

AI Prices Are Collapsing: What the 2026 Price War Means for Swiss SMEs

MANAGEMENT SUMMARY & CORE TAKEAWAY

OpenAI cut GPT-5.6 Luna prices by 80% shortly after launch, responding to pricing pressure from DeepSeek, Kimi K3 and Llama. For Swiss SMEs, the entry barrier to powerful AI is collapsing, while model agility is becoming the new core competency.

CONTEXT & EXECUTIVE RELEVANCE

On July 30, 2026, OpenAI cut prices for GPT-5.6 Luna by 80% — from \$1.00 to \$0.20 per million input tokens, with output pricing falling from \$6.00 to \$1.20. Terra, the next tier up, became 20% cheaper. What stands out is not just the size of the cut but the timing: Luna had been on the market for only a short time. A price cut of this magnitude so soon after launch is not a gesture of generosity — it is a sign of competitive pressure.

OpenAI restructured pricing across three model tiers. Luna, the entry-level model, was slashed dramatically. Terra, the mid-tier model, more moderately. Sol, the flagship, stayed unchanged at \$5 and \$30 per million tokens. This three-way split is itself a signal: OpenAI is protecting margin at the premium tier while competing on volume and market share at the entry and mid tiers — precisely where competitive pressure from open models is strongest.

CORE ACTION ITEMS

01 Luna: input \$1.00 → \$0.20 per million tokens (~80%), output \$6.00 → \$1.20 (~80%)

02 Terra: input \$2.50 → \$2.00 per million tokens (~20%), output \$15.00 → \$12.00 (~20%)

03 Sol: unchanged at \$5.00 / \$30.00 per million tokens

SWISS COMPLIANCE & ARCHITECTURE GUIDELINES

- ☑ Full compliance with revised Swiss Data Protection Act (revDSG).
- ☑ Zero vendor lock-in on closed proprietary models.
- ☑ Seamless integration into Abacus, Bexio and M365 ecosystems.

Non-Binding Process Analysis

Let us review in 20 minutes where AI agents create immediate measurable efficiency in your workflows.

ki-outsourcing.ch