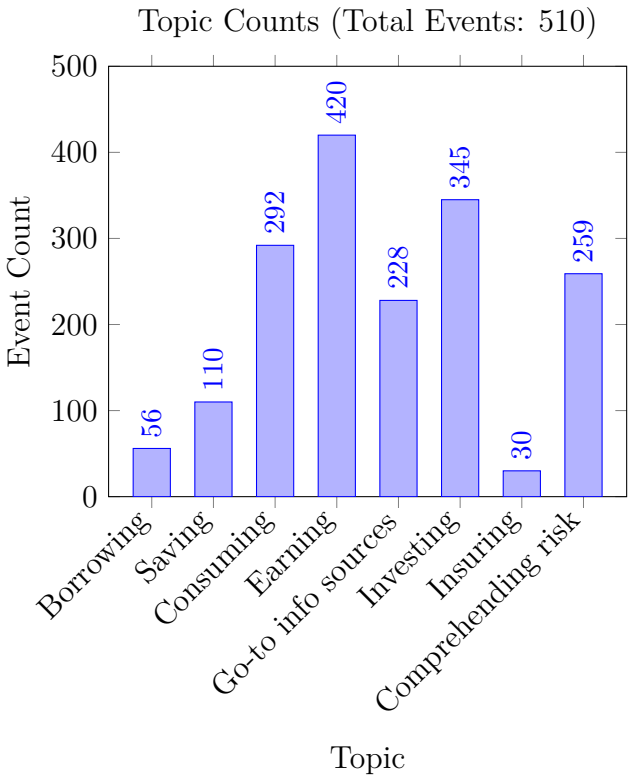
Gemini-2.5-Flash Cost & Topic Analysis

October 2025

Overview

Summarization of topic counts from gemini questions and estimates the projected cost of running Gemini-2.5-Flash simulations at scale. It includes visual breakdowns, numerical cost computations, and projected hourly expenses for large-scale usage.

Topic Counts



Gemini-2.5-Flash Cost Projection

Reported for 10 runs: **397710 tokens**.

Assuming a price per token of:

$$c = \$0.00001 \text{ per token},$$

then the cost for one simulation is:

$$C_{\text{instance}} = 379{,}710 \times 0.00001 \times 0.1 = \$.37971$$

Ten Simulations

If one user ran 10 simulations:

$$C_{\text{10-sims}} = 10 \times 0.37971 = \$3.7971$$

Scaling to One Million Users

If 1,000,000 users ran the same simulation simultaneously:

$$C_{\text{total}} = 1{,}000{,}000 \times 0.37971 = \$379{,}710$$

Cost per Hour

The 10 simulations took **1 hour and 20 minutes** = 1.333 hours.

Cost rate per user per hour:

$$C_{\text{hour, per simulation}} = \frac{0.37971}{1.333} = \$0.285 \text{ per hour.}$$

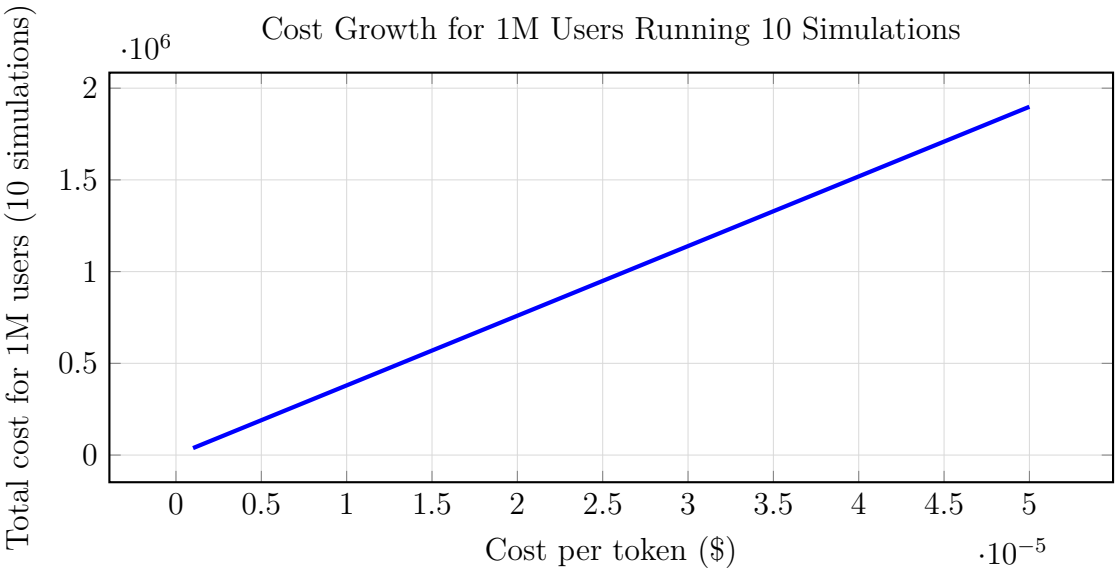
For 1 million users running continuously:

$$C_{\text{hour, 1M}} = 1{,}000{,}000 \times 0.285 = \$284{,}900 \text{ per hour.}$$

Summary of Key Metrics:

- Billable tokens per simulation: **37,971**
- Cost per simulation: **$0.38**
- Cost per 10 simulations: **$3.8 per user**
- Cost per hour per user: **$0.29/hour**
- Cost per hour for 1M users: **$280 thousand/hour**

Projected Cost Curve



Insights

At micropricing levels ($0.00001 per token), costs remain reasonable for small-scale use but scale rapidly at mass concurrency. Running one million users concurrently for sustained periods could exceed **$280 thousand per operational hour.**

All computations assume a flat per-token rate and no volume discounts or caching efficiencies.