Ihr Partner für IT

Pressemitteilung
05. November 2026

Observability in KI-Systemen:
Wenn Dashboards grün bleiben, obwohl das System lügt

Klassisches Monitoring reicht bei KI-Systemen längst nicht mehr aus: Selbst, wenn alle Dashboards grünes Licht geben, können Modelle unbemerkt an Qualität verlieren, falsche Antworten liefern oder Sicherheitslücken aufweisen. Der Beitrag beleuchtet die vier zentralen Herausforderungen der Observability in KI-Systemen und zeigt, mit welchen neuen Ansätzen sich Model Drift, fehlerhafte Ergebnisse und Sicherheitsrisiken frühzeitig erkennen und besser kontrollieren lassen.

Jahrzehntelang haben Software-Teams gelernt, ihre Systeme zu beobachten: Logs zeigen den Pfad, Metrics den Zustand, Traces die Ursache. Dieses Prinzip funktioniert – solange die zugrunde liegende Software deterministisch und inspizierbar ist. Bei neuronalen Netzen und Large Language Models gilt beides nicht mehr. Und das ist kein Detail am Rande, sondern ein fundamentaler Bruch mit den Grundannahmen, auf denen unsere gesamte Observability-Praxis aufbaut.

Grüne Kurven, kaputtes Modell

Wer heute KI-Komponenten in produktive Workflows integriert, kennt das Problem: Die Latenz stimmt, Fehlerraten sind niedrig, alle klassischen Dashboards zeigen Grün – und trotzdem liefert das Modell im Hintergrund zunehmend Unsinn. Der Fehler ist nicht technisch, sondern semantisch. Und dafür hat unser klassisches Observability-Werkzeug keine Antwort.

Praktisch jedes Team, das KI-Komponenten produktiv einsetzt, stößt früher oder später auf vier verwandte Problemfelder:

- **Model Drift ohne sichtbaren Trigger**: die Qualität sinkt schleichend, ohne Deployment, ohne Incident, ohne Alarm.
- **Halluzinationserkennung in Echtzeit**: Wie erkennt man plausibel klingende, aber falsche Aussagen, bevor sie Nutzer*innen erreichen?
- **Adversariale Inputs ohne klassische Anomalie**: Prompt-Injections und Jailbreaks sehen statistisch oft völlig harmlos aus.
- **Das strukturelle Problem "LLM-as-a-Judge"**: Ein Modell, das man nicht versteht, bewertet ein anderes, das man nicht versteht.

Gerade über den letzten Punkt lohnt es sich nachzudenken: Viele Teams nutzen mittlerweile LLMs, um die Qualität anderer LLM-Outputs zu bewerten – ein pragmatischer Weg, menschliches Review zu skalieren. Nur schiebt das die Blackbox-Problematik nicht weg, sondern verdoppelt sie.

Ihr Partner für IT

Pressemitteilung
05. November 2026

Was tatsächlich funktioniert

Die gute Nachricht: Es gibt längst Ansätze, die sich in echten Produktionsumgebungen bewährt haben, auch wenn sie noch nicht so etabliert sind wie klassisches Monitoring.

- **Embedding-Monitoring:** Frühwarnsystem für Distributional Shift, lange bevor sich Qualitätsprobleme in harten Metriken zeigen.
- **Behavioral Testing und Metamorphic Testing**: Ersatz für klassische Unit Tests, die bei fehlendem festen Erwartungswert ins Leere laufen.
- **Evals als Engineering-Disziplin**: versioniert, reviewed und kontinuierlich erweitert, statt einmalig vor dem Deployment abgehakt zu werden.

Ebenso wichtig ist die Ehrlichkeit über die Grenzen: Keine dieser Methoden löst das Grundproblem der fehlenden Inspizierbarkeit – sie machen KI-Systeme beherrschbarer, nicht durchsichtig. Wer sich darauf verlässt, dass ein Eval-Set oder ein Judge-Modell alle Probleme abfängt, wird überrascht werden.

Warum die Problematik so relevant ist

Ob Chatbots, RAG-Systeme, automatisierte Entscheidungsunterstützung oder Code-Generierung – KI-Komponenten stecken längst in kritischen Workflows, oft ohne, dass die bestehende Observability-Infrastruktur mitgewachsen ist. Die drei klassischen Säulen bleiben wichtig, reichen aber allein nicht mehr aus. Es lohnt sich, jetzt eine Haltung dazu zu entwickeln, bevor der nächste Incident sie erzwingt.

Interessiert am Thema?

Dann erfahren Sie noch mehr bei den Vorträgen von **Sebastian Lunemann**, Consultant TEAM GmbH:

- KI-Navigator – Do. 19.11.26 | 12:00 – 12:45 Uhr | Raum Helsinki
- UKOUG Conference Discover – 23.11.2026 | 15:00 – 15:45 Uhr |Mezzanine - Muscadet

Mehr Informationen unter:
www.team-pb.de/doag-konferenz

Ergänzend stehen unsere Expert*innen unter vertrieb-oracle@team-pb.de oder unter 05254 8008-50 jederzeit für alle Fragen zur Verfügung.

(4.043 Zeichen inkl. Leerzeichen)

Ihr Partner für IT

Pressemitteilung
05. November 2026

Über TEAM

*Die TEAM GmbH aus Paderborn ist auf die Konzeption und Realisierung moderner IT-Lösungen spezialisiert. Auf Basis von Standardsoftware oder als Individualentwicklung realisieren unsere Expert*innen zukunftsorientierte Lösungen und stellen diese passgenau für Unternehmen aus diversen Branchen bereit.*

- *TEAM entwickelt Datenbank-Konzepte auf höchstem Niveau, bieten Managed Services, Schulungen, Cloud-Strategien, individuelle Software-Entwicklungen und maßgeschneiderte KI-Lösungen. TEAM ist einer der führenden Oracle-Partner Deutschlands.*

- *In unserem Geschäftsbereich Intralogistik-Software machen wir Lager zukunftsfähig: Mit dem Warehouse Management System ProStore® digitalisieren, automatisieren und optimieren wir sämtliche Lagerprozesse – transparent, skalierbar und bereit für die Anforderungen moderner Logistik.*

*Die TEAM GmbH wurde 1982 in Paderborn gegründet und beschäftigt rund 100 Mitarbeiter*innen. TEAM ist ein Mitglied der Materna-Gruppe. Renommierte Unternehmen aus Industrie, Handel und Dienstleistung gehören zu unseren langjährigen Kunden.*

Pressekontakt

Clarissa Lorz
Senior Marketing Managerin

TEAM GmbH
Hermann-Löns-Straße 88
33104 Paderborn

Fon +49 52 54 80 08-86
Fax +49 52 54 80 08-19
Mail marketing@team-pb.de
Web www.team-pb.de