

# Warum Generative KI halluzinieren muss:

	Regelbasierte KI	Lernbasierte KI	Generative KI
Vergleich	Zettelkastenprofi	Bibliothekar*in	Professor*in
Merkmale	Sucht nach „Wenn – dann“-Regeln	Findet bekannte Infos wieder	Erfindet Antworten aus dem Gelernten
Nachvollziehbar?	Ja, jede Entscheidung ist erklärbar	Bedingt, abhängig vom Algorithmus + Daten	Nein, Antworten entstehen in einer “Black Box”

## 1. Was macht Generative KI anders?

Generative KI wie ChatGPT hat als Großes Sprachmodell (LLM = Large Language Modell) Millionen Texte gelesen und in Bruchstücke von Wörtern (sog. Token) zerlegt. Dabei wurden diese in Wahrscheinlichkeiten zueinander angeordnet (“gewichtet”). Ein LLM erinnert sich nicht an einzelne Sätze. Stattdessen „baut“ es neue Sätze, basierend auf den erlernten Wahrscheinlichkeiten der Wörter zueinander + ein bisschen Zufall.

## 2. KI halluziniert, d.h. sie erfindet Antworten inkl. Fakten

Generative KI „weiß“ nichts und schlägt auch nicht nach. Es sagt als prädiktive (vorhersagende) KI Texte voraus. Das klappt bei kurzen, oft wiederholten Infos mit klaren Wahrscheinlichkeiten in den Datenverbindungen meist sehr gut: „Paris ist die Hauptstadt von Frankreich.“  Bei komplexen oder seltenen Aussagen „räts“ sie dagegen öfter daneben – und es klingt trotzdem immer überzeugend. Das nennen wir **Halluzination**. Auch Quellen, kann sie sehr 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 im Rollstuhl und leiden. Frauen sind jung, weiß, extrem schlank und gutaussehend. Mit klaren Prompts lässt sich dagegen teilweise ankämpfen, dafür braucht es 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.
- Klares Prompting:** Formuliere deine Fragen offen, aber präzise.
- Nach Unsicherheiten fragen:** Bitte die KI, zu sagen, wenn sie spekuliert. Bitte um Belege und Begründungen.
- Schrittweise vorgehen:** Stelle eine komplexe Frage in Etappen (erst zusammenfassen, dann ausformulieren etc.), sog. “Chain of Thought”.
- Lass die KI abwägen:** Lass dir bewusst mehrere Antworten ausgeben. Frage, welche am plausibelsten ist („Tree of Thoughts“).
- Doppelt prüfen:** Führe bei kritischen Themen immer einen Faktencheck mit echten Quellen durch.

