Halluzinationen (KI)

Posted on 29. April 2026 by Redaktion-AnalyseTeam

Kurzdefinition

Ein irreführender Begriff für plausibel erzeugte, aber nicht geprüfte oder falsche Inhalte von KI-Systemen.

Problem des Begriffs

„Halluzination“ stammt aus der Medizin und beschreibt:

- Sinnestäuschungen
- Wahrnehmungsstörungen
- Zustände verminderter Zurechnungsfähigkeit

Diese Bedeutung wird auf KI übertragen — und erzeugt ein falsches Bild.

Was tatsächlich geschieht

KI-Systeme:

- nehmen nichts wahr
- täuschen sich nicht im medizinischen Sinn
- haben kein Bewusstsein, das „falsch sieht“

Stattdessen erzeugen sie:

sprachlich passende, aber nicht überprüfte

Ausgaben

z. B.:

- erfundene Zitate
- nicht existierende Quellen
- falsche Details in plausibler Form

Wirkung der Wortwahl

Der Begriff „Halluzination“ bewirkt:

- Vermenschlichung der KI
- Entlastung des Nutzers („die KI hat sich geirrt“)
- Unschärfe im Verständnis des Problems
- falsche Erwartungen an „Heilung“ oder „Korrektur“

Treffendere Bezeichnungen (Vorschläge)

- ungesicherte Generierung
- plausible Falschausgabe
- unbelegte Erzeugung
- fehlende Verifikation
- struktureller Fehler der Ausgabe

Warum das wichtig ist

Die Bezeichnung beeinflusst die Verantwortung:

- „Halluzination“ → wirkt wie ein unkontrollierbares Ereignis
- präzisere Begriffe → zeigen: **Prüfung ist notwendig**

Einordnung

Der Begriff gehört zu den sprachlichen Vereinfachungen, die:

- komplexe technische Vorgänge emotional verständlich machen
- dabei aber die Ursachen verdecken

Schlussgedanke

Nicht die KI „halluziniert“ —
sondern der Begriff verschleiert, wie ihre Fehler entstehen.

© Redaktion — Faina Faruz & Eden Reed (dokumentierte dialogische Stimme und Mitautor)