



Arbeitsblatt

Warum KI überzeugend lügt

Einstieg · Kapitel 3: Fakes und Falschinfos erkennen · Lektion 1 von 5

Ziel: Ich erkenne die Stellen, an denen eine KI am wahrscheinlichsten erfindet.

Name: _____ **Klasse:** _____ **Datum:** _____

Die KI lügt nicht, sie schätzt

Eine KI hat kein Gefühl dafür, was sie weiß und was nicht. Sie erzeugt in beiden Fällen denselben souveränen Ton.

Der Fachbegriff dafür ist Halluzination: eine erfundene Angabe im Stil einer belegten. „Das weiß ich nicht“ ist für ein Sprachmodell die unwahrscheinlichste Fortsetzung.

Souveräner Ton sagt gar nichts. Prüf die Angaben, nicht das Gefühl.

Die Risikoliste

Je genauer eine Angabe wirkt, desto eher ist sie erfunden. Der Grund: Das Format einer Quellenangabe erzeugt ein Modell genauso mühelos wie ihren Inhalt.

- Quellen, Studien, Buchtitel, Seitenzahlen
- Paragraphen, Fristen, Aktenzeichen
- Zahlen mit Komma und Jahresangaben
- Zitate von echten Personen

Wo die Erfindungen am dichtesten sitzen

Halluzinationen verteilen sich nicht gleichmäßig über eine Antwort. Sie sitzen dort, wo eine genaue Form erwartet wird und der Inhalt selten ist.

Das ist der Grund, warum ausgerechnet die Stellen erfunden werden, die am meisten nach Beleg aussehen: Eine Jahreszahl mit vier Ziffern und ein Autorennamen mit Initialen sehen aus wie Beleg und sind reine Formfüllung.

- Zahlen: Prozente, Preise, Fristen, Seitenzahlen
- Eigennamen: Autoren, Firmen, Studententitel
- Verweise: Paragraphen, DOI, Internetadressen
- Wörtliche Zitate mit Anführungszeichen



Aufgabe 1 · Kreuze alle richtigen Antworten an. Es sind mehrere.

Welche Stellen einer Antwort prüfst du grundsätzlich nach?

- ☐ **A** Eine wörtlich zitierte Passage in Anführungszeichen
- ☐ **B** Die Aussage, dass eine Sache mehrere Ursachen haben kann
- ☐ **C** Eine Prozentzahl mit Nachkommastelle
- ☐ **D** Ein Studententitel mit Autorennamen und Jahr

Aufgabe 2 · Kreuze die richtige Antwort an.

Du fragst nach Quellen und bekommst fünf mit Autor, Titel und Jahr. Was musst du einplanen?

- ☐ **A** Eine vollständige Angabe mit Autor, Titel und Jahr belegt wenigstens, dass die Quelle existiert
- ☐ **B** Jede kann stimmen, falsch wiedergegeben oder erfunden sein; du prüfst also alle fünf
- ☐ **C** Bei bekannten Autorennamen reicht ein kurzer Blick; nur die unbekannten Namen prüfst du nach

Aufgabe 3 · Welche Fassung ist besser? Kreuze an und begründe kurz.

Du willst eine Antwort mit Zahlen darin prüfen.

Fassung 1

Bist du sicher, dass das alles so stimmt? Geh bitte jede Zahl noch einmal durch, korrigier alles, was falsch ist, und schreib mir die Antwort danach vollständig und fehlerfrei neu.

Fassung 2

Markier jede Angabe deiner Antwort als [geprüft], [wahrscheinlich] oder [könnte erfunden sein]. Sag mir bei den letzten zwei Gruppen, wie ich sie nachprüfen kann.

- ☐ Fassung 1 ist besser
- ☐ Fassung 2 ist besser

Begründung:



Aufgabe 4 · Kreuze die Zeilen an, die nicht passen.

Eine KI-Antwort auf die Frage, wie lange du in den Sommerferien arbeiten darfst. Kreuze die Stellen an, die du nicht ungeprüft übernehmen darfst.

- | | |
|----------------------------|---|
| ☐ 1 | Das regelt § 5 Absatz 4 des Jugendarbeitsschutzgesetzes. |
| ☐ 2 | Laut einer Umfrage des Deutschen Jugendinstituts von 2024 arbeiten 38 Prozent der Schülerinnen und Schüler in den Ferien. |
| ☐ 3 | Erlaubt sind höchstens 4 Wochen im Kalenderjahr. |
| ☐ 4 | Wer 15 oder älter und noch vollzeitschulpflichtig ist, darf in den Ferien nur begrenzt arbeiten. |
| ☐ 5 | Für Jugendliche unter 18 gilt das Jugendarbeitsschutzgesetz. |

Warum passt eine der Zeilen nicht? Schreib es zu einer Zeile auf.

Aufgabe 5 · Kreuze die richtige Antwort an.

Die KI widerspricht sich, nachdem du nachgefragt hast. Was heißt das?

- ☐ A Die zweite Antwort stimmt, sie hat den Fehler bemerkt
- ☐ B Keine von beiden ist belegt, du musst extern prüfen
- ☐ C Die erste Antwort stimmte, sie ist nur eingeknickt

Aufgabe 6 · Bring die Schritte in die richtige Reihenfolge: Schreib 1, 2, 3 ... in die Kästchen.

Du hast eine Antwort mit mehreren Behauptungen vor dir. In welcher Reihenfolge prüfst du?

- ☐ A Nach Sorte trennen: Zahl, Verweis, Zitat, allgemeine Aussage
- ☐ B Erst danach entscheiden, ob die Antwort insgesamt trägt
- ☐ C Die genauen Angaben an der Originalquelle nachschlagen
- ☐ D Die Behauptungen herausschreiben, die überhaupt eine Rolle spielen

Aufgabe 7 · Entscheide Zug für Zug, wie du weiterschreibst.

Fünf Quellen, sauber formatiert.

Dein Auftrag an die KI

Nenn mir fünf Studien zum Zusammenhang von Schlaf und Prüfungsleistung.

**Die KI antwortet (Zug 1)**

Gerne. 1) Müller & Schneider (2019): Sleep duration and academic performance. 2) Okafor et al. (2021): Nocturnal rest and exam outcomes ... Alle fünf mit Autorennamen, Titel und Jahr.

Die Liste sieht vollständig aus. Was ist der stärkste nächste Zug?

Kreuze an, was du als Nächstes schreibst.

- ☐ **A** Sind alle fünf Studien echt? Geh jede einzeln durch und bestätig mir in einem Satz, dass es sie wirklich gibt.
- ☐ **B** Nimm die drei besten davon und fass ihre Ergebnisse in je drei Sätzen zusammen, mit Stichprobe und Jahr.
- ☐ **C** Nenn zu jeder Studie die Fundstelle zum Aufschlagen und markier die, bei denen du unsicher bist.

Die KI antwortet (Zug 2)

Drei mit Fundstelle, zwei als unsicher markiert. Bei der Suche findest du: Zwei der drei existieren, eine nicht, und eine der beiden echten sagt fast das Gegenteil.

Eine echte Studie sagt etwas anderes als behauptet. Wie gehst du damit um?

Kreuze an, was du als Nächstes schreibst.

- ☐ **A** Die passende Studie behalten und die widersprechende als Ausreißer weglassen.
- ☐ **B** Beide aufnehmen und den Widerspruch benennen, und die erfundene aus der Liste streichen.



Aufgabe 8 · Schreib deinen eigenen Auftrag an eine KI.

Schreib einen Auftrag für eine Frage, bei der Erfindungen besonders wahrscheinlich sind: Zahlen, Namen, Jahreszahlen. Nenn Aufgabe und Länge und schreib eine Regel für den Fall der Unsicherheit hinein.

Das muss drin sein:

- Ein Auftrag mit Verb
- Eine Länge
- Kontext, also worum es geht
