



DER SICHERE WEG ZU PRIVATE GPT

Das neue Angebot der M&L AG für Betreiber und Entwickler produktiver KI-Lösungen.

Der nächste logische Schritt für professionelle GenAI-Systeme ist nicht das Modell, sondern die Infrastruktur dahinter. Mit unserem Private-GPT-Angebot schaffen wir die technische und wirtschaftliche Grundlage für produktionsreife LLM- und Agentensysteme – souverän, skalierbar und kontrollierbar.

1 STABILE KOSTEN & KOSTENTRANSparenZ

- Eigener KI-Server statt Token-Abrechnung und Preissprüngen.
- Planbare Betriebskosten – ideal für SLA- und Paketmodelle.
- Saubere Kalkulation für Vermittler, Entwickler und Endkunden.

2 DATENHOHEIT & DATENSICHERHEIT

- Öffentliche KI-Clouds können rechtlich dem USA Cloud Act unterliegen – auch bei EU-Hosting.
- On-Premises Betrieb: Daten bleiben im Unternehmen und verlassen die Trusted Infrastruktur nicht.
- Perfekt für sensible Bereiche: HR, Finanzen, Recht, Produktion, internes Wissen.

3 SCHNELLE INTEGRATION IN BESTEHENDE GenAI IMPLEMENTIERUNGEN

- Keine Insellösung: Private GPT ergänzt bestehende Architekturen.
- OpenAI kompatible API & MCP: Auch für Multi-Agenten-Portfolios.
- Weniger Umbau, schneller Go-live, bessere Projektmargen.

Ihr Ansprechpartner M&L AG:

Johann Schippers ■ Head of Innovation & Solution Development

Telefon: +49 69 963 632 37 ■ E-Mail: jschippers@mlconsult.com

**MACHEN SIE PRIVATE GPT ZUM FUNDAMENT IHRER KI-ANGEBOTE.
JETZT ANGEBOT ODER ARCHITEKTURGESPRÄCH ANFORDERN.**

TECHNISCHE ARCHITEKTUR & BETRIEBSMODELL

Eine kontrollierbare, performante und compliance-fähige LLM-Infrastruktur für professionelle KI- und Agentenlösungen – ohne Abhängigkeit von Public-Cloud-Anbietern.

■ Datensouveränität & KRITIS-Fähigkeit

Keine Daten verlassen das Unternehmen permanent oder werden in einer Public Cloud gespeichert. Ideal für sensible Daten, regulierte Branchen und KRITIS-Infrastrukturen. Rollenbasiertes Zugriffsmanagement (RBAC). User-based Databases für maximale Mandantentrennung und Vertraulichkeit.

■ Integration bestehender GenAI-Implementierungen

Standardisierte MCP-Schnittstelle für modulare Systemintegration. OpenAI-kompatible API – bestehende Integrationen können nahezu ohne Code-Anpassung migriert werden. Kompatibel mit bestehenden Bot-Frameworks, Agent-Architekturen und Orchestrierungs-Engins. Kein tokenbasiertes Abrechnungsmodell.

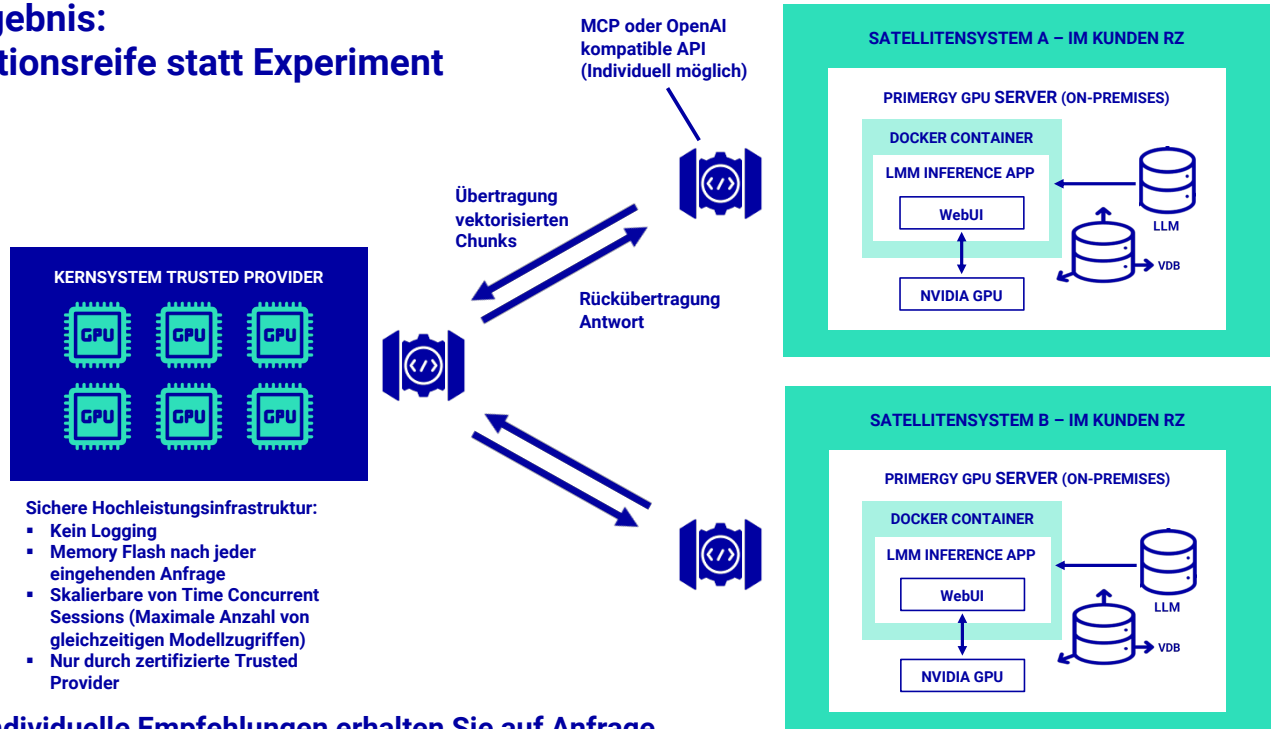
■ Stabiles & skalierbares Preismodell

Kein Token basiertes Abrechnungsmodell, sondern transparente Hardware und SaaS Kostenstruktur basierend auf verfügbaren Time Concurrent Sessions (TCS). Bedarfsgerechte Infrastruktur. Technische Skalierung über die maximale Anzahl gleichzeitiger Modellzugriffe (TCS).

■ Sichere Hochleistungsinfrastruktur von FSAS TECHNOLOGIES, dem Pionier für souveräne KI-Architekturen im europäischen Markt

Kein Logging sensibler Inhalte. Memory Flash nach jeder eingehenden Anfrage. Skalierbar über definierte Time Concurrent Sessions. Betrieb ausschließlich durch FSAS zertifizierte Trusted Provider.

■ Das Ergebnis: Produktionsreife statt Experiment



Preise und individuelle Empfehlungen erhalten Sie auf Anfrage.