

AI-PLATFORM

Sovereign AI Platform Engineering

Fünf Tage, nach denen Ihr Team eine Inferenz-Plattform nicht nur beschreiben kann, sondern eine gebaut, vermessen und übergeben hat.

DAUER	GRUPPE	FORMAT
5 Tage	bis 12 Personen	Inhouse · Offen · Im Lab · Deutsch oder Englisch

DAS EXISTIERT DANACH

- ▶ Ein Referenzarchitektur-Dokument über alle Schichten: Compute, Netzwerk, Storage, Identity, Registry, Inference-Stack
- ▶ Ein GPU- und Kapazitätsmodell mit nachvollziehbarer Sizing-Rechnung von L40S bis H100 für Ihre Lastannahmen, gestützt auf eigene Messwerte aus dem Lab
- ▶ Ein lauffähiger Inference-Stack, den Ihr Team selbst aufgebaut hat: Modell-Registry, Inference-Server, API-Gateway, Monitoring
- ▶ Ein Betriebsübergabe-Paket mit Runbook-Gerüst, Update-Pfad und Störungsprotokollen aus den Übungen
- ▶ Ein Entscheidungslog zu den strittigen Architekturfragen Ihres Hauses mit dokumentierter Begründung

AGENDA

Tag 1	Ihre Lastannahmen auf dem Prüfstand · Die Referenzarchitektur Schicht für Schicht · GPU-Kapazitätsmodell von L40S bis H100 · Übung: Sizing-Rechnung für Ihr Szenario · Tagesergebnis: dokumentiertes Kapazitätsmodell
Tag 2	Azure Local als Fundament · Identity durchgängig gedacht · Netzwerkdesign und Egress-Kontrolle · Übung: Fundament-Design im Lab umsetzen · Tagesergebnis: freigegebenes Fundament
Tag 3	Registry: Herkunft und Signatur von Modellen · Inference-Server im ehrlichen Vergleich · Aufbau: vom Modell zum API-Endpunkt · Lasttest und Tuning · Tagesergebnis: lauffähiger Stack mit Messprotokoll
Tag 4	Observability für GPU-Plattformen · Update-Pfad: Modelle, Treiber, Firmware, Container · Mandanten, Quoten und interne Verrechnung · Störungsübung: der Node fällt aus · Tagesergebnis: Betriebs-Dashboard und Störungsprotokoll
Tag 5	Das Runbook wird vollständig · Architektur-Review gegen Ihre Anforderungen · Abnahme-Simulation: Übergabe an den Betrieb · Transferplan für Montag · Abschlussergebnis: referenzierbare Blaupause



IHR TRAINER

Dino Bordonaro

8x Microsoft MVP in Folge (Microsoft Azure · Cloud and Datacenter Management), Microsoft Certified Trainer, fast 30 Jahre IT. Baut souveräne KI-Systeme auf eigener Hardware (Azure Local, Foundry Local, NVIDIA-GPUs) und schult genau darauf. Über 800 Professionals ausgebildet, jedes Konzept im eigenen Enterprise-Lab im TÜV-geprüften Rechenzentrum erprobt.

39.500 €

 pauschal, zzgl. MwSt.

Der Kurs baut Kompetenz und eine Blaupause auf, keine Produktionsumgebung in Ihrem Rechenzentrum. Implementierung und Managed Service sind eigenständige Leistungen von BORDONARO und werden separat beauftragt.

AI-PLATFORM