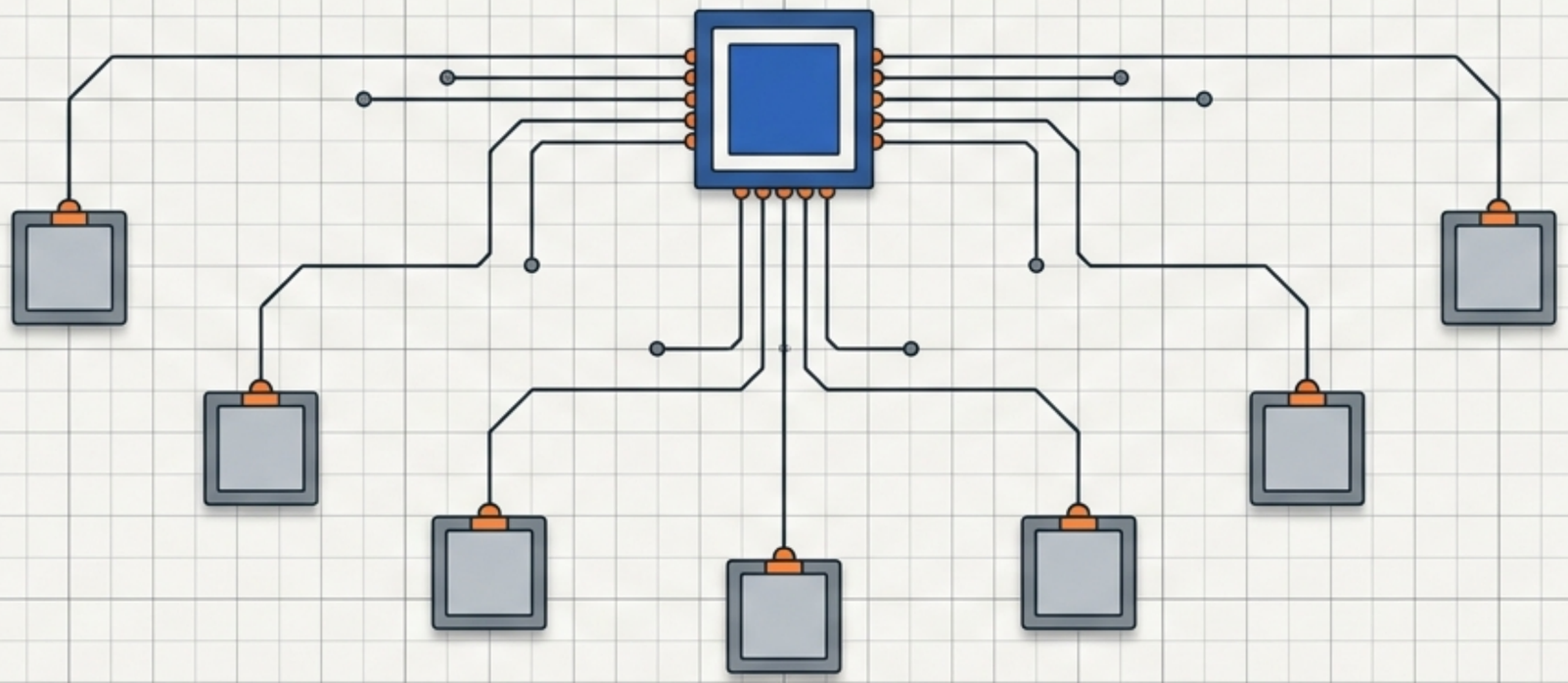
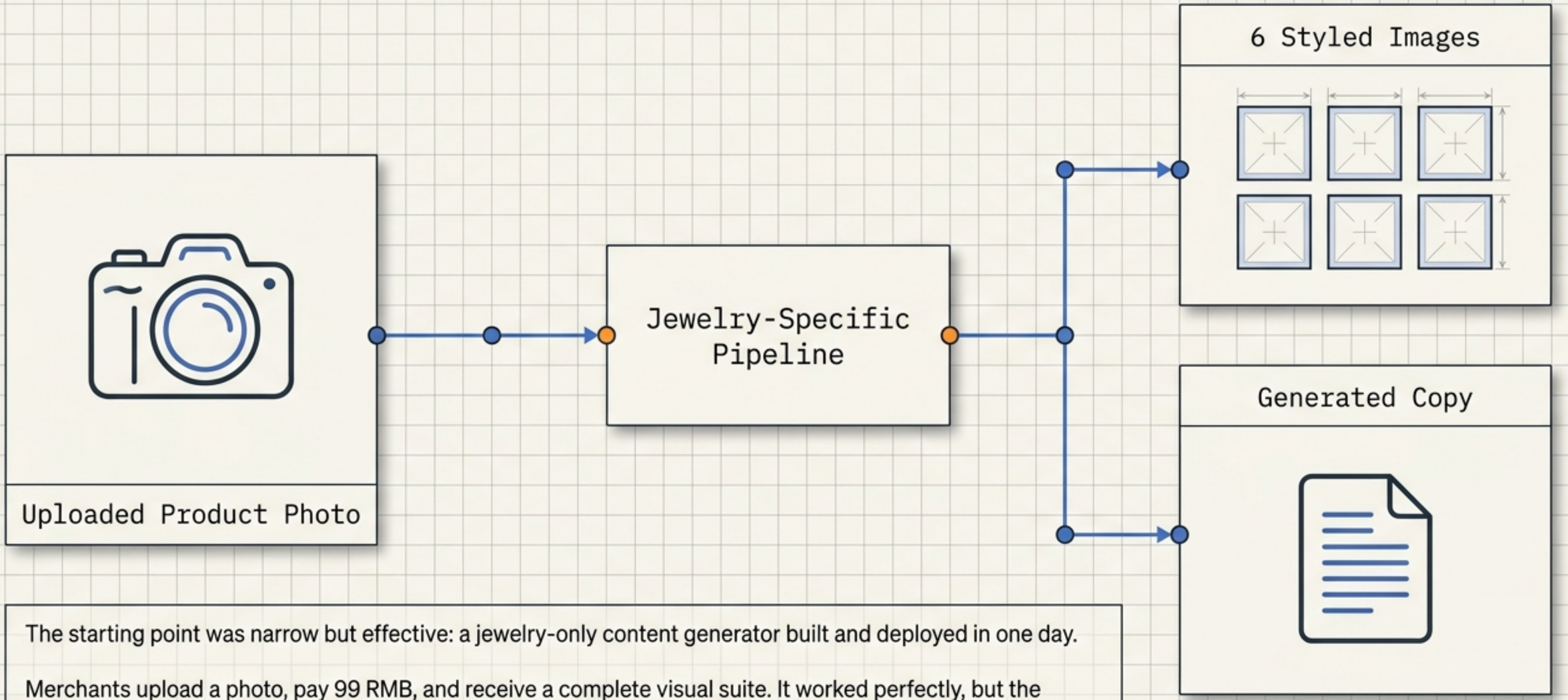
Building a Full-Category Content Engine

Abstracting a niche MVP into a high-margin, universal AI platform.

An Architectural Case Study REF: ARCH-2024-42.1 // PLATFORM_ENGINE



The Baseline: A Profitable Micro-MVP



The starting point was narrow but effective: a jewelry-only content generator built and deployed in one day.

Merchants upload a photo, pay 99 RMB, and receive a complete visual suite. It worked perfectly, but the architecture was entirely domain-locked.

The Question That Rewired the Architecture

> “Why don’t all categories get the full pipeline? Isn’t the infrastructure universal?”

✕ The Flawed Mental Model

Jewelry prompts are polished, food prompts are rough. Therefore, build fewer features for food.

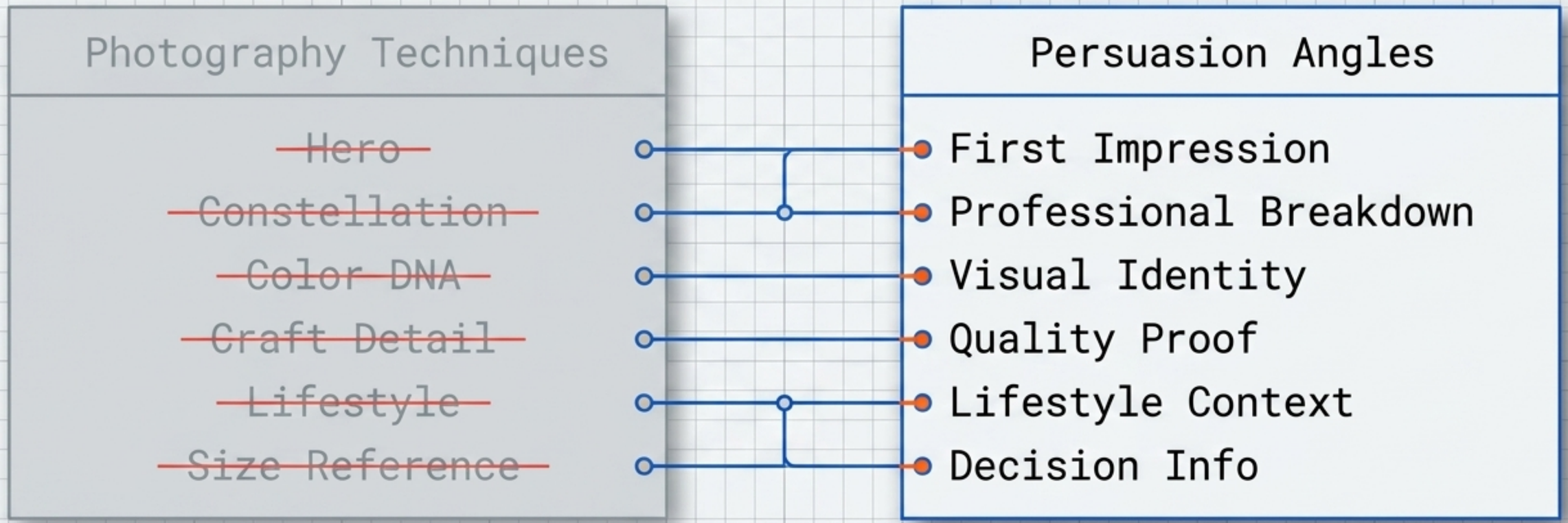
✓ The Correct Mental Model

The pipeline is the same. The templates are the same. Just change the words inside them.

Abstracting Photography into Persuasion

The breakthrough: the 6 template slots are not photography techniques. They are the 6 fundamental angles required to persuade a consumer.

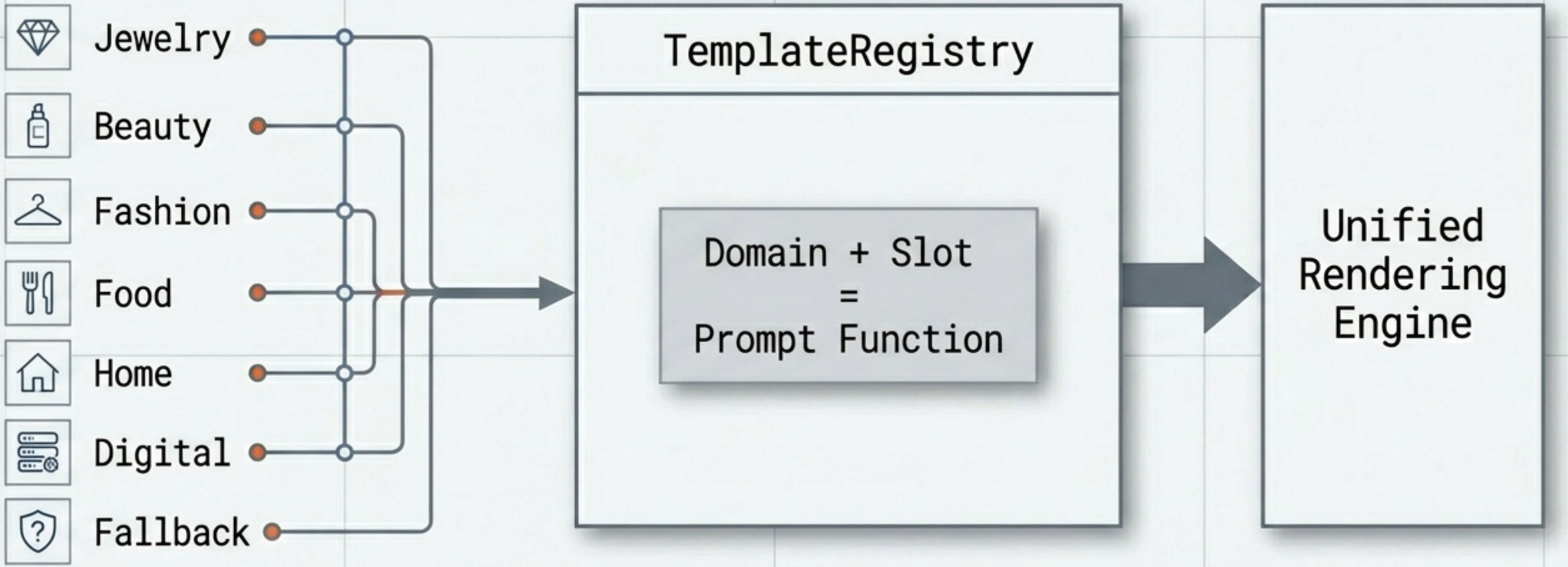
Once abstracted, the system can sell anything.



The Universal Persuasion Matrix

Slot & Angle	Jewelry	Food	Digital Product
1. First Impression	Studio hero shot	Plated dish beauty shot	Course cover mockup
2. Professional Breakdown	Gem constellation	Ingredient deconstruction	Curriculum structure
3. Visual Identity	Color DNA swatches	Recipe step progression	Brand mood board
4. Quality Proof	Craft macro detail	Texture close-up	Student results showcase
5. Lifestyle Context	Wearing scene	Table setting / sharing	Workspace / study scene
6. Decision Info	Size with coin	Nutrition / serving specs	Pricing comparison

The Category-Agnostic Infrastructure



The intelligence lives exclusively in the prompt layer.

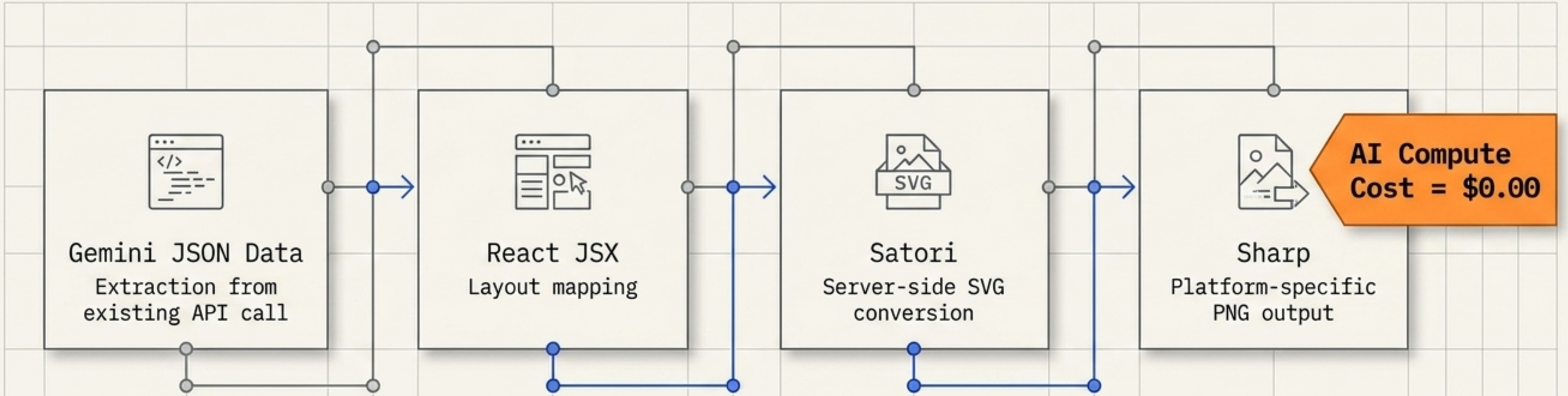
To add a new domain, we simply write 6 new prompt functions and register them. The underlying pipeline, streaming, storage, and UI remain completely untouched.

The Missing Variable in AI Content Tools

Every AI competitor focuses exclusively on generating product photos.
But studying high-converting posts reveals a different reality.

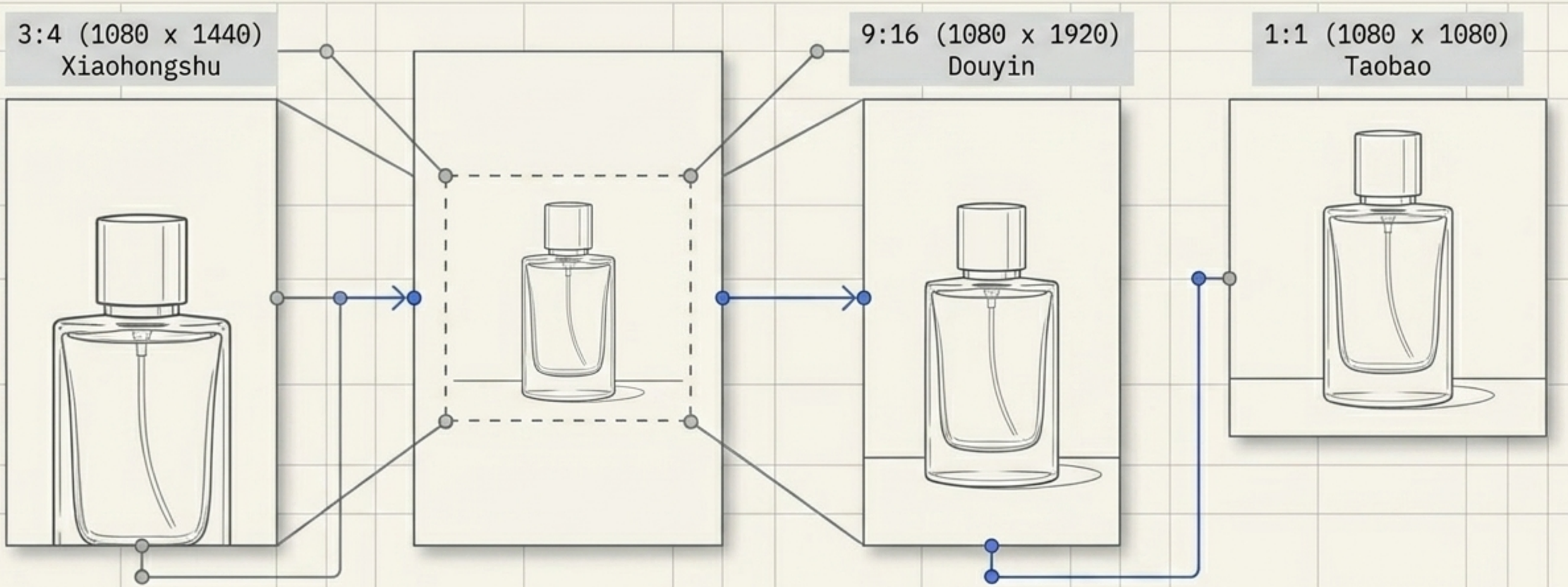
The 9-image carousel relies heavily on text cards. Without them, the carousel doesn't convert.

The Zero-Cost Rendering Pipeline



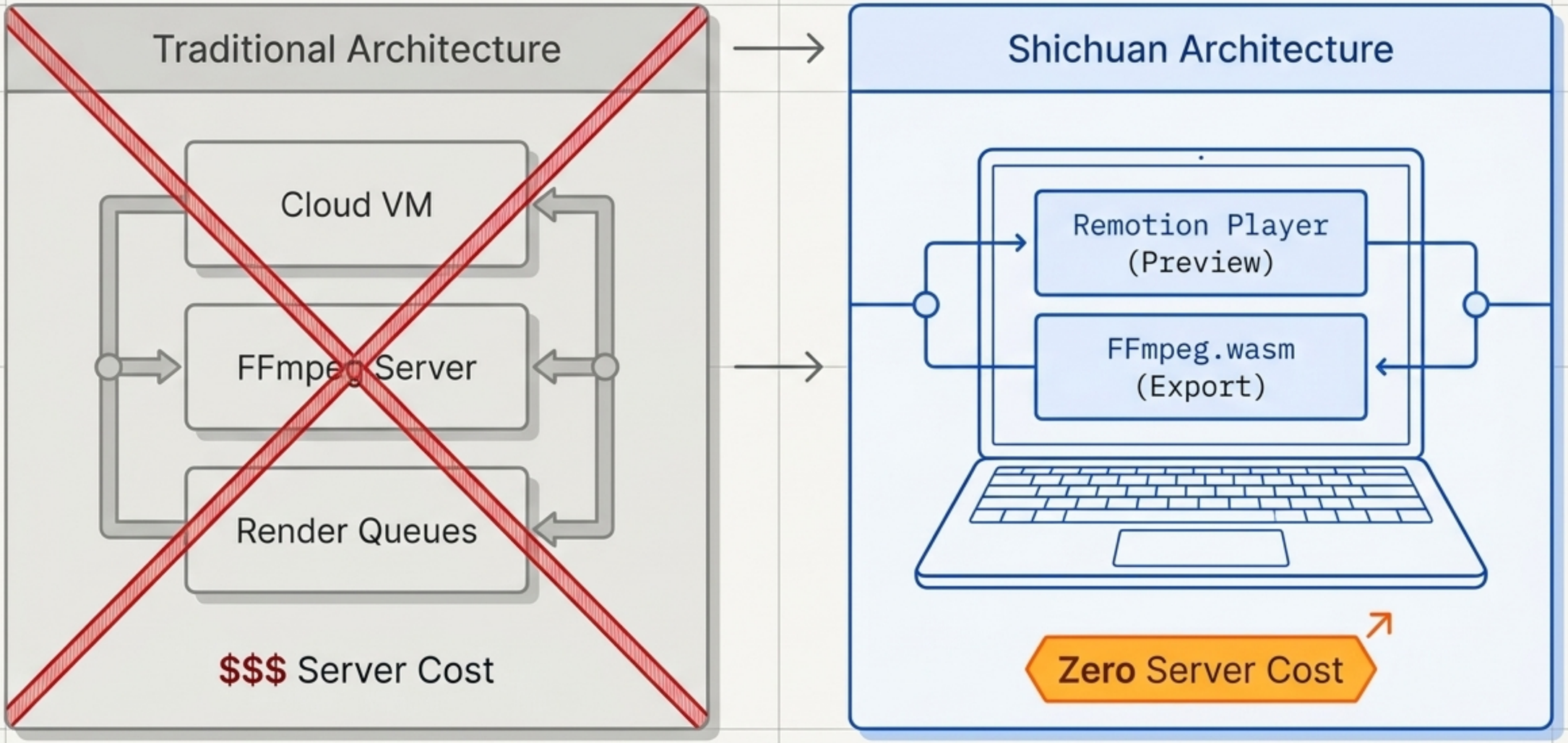
The analysis step already extracts selling points, materials, and specs. We feed that structured data into pure server-side rendering. Free by-products. Zero marginal cost. Total pipeline completion.

The Aspect Ratio Auto-Cropper



Sharp resizes and extracts with gravity-center cropping.
One generation run yields complete, platform-native image sets instantly.

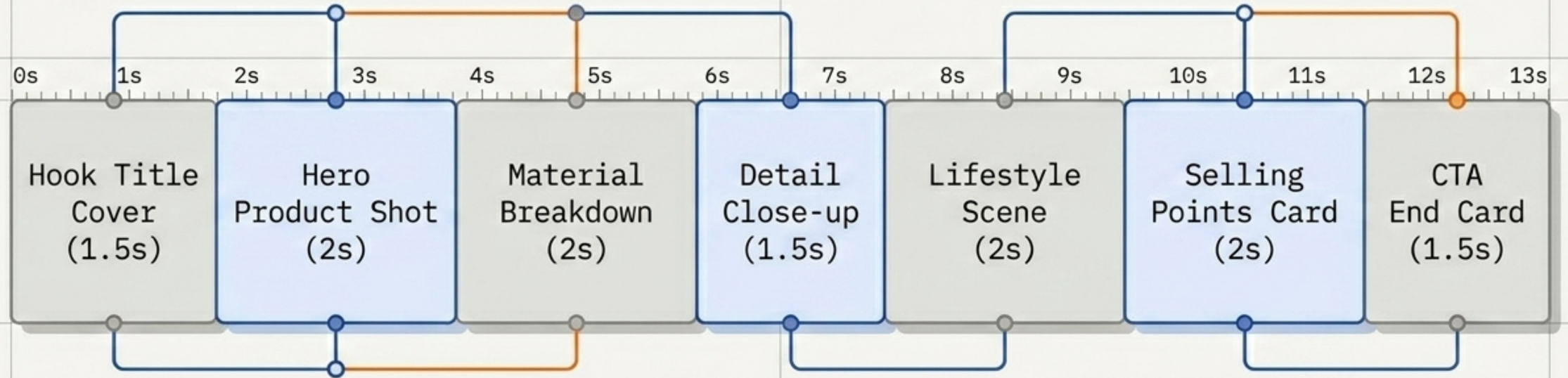
The Compute Shift: Client-Side Video



Server-side video rendering requires infrastructure, queuing, and ongoing overhead.
By moving the encoding entirely to the client's browser via FFmpeg.wasm, we eliminate server costs, farm queues, and infrastructure complexity.

The 13-Second Video Sequence

A proven short-form structure for Douyin and XHS. Rendered with Ken Burns pan/zoom effects and cross-fades entirely via a React component playing back the already-generated images.



Edge Case Validation: Mixed Materials

Input: Mid-Range Fashion Tote

Output: Constellation Slot

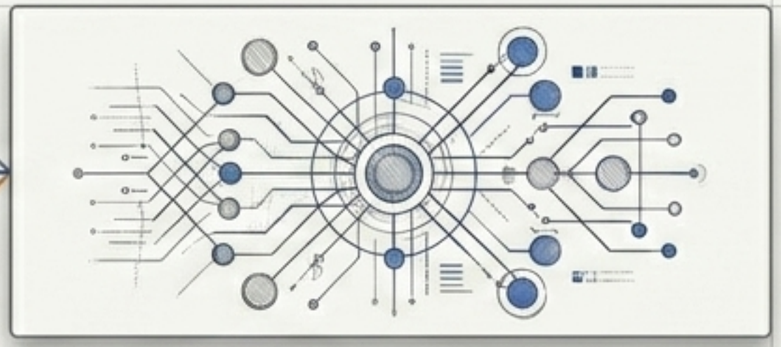
The **Constellation slot** was originally designed to arrange gem specimens in a museum vitrine.

When fed a mid-range bag, the AI perfectly generalized the underlying concept: it deconstructed the product and presented the raw materials as high-end specimens. The aesthetic transferred flawlessly.

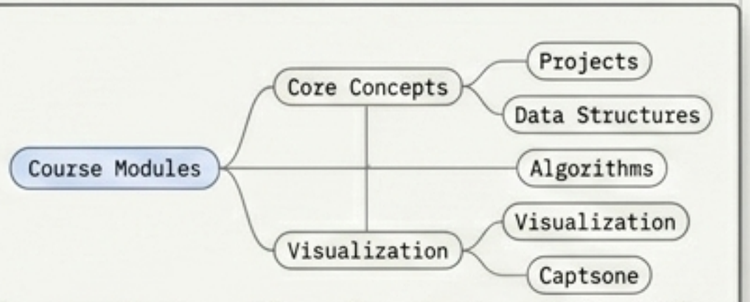
Edge Case Validation: Digital Products

Text Input Only

Python data analysis course, 40 hours...



Hero Slot: Abstract Concept Art

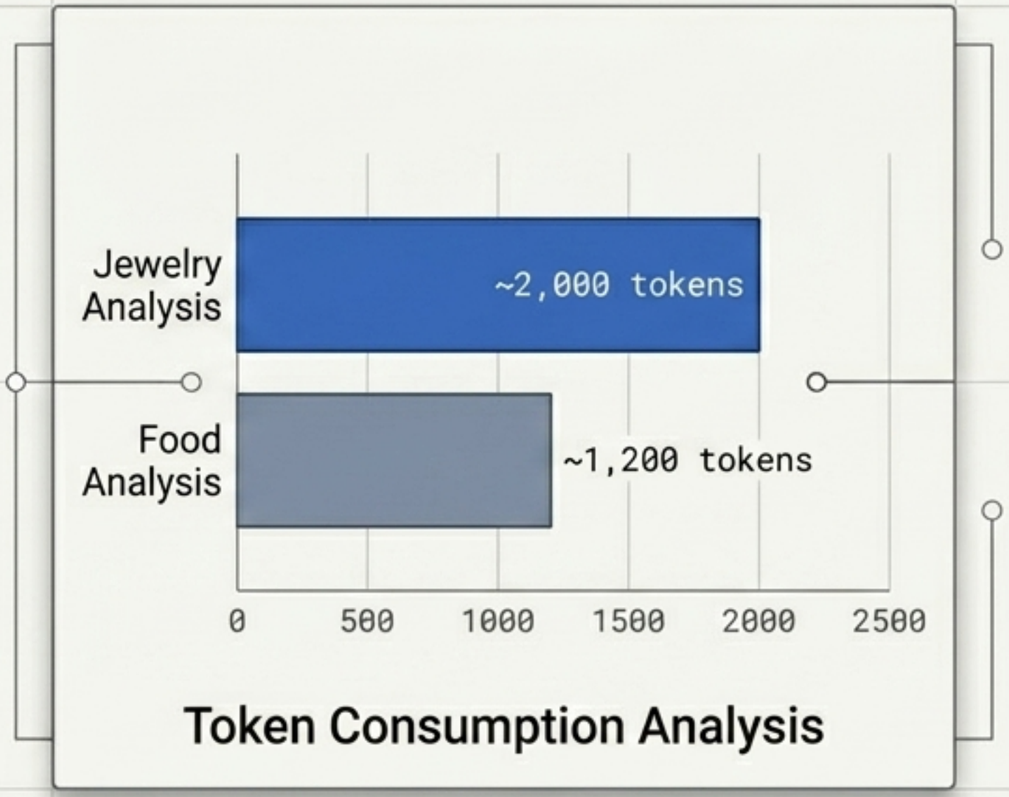


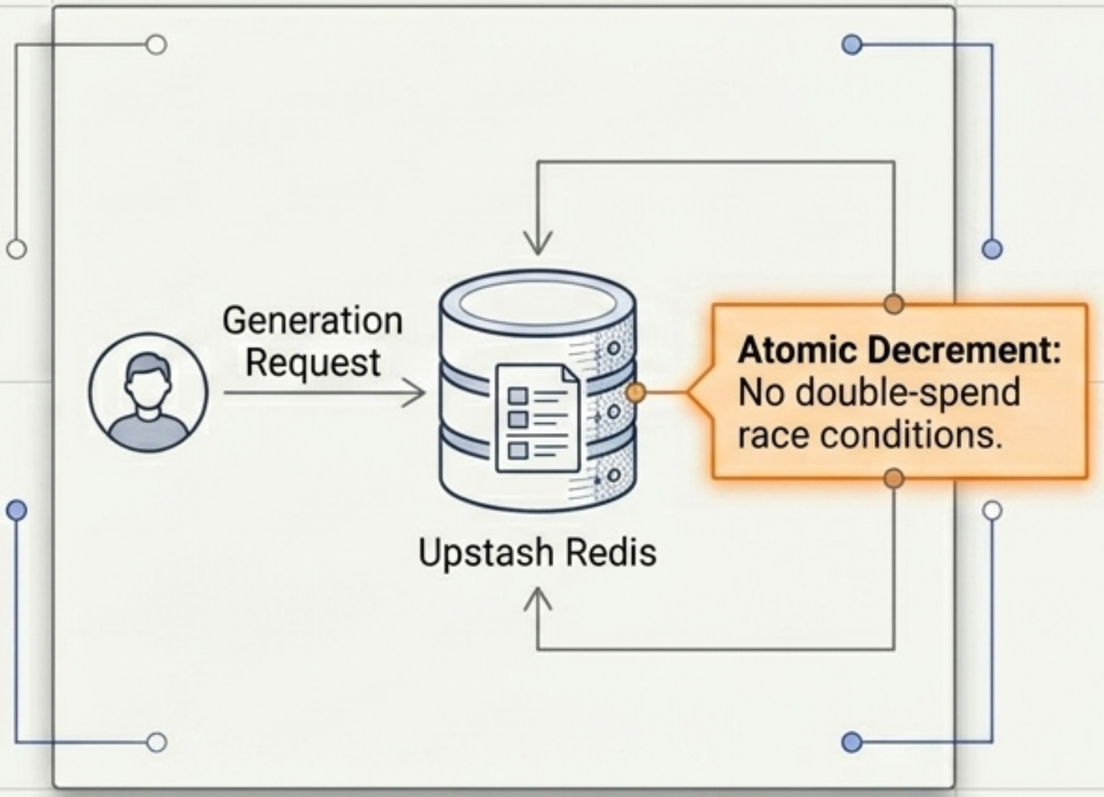
Constellation Slot: Curriculum Mind Map

What happens when a product has no physical form?

The analysis step generates a detailed product profile from pure text. The system swaps precise physical arrangements for conceptual, artistic representations. Zero input photo required.

Precision Infrastructure and Token Economics



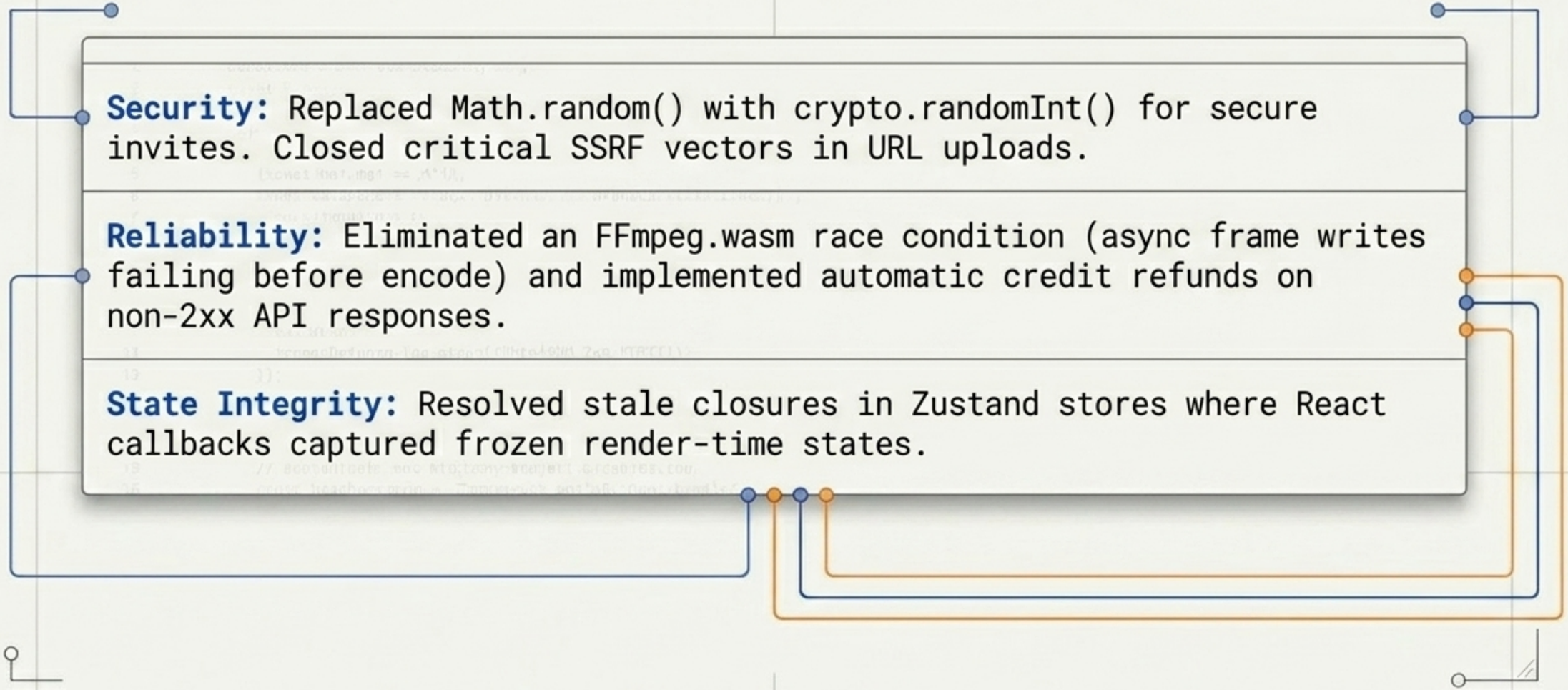


Estimating API costs with rough averages is dangerous when prompt lengths vary so dramatically by domain.

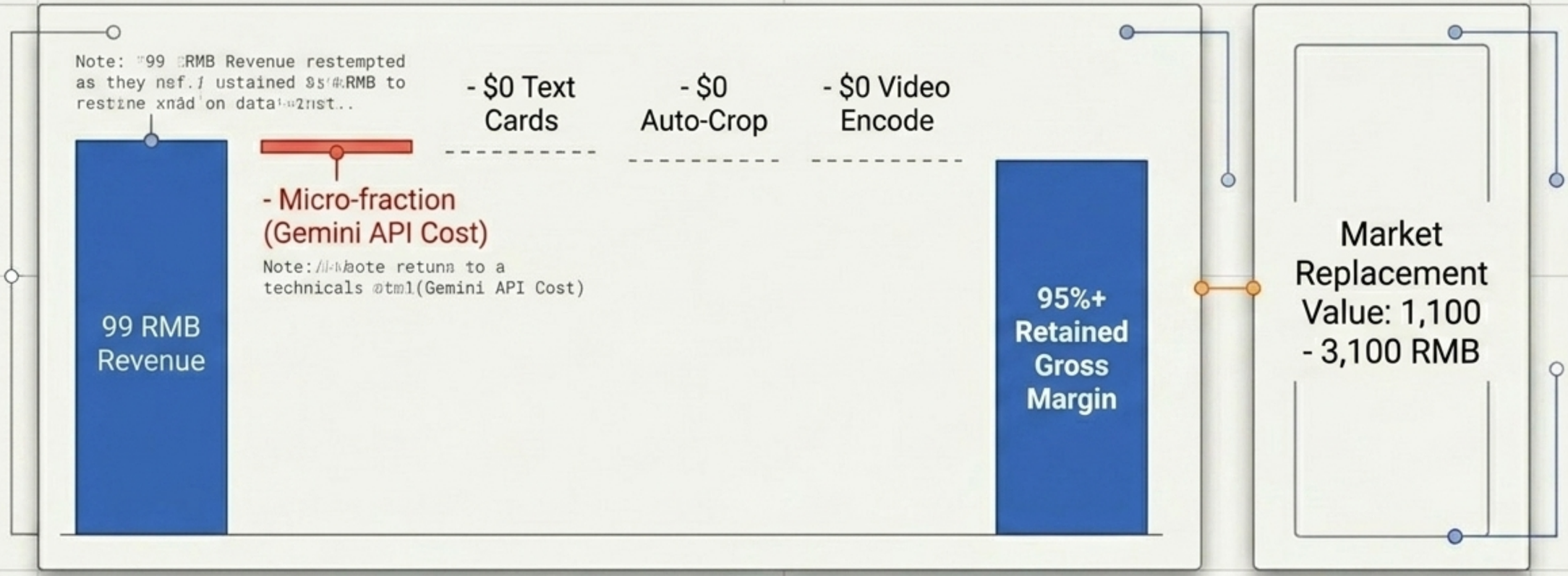
Switching to Gemini's `usageMetadata` allowed exact input/output token tracking per request, seamlessly tied to a robust Redis credit system.

Bulletproofing the Codebase

A systematic audit yielded 22 fixes to transition from demo to enterprise-grade:

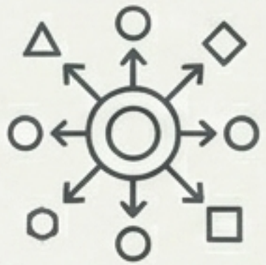


The Margin Engine



Priced at 1/10th to 1/30th of the replacement value. Cheap enough that merchants don’t hesitate; margins high enough to autonomously fund growth. The blended cost including text cards is lower than AI-only generation suites.

The Architecture of AI Platforms



1. Vary the Prompts, Not the Pipeline

Treat domains as configuration, not custom builds. Infrastructure must remain strictly category-agnostic. Intelligence belongs in the prompt layer.

2. The Best Features Can Cost Zero AI Compute

Stop competing solely on image generation. High-conversion assets (like text cards) can be generated via pure server-side rendering.

3. Move Compute to the Client

Eliminate server overhead wherever possible. Browser-side tools like FFmpeg.wasm ship powerful features with zero marginal infrastructure cost.