

KI-Preisverfall 2026: Warum sinkende Modellkosten die Einstiegshürde für Schweizer KMU jetzt senken

MANAGEMENT SUMMARY & KERNAUSSAGE

Die Kosten für KI-Modelle sind seit 2022 um über 90% gefallen – schneller als PC-Preise über 15 Jahre hinweg. Gleichzeitig investieren Hyperscaler 757 Milliarden Dollar in 2026 und treiben die Infrastruktur massiv aus. Für Schweizer KMU bedeutet dieser Preisverfall: KI-Outsourcing wird erstmals wirklich erschwinglich, während die strategische Frage nicht mehr lautet «Können wir uns KI leisten?», sondern «Welche Modelle, welche Anbieter und welches Timing maximieren unseren ROI?». Die Make-vs-Buy-Entscheidung, Vendor-Diversifikation und qualitätsadjustierte Kostenvergleiche rücken in den Fokus – und wer jetzt die richtigen Weichen stellt, verschafft sich einen messbaren Wettbewerbsvorteil.

KONTEXT & RELEVANZ FÜR ENTSCHEIDER

Kein Technologiezyklus der letzten 40 Jahre hat Kosten schneller gesenkt als generative KI. Goldman Sachs dokumentiert einen Preis-Drop von über 90% bei KI-Modellen seit 2022. Zum Vergleich: PC-Preise brauchten 15 Jahre für einen vergleichbaren Verfall. Ein qualitätsadjustierter AI-Preis-Index der Stanford University (SSRN, März 2026) zeigt einen Rückgang von 78% allein zwischen 2024 und 2026 – nach Bereinigung um Leistungssteigerungen. Das heisst: Du bekommst heute nicht nur billigere, sondern massiv leistungsfähigere Modelle für einen Bruchteil des Budgets von vor drei Jahren.

Die Zahlen sprechen für sich: DeepSeek V4-Pro kostet 0,86 Dollar pro Million Output-Tokens und ist damit 28-mal günstiger als Claude Opus 4.7, während es auf dem SWE-bench-Score nur 0,2 Punkte dahinter liegt (80,6 vs. 80,8). Claude Opus 4.6 selbst kam 67% günstiger auf den Markt als sein Vorgänger Opus 4. GPT-5 Nano startet bei 0,05 Dollar Input und 0,40 Dollar Output pro Million Tokens. Gemini 2.5 Flash-Lite bietet ein 1-Million-Token-Context-Window für 0,10/0,40 Dollar. Diese Preisspanne verschiebt die Wirtschaftlichkeit von KI-Projekten fundamental – und zwar nicht nur für Grosskonzerne, sondern gerade für [Schweizer KMU, die strategisch in KI investieren](<https://www.ki-podcast.ch/ki-standort-schweiz-kmu-strategie-und-globales-rennen>), aber bisher an den Kosten gescheitert sind.

KERN-HANDLUNGSEMPFEHLUNGEN

- 01 API-First statt Model-Ownership: Nutze spezialisierte Modelle über API, statt eigene zu hosten.
- 02 Fokus auf Orchestrierung: Deine Wertschöpfung liegt in der Integration, nicht im Modell selbst.
- 03 Partner mit Expertise: Ein externer KI-Dienstleister amortisiert sich oft schon im ersten Quartal.
- 04 Pilot vor Scale: Starte mit Low-Cost-Modellen, validiere den Use-Case, skaliere dann mit Premium-Anbietern.

SCHWEIZER UMSETZUNGSKRITERIEN (REVD SG & IT-ARCHITEKTUR)

- ☑ Vollständige Konformität mit dem neuen Schweizer Datenschutzgesetz (revDSG).
- ☑ Keine Abhängigkeiten: Zero Lock-In auf proprietäre geschlossene Modelle.
- ☑ Schnittstellen zu Schweizer ERPs (Abacus, Bexio, M365) im operativen Betrieb.

Unverbindliche Prozess-Analyse

Lass uns in 20 Minuten prüfen, wo KI-Agenten in deinem Unternehmen sofort messbare Entlastung schaffen.

ki-outsourcing.ch