Warum ChatGPT halluzinieren muss

DREI ARTEN VON KI	REGELBASIERTE KI	LERNBASIERTE KI	GENERATIVE KI
Vergleichbar mit	**Zettelkastenprofi** 	**Bibliothekar*in** 	**Professor*in** 
Beispiele	ELIZA, Deep Blue	Google-Suche	ChatGPT, Mistral, Claude
Merkmale	Sucht nach „Wenn – dann“-Regeln	Findet bekannte Infos wieder	Erfindet Antworten aus dem Gelernten
Nachvollziehbar?	Ja, jede Entscheidung ist erklärbar	Bedingt, abhängig von Algorithmus und Daten	Nein, Antworten entstehen in einer „Black Box“

1. Was macht generative KI anders?

Generative KI wie ChatGPT ist ein **Large Language Modell (kurz LLM)**. Sie hat Millionen Texte gelesen und in Wortbruchstücke zerlegt, sogenannte Token. Diese Token wurden in Wahrscheinlichkeiten zueinander angeordnet („gewichtet“). Ein LLM erinnert sich nicht an Sätze und auch nicht an Daten hinter den Worten. Es sagt neue Sätze nur voraus, basierend auf den gelernten Wahrscheinlichkeiten der Wörter zueinander, plus ein bisschen Zufall.

2. KI halluziniert: Sie erfindet Antworten, auch Fakten

Generative KI „weiß“ nichts und schlägt auch nicht nach. Als prädiktive (vorhersagende) KI sagt sie Texte voraus. Das klappt sehr gut bei kurzen, oft wiederholten Infos mit klaren Wahrscheinlichkeiten: „Paris ist die Hauptstadt von … Frankreich.“ Bei komplexen oder seltenen Aussagen rät sie öfter daneben, und es klingt immer überzeugend. Das nennen wir **Halluzination**. Auch Quellen kann sie überzeugend erfinden.

3. KI ist nicht neutral

KI lernt Verzerrungen (Bias) aus den Trainingsdaten mit. Bilder von „doctor“ zeigen weiße Männer. Menschen mit Behinderung sitzen leidend im Rollstuhl. Frauen sind jung, weiß, zu schlank und gutaussehend. Mit klaren Prompts kannst du dagegen ein Stück weit ankämpfen. Das braucht aber viel Sensibilität und Bewusstsein.

4. Das hilft gegen Halluzinationen

- **Relevanten Kontext liefern:** Beschreibe Kontext und Anlass. Hänge Dateien an, lass die KI danach suchen oder gib sie als Text mit.
- **Nach Unsicherheiten fragen:** Bitte die KI, zu sagen, wenn sie spekuliert. Fordere von Beginn an Belege und Begründungen.
- **Die KI abwägen lassen:** Lass dir bewusst mehrere Antworten ausgeben und frage, welche am plausibelsten ist („Tree of Thoughts“).
- **Klar prompten:** Formuliere deine Fragen offen, aber präzise.
- **Schrittweise vorgehen:** Stelle eine komplexe Frage in Etappen: erst zusammenfassen, dann ausformulieren usw. („Chain of Thought“).
- **Doppelt prüfen:** Mach bei kritischen Themen immer einen Faktencheck mit den echten Quellen.

Autorin: Julia Junge · Stand: 13. Oktober 2026 · keine Gewähr

CC BY SA