Artikelzusammenfassung

Chatbot-Antworten prüfen: Wann ist der Einsatz vertretbar?

Ein passender Quellenlink macht eine KI-Antwort noch nicht verlässlich. Entscheidend sind die **richtigen Informationen**, fachliche Prüfung und die Bedingungen, unter denen Menschen eine Antwort verwenden.

FÜNF PRÜFFELDER – EINE FALLBEZOGENE SYNTHESE

1	**Informationsstand**	Gelten die Dokumente und Versionen? Werden Ausnahmen bei der Suche gefunden?
2	**Antwort & Beleg**	Sind Quellenbezug, Richtigkeit und alle relevanten Bedingungen geprüft?
3	**Grenzverhalten**	Fragt der Chatbot bei Lücken nach und vermeidet er unzulässige Freigaben?
4	**Berechtigung & Kontrolle**	Können Fachleute Antworten prüfen, widersprechen und rechtzeitig handeln?
5	**Betriebliche Eignung**	Funktioniert der Prüfablauf – auch mit Aufwand, Fehlern und Rückfallweg?

EINE GUTE GESAMTQUOTE KANN KRITISCHE FEHLER VERDECKEN

Fiktiver Fall: 57 von 60 Antworten sind fachlich ausreichend. In der Gruppe mit Sonderregeln und Versionskonflikten fehlen jedoch zweimal wichtige Bedingungen. Routineerfolge gleichen kritische Fehler nicht aus.

NÄCHSTE UNSICHERHEIT GEZIELT PRÜFEN

Prozess-Proof-of-Concept (PoC): Mit abgeschlossenen Fällen prüfen, ob Fachleute KI-Antworten und Originalbelege im vorgesehenen Ablauf tatsächlich gegenprüfen können. Noch keine Freigabe für einen betrieblichen Pilot.

VOR PILOT UND BETRIEB

Vor dem Pilot: Kritische Fehler beheben und erneut testen; Kenntnisse, Rechte, Zeit und Kapazität der Prüfenden klären.

Im Betrieb: Änderungen an Dokumenten, Suche und Modell überwachen; Fehlantworten fachlich bewerten, korrigieren und erneut testen.

TECHNISCH PLAUSIBEL HEISST NICHT BETRIEBLICH VERLÄSSLICH

Die nächste Entscheidung muss zur vorhandenen Evidenz passen. Erst wenn Antwortqualität und Kontrollweg für die vorgesehene Nutzung tragen, kommt eine erweiterte Verwendung infrage.

Albert Leins · albertleins.de

Vollständiger Fachbeitrag und Quellen

Grundlagen: NIST AI RMF · OWASP LLM09 · Microsoft Azure RAG Evaluation