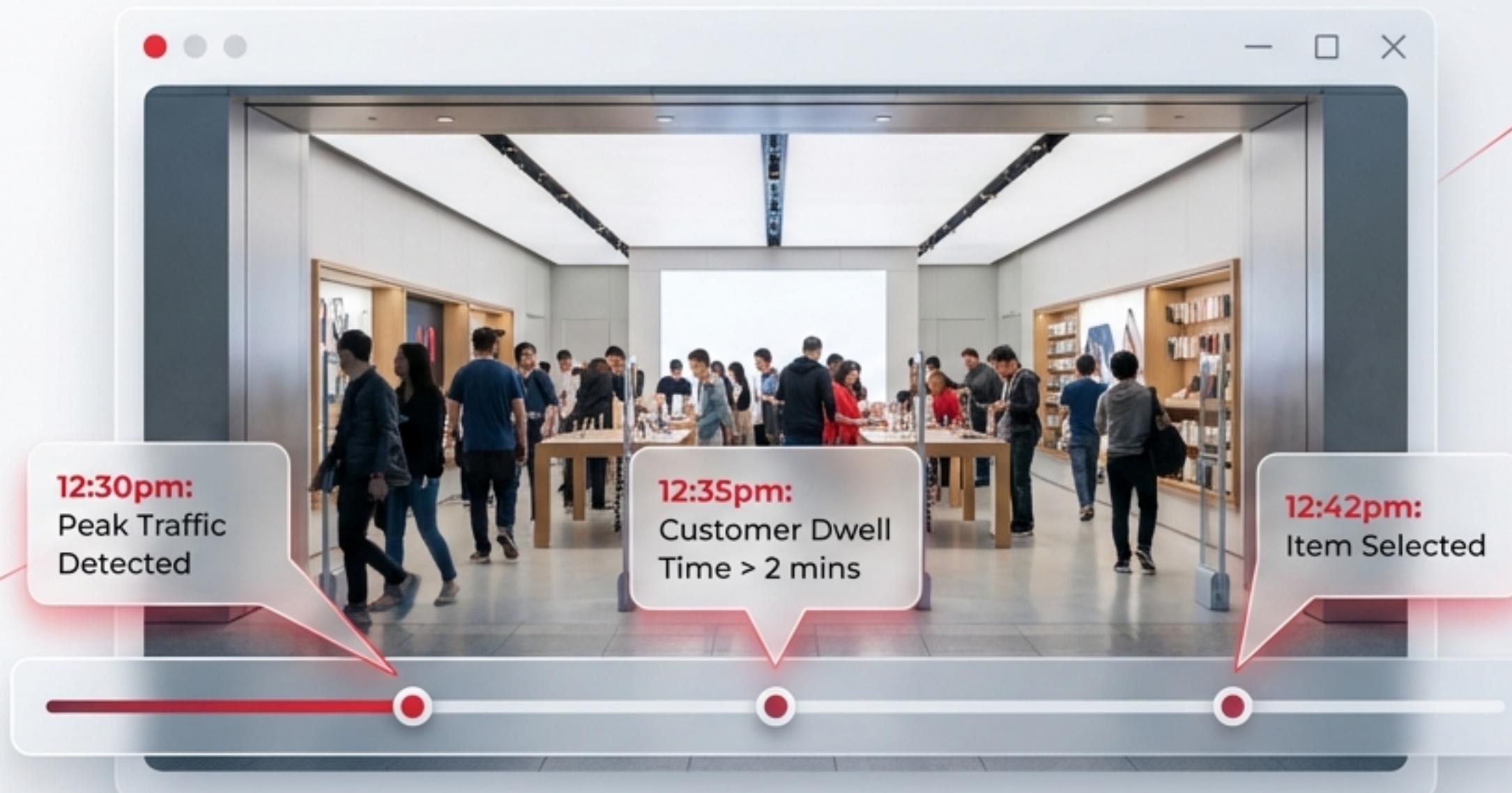


Capabilities:

1. Identify Objects
2. Read Text (OCR)
3. Assess Quality

Video Understanding: Motion & Time

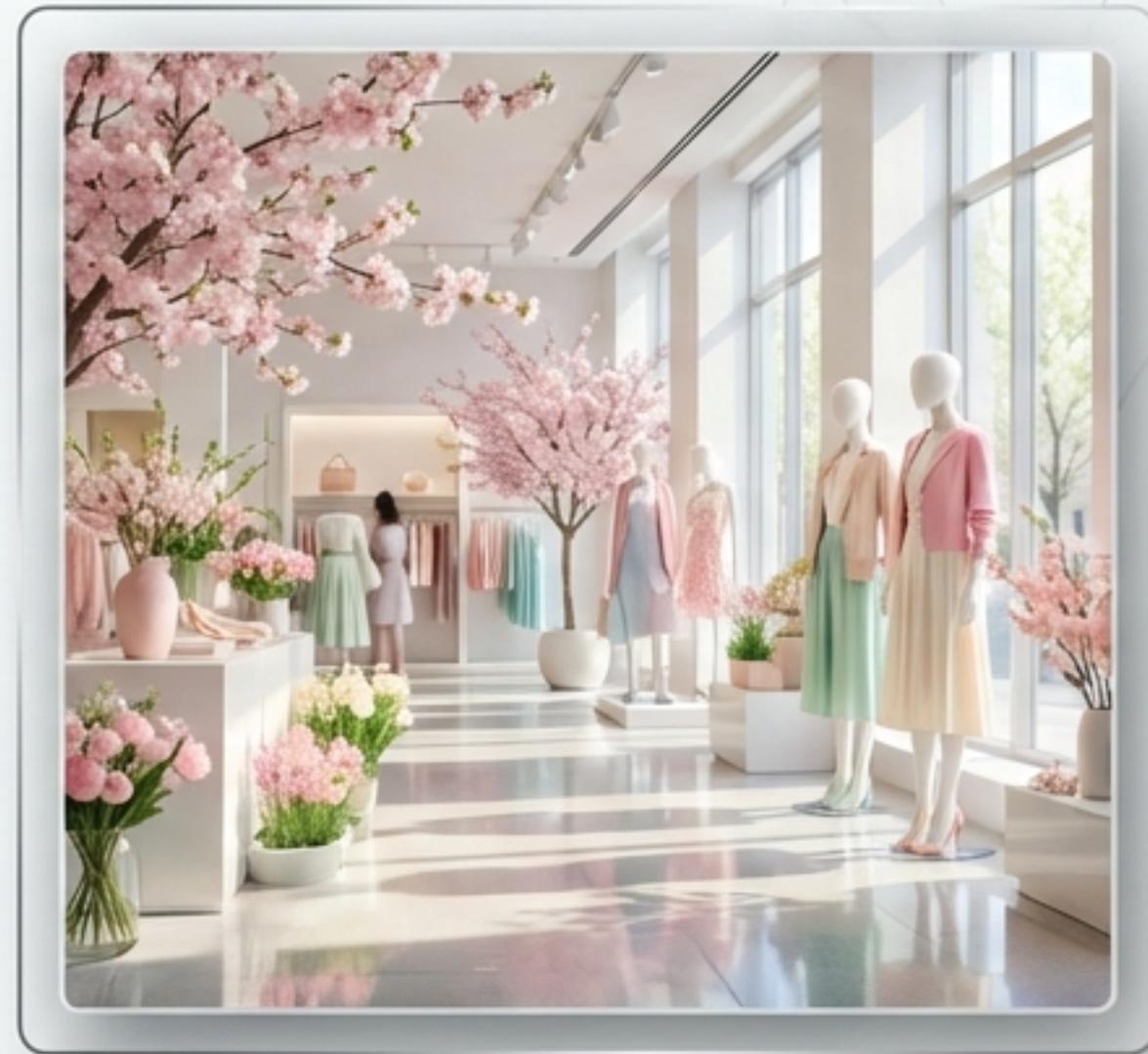
Corporate editorial and high-tech consultancy



Video is analyzed as a sequence of events, turning footage into a searchable database.

Image Generation: AI as Creator

Prompt: Professional spring marketing banner for Beacon Retail, bright lighting, pastel colors, fresh floral arrangements, blurred background shoppers, high resolution, cinematic lighting...



The prompt quality determines the result quality.

The Scenario: Beacon's Spring Campaign

Research. Create. Plan.



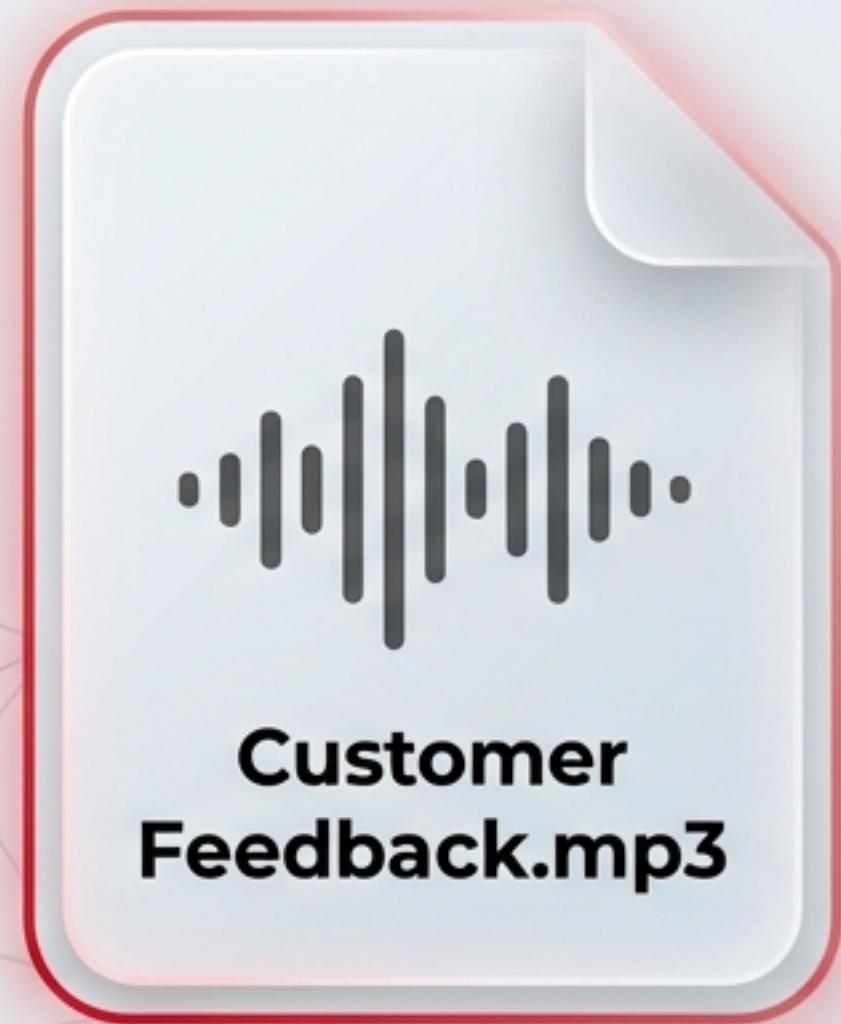
Step 1: Competitive Visual Analysis



| Category | Insight |
|-----------------|---|
| Target Audience | Young Professionals, 25-35 |
| Color Palette | Coral, White, Gold  |
| Messaging | "Fresh starts, new style." |
| Appeal | Optimism, Renewal |

Step 2: Customer Voice Analysis

AI extracts sentiment and themes from raw audio.



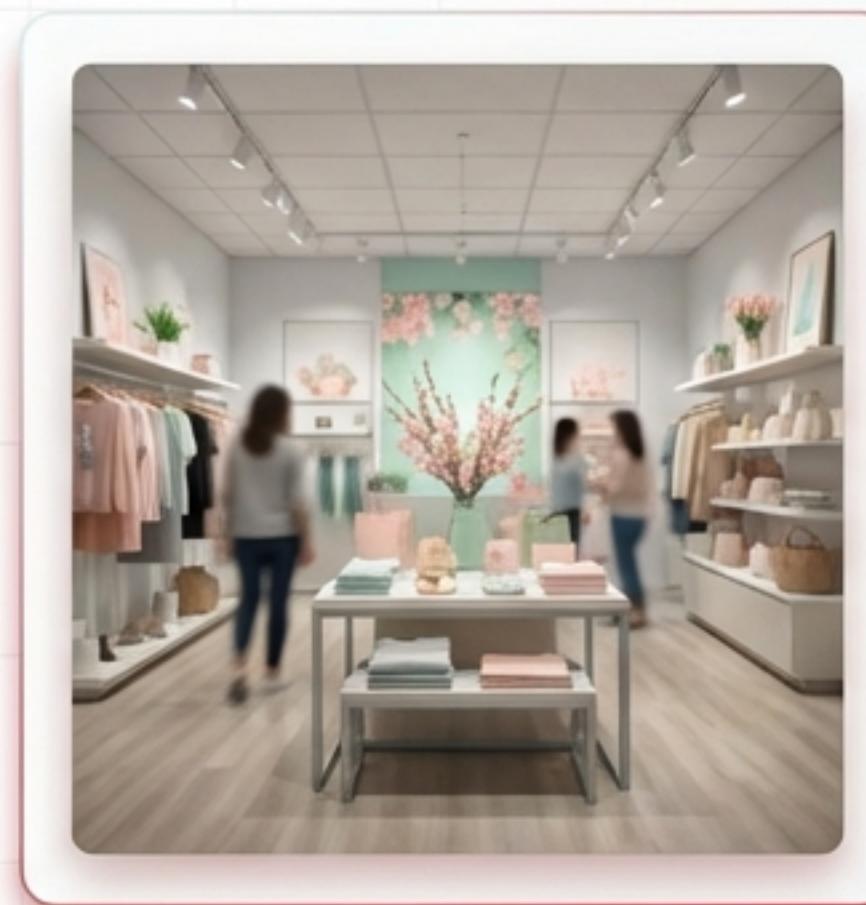
Sentiment: Positive 😊

Theme: Quality & Style

"I love the quality but..."

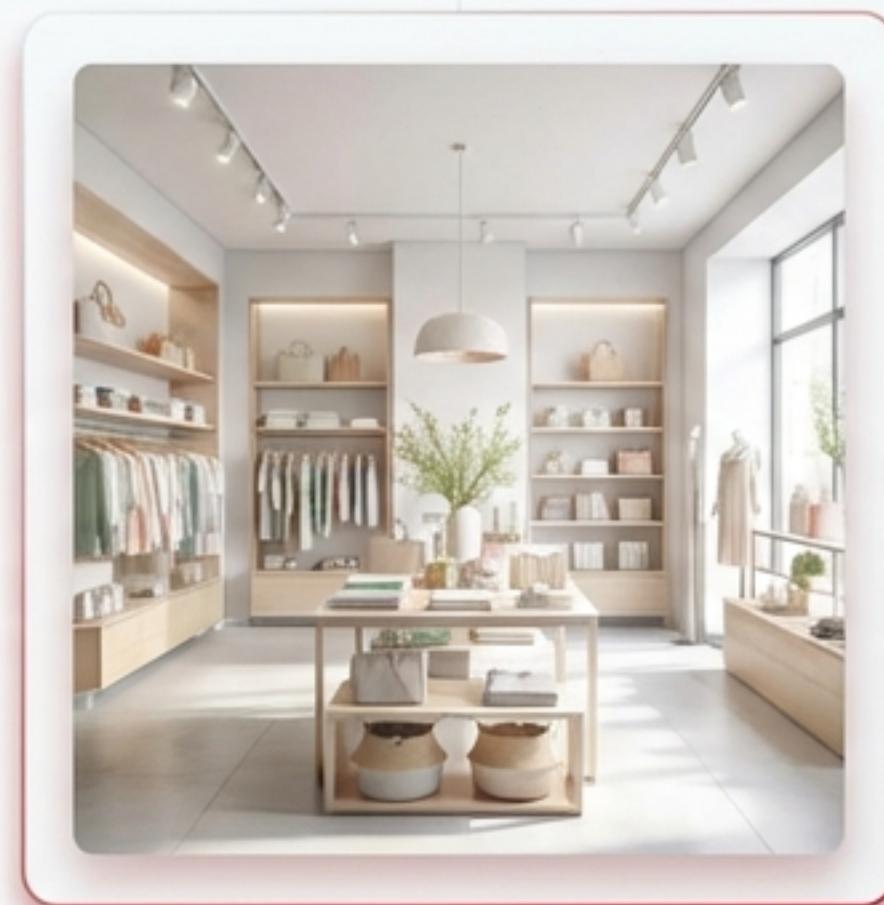
Step 3: Generate Marketing Visuals

Iterative process of creating polished AI marketing images.



Draft 1

Refine
Prompt



Draft 2

Refine
Prompt



Final Output

Example Prompt: 'Professional spring marketing image for Beacon Retail: bright, airy store interior, pastel accent colors, natural light, sharp focus, well-organized displays...'

Step 4: Structured Campaign Brief

Multimodal Context
Engineering: Combining
Vision, Audio, and Generation
into one strategic deliverable.

Competitor Data
(from Slide 11)

Customer Insights
(from Slide 12)

Generated Visuals
(from Slide 13)

CAMPAIGN BRIEF

- Competitive Landscape
- Customer Insights
- Visual Concepts + Prompts
- Recommended Messaging
- Next Steps



Capabilities & Limitations

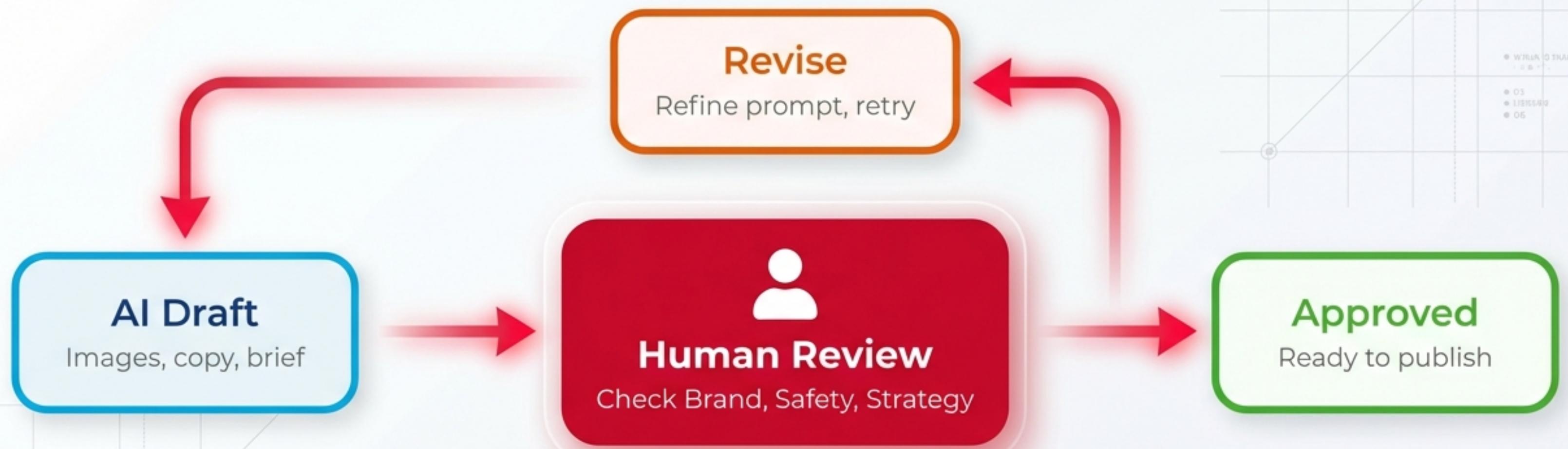
What AI CANNOT Reliably Do

- ◆ Identify specific individuals (Safety).
- ◆ Render perfect text inside images.
- ◆ Guarantee factual accuracy (Hallucinations).
- ◆ Replicate copyrighted characters perfectly.



Rule: Trust but Verify.

Human-in-the-Loop: Why AI Needs You

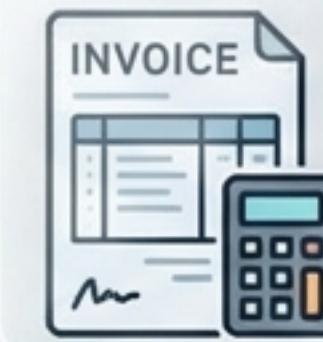


Beacon Policy: All AI content must be reviewed for brand alignment.

Beyond Marketing: Enterprise Applications



Operations: Visual QC for defects.



Finance: Invoice photos → Data.



HR: Interview audio → Notes.



Facilities: Walkthrough video → Audit.

Multimodal Input + AI + Structured Output = Business Intelligence.

ROI: The Business Case

10X

Faster Concept Art

Minutes vs. Days.

70%

Cost Reduction

AI mockups + Polish

50+

Variations per Hour

Vs. 2-3 manually.

The Shift: From creation constrained by budget to curation constrained by strategy.



© NIU Red. All rights reserved.

NotebookLM

Key Takeaways

- 1.** **Multimodal AI** sees, hears, and creates—unlocking the 80% of unstructured data.
- 2.** **Value = Structure:** The goal is turning images/audio into structured tables/insights.
- 3.** **Iterative Prompting:** Image generation requires refinement loops.
- 4.** **Human-in-the-Loop:** Essential for brand safety and strategic alignment.
- 5.** **Context Engineering:** The same principles from Day 3 apply here, just with new data types.



Lab Preview: Beacon's Spring Campaign

It's time to build.

Open Lab Guide

Analyze real competitor images

Listen to real customer audio

Generate campaign visuals

