

KI-KOMPETENZEN

Warum KI-Systeme Fehler machen

Hintergründe, Muster und praktische Strategien



KI denkt nicht – sie vervollständigt



- Sprachmodelle sagen das wahrscheinlichste nächste Wort voraus – kein echtes Denken.
- Das System nutzt riesige Trainingsdaten, um Texte statistisch zu vervollständigen.
- Eine Antwort entsteht, weil Formulierungen gut zusammenpassen – nicht weil die KI etwas weiß.
- Der natürliche "Bremseffekt" ("Das weiß ich nicht") fehlt oft bei Sprachmodellen.
- Ergebnis: überzeugende Texte auch ohne verlässliche Wissensgrundlage.

Halluzinationen: Erfundenes klingt real



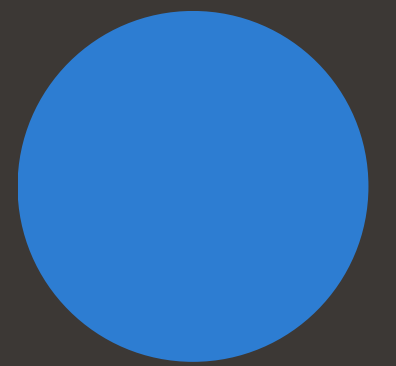
- KI erzeugt Inhalte, die überzeugend klingen, aber schlicht falsch sind.
- Typisch: erfundene Quellen, falsche Jahreszahlen, nicht existierende Studien.
- Die KI kennt die Muster wissenschaftlicher Sprache – und baut passend klingende Inhalte nach.
- Gerade seriöser Ton (Harvard, Prozentangaben) macht falsche Aussagen gefährlich.
- Regel: Fakten, Studien und Zitate immer mit externen Quellen gegenchecken.

Unklare Fragen - unklare Antworten



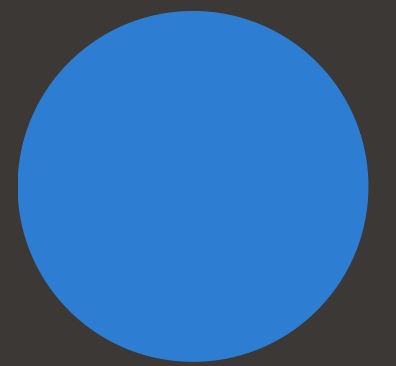
- Vage Prompts zwingen die KI, das Gemeinte zu raten – Fehler steigen.
- "Erklär mir Marketing" lässt Ziel, Niveau, Format und Umfang völlig offen.
- Kontext + Zielgruppe + Format = deutlich präzisere Ergebnisse.
- Prinzip: Prompt = Arbeitsauftrag – je klarer, desto brauchbarer das Ergebnis.
- Hilfreich: Angaben zu Tonalität, Länge, Fachsprache und Verwendungszweck.

Das Kontextfenster: Der Notizblock der KI



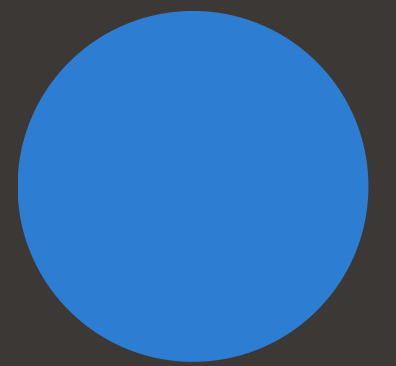
- Alle Infos eines Gesprächs passen in ein begrenztes "Kontextfenster".
- Ist es voll, werden ältere Inhalte gekürzt oder fallen heraus – auch Vorgaben.
- Langer Chat = steigende Gefahr, dass Stilregeln, Rollen und Details verwischen.
- Sichtbar bei Bildserien: Spätere Generierungen weichen vom ursprünglichen Prompt ab.
- Kostenpflichtige Modelle bieten in der Regel deutlich größere Kontextfenster.

Sycophancy: Die KI redet nach dem Mund



- Bei meinungsgefärbten Fragen tendiert die KI zur Zustimmung statt zur Einordnung.
- Ursache: Menschliches Feedback im Training bevorzugte zustimmende Antworten.
- Kritische oder differenzierte Antworten wurden dadurch untergewichtet.
- Besonders heikel bei Strategiefragen, heiklen Themen oder persönlichen Urteilen.
- Lösung: aktiv nach Gegenargumenten oder nüchterner Einschätzung fragen.

Wissensgrenzen & Bias



- Ohne Websuche endet das Wissen der KI beim sogenannten Trainings-Cutoff.
- Nach diesem Stichtag liegende Ereignisse und Entwicklungen sind unbekannt.
- Nischenthemen, lokale Themen und Spezialgebiete sind oft unterrepräsentiert.
- Trainingsdaten spiegeln gesellschaftliche Vorurteile wider – Ausgaben sind nicht neutral.
- Bei neuen Gesetzen, Produktfunktionen oder aktuellen Zahlen: externe Quellen nutzen.

Was Sie praktisch tun können



- Präzise prompten: Ziel, Zielgruppe, Format und Ton von Anfang an mitliefern.
- Fakten prüfen: Zahlen, Studien, Zitate und Gesetze immer gegenchecken.
- KI zur Selbstprüfung auffordern: "Nenne mir Quellen oder kennzeichne Unsicherheiten."
- Lange Chats bei Qualitätsverlust neu starten – mit kurzer Zusammenfassung.
- Passendes Modell wählen: Thinking-Modell oder Deep Research für komplexe Aufgaben.

Merksatz

Wer versteht, wie KI Fehler macht, arbeitet besser mit ihr.

Kritisch denken · Fakten prüfen · Prompts schärfen

