

AI Gets Cheaper and Smarter – Why Your Cost Model No Longer Works

MANAGEMENT SUMMARY & CORE TAKEAWAY

The cost of a given level of AI performance has dropped about 47% per quarter since 2023 while capability rises. Your 2023/24 cost model is obsolete. The real question is not whether AI is a bubble, but whether your calculation keeps pace with deflation and whether you use model routing or outsourcing now.

CONTEXT & EXECUTIVE RELEVANCE

Since 2023, the cost of a given level of AI performance has fallen by about 47% per quarter. That's not a dry statistic – it's the most important cost driver in your business. What costs 100 dollars today will cost a small fraction of that in a year, at equal or better quality. This deflation outpaces DNA sequencing, classical compute, and batteries.

Andreessen Horowitz measured the effect back in 2024: reaching GPT-3-level capability fell from about \$60 to \$0.06 per million tokens – a 1000x drop in three years. HFS Research confirmed the dynamic in October 2026: effective frontier intelligence costs fell 59% in two weeks without any provider lowering list prices. Prices fall through new model generations, not discounts.

COMPARISON & BENCHMARK MATRIX

In-house team (3–6 months build-up)	High fixed costs, slow adaptation to price curve, full control
Direct API usage without orchestration	Low model costs, but 70–90% of project costs fall in integration and operations
AI outsourcing (external unit)	Scales with price curve, delivers rapid productivity, minimal fixed costs

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