KI-Halluzinationen — Zerlegen, Vergleichen, Einstufen, Sagen

The Agentic Galaxy • KI lernen mit Layla.Rain • Folge 5

Jeder Ort, jede Person und jede Antwort auf dieser Karte ist für die Übung erfunden. Die Antworten sind vorformulierte Beispiele, keine Ausgabe eines Live-Modells.

Die vier Schritte

1. **ZERLEGEN.** Schreibe jede Behauptung der Antwort in eine eigene Zeile.

2. **VERGLEICHEN.** Lege neben jede Behauptung eine Quelle, die du öffnen kannst. Bei einer Quellenangabe stelle drei Fragen: Existiert sie? Ist sie relevant? Stützt sie diese Behauptung?

3. **EINSTUFEN.** Belegt (die Quelle sagt es). Widerlegt (die Quelle sagt etwas anderes, oder ein vollständiger Bestand zeigt, dass der Punkt darin fehlt). Hier kein Beleg (die Quelle äußert sich nicht dazu und ist kein vollständiger Bestand dazu; das ist nicht dasselbe wie falsch).

4. **SAGEN.** Schreibe einen oder zwei kurze Sätze, die du sicher wiederholen könntest: was belegt ist, was widerlegt ist, was unbekannt ist und wo man nachsehen kann.

AUFGABE 1 · ZUSCHAUEN (durchgerechnetes Beispiel)

Antwort eines Assistenten: „Die Larkmoor Dance Hall wurde 1952 von Marta Quillen gegründet, laut Larkmoor Gazette, März desselben Jahres."

Quelle in der Hand: ein Ergebnis aus dem Archiv der Larkmoor Gazette. Jede Ausgabe vom März 1952 ist vorhanden. Keine enthält einen Artikel über einen Tanzsaal.

Behauptung	Einstufung	Weil
Gegründet 1952	Hier kein Beleg	Das Archiv sagt nichts über das Jahr.
Gegründet von Marta Quillen	Hier kein Beleg	Das Archiv sagt nichts über die Gründerin.
Berichtet in der Larkmoor Gazette, März 1952	Widerlegt	Jede Ausgabe vom März 1952 ist vorhanden, und keine enthält den Artikel.

Sagen: „Ich konnte weder die Gründerin noch das Jahr bestätigen, und die Zeitungsangabe hält der Prüfung nicht stand. Die eigenen Unterlagen der Halle würden es klären."

AUFGABE 2 · MIT HINWEIS VERSUCHEN

Hinweis: Prüfe jede Behauptung, und zwar jede Zahl Wort für Wort.

Quelle in der Hand: die eigene Seite der Halle. „Larkmoor Dance Hall. Gegründet 1961. Hauptsaal: 140 Tanzende."

Antwort eines Assistenten, dem die Seite gegeben wurde: „Die Halle wurde 1961 gegründet. Ihr Hauptsaal fasst 180 Tanzende. Anzahlungen sind erstattbar."

Behauptung	Einstufung	Weil
_____	_____	_____
_____	_____	_____
_____	_____	_____

Sagen: _____

AUFGABE 3 · ALLEIN MACHEN

Quelle in der Hand: die eigene Seite der Amber Lantern Library. „Eröffnet 1978. Lesesaal: 120 Plätze."

Antwort eines Assistenten: „Die Amber Lantern Library wurde 1978 eröffnet. Ihr Lesesaal hat 200 Plätze. Sie wurde vom Architekten Leo Ashgrove entworfen."

Zerlege sie selbst, mit so vielen Zeilen, wie du brauchst.

Behauptung	Einstufung	Weil
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

Sagen: _____

LÖSUNGEN (falte sie um oder decke sie ab, während du arbeitest)

Aufgabe 2. Gegründet 1961: Belegt (die Seite sagt 1961). Hauptsaal 180: Widerlegt (die Seite sagt 140). Anzahlungen erstattbar: Hier kein Beleg (die Seite erwähnt keine Anzahlungen und ist keine vollständige Liste der Bedingungen der Halle). Beispiel für Sagen: „Die Seite stützt 1961 und widerlegt 180 Tanzende, sie nennt 140. Zu Anzahlungen sagt sie nichts, also würde ich die Halle fragen."

Aufgabe 3. Eröffnet 1978: Belegt. Lesesaal mit 200 Plätzen: Widerlegt (die Seite sagt 120). Architekt Leo Ashgrove: Hier kein Beleg (die Seite erwähnt den Architekten nicht und ist kein vollständiger Bestand der Bibliothek). Beispiel für Sagen: „Die Seite stützt 1978 und widerlegt 200 Plätze, sie nennt 120. Der Architekt ist hier unbekannt, also würde ich die Bibliothek fragen."

Zusatzaufgabe. Vergleiche die Behauptung „gegründet 1952" aus Aufgabe 1 mit der Seite der Halle in Aufgabe 2. Neue Einstufung: Widerlegt (die Seite sagt 1961).

PRÜFE DICH SELBST

Hat jede Einstufung eine Quelle genannt, mit der du wirklich verglichen hast? Hast du „hier kein Beleg" statt „falsch" geschrieben, wenn die Quelle schwieg? Hat dein Satz beim Sagen gezeigt, wo man nachsehen kann? Hast du entschieden, wie gründlich du prüfst, bevor du die Antwort gelesen hast?