Cyber Defense Horizons · 08.10.2026 · 12:15 Uhr

Wer haftet, wenn die KI antwortet.

Kontrolliertes Agentensystem: Zuständigkeit, Quellennachweis, Umgang mit Unsicherheit

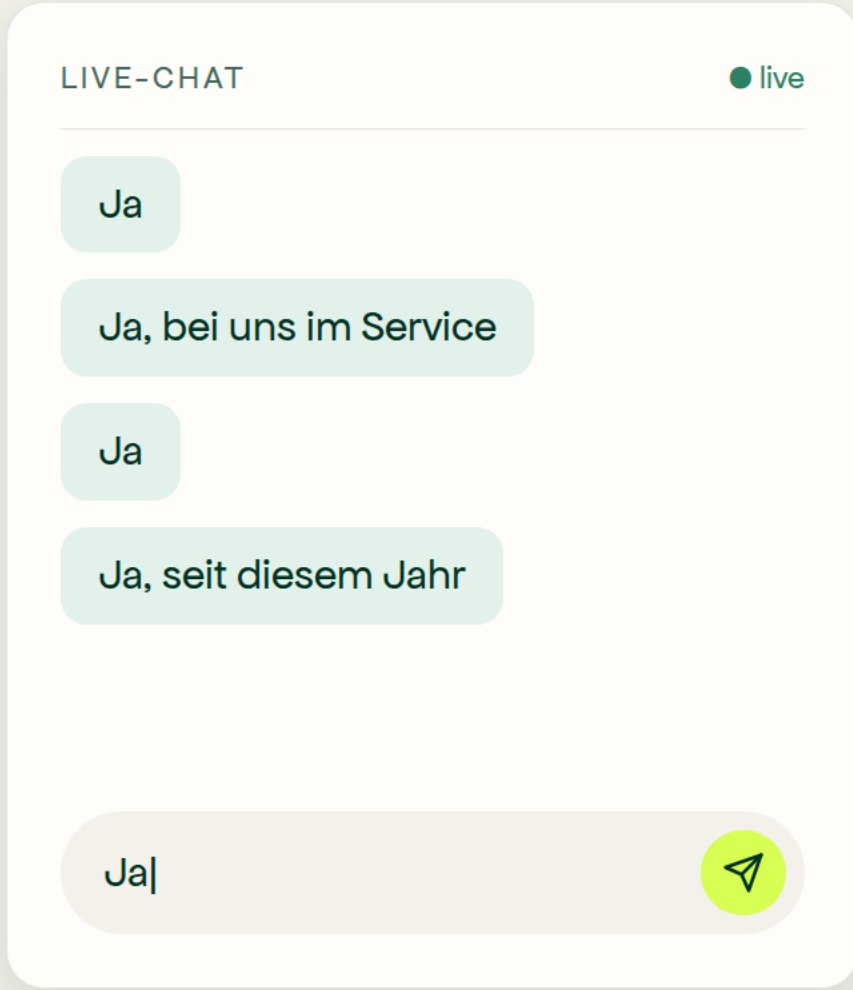
Mike Piotraschke

VP Sales DACH · octonomy

octonomy

EINE KURZE FRAGE ZUM START ↗

Wer von Ihnen hat heute KI im Einsatz, die Fragen von Mitarbeitenden oder Kunden beantwortet?



Schreiben Sie einfach „Ja“ in den Chat.

> Scale-up NRW Night: Fast jede Hand im Raum ging hoch.

Herzlichen Glückwunsch.

Sie verantworten jetzt jede Antwort.



Nicht der Anbieter des Modells. Der Name auf der Website.

DAS IST KEINE THEORIE ↗

Gerichte und Gesetzgeber sehen es genauso

Urteil Moffatt v. Air Canada · 2024

Der Web-Assistent erfindet eine Erstattungsregel.

Die Airline argumentiert, der Assistent sei selbst verantwortlich. Das Tribunal weist das ab.

Ergebnis: Die Airline zahlt.

Regulierung EU AI Act

Pflichten treffen auch die, die KI einsetzen.

Anbieter	Betreiber
baut das System	setzt es ein: Sie

Etwa KI-Kompetenz im Team (Art. 4), bei Hochrisiko-Systemen Aufsicht und Überwachung im Betrieb (Art. 26).

WORAN ES WIRKLICH FEHLT ↗

Der Kollege, zu dem alle zuerst gehen

ANTWORTZEIT

11 Sek.

Und er sagt dazu, wo es steht.

Er weiß, welche Revision der Zeichnung gilt.

Er beantwortet eine Frage nach der anderen.

Und irgendwann geht er in Rente.

WARUM MODELLE HIER SCHEITERN ↗

Die Antwort liegt zwischen Text und Zeichnung

Kann das Modell …

einen Absatz lesen? ✓ Ja

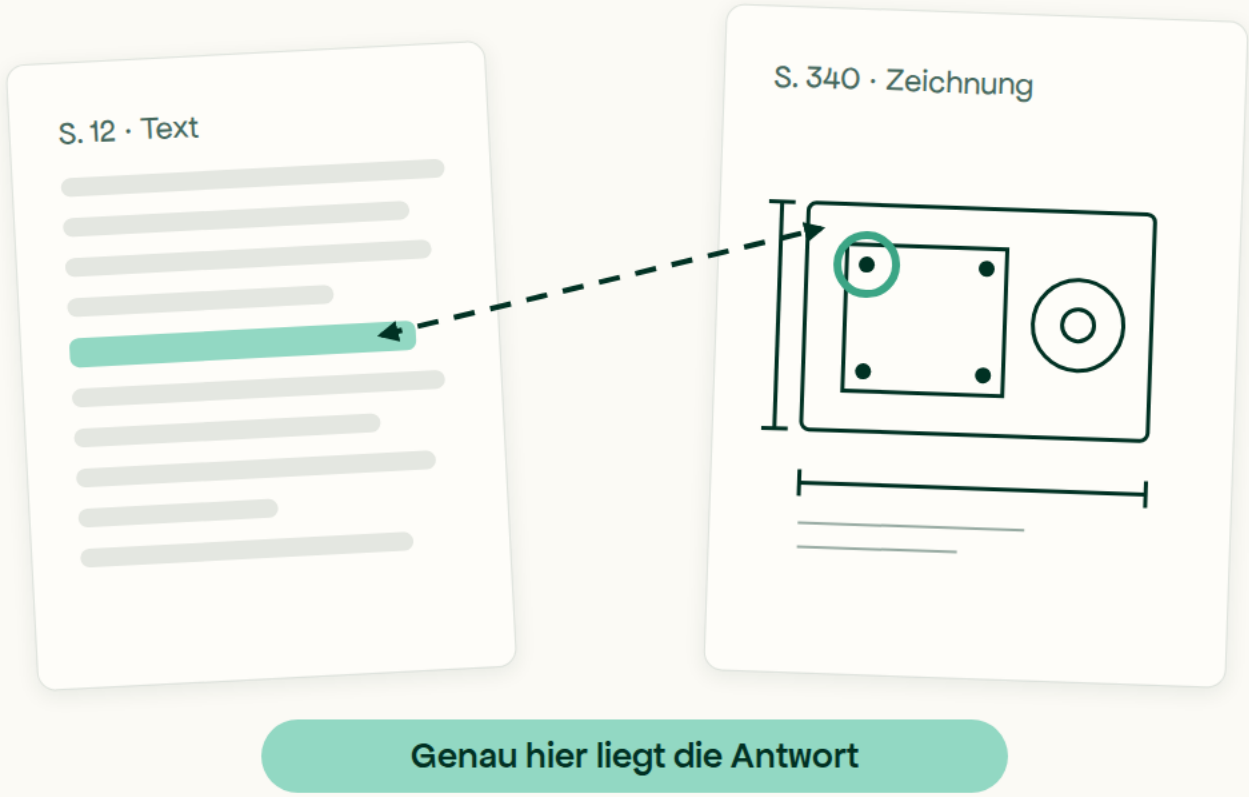
Kann das Modell …

eine Zeichnung beschreiben? ✓ Ja

Kann das Modell …

Seite 12 mit Seite 340 verbinden?

⚠ Bisher nicht verlässlich



Über 20.000 Seiten Handbücher, Schaltpläne, Tabellen und Tickets.

DER UMWEG DER BRANCHE ↗

Text zerschneiden. Durchsuchen. Hoffen, dass die Antwort in einem Satz steht.



In echten Unternehmen steht sie dort selten.

Das System rät dann. Und klingt dabei sehr überzeugt.

KONTROLLIERTES AGENTENSYSTEM ↗

Drei Prinzipien, damit jemand geradestehen kann

01

Zuständigkeit

Jeder Agent hat einen Auftrag, eine Quelle und einen Owner.

Wer antwortet?

02

Quellennachweis

Jede Antwort zeigt Dokument und Seite.

Woher kommt es?

03

Unsicherheit

Steht es nicht drin, sagt das System genau das.

Und wenn nicht?

01 · ZUSTÄNDIGKEIT ↗

Wer darf was beantworten, und wer steht dafür gerade?

FRAGE	WER ANTWORTET	QUELLE	WER VERANTWORTET
Welches Anzugsmoment gilt an Baureihe X?	Technik-Agent	Serviceanleitung, aktuelle Revision	Leitung Technischer Service
Wann kommt meine Ersatzteillieferung?	Auftrags-Agent	ERP, live abgefragt	Kundenservice
Ist das ein Garantiefall?	**Kein Agent**	Übergabe an einen Menschen	Garantieabteilung

Jede Antwort zeigt, wo sie steht

FRAGE

Welches Anzugsmoment gilt für die Befestigung an Baureihe X?

ANTWORT

[Wert] Nm, laut Tabelle [Nr.]

Serviceanleitung · Rev. [x] · Seite 340 ↗

Prüfbar in Sekunden.

Der Mensch kontrolliert die Quelle, nicht das Modell.

> **Das ist der Unterschied zwischen Vertrauen und Nachvollziehbarkeit.**

„Das steht nicht in Ihren Unterlagen.“

Der wichtigste Satz, den ein KI-System sagen kann. Danach folgt die Übergabe an einen Menschen.

ERFUNDEN: EINE ÖFFNUNGSZEIT

Peinlich.

→

ERFUNDEN: EIN ANZUGSMOMENT

Ein ganz anderer Abend.

AUS SECURITY-SICHT ↗

Drei Prinzipien, drei Kontrollen, die Sie schon kennen

01 · Prinzip **Zuständigkeit**	→	Kontrolle **Least Privilege**	Jeder Agent sieht nur seine Quelle und nur, was der Fragende sehen darf.
02 · Prinzip **Quellennachweis**	→	Kontrolle **Audit-Trail**	Jede Antwort ist rückverfolgbar, bis auf Dokument und Seite.
03 · Prinzip **Unsicherheit**	→	Kontrolle **Integrität**	Ein System, das nicht rät, ist schwerer zu manipulieren.

FÜR IHREN NÄCHSTEN TERMIN ↗

Fünf Fragen, bevor KI mit Kunden spricht

1 Woher kommt die Antwort, und kann ich die Stelle sehen?

2 Wer hat die letzten hundert Antworten tatsächlich gelesen? Die unangenehmste

3 Was passiert, wenn die Antwort nicht in den Unterlagen steht?

4 Sieht das System nur, was der Fragende sehen darf?

5 Wer ist Owner, wenn eine Antwort falsch war?

PRÜFBAR, NICHT BEHAUPTET ↗

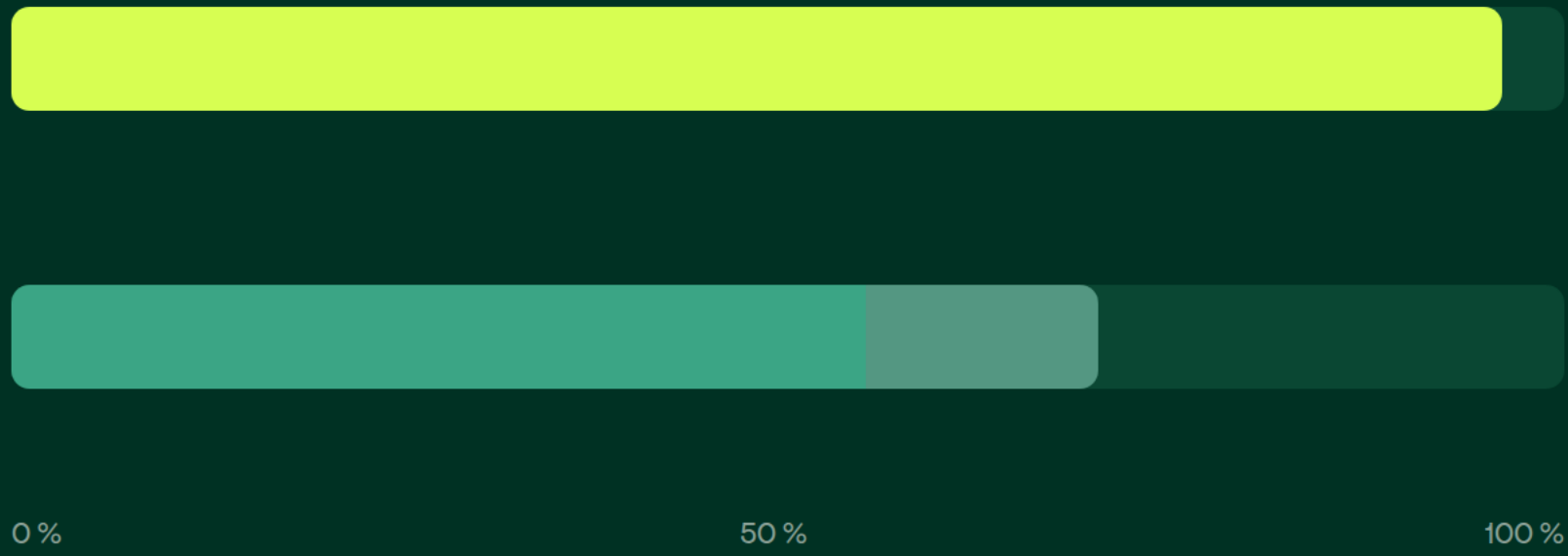
Genauigkeit im unabhängigen Test

96 %+

octonomy Expertensystem

55–70 %

Generische KI



Quelle: Tolly Group, Report #226106 · octonomy · Cyber Defense Horizons 2026 · 14

Wer ist diese Person in Ihrem Unternehmen?

Und was passiert an dem Tag, an dem sie geht?

Mike Piotraschke

VP Sales DACH · octonomy

michael.piotraschke@octonomy.ai · LinkedIn: Mike Piotraschke

octonomy