

# Private-AI-TCO-Studie

Vollkosten-Vergleich Private AI on-premise vs. Sovereign Cloud vs. Public Cloud (EU-Tenant) über 36 Monate.

18 Seiten | Zielgruppe: CTO, IT-Leitung, CFO

KBD KI-Beratung Deutschland | [ki-beratung-unternehmen.de](https://ki-beratung-unternehmen.de)

## Inhaltsverzeichnis

1. Drei Deployment-Modelle
2. Profil A: 200 Mitarbeiter
3. Profil B: 500 Mitarbeiter
4. Profil C: 1.500 Mitarbeiter
5. GPU-Empfehlung
6. Empfehlung

# 1. Drei Deployment-Modelle

1. On-Premise Private AI: Eigene Hardware, volle Kontrolle, höchste Anfangsinvestition.
2. Sovereign Cloud: EU-Rechenzentren, deutscher Betreiber, mittlere Kosten.
3. Public Cloud (EU-Tenant): AWS/Azure/GCP mit EU-Datenresidenz, niedrigste Einstiegskosten.



## 2. Profil A: 200 Mitarbeiter

On-Premise: 85.000 EUR Setup + 2.800 EUR/Monat = 185.800 EUR / 36 Monate

Sovereign Cloud: 12.000 EUR Setup + 3.200 EUR/Monat = 127.200 EUR / 36 Monate

Public Cloud: 5.000 EUR Setup + 4.100 EUR/Monat = 152.600 EUR / 36 Monate

Break-Even On-Prem vs Cloud: Monat 28



### 3. Profil B: 500 Mitarbeiter

On-Premise:  $165.000 \text{ EUR Setup} + 4.500 \text{ EUR/Monat} = 327.000 \text{ EUR} / 36 \text{ Monate}$

Sovereign Cloud:  $22.000 \text{ EUR Setup} + 6.800 \text{ EUR/Monat} = 266.800 \text{ EUR} / 36 \text{ Monate}$

Public Cloud:  $8.000 \text{ EUR Setup} + 8.900 \text{ EUR/Monat} = 328.400 \text{ EUR} / 36 \text{ Monate}$

Break-Even On-Prem vs Cloud: Monat 22



## 4. Profil C: 1.500 Mitarbeiter

On-Premise: 320.000 EUR Setup + 8.200 EUR/Monat = 615.200 EUR / 36 Monate

Sovereign Cloud: 45.000 EUR Setup + 14.500 EUR/Monat = 567.000 EUR / 36 Monate

Public Cloud: 15.000 EUR Setup + 19.800 EUR/Monat = 727.800 EUR / 36 Monate

Break-Even On-Prem vs Cloud: Monat 18



## 5. GPU-Empfehlung

Bis 200 MA: NVIDIA A30 (24 GB) – ausreichend für 7B-Modelle.

200–500 MA: NVIDIA A100 (40/80 GB) – gute Balance Leistung/Kosten.

500+ MA: NVIDIA H100 (80 GB) – für große Modelle und hohen Durchsatz.

Alternativ: AMD MI300X für kosteneffiziente Skalierung.



## 6. Empfehlung

Bis 500 MA: Sovereign Cloud bietet das beste Kosten-Risiko-Verhältnis.

Ab 500 MA: On-Premise wird durch niedrigere laufende Kosten attraktiv.

Sensible Branchen (Finanzen, Gesundheit): On-Premise bevorzugen.

Startphase: Public Cloud für schnelle Validierung, dann migrieren.



# Bereit für den nächsten Schritt?

Vereinbaren Sie ein kostenfreies 30-Minuten-Erstgespräch mit unseren KI-Experten.

<https://www.ki-beratung-unternehmen.de/termin-buchen>

KBD KI-Beratung Deutschland  
DGSVO-konforme KI-Beratung & Implementierung für den Mittelstand  
[www.ki-beratung-unternehmen.de](http://www.ki-beratung-unternehmen.de)