

AI - EVAL

Evaluation, Observability und Guardrails

Ohne Golden Dataset und Release-Gates ist jedes KI-Release ein Experiment an Ihren Nutzern.

DAUER**2 Tage****GRUPPE****bis 12 Personen****FORMAT**

Inhouse · Remote · Im Lab · Deutsch oder Englisch

DAS EXISTIERT DANACH

- ▶ Ein Golden Dataset für das eigene System: kuratierte Fälle mit Erwartungsergebnissen, Randfällen und bewusst schwierigen Anfragen
- ▶ Implementierte Qualitätsmetriken für Groundedness, Relevanz und Vollständigkeit, inklusive kalibrierem LLM-as-Judge mit Stichproben-Gegenprüfung
- ▶ Tracing über die gesamte Kette von der Anfrage über Retrieval und Modellaufruf bis zur Antwort, mit Kosten und Latenz je Schritt
- ▶ Ein Kosten- und Latenzbudget je Anfragetyp mit Alarmschwellen
- ▶ Definierte Release-Gates: welche Messwerte ein Deployment passieren muss und wann ein Rollback ausgelöst wird

AGENDA**Tag 1**

Der Blindflug-Befund · Das Golden Dataset bauen · Metriken implementieren: Groundedness, Relevanz, Vollständigkeit · Der erste Evaluationslauf · Tagesergebnis: eine belastbare Baseline

Tag 2

Tracing: die Kette sichtbar machen · Kosten- und Latenzbudget · Guardrails vor und nach dem Modell · Release-Gates und Regression-Erkennung · Tagesergebnis: das Gate hält

**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.

15.800 € pauschal, zzgl. MwSt.

Der Track befähigt Ihr Team, das eigene Messsystem zu bauen und zu betreiben. Er ist kein Audit Ihres Produktsystems und kein Managed-Monitoring-Service, dafür sprechen wir über ein Projekt.

AI - EVAL