

# HALLUZINATIONEN ERKENNEN FAKTENCHECK MIT KI

Für Schüler:innen – wie du der KI nicht alles glaubst

KI-AKADEMIE  
FÜR LEHRKRÄFTE  
by **educaite**

Material der KI-Akademie für Lehrkräfte von educaite

## 📄 DAS WICHTIGSTE AUF EINER SEITE

**KI klingt immer überzeugt – auch wenn sie Quatsch erzählt.** „Halluzinationen“ sind erfundene Antworten, die echt klingen, aber falsch sind. Wer das Muster kennt, fällt nicht darauf rein. Hier das Wichtigste, Beispiele und ein Faktencheck folgen ab Seite 2.

## KI-HALLUZINATIONEN – AUF EINEN BLICK

### ① Was ist das?

Eine selbstbewusst vorgetragene Falschantwort – klingt super, stimmt aber nicht.

### ② Warum passiert das?

KI baut *wahrscheinliche* Sätze, keine geprüften Fakten. Hauptsache, es klingt plausibel.

### ③ Wo besonders oft?

Bei Zahlen, Daten, Zitaten, Quellen, Links und Nischenwissen.

### ✓ So bist du sicher

- Quellen einfordern und selbst öffnen
- Wichtiges mit 2. Quelle prüfen
- Konkret fragen, damit es prüfbar ist

### ④ Warnsignale

Sehr sicher, sehr genau, aber keine Quelle – oder eine Quelle, die du nicht findest.

### ⑤ Gegencheck

Zweite Quelle, zweites KI-Modell oder selbst nachschlagen.

### ⑥ Goldene Regel

Bei Fakten nie blind vertrauen – kurz prüfen, bevor du es übernimmst.

### ✗ Riskant

- Zahlen/Zitate ungeprüft abschreiben
- Erfundene Quellen einfach glauben
- KI für aktuelle Ereignisse blind nutzen

## SCHLAUER PROMPT ZUM MITNEHMEN

„Nenne mir für jede Behauptung eine Quelle und sag, wie sicher du dir bist. Wenn du etwas nicht weißt, sag es – erfinde nichts.“ Dann die Quellen wirklich öffnen.

## Was Halluzinationen sind – und warum sie passieren

Eine KI-Halluzination ist eine frei erfundene Antwort, die der Chatbot mit voller Überzeugung ausgibt. Das passiert nicht nur selten und nicht nur bei einem Tool – alle großen Modelle (ChatGPT, Claude, Gemini & Co.) tun das gelegentlich. Der Grund liegt in ihrer Bauweise: Sprachmodelle sind darauf trainiert, **wahrscheinlich klingende** Texte zu erzeugen, nicht zwingend **korrekte**. Ob der Inhalt stimmt, ist für das Modell erst einmal zweitrangig.

Zwei echte Beispiele:

- 2023 behauptete ein KI-Chatbot bei einer großen Live-Präsentation, ein Weltraumteleskop habe „das erste Bild eines Exoplaneten“ gemacht – das stimmte nicht.
- Zwei Anwälte reichten 2023 vor Gericht sechs Urteile als Beleg ein, die ChatGPT erfunden hatte. Keines existierte wirklich.

Merke: Wenn sogar Profis darauf hereinfliegen, lohnt sich der Faktencheck immer.

## Wo die KI besonders oft daneben liegt

- **Zahlen & Daten:** Jahreszahlen, Statistiken, Geburtstage – gern leicht (oder ganz) falsch.
- **Zitate:** „Wer hat das gesagt?“ – die KI legt Personen oft Sätze in den Mund, die sie nie gesagt haben.
- **Quellen & Links:** erfundene Buchtitel, Studien oder URLs, die es nicht gibt.
- **Nischen- & Lokalwissen:** je spezieller das Thema, desto wackliger.
- **Aktuelles:** Ereignisse nach dem Trainingsstand kennt die KI nicht zuverlässig.

# Der 5-Schritte-Faktencheck

## 1. Quellen einfordern

Frag direkt: „Woher hast du das? Nenne mir die Quelle.“ Schon das entlarvt manche Erfindung.

## 2. Quellen selbst öffnen

Wichtig: Die genannte Quelle wirklich aufrufen und nachsehen. Existiert sie? Steht da, was die KI behauptet? Erfundene Links fallen hier auf.

## 3. Zweite Meinung holen

Stell dieselbe Frage einem zweiten KI-Modell oder nutze eine KI mit echter Websuche. Widersprechen sich die Antworten, ist Vorsicht angesagt.

## 4. Klassisch gegenprüfen

Bei wichtigen Fakten ein Schulbuch, ein Lexikon oder eine seriöse Website checken – gerade für Referate und Hausarbeiten.

## 5. Spezifisch fragen

Je genauer deine Frage, desto prüfbarer die Antwort. „Nenne 3 Belege mit Jahr und Quelle“ ist besser als „Erzähl mir was über ...“.

## Clever fragen – und kurz üben

- **Unsicherheit abfragen:** „Wie sicher bist du? Sag es, wenn du es nicht genau weißt.“ Gute Modelle geben das zu.
- **Erfinden verbieten:** „Erfinde keine Quellen. Wenn du keine kennst, schreib das.“
- **Mini-Übung:** Frag die KI nach einem Zitat einer berühmten Person und prüfe es online. Oft erlebst du live, wie eine Halluzination entsteht.

### ✓ Dein Faktencheck-Reflex

Mach es dir zur Gewohnheit: Bevor eine KI-Info in dein Heft, Referat oder deine Hausarbeit wandert, prüfst du sie kurz. Dieser Reflex unterscheidet jemanden, der KI *nutzt*, von jemandem, der ihr einfach *glaubt* – und er schützt deine Noten.



### Fazit

KI ist ein starkes Werkzeug, aber kein Lexikon: Sie klingt sicher, auch wenn sie irrt. Wenn du Quellen einforderst, sie selbst öffnest und Wichtiges gegenprüfst, nutzt du KI clever – statt ihr blind zu vertrauen. Dein Kopf bleibt der Faktenchecker.