

KI wird billiger und klüger – deine Kostenrechnung stimmt nicht mehr

MANAGEMENT SUMMARY & KERNAUSSAGE

Die Kosten für ein bestimmtes KI-Leistungsniveau fallen seit 2023 um rund 47 Prozent pro Quartal – die Leistung steigt gleichzeitig. Deine Kostenrechnung von 2023/2024 ist damit Makulatur. Entscheidend ist nicht die Blase-Frage, sondern ob deine Kalkulation mit der Deflation Schritt hält und ob du Modell-Routing oder KI-Outsourcing jetzt einsetzt.

KONTEXT & RELEVANZ FÜR ENTSCHEIDER

Seit 2023 fällt der Preis für ein bestimmtes KI-Leistungsniveau um rund 47 Prozent pro Quartal. Das klingt nach einer trockenen Kennzahl, ist aber der wichtigste Kostenfaktor deines Unternehmens: Würde man die genannte Rate zugrunde legen, würde dieselbe Leistung in einem Jahr deutlich weniger kosten – nur noch einen Bruchteil der heutigen Summe, bei gleicher oder besserer Qualität. Die Deflation verläuft schneller als bei DNA-Sequenzierung, klassischer Rechenleistung oder Batterien.

Andreessen Horowitz hat den Effekt bereits 2024 gemessen: Um GPT-3-Niveau zu erreichen, fiel der Preis von rund 60 USD auf 0,06 USD pro Million Tokens – ein Faktor 1000 in drei Jahren. HFS Research bestätigt die Dynamik: Im Oktober 2026 fielen die effektiven Kosten für Frontier-Intelligenz um 59 Prozent in zwei Wochen, ohne dass ein Anbieter einen Listenpreis senkte. Die Preise purzeln durch neue Modell-Generationen, nicht durch Rabatte.

VERGLEICH & SCHWEIZER MARKTPRAXIS

Internes Team (3–6 Monate Aufbau)	Hohe Fixkosten, langsame Anpassung an Preiskurve, dafür volle Kontrolle
Direkte API-Nutzung ohne Orchestrierung	Tiefe Modellkosten, aber 70–90 % der Projektkosten fallen in Integration und Betrieb an
KI-Outsourcing (externe Einheit)	Skaliert mit Preiskurve, liefert schnelle Produktivität, kaum Fixkosten

SCHWEIZER UMSETZUNGSKRITERIEN (REVDSG & 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