

Masking a Four-Minute Compute Wait with Server-Sent Events

The Front-End Illusion



The Back-End Reality

The Challenge:
4 high-res AI photos^{\$\$} take 15–60 seconds each.
Static loading bars feel broken.

Parallel Generation: 4 heavy
AI models run simultaneously.

Unidirectional Flow: No heavy
WebSocket overhead; client only
receives continuous events.

Event Stream:
[Started] -> [Photo_Completed]
-> [Completed]

Instead of a progress
bar, users see a
slideshow of their
photos "arriving"
one by one. The wait
becomes part of
the luxury
experience!

Engineering the Illusion of Inadvertent Luxury



If you tell the AI 'Hermes', you get a trademark violation. If you tell it 'Silk Sarong' but calibrate the scene aesthetic to luxury in the backend... you get the vibe without the lawsuit!

Rule 1:
Never Centered.
Luxury items are the supporting cast. Override the AI's instinct to center the product.

Rule 2:
Partially Visible.
A fully visible logo screams "Ad." Obscured logos read as "Real Life."

Rule 3:
Candid Moments.
Perfect lighting looks produced. Force the prompt to feel like a "normal Tuesday."

Brute-Forcing Facial Consistency with a Text-Only QC Loop

Step 1: The Heavy Lift.
Gemini creates the 4-photo set.

Step 2: The Interrogation.
A fast, text-only Flash model compares output to reference.

Step 3: The Ultimate Question.
"Is this the exact same person?"

Step 4: The Loop.
If False, retry up to 2 times.

Why this matters: Users know their own faces intimately. A slightly sharper jawline breaks the magic. This text-only check adds ~3 seconds and a fraction of a cent, but catches the worst 15-20% of facial hallucinations before the user ever sees them.



Generating the Perfect "Humble-Brag" Caption

The Versailles

Rule: Describe a mundane task; let the photo show the wealth.
Output: "Said I'd wake at nine, slept two more hours."

The Poetic

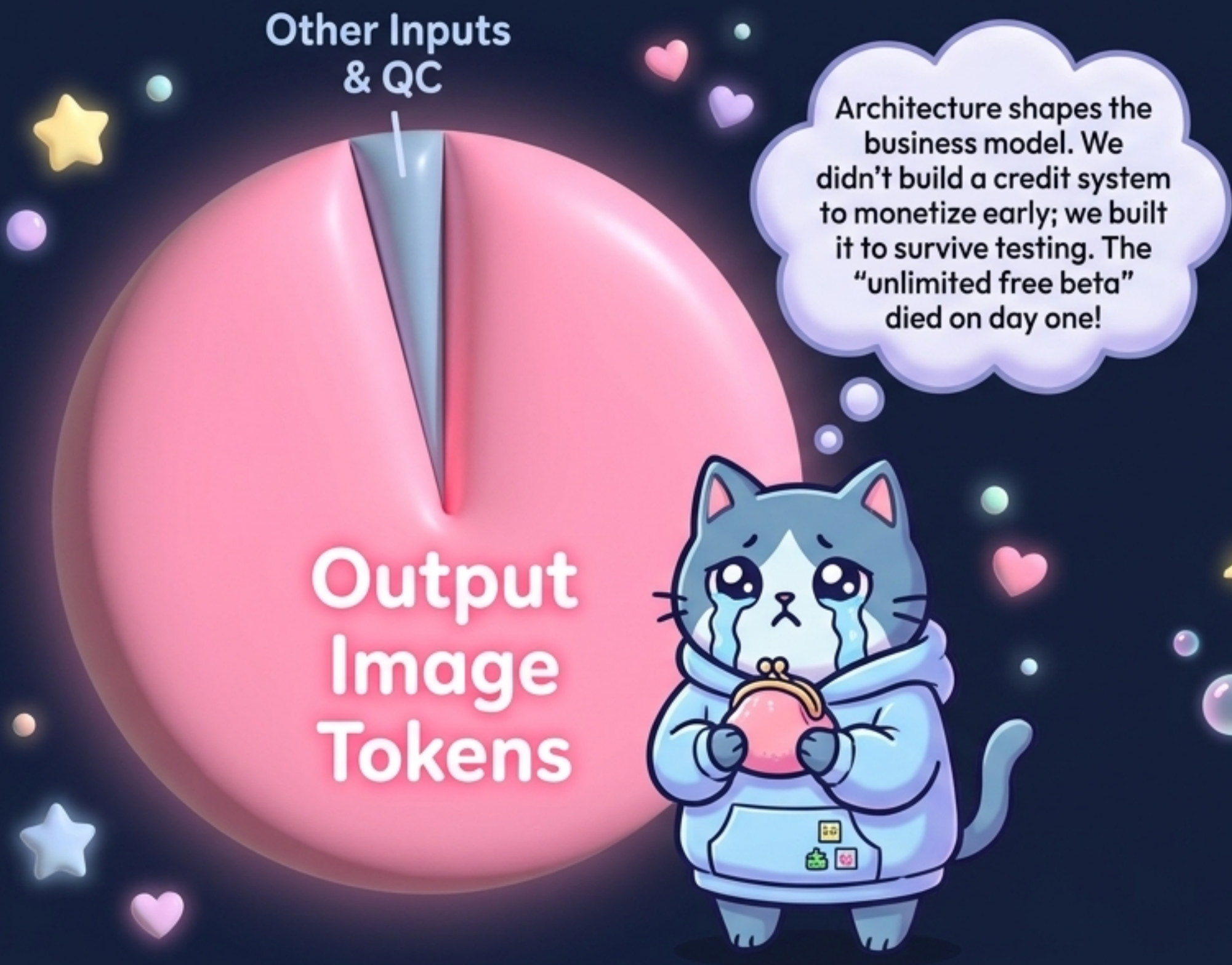
Rule: Focus on feelings and imagery. Zero narration.
Output: "Water and sky became one, heart went transparent too."

The Minimal

Rule: Maximum whitespace. Under 50 characters.
Output: "Soaking. Don't want to move."

The Golden Rule: Absolutely NO brand names, hotels, or locations.
Mentioning "The Four Seasons" immediately destroys the illusion of nonchalance!

Output Image Tokens Are 96% of the API Bill



- **The Pixel Penalty:** Generating images is ~120x more expensive per token than generating text.
- **The 4K Multiplier:** Scaling resolution from 1K to 4K increases costs by 2.3x per image.
- **The QC Wildcard:** If a photo fails the drift check and retries twice, that single shot costs 3x the baseline.
- **The Rounding Error:** Prompts, reference inputs, captions, and QC checks make up < 4% of the total cost.