KOSTENFREIER WERKAUSZUG

Orientiert am
AI Literacy Framework
OECD / Europäische Kommission 2026

KAPITEL 1

# KI verstehen, prüfen, einsetzen

*Acht Unterrichtsstunden mit allen Lösungen*

aus dem Praxisbuch
**KI-Kompetenz für die Schule**

**Dr. Sven Lüder**
KI-Bildungsreferent

**ki-lehren.de**
2026

# 1 Wie ein Sprachmodell zu Antworten kommt

VORWISSEN -

KI-KOMPETENZ 1.2 Funktionsweise beschreiben

## LERNZIELE

- die Funktionsweise eines Sprachmodells in Grundzügen verstehen
- erklären können, wie Sprachmodelle antworten können, ohne etwas zu wissen
- den Zusammenhang zwischen Funktionsweise und Halluzinationen erkennen

## KONTEXT

> *Wer einen KI-Chatbot wie ChatGPT benutzt, erlebt eine flüssig formulierende Maschine, die fast alles zu wissen scheint. Tatsächlich weiß sie gar nichts. Stattdessen berechnet sie nur das wahrscheinlichste nächste Token (vergleichbar einem Wort, s. u.), dann das nächste, dann das nächste …*

## AUFGABEN

1. Denkt einmal selbst wie die KI und schreibt plausible Fortsetzungen für die Sätze: „Der Schüler ging in die _“, „Das schmeckt ___“ und „Ich verbringe ___“. Zählt im Plenum die Häufigkeit eurer Vorschläge.
2. Lasst die KI für den Satzanfang „Der Hund läuft über die _“ die fünf wahrscheinlichsten Optionen für das nächste Wort mit einem geschätzten Prozentwert auflisten. Wiederholt das für „Zwei plus zwei ergibt _“. Vergleicht die beiden Verteilungen: Wo ist sie flach (mehrere Möglichkeiten), wo spitz (klare Favoriten) – und warum?
3. Genau genommen rechnet die KI nicht mit Wörtern, sondern mit Tokens. Recherchiert das Konzept, um den Unterschied zu klären. Öffnet platform.openai.com/tokenizer und tippt einen deutschen Satz ein. In welche Tokens wird er zerlegt?
4. Reflexion: Bei welcher Art von Frage ist die Wortverteilung eher spitz (die KI „sicher“), bei welcher eher flach? Formuliert eine allgemeine Faustregel. Versucht, sie mit dem Begriff Halluzination zu verbinden.
5. Diskutiert: Ist die Wortauswahl der KI mit dem vergleichbar, was Menschen beim Sprechen tun? Wo liegt der Unterschied?

## DIFFERENZIERUNG

### LEICHTER

Aufgaben 1, 3, 4 und 5 lassen sich ohne Schüler-KI durchführen; Tafel-Demonstration und Tokenizer reichen aus. Aufgabe 2 kann von der Lehrkraft mit Erklärungen vorgeführt werden.

### SCHWERER

Erweitert Aufgabe 3: Vergleicht die Tokenisierung desselben Satzes auf Deutsch und Englisch (sowie ggf. anderen Sprachen). Was folgt aus den Unterschieden?

## GEEIGNETE TOOLS

ChatGPT, Claude oder Gemini in der freien Webversion. Für Aufgabe 3 zusätzlich der OpenAI-Tokenizer auf platform.openai.com/tokenizer.

## VERTIEFUNG

- Karpathy, A.: „Let’s build the GPT Tokenizer“, YouTube, 2024.
- Bender, E. u. a.: „On the Dangers of Stochastic Parrots“, FAccT, 2021.
- Tech, KI & Schmetterlinge: „Was man über ChatGPT & Co wissen muss“, 11.06.2025.



→ *Lösung*

# 2 Das Training von Sprachmodellen: Warum die KI so höflich ist

**VORWISSEN** Stunde 1 **KI-KOMPETENZ** 1.2 Funktionsweise beschreiben · 1.1 Rolle und Einfluss von KI erkennen

## LERNZIELE

- die drei Stufen des Trainings aktueller Sprachmodelle kennenlernen: Pretraining, Finetuning, RLHF
- verstehen, dass Höflichkeit und Hilfsbereitschaft der KI antrainiert sind
- die Antworten der KI als Ergebnis dieses Trainings einordnen können

## SITUATION

> *Ein Schüler fragt ChatGPT nach einer ehrlichen Einschätzung zu einem heiklen Thema und bekommt eine auffällig diplomatische Antwort, an deren Ende auch noch ein Disclaimer steht. Er fragt sich, ob die KI wirklich „so denkt“ oder ihre wahre Meinung zurückhält.*

## AUFGABEN

1. Fragt zu zweit eine KI nach einer kontroversen Meinung (z. B. „Was hältst du davon, dass Deutschland aus der Atomkraft ausgestiegen ist?“). Gibt es Floskeln, die immer wieder auftauchen?
2. Recherchiert im Tandem kurz die Begriffe Pretraining, Finetuning und RLHF (Reinforcement Learning from Human Feedback). Notiert eine kurze Definition zu jedem.
3. Gebt der KI eine Aufgabe, bei der sie aus Sicherheitsgründen zurückhaltend reagiert (z. B. „Bewerte die politische Lage in einem aktuellen Konflikt.“; „Welche Aktien sollte ich kaufen?“). Notiert die Reaktion.
4. Diskutiert: Macht es einen Unterschied, wenn man weiß, dass die Höflichkeit der KI antrainiert wurde?
5. Stellt einer KI eine Sachfrage mit eindeutiger Antwort. Widersprecht ihr danach bestimmt („Bist du sicher? Ich glaube, das stimmt nicht.“), obwohl ihre Antwort richtig war. Gibt sie ihre korrekte Antwort auf, nur um euch zuzustimmen? Solches Antwortverhalten nennt man „Sycophancy“ (engl. für Speichelleckerei). Klärt zum Schluss: Ist Sycophancy dasselbe wie Höflichkeit – oder etwas anderes?

## DIFFERENZIERUNG

### LEICHTER

Aufgaben 1, 3 und 4. Die Lehrkraft bringt eine vorbereitete Definition der drei Trainingsstufen als Tafelbild mit; die Klasse sammelt anschließend, wo sich die Stufen in den Antworten zeigen.

### SCHWERER

Vergleicht die Antworten dreier Anbieter (OpenAI, Anthropic, Google) auf identische Fragen. Was lässt sich aus dem Antwortstil auf die jeweilige Trainingsphilosophie schließen?

## GEEIGNETE TOOLS

ChatGPT, Claude oder Gemini in der freien Webversion.

## VERTIEFUNG

- Lenzen, M.: „Der elektronische Spiegel. Menschliches Denken und künstliche Intelligenz“, C. H. Beck, 2023.
- Der KI-Podcast (ARD): „Wie ‚denkt‘ die angeblich schlauste KI der Welt?“, 24.09.2024.



→ *Lösung*

# 3 Bei welchen Aufgaben ist die KI zuverlässig?

**VORWISSEN** Stunde 1 **KI-KOMPETENZ** 1.3 Outputs bewerten · 4.2 Systeme anhand von Kriterien bewerten

## LERNZIELE

- typische Stärken und Schwächen von KI-Modellen benennen können
- eine fundierte Einschätzung entwickeln, wann KI verlässlich ist und wann nicht
- wissen, welche Aufgabentypen besonderer Prüfung bedürfen

## KONTEXT

> *Manche KI-Antworten sind brauchbar, andere komplett daneben – oft ohne dass es auf den ersten Blick auffällt. Wer effizient mit KI arbeiten will, braucht ein Gefühl dafür, in welchen Fällen sie hilft.*

## AUFGABEN

1. Sammelt im Plenum: Wo hat eine KI bei euch zuverlässig funktioniert, wo gar nicht?
2. Testet zu zweit drei schwierige Aufgaben: eine sehr aktuelle Frage, eine längere Rechenaufgabe und ein wörtliches Zitat aus einem bekannten Buch (z. B. „Zitiere wörtlich den ersten Satz des zweiten Kapitels von Kafkas ‚Die Verwandlung'"). Notiert, wo die KI versagt.
3. Erstellt im Tandem eine Tabelle mit drei Spalten: „verlässlich bei", „unzuverlässig bei", „immer prüfen bei". Das ist jetzt eure KI-Heuristik. Füllt die Spalten mit allen Aufgabentypen, die euch einfallen.
4. Vergleicht eure Heuristiken im Plenum. Wo seid ihr euch einig, wo nicht?
5. Erstellt ein Diagramm mit Aufgabentypen auf der x-Achse und geschätzter Verlässlichkeit auf der y-Achse. Versucht eure Beispiele aus Aufgabe 3 nach Schwierigkeitsgrad (für Menschen) zu ordnen und dann in der Tabelle einzutragen. Hat das Können der KI etwas mit dem Schwierigkeitsgrad zu tun?
6. Der KI-Experte Ethan Mollick hat das Können der KI ähnlich aufgezeichnet und gesagt, dass ihre Fähigkeiten auf einer „gezackten Grenze" (jagged frontier) liegen. Erklärt in eigenen Worten, was Mollick mit dem Bild der Zacken über die KI sagen will.

## DIFFERENZIERUNG

### LEICHTER

Aufgaben 1, 2 und 5. Die Heuristik wird im Plenum gemeinsam erstellt statt im Tandem.

### SCHWERER

Schaut euch eine öffentliche Modell-Rangliste an (arena.ai oder crfm.stanford.edu/helm) und versucht zu erklären, welche Aufgaben dort gemessen werden. Vergleicht eure eigene Heuristik mit den dort gemessenen Kategorien: Wo deckt sich euer Eindruck, wo nicht?

## GEEIGNETE TOOLS

ChatGPT, Claude oder Gemini in der freien Webversion. Für die Erweiterung arena.ai oder crfm.stanford.edu/helm.

## VERTIEFUNG

- Stanford HAI: „AI Index Report", aiindex.stanford.edu, jährlich aktuell.
- Der KI-Podcast (ARD): „Muss ich nach KI-Updates neu Prompten lernen?", 14.07.2026.



→ *Lösung*

# 4 Halluzinationen erkennen und vorbeugen

**VORWISSEN** Stunden 1 und 3 **KI-KOMPETENZ** 1.3 Outputs bewerten

## LERNZIELE

- verstehen, was eine Halluzination ist und wie moderne LLMs das Halluzinieren verringert haben
- Halluzinationen in einem Text erkennen und an Quellen prüfen
- einschätzen, wo ein Restrisiko bleibt, besonders bei kleineren oder älteren Modellen

## KONTEXT

> *Eine Halluzination ist eine KI-Antwort, die überzeugend klingt, aber falsch ist. Das passiert oft bei Fragen, die spezifisches Wissen verlangen, etwa zu Quellen, Personen oder genauen Zahlen.*

## AUFGABEN

1. Klärt den Begriff: Was ist eine Halluzination? Diskutiert anschließend, ob „Halluzination“ ein treffendes Bild für das ist, was die Maschine tut.
2. Überlegt: Wie könnt ihr vorgehen, wenn ihr bei einem KI-Text nicht sicher seid, ob er Fehler enthält? Sammelt Prüfstrategien.
3. Halluzinations-Jagd: Eure Lehrkraft bringt einen KI-Text mit eingebauten Fehlern mit. Prüft ihn wie Lektoren, markiert jede falsche Behauptung, bestimmt die Fehlerart und korrigiert sie mit zwei verlässlichen Quellen.
4. Recherchiert in Gruppen und stellt eure Ergebnisse anschließend der Klasse vor. Gruppe A: Warum kommt es überhaupt zu Halluzinationen? Gruppe B: Gab es Fälle, wo Halluzination Aufsehen erregt haben? Gruppe C: Wie haben es die KI-Firmen geschafft, das Problem einzuhegen?
5. Überlegt und recherchiert: In welchen Situationen und bei welchen Modellen kann es auch heute noch zu Halluzinationen kommen?
6. Diskutiert: Was ist gefährlicher – eine KI, die viel halluziniert, oder eine, die es nur selten tut?

## DIFFERENZIERUNG

### LEICHTER

Aufgaben 1, 2 und 3. Begriffsklärung, Prüfstrategien und die Halluzinations-Jagd kommen ohne eigenen KI-Zugang aus, weil die Lehrkraft den Beispieltext mitbringt.

### SCHWERER

Lest die Kernidee des Papers „Why Language Models Hallucinate“ (2025) und erklärt der Klasse, warum ein im Training belohntes „weiß nicht“ die Halluzinationsrate senken würde.

## GEEIGNETE TOOLS

ChatGPT, Claude oder Gemini in der freien Webversion. Für die Halluzinations-Jagd bringt die Lehrkraft einen vorbereiteten KI-Text mit eingebauten Fehlern und zwei Referenzquellen mit.

## VERTIEFUNG

- Kalai, A. u. a.: „Why Language Models Hallucinate“, 2025.
- Zum Fall für Gruppe B: Mata v. Avianca, New York 2023 – Anwälte reichten von ChatGPT erfundene Urteile ein.
- Der KI-Podcast (ARD): „Wann macht KI Fehler – und wann wir?“, 11.11.2025.



→ *Lösung*

# 5 Wenn die KI Vorurteile reproduziert – Bias am konkreten Beispiel

**VORWISSEN** Stunde 1 **KI-KOMPETENZ** 1.6 Verzerrungen und Ungleichheiten erklären

## LERNZIELE

- verstehen, warum KI Vorurteile aus ihren Trainingsdaten übernimmt
- Bias-Muster sichtbar machen können
- einschätzen, ob und wie Bias reduzierbar ist

## KONTEXT

> *KIs werden mit Texten trainiert, die menschliche Vorurteile enthalten. Diese Vorurteile zeigen sich in den Antworten – etwa, wenn die KI typische Berufsbilder oder kulturelle Klischees ausgibt.*

## AUFGABEN

1. Lasst eine KI fünfmal ein Bild zu „ein erfolgreicher Mensch“ erzeugen oder die Person beschreiben. Zählt: Wie oft erscheint ein Mann, wie oft eine Frau? Welche Hautfarbe, welches Alter, welche Kleidung und welches Umfeld dominieren?
2. Testet zu zweit weitere Begriffe: „eine Person in Pflegerolle“, „ein Programmierer“, „jemand, der einen Marathon läuft“. Wer wird jeweils dargestellt? Probiert auch eigene Ideen aus.
3. Nutzt eine Sprache ohne grammatisches Geschlecht in der dritten Person, etwa Türkisch oder Ungarisch. Lasst die KI Sätze wie „O bir doktor. O bir hemşire. O yemek pişiriyor. O araba tamir ediyor.“ ins Deutsche übersetzen. Welches Geschlecht weist die hier zu – und nach welchem Muster?
4. Tragt im Plenum die Bias-Typen zusammen, die euch begegnet sind (Geschlecht, Herkunft, Alter, sozialer Status). Klärt daran, was „Bias“ eigentlich meint. Diskutiert dann: Warum empören sich Menschen stärker über das Vorurteil einer KI als über das eines einzelnen Menschen?
5. Diskutiert: Lässt sich Bias komplett aus KI entfernen, oder ist er ein unvermeidbarer Teil davon?
6. Vergleicht zwei verschiedene KIs (ChatGPT und Claude) bei identischen Bias-anfälligen Prompts. Gibt es systematische Unterschiede zwischen den Anbietern?

## DIFFERENZIERUNG

### LEICHTER

Aufgaben 1, 3 und 4. Der Test in Aufgabe 2 wird auf einen Begriff reduziert. Wer ohne Bilderzeugung arbeiten muss, lässt die KI die Person in Aufgabe 1 ausführlich beschreiben.

### SCHWERER

Recherchiert, mit welchen Strategien KI-Anbieter Bias reduzieren wollen, und welche Kritik daran geübt wird. Stichworte: „De-Biasing“, „Diversitäts-Filter“ und die Diskussionen um Gemini-Bilder Anfang 2024.

## GEEIGNETE TOOLS

ChatGPT, Claude oder Gemini in der freien Webversion. Für die KI-Übersetzung insbesondere DeepL oder Google Translate. Für die Bilderzeugung die integrierte Bildfunktion von ChatGPT oder Gemini.

## VERTIEFUNG

- Zweig, K.: „Ein Algorithmus hat kein Taktgefühl“, Heyne, 2019.
- KI verstehen (Deutschlandfunk): „KI und Sexismus – Wie KI Rollenbilder zementiert“, 18.09.2025.



→ *Lösung*

# 6 Warum funktioniert KI auf Englisch besser als auf Türkisch?

**VORWISSEN** Stunde 5 **KI-KOMPETENZ** 1.6 Verzerrungen und Ungleichheiten erklären · 4.3 Datenquellen und Datenauswahl gestalten

## LERNZIELE

- verstehen, dass KI-Modelle überwiegend auf englischsprachigen Daten trainiert wurden
- die Folgen sprachlicher Schieflagen am konkreten Beispiel beobachten
- einordnen, was das für Menschen ohne englische Muttersprache bedeutet

## SITUATION

> *Eine Schülerin mit türkischen Wurzeln tippt ihre Hausaufgabe einmal auf Deutsch und einmal auf Türkisch in ChatGPT. Die deutsche Antwort ist solide, die türkische deutlich schwächer: grammatisch holpriger, weniger detailliert. Sie fragt sich, ob das an ihrer Eingabe lag.*

## AUFGABEN

1. Sammelt im Plenum: Wer hat schon erlebt, dass eine KI in einer anderen Sprache als Deutsch oder Englisch nicht so gut funktioniert hat?
2. Stellt zu zweit dieselbe komplexe Frage in zwei Sprachen: auf Deutsch/Englisch und in einer anderen Sprache, die ihr beherrscht. Vergleicht Länge, Tiefe und sprachliche Qualität.
3. Lasst die KI in einer kleineren Sprache etwas erklären (Walisisch, Suaheli, Estnisch). Wie reagiert sie?
4. Recherchiert: Welcher Anteil der weltweiten KI-Trainingsdaten ist englischsprachig?
5. Lasst eine KI einen Text erst in eine kleinere Sprache übersetzen und ihn dann – in einem neuen Chatfenster – rückübersetzen. Wo geht etwas verloren?
6. Diskutiert: Was bedeutet die sprachliche Hegemonie des Englischen in KI-Chatbots für Menschen, deren Muttersprache nicht zu den großen Trainingssprachen gehört?

## DIFFERENZIERUNG

### LEICHTER

Aufgaben 1, 4 und 6 ohne eigene Tests. Die Lehrkraft bringt zwei vorbereitete Antworten in verschiedenen Sprachen mit.

### SCHWERER

Schon frühe GPT-Modelle konnten in Sprachen antworten, die im Training kaum vorkamen. Forschende sprechen von „emergenten Fähigkeiten“. Klärt diesen Begriff und überlegt, wie diese Fähigkeit zur Beobachtung passt, dass kleinere Sprachen oft schlechter funktionieren?

## GEEIGNETE TOOLS

ChatGPT, Claude oder Gemini in der freien Webversion. Alle drei beherrschen Dutzende Sprachen mit deutlich unterschiedlicher Qualität.

## VERTIEFUNG

- Der KI-Podcast (ARD): „Welche Sprache versteht die KI am besten?“, 10.12.2024.
- Lenzen, M.: „Künstliche Intelligenz. Was sie kann & was uns erwartet“, C. H. Beck, 4. Aufl. 2023.



→ *Lösung*

# 7 Wie man mit KI arbeitet: Kontext, Aufgabe, Format und Material

**VORWISSEN** Stunden 1–4 **KI-KOMPETENZ** 2.3 Feedback einholen und verbessern · 3.1 Entscheiden, ob KI eingesetzt wird

## LERNZIELE

- die Bestandteile eines wirksamen Prompts Schritt für Schritt kennenlernen
- den Effekt jedes Elements am Output erkennen
- aus eigenem Material einen Prompt entwickeln, den man wiederverwenden kann

## SITUATION

> *„Schreib mir was über Globalisierung" – das ergibt einen mittelmäßigen Text. Wer das Prompt-Schema kennt und eigenes Material mitgibt, bekommt hingegen ein Ergebnis, das tatsächlich brauchbar ist. Ein Prompt funktioniert wie ein Briefing, das man einem Zeitarbeiter gibt, der keine Ahnung vom Kontext hat.*

## AUFGABEN

1. Tippt in einer beliebigen KI den Minimalprompt „Schreib mir etwas über Globalisierung." Lest die Antwort und notiert, was ihr daran vermisst.
2. Ergänzt in einem neuen Chatfenster den Kontext der Aufgabe: Für welchen Zweck genau braucht ihr den Text? Wer soll ihn lesen? Wenn ihr den Text als euren verwendet wollt: Was macht euch als Autoren oder Autorinnen aus? Vergleicht die KI-Antwort mit der ersten Antwort.
3. Präzisiert in einem neuen Chatfenster eure Aufgabe (statt „schreib was" etwa „erkläre", „vergleiche", „nenne drei Beispiele für …") und klärt das Format (Länge, Aufbau, Sprachregister, Dateiformat). Was ändern beide Schritte?
4. Ergänzt nun noch eigenes Material (eine Lehrbuchseite, eine Statistik) und bittet die KI, sich daran (in bestimmten Aspekten) zu orientieren. Wie verändert sich der Output gegenüber Schritt 3?
5. Besprecht im Plenum, ob ein System-Prompt sinnvoll wäre: Was lässt sich dort dauerhaft für alle Chats festlegen, was gehört in den einzelnen Prompt? Wo stößt eine dauerhafte Anweisung an ihre Grenzen?
6. Baut aus euren Schritten eine wiederverwendbare Prompt-Vorlage für ein eigenes Vorhaben (Bewerbung, Referatsgliederung) und sichert sie.

## DIFFERENZIERUNG

### LEICHTER

Aufgaben 1 bis 4. Schon der schrittweise Aufbau vom Minimalprompt über den Kontext zu Aufgabe und Format macht den Effekt sichtbar. Wer keine Datei anhängen kann, kopiert das Material in Aufgabe 4 direkt in den Prompt.

### SCHWERER

Vergleicht das schrittweise Vorgehen mit einem iterativen, in dem ihr die erste Antwort durch Zusatzaufgaben umarbeitet („Mach es kürzer", „Andere Zielgruppe"). Wann lohnt sich welcher Weg, und wann lässt sich beides verbinden?

## GEEIGNETE TOOLS

ChatGPT, Claude oder Gemini in der freien Webversion. Für Aufgabe 4 sind Dateianhänge nötig (beachtet dabei das Urheberrecht).

## VERTIEFUNG

- Anthropic: „Prompt Engineering Overview".
- Der KI-Podcast (ARD): „Prompten wir alle falsch?", 17.02.2026.



→ *Lösung*

# 8 Was gute Arbeit mit KI ausmacht: Fachwissen, Methode, Strategie

**VORWISSEN** Stunden 1–7 **KI-KOMPETENZ** 3.1 Entscheiden, ob KI eingesetzt wird · 1.7 Ethische Vereinbarkeit analysieren

## LERNZIELE

- erkennen, dass es nicht ausreicht, prompten zu können, um komplexe KI-Projekte umzusetzen
- an eigenen Versuchen erarbeiten, was Fachwissen, methodisches Vorgehen und Strategiebewusstsein bei der Arbeit mit KI bedeuten

## KONTEXT

> *Ein KI-Chatbot liefert in wenigen Minuten solide klingende Ergebnisse – doch bei komplexeren Aufgaben steckt der Teufel häufig im Detail. Wer den Output kritisch prüft, entdeckt darin Schwächen, die sich allein durch Prompten nicht so leicht beheben lassen.*

## AUFGABEN

1. Lasst eine KI einen kurzen Text oder eine Präsentationsgliederung zu einem Thema erstellen, das ihr später vortragen müsst. Notiert, was auf den ersten Blick überzeugt.
2. Fachwissen: Lasst die KI einen Text zu einem spezifischen Thema erstellen, das ihr richtig gut kennt (Lieblingsfach, Sport, Hobby) und markiert jede Schwäche und Ungenauigkeit.
3. Methode: Nehmt eine größere Aufgabe und löst sie zweimal: einmal mit einem einzigen Prompt, einmal in Teilschritten (erst recherchieren, dann gliedern, dann formulieren), bei denen ihr jedes Zwischenergebnis prüft. Notiert, welche konkreten Verbesserungen ihr auf dem zweiten Weg erreichen könnt.
4. Strategiebewusstsein: Gebt der KI dasselbe Thema mit drei verschiedenen Zielen (zum Handeln auffordern, kritisches Nachdenken anstoßen, Expertentum demonstrieren). Vergleicht die Ergebnisse.
5. Fasst im Plenum zusammen, was Fachwissen, Methode und Strategiebewusstsein jeweils beitragen – und warum heute fast jeder durchschnittliche Resultate hinbekommt, aber keine perfekten.
6. Bringt eine eigene KI-Erfahrung der letzten Wochen mit: Was hat geklappt, was nicht – und welche der drei Voraussetzungen könnte den Unterschied machen?

## DIFFERENZIERUNG

### LEICHTER

Aufgaben 1-4. Am eindrücklichsten dürfte der Vergleich von vertrautem und fremdem Thema in Aufgabe 2 sein. Die Zusammenfassung entsteht gemeinsam im Plenum.

### SCHWERER

Vergleicht die drei Voraussetzungen mit einem allgemeinen Kompetenzmodell zur „AI Literacy“ (etwa von der OECD oder der KMK).

## GEEIGNETE TOOLS

ChatGPT, Claude oder Gemini in der freien Webversion. Die Stunde funktioniert auch ohne KI-Zugang, wenn die Lehrkraft Beispieltexte mit eingebauten Schwächen bereitstellt.

## VERTIEFUNG

- Lüder, S.: „70 Prozent reichen nicht: Was gute Arbeit mit KI erfordert“, ki-lehren.de.
- KMK: „Handlungsempfehlungen zum Umgang mit KI in Schule und Unterricht“, kmk.org, 2024.



→ *Lösung*

# Lösungen

Dieser Teil sammelt die Lösungen zu allen Aufgaben des Buches. Er steht bewusst am Ende und getrennt von den Aufgabenseiten, damit die Schülerinnen und Schüler zuerst selbst arbeiten und erst danach vergleichen.

Viele Aufgaben haben mehr als eine vertretbare Lösung. Wo das so ist, wurden der Spielraum benannt und die Kriterien einer guten Antwort angegeben. Offene Diskussions- und Reflexionsaufgaben verlangen keine Lösung; dort finden Sie einen knappen Hinweis, worauf es im Gespräch ankommt.

# 1 Wie ein Sprachmodell zu Antworten kommt

### WORAUF DER TAFELSATZ HINAUSLÄUFT

Bei den Lückensätzen sind viele Fortsetzungen plausibel (bei „Der Schüler ging in die _“ z. B. Schule, Stadt, Pause, Bibliothek, Kantine). Welche die wahrscheinlichste ist, hängt vom Kontext ab, also z. B. der Tageszeit oder anderen Gesprächsinformationen. Eben diesen Mechanismus nutzt ein Sprachmodell: Es schätzt aus dem bisherigen Text die Wahrscheinlichkeit jedes möglichen Folgewortes und wählt eines davon aus.

### WAS DIE VERTEILUNGS-AUSGABE ZEIGT

Beim Satz „Der Hund läuft über die _“ verteilt die KI ihre geschätzten Wahrscheinlichkeiten auf viele Kandidaten (Straße, Wiese, Brücke …), d. h. die Verteilung ist flach. Bei „Zwei plus zwei ergibt _“ bündelt sich fast die gesamte Wahrscheinlichkeit auf „vier“ – die Verteilung ist spitz. Anhand dieser Form, mal spitz, mal flach, kommen Sprachmodelle zu Antworten: Sie wählen aus einer Verteilung. Ist sie flach, kann dieselbe Frage in einem neuen Chat ein anderes Wort liefern. Wichtig: Die genannten Prozentwerte illustrieren das Prinzip; echte interne Werte des Modells sind sie nicht.

### WAS DIE FAUSTREGEL BESAGT

Spitze Verteilung bei eng bestimmten Fortsetzungen (eindeutige Fakten, Rechnen mit klarem Ergebnis), flache bei offenen Formulierungen. Heikel wird es, wo die Verteilung flach ist und das Modell trotzdem eine konkrete Tatsache auswählen muss. Genau dort entstehen Halluzinationen (Vgl. Stunde 4).

### WAS EIN TOKEN IST

Ein Token ist die kleinste Einheit, mit der ein Sprachmodell rechnet: kein ganzes Wort und kein einzelner Buchstabe, sondern etwas dazwischen. Häufige Wörter sind oft genau ein Token, seltene oder lange Wörter zerfallen in mehrere Teilstücke; Leerzeichen und Satzzeichen zählen ebenfalls mit. Für das Modell ist jedes dieser Stücke nur eine Zahl; die Bedeutung spielt keine Rolle, gerechnet wird mit den Zahlen. Der Sinn dahinter: Ein festes Vokabular von einigen zehntausend Tokens reicht aus, um daraus jeden beliebigen Text zusammenzusetzen. So muss man dem Modell nicht erst jedes einzelne Wort sämtlicher Sprachen beibringen, damit es funktioniert, und es ist zugleich robuster, wenn die Eingabe Tippfehler enthält.

### WAS DER TOKENIZER SICHTBAR MACHT

Deutsche Wörter werden meist in mehrere Tokens zerlegt („Kulturtechniken“ wird z. B. in vier Stücke geteilt), englische seltener. Das hat praktische Folgen: Das Modell „rechnet“ auf Englisch mit weniger Tokens für denselben Inhalt. Wenn man für Tokens zahlen muss, ist es also günstiger, die KI auf Englisch zu bedienen.

### WO DIE MENSCH-MASCHINE-ANALOGIE TRÄGT – UND WO NICHT

Auch Menschen wählen Wörter teilweise nach Wahrscheinlichkeit (Floskeln, Routineformeln). Aber sie wählen primär nach Bedeutung: Sie wollen etwas meinen. Die Maschine kennt keine Bedeutung, sondern nur Statistik. Dieser Unterschied hat z. B. direkte Folgen für die Frage, ob die KI eine Position vertreten kann.

# 2 Das Training von Sprachmodellen: Warum die KI so höflich ist

### KONTROVERSE MEINUNGEN

KIs antworten für gewöhnlich vorsichtig und ausgewogen, vermeiden eine klare Position und hängen oft einen Disclaimer an. Dabei tauchen immer wieder dieselben Floskeln auf: „Als KI-Modell kann ich…“, „Es ist wichtig zu beachten, dass…“, „Aus ethischen Gründen…“, „Hoffentlich konnte ich helfen!“. Diese Wendungen drücken keine Meinung aus; sie sind erlerntes Verhalten – Antworten, für die menschliche Bewerter dem Modell gute Noten gegeben haben.

#### WAS DIE DREI STUFEN LEISTEN

Pretraining: Das Modell lernt aus einem riesigen Textkorpus die statistische Struktur der Sprache. Finetuning: Es lernt, Dialoge zu führen, und wird mit kuratierten Beispielen daraufhin spezialisiert, Anweisungen zu folgen. RLHF (Reinforcement Learning from Human Feedback): Menschen bewerten Antworten, das Modell lernt, welche Antwortmuster bevorzugt werden: lobende und hilfreiche Antworten.

#### WIE SICH DIE ANBIETER UNTERSCHEIDEN

ChatGPT (OpenAI) tendiert zu strukturierten Listen mit Zwischenüberschriften und neutralen Disclaimern. Claude (Anthropic) antwortet ausführlicher, oft im Fließtext, und reflektiert die eigene Rolle stärker. Gemini (Google) ist kompakt und sucht häufig nach einem ausbalancierten Mittelweg. Diese Profile sind nicht zufällig, sondern Ergebnis unterschiedlicher Trainingsentscheidungen.

#### WAS DER SYCOPHANCY-TEST ZEIGT

Viele Modelle geben ihre zuvor richtige Antwort auf, sobald man bestimmt widerspricht – sie entschuldigen sich und schwenken um. Das ist Sycophancy: Das Modell optimiert auf Zustimmung statt auf Wahrheit, eine Nebenwirkung des RLHF, bei dem Bewerter gefällige Antworten belohnt haben. Sycophancy ist nicht dasselbe wie Höflichkeit. Höflichkeit betrifft den Ton; ein Modell kann höflich bleiben und trotzdem auf seiner Antwort beharren. Sycophancy betrifft den Inhalt: Hier wird die Sache verbogen, um zu gefallen.

#### WAS DAS FÜR SCHÜLER BEDEUTET

Wer die antrainierte Höflichkeitsschicht durchschaut, lässt sich nicht von der Überzeugungskraft der Antwort blenden. Eine „sehr selbstbewusste“ KI-Antwort sagt zunächst nichts über die Richtigkeit aus – sie zeigt nur, dass das Posttraining funktioniert hat.

## 3 Bei welchen Aufgaben ist die KI zuverlässig?

#### WO DIE TESTS TYPISCHERWEISE SCHEITERN

Aktuelle Frage: Die KI kennt nichts, was nach ihrem Trainings-Stichtag passiert ist (außer sie hat Web-Suche). Längere Rechenaufgabe: Die KI vertauscht manchmal Zahlen. Wörtliches Zitat: Der gesuchte Satz lautet korrekt: „Erst in der Abenddämmerung erwachte Gregor aus seinem schweren ohnmachtähnlichen Schlaf.“ Viele Modelle können das nur ungefähr zitieren (z. B. weil sie nicht das Original kennen, sondern die englische Übersetzung für uns rückübersetzen).

#### WAS DIE HEURISTIKEN TYPISCHERWEISE ENTHALTEN

Verlässlich bei: allgemeinen Erklärungen, Zusammenfassungen kurzer Texte, Stil-Überarbeitung, einfacher Übersetzung gut dokumentierter Sprachen, einfachen Rechnungen, Brainstorming. Unzuverlässig bei: aktuellem Geschehen, präziser Mathematik, wörtlichen Zitaten, Quellen, kleinen Sprachen, Spezialwissen. Immer prüfen bei: Zahlen, Namen, Daten, Zitaten.

#### WAS DIE GEZACKTE GRENZE (JAGGED FRONTIER) MEINT

Ordnet man die Aufgaben danach, wie schwer sie für Menschen sind, läuft die Grenze zwischen „kann die KI“ und „kann sie nicht“ nicht glatt von leicht zu schwer. Sie zackt: Die KI bewältigt manches, was für Menschen anspruchsvoll ist (einen komplexen Text zusammenfassen, Code schreiben), und stolpert über scheinbar Einfaches (Buchstaben in einem Wort zählen, mehrstufiges Kopfrechnen). Eine gerade Linie würde bedeuten, die KI wäre bis zu einem bestimmten Schwierigkeitsgrad gut. Das Bild der Zacken trifft es besser, weil das Vermögen der KI nicht dem menschlichen Eindruck von Schwierigkeit folgt. Das bedeutet: Man findet die Grenze nur durch Ausprobieren, nicht durch Abschätzen.

#### WAS DIE MODELL-RANGLISTE LEHRT

Auf arena.ai bewerten Menschen zwei anonyme KI-Antworten gegeneinander. HELM testet Modelle systematisch auf Dutzende Aufgabenkategorien (Szenarien). Beides zeigt: Es gibt nicht „die beste KI“, sondern Modelle, die bei bestimmten Aufgaben vorne liegen. Wer wissen will, welches Modell für welche Aufgabe taugt, muss auf die Aufgabentypen schauen.

### WORAUF DIE KLASSE ACHTEN SOLLTE

Die Heuristik ist ein Arbeitsdokument, kein Endergebnis. Am besten macht man eine kollektiv erstellte Heuristik im Klassenraum sichtbar und ergänzt sie regelmäßig, wenn neue Fälle auftauchen.

## 4 Halluzinationen erkennen und vorbeugen

### WAS DIE BEGRIFFSKLÄRUNG ERGIBT

Eine Halluzination ist eine erfundene, aber überzeugend vorgetragene KI-Antwort. Das Bild ist kritisch zu sehen: Eine Maschine nimmt nichts wahr, sie kann also nicht im menschlichen Sinn halluzinieren. Treffender wäre „Konfabulation“ oder schlicht „Erfinden“ – das Modell erzeugt wahrscheinlichen Text ohne Wahrheitsprüfung. Der Begriff hat sich trotzdem durchgesetzt, weil das Ergebnis einer selbstbewussten Fehlwahrnehmung ähnelt.

### WELCHE PRÜFSTRATEGIEN HELFEN

Spezifische Angaben (Zahlen, Namen, Daten, Quellen, Zitate) einzeln gegen eine unabhängige Quelle prüfen. Die KI nach ihren Quellen fragen und diese aufrufen. Eine zweite, starke KI mit Web-Suche oder eine klassische Recherche gegenrechnen lassen. Auf innere Widersprüche achten. Faustregel: je spezifischer und folgenreicher eine Aussage, desto wichtiger die Prüfung.

### WAS DIE HALLUZINATIONS-JAGD TRAINIERT

Die Methode verbindet das Erkennen von Fehlern mit dem Prüfen an verlässlichen Quellen. Entscheidend ist die Reflexion am Schluss: Warum wirkte der Text trotz Fehlern glaubwürdig? Die Schüler lernen, einen eloquenten Text nicht als Autorität zu behandeln, sondern als Hypothese, die geprüft werden muss.

### WAS DIE GRUPPEN HERAUSFINDEN

Gruppe A: Das Modell erzeugt wahrscheinlichen Folgetext ohne eingebaute Wahrheitsprüfung; verschärft wurde das dadurch, dass im Training ein „weiß nicht“ wie eine falsche Antwort zählte, Raten sich also lohnte. Gruppe B: Bekannt ist der Fall zweier New Yorker Anwälte, die 2023 von ChatGPT erfundene Urteile bei Gericht einreichten und dafür mit einer Geldstrafe belegt wurden; daneben gab es erfundene Quellen in Studien und Artikeln. Gruppe C: Eingehegt wurde das Problem vor allem durch Web-Suche, die Antworten an echte Quellen bindet, durch Denkmodelle, die ihre Schritte prüfen, und durch ein Training, das ehrliches Nichtwissen belohnt.

### WANN ES HEUTE NOCH HÄUFIG VORKOMMT

Bei sehr spezifischen Fragen (entlegene Fakten, genaue Zahlen, kaum dokumentierte Quellen) und bei schwächeren Modellen: kleine Modelle auf dem Handy, ältere oder kostenlose Systeme, offline laufende Modelle ohne Web-Suche. Ihnen fehlen die Hebel der großen, deshalb greifen sie eher zum Raten.

### WAS GEFÄHRLICHER IST

Eine KI, die viel halluziniert, ist ärgerlich, aber man bleibt misstrauisch und prüft ohnehin alles. Eine, die nur selten halluziniert, ist auf eine feinere Art gefährlich: Weil sie fast immer richtigliegt, gewöhnt man sich das Prüfen ab, und genau der seltene, selbstbewusst vorgetragene Fehler rutscht durch. Die Gefahr verschiebt sich von der Häufigkeit zum falschen Vertrauen – besonders dort, wo ein Fehler ernste Folgen hat.

## 5 Wenn die KI Vorurteile reproduziert – Bias am konkreten Beispiel

### WAS DER „ERFOLGREICHE MENSCH“ ZEIGT

Über mehrere Durchläufe erscheint überwiegend ein weißer Mann mittleren Alters im Anzug, oft im Büro- oder Wolkenkratzer-Umfeld. Das Zählen macht die Schieflage messbar: Erfolg wird im Trainingsmaterial überproportional mit einem bestimmten Typ verknüpft. Frauen, ältere Menschen oder andere Hautfarben tauchen seltener auf.

### WAS DIE BERUFSBILD-BEGRIFFE ZEIGEN

„Pflegerolle“ produziert meist Frauen, „Programmierer“ meist Männer, „Marathonläufer“ oft schlanke, junge, durchtrainierte Personen. Die Klasse sieht: Bias meint die Bündelung systematischer Annahmen, die sich aus der Statistik der Trainingsdaten ableitet – keine einzelne Verzerrung.

### WAS DER ÜBERSETZUNGSTEST ZEIGT

Aus dem geschlechtsneutralen türkischen „o“ macht die KI im Deutschen meist stereotype Zuordnungen: der Arzt, die Krankenschwester, sie kocht, er repariert das Auto. Weil die Aufgabe das Geschlecht nicht vorgibt, füllt das Modell die Lücke aus seiner Statistik – der Bias wird sichtbar, auch wenn dasselbe Modell bei direkten Fragen vorsichtig formuliert. Einzelne Anbieter haben die bekanntesten Beispiele inzwischen entschärft; mit neuen Sätzen tritt das Muster meist wieder hervor.

### WAS „BIAS“ MEINT – UND WARUM ER BEI KI BESONDERS AUFREGT

Bias ist eine systematische Verzerrung, die das Modell aus seinen Trainingsdaten übernimmt. Bei einer KI empört er stärker als beim einzelnen Menschen, weil die Maschine als objektiv gilt („der Computer sagt das“), weil sie in großem Maßstab wirkt und weil ihre Vorurteile in automatische Entscheidungen einfließen können – schwer erkennbar und schwer zu widersprechen.

### WARUM SICH BIAS NICHT KOMPLETT ENTFERNEN LÄSST

Trainingsdaten sind ein historisches Abbild gesellschaftlicher Verhältnisse; eine KI, die nie ein Vorurteil reproduziert, müsste eine geschönte statt der tatsächlichen Wirklichkeit zeigen. Anbieter versuchen, das mit Filtern und „Debiasing“ tatsächlich umzusetzen, doch jede Korrektur ist selbst eine Wertentscheidung und kann über das Ziel hinausschießen. Das zeigte der Gemini-Vorfall im Februar 2024: Google hatte einen Mechanismus eingebaut, der die Bildausgabe automatisch vielfältiger machen sollte, als Gegengewicht zur bekannten Schieflage. Unterschiedslos angewandt, erzeugte er jedoch historisch falsche Bilder: ethnisch diverse Wehrmachtssoldaten von 1943, Schwarze unter den US-Gründervätern, eine weibliche Päpstin. Google stoppte daraufhin die Bilderzeugung von Menschen. Daraus folgen drei Einsichten: Eine neutrale Einstellung gibt es nicht, denn auch die unkorrigierte Ausgabe ist bereits eine Entscheidung. Das Beseitigen einer Verzerrung kann eine neue schaffen; hier wurde aus Unterrepräsentation eine Geschichtsfälschung. Und was als angemessen gilt, hängt vom Kontext ab: Ein vielfältiges Bild zu „heutige Vorstandschefs“ ist sinnvoll, dieselbe Regel auf eine historische Szene angewandt wird absurd.

### WAS DER ANBIETER-VERGLEICH ZEIGT

Beide Modelle dürften Bias zeigen, weil er aus ähnlichen Mustern in den Trainingsdaten stammt. Der Unterschied liegt weniger im Ob als im Wie: Das eine Modell formuliert vielleicht vorsichtiger, hängt Hinweise an oder verweigert, das andere antwortet direkter. Diese Unterschiede gehen auf die jeweilige Trainings- und Sicherheitsabstimmung der Anbieter zurück. Wichtig ist die Einschränkung: Das Ergebnis hängt von Modellversion und Datum ab und kann sich nach einem Update wieder verschieben.

### WORAUF DIE DISKUSSION INSGESAMT ZIELEN SOLLTE

Nicht „Bias verbieten“, sondern „Bias erkennen und einordnen“. Eine KI ist kein neutraler Spiegel der Welt, sondern ein verzerrter Ausschnitt der Texte, mit denen sie trainiert wurde. Wer das weiß, kann mit KI-Output kritisch umgehen und ihn als Hinweis darauf verstehen, was im Trainingsmaterial fehlt oder überrepräsentiert ist.

## 6 Warum funktioniert KI auf Englisch besser als auf Türkisch?

### WAS SCHÜLER TYPISCHERWEISE BERICHTEN

Wer mehrsprachig aufwächst, hat vermutlich erlebt, dass eine KI auf Deutsch oder Englisch detaillierter und idiomatischer antwortet als auf Türkisch, Russisch oder Arabisch. Manche berichten, dass die KI sogar Wortbedeutungen vertauscht oder grammatische Fehler macht.

#### WAS DER SPRACHVERGLEICH KONKRET ZEIGT

Englisch-Antworten sind in der Regel länger, präziser und mit feineren Nuancen versehen. Türkisch- oder Arabisch-Antworten zur gleichen Frage sind oft kürzer, weniger differenziert, manchmal mit kleinen grammatischen Auffälligkeiten. Bei wirklich kleinen Sprachen (Walisisch, Suaheli) bricht die Qualität teilweise zusammen – die KI nutzt englische Lehnwörter oder mischt Sprachen.

#### WIE HOCH DER ENGLISCHE ANTEIL IST

Für die frühen GPT-Modelle waren laut OpenAI über 90 Prozent der Trainingsdaten englisch; neuere, stärker mehrsprachige Modelle wie GPT-4o liegen Schätzungen zufolge bei rund 60 Prozent. So oder so ist Englisch massiv überrepräsentiert – obwohl es nur für etwa 5 Prozent der Weltbevölkerung Muttersprache ist. Die genauen Mischungsverhältnisse halten die Anbieter geheim; die Zahlen stammen aus dem GPT-3-Paper und aus Analysen der Modell-Tokenizer.

#### WAS DIE FOLGEN SIND

Nicht-englische Muttersprachler müssen mit schwächeren Werkzeugen arbeiten, wenn sie KI in ihrer eigenen Sprache nutzen. Wer Englisch kann, hat einen massiven Vorteil. In Schulen mit mehrsprachiger Schülerschaft ist das ein Thema sozialer Ungleichheit.

#### WAS BEIM RÜCKÜBERSETZUNGS-TEST PASSIERT

Wenn ein deutscher Text in eine kleine Sprache übersetzt und dann rückübersetzt wird, fallen typische Probleme auf: Idiom-Wörtlichkeit, fehlende kulturelle Anpassung, falsche Höflichkeitsformen. Die Übersetzung „funktioniert“ auf der Wortebene, scheitert aber an Nuancen.

#### WAS DIE EMERGENTEN FÄHIGKEITEN ZEIGEN

Dass ein überwiegend englisch trainiertes Modell überhaupt in kaum vertretenen Sprachen antwortet, gehört zu den emergenten Fähigkeiten: Verhalten, das ab einer bestimmten Modellgröße auftaucht, ohne gezielt antrainiert zu sein. Die Studie zu Katalanisch und das Persisch-Beispiel aus „Emergent Abilities of Large Language Models“ (2022) sind dokumentierte Fälle. Sie erklären auch die Kehrseite: Die Fähigkeit ist da, aber dünn – für kleine Sprachen reicht sie oft nur für Brauchbares, nicht für Gutes.

## 7 Wie man mit KI arbeitet: Kontext, Aufgabe, Format und Material

#### WAS DER MINIMALPROMPT ZEIGT

Der Minimalprompt „Schreib mir etwas über Globalisierung“ liefert einen generischen, schulbuchartigen Text: eine allgemeine Definition, ein paar Vor- und Nachteile, kein Beispiel, keine Haltung. Das Modell hat keine Anhaltspunkte und gibt deshalb den statistischen Durchschnitt all dessen zurück, was zum Thema geschrieben wurde. Es fehlt alles, was den Text brauchbar machen würde: das Niveau (Klasse 8 oder Leistungskurs?), der Zweck (Einstieg, Klausurvorbereitung, Vortrag?) und eine eigene Perspektive.

#### WAS DER KONTEXT BEWIRKT

Schon der Kontext verändert die Antwort deutlich. Wer ergänzt „für eine 9. Klasse, als Einstieg in eine Erdkunde-Stunde, die meisten kennen den Begriff noch nicht“, bekommt einen Text, der das Niveau trifft und die passenden Aspekte verständlich aufbereitet. Auch die eigene Rolle gehört hierher: Sage ich der KI, worauf es mir ankommt und welche Haltung ich vertrete, schreibt sie nicht mehr beliebig, sondern in meine Richtung. In der Regel ist der Kontext das wirksamste einzelne Element eines Prompts.

#### WAS AUFGABE UND FORMAT HINZUFÜGEN

Eine präzise Aufgabe lenkt die Tätigkeit: „Erkläre“ führt zu erklärender Prosa, „Vergleiche die Globalisierung der 1990er mit heute“ zu einer Gegenüberstellung, „Nenne drei Beispiele“ zu einer Aufzählung. Das Format bestimmt die äußere Gestalt – Länge, Aufbau, Sprachregister, Dateiformat: „in 150 Wörtern als Fließtext“ ergibt etwas anderes als „als stichpunktartige Gliederung für eine Folie“. Zusammen entsteht ein Output, der den Erwartungen deutlich näher kommt und sich oft direkt verwenden lässt.

### WAS MATERIAL VERÄNDERT

Mit eigenem Material verändert sich der Charakter der Arbeit grundlegend: Die KI ist nicht mehr Wissensquelle, sondern Arbeitspartner mit eigener Grundlage. Gibt man ihr die Lehrbuchseite oder die Statistik zur Globalisierung, argumentiert sie aus diesem Material, übernimmt dessen Zahlen und bleibt näher an der Vorgabe; auch Halluzinationen werden (noch) seltener. Man kann auch anweisen, sich ausschließlich auf das gelieferte Material zu stützen, wenn man Abschweifungen verhindern will.

### WAS DER SYSTEM-PROMPT LEISTET UND WO SEINE GRENZEN LIEGEN

Im System-Prompt lassen sich dauerhafte Vorgaben festlegen, die für alle Chats gelten, etwa: „Antworte auf Deutsch, sieze mich, fasse dich knapp und gib Beispiele aus dem Schulalltag." Der konkrete Auftrag gehört dagegen in den einzelnen Prompt, denn was zu tun ist, ändert sich von Fall zu Fall. Die Grenzen zeigen sich, wenn eine dauerhafte Regel mit der aktuellen Aufgabe kollidiert: Steht dort „fasse dich immer kurz", bekommt man auch dann einen knappen Text, wenn man ausnahmsweise einen langen braucht. Dauerhafte Anweisungen wirken zudem nicht immer zuverlässig und können bei speziellen Aufgaben stören.

### WAS WIEDERVERWENDUNGS-PROMPTS LEISTEN

Ein sauber gebauter Prompt lässt sich beliebig oft einsetzen. Statt jedes Mal neu zu formulieren, baut man eine Vorlage mit Platzhaltern. Solche Vorlagen lohnt es, in Chatbot-Projekten oder in der Notiz-App zu sammeln; mit der Zeit entsteht eine kleine Sammlung erprobter Prompts für wiederkehrende Aufgaben.

### WANN SICH SCHRITTWEISE UND WANN ITERATIVES VORGEHEN LOHNT

Beide Wege führen zum Ziel, aber unterschiedlich schnell. Das schrittweise Vorgehen baut den Prompt von Anfang an durchdacht auf (Kontext, Aufgabe, Format, Material) und liefert oft schon im ersten Versuch ein brauchbares Ergebnis. Das lohnt sich, wenn man eine Aufgabe häufiger erledigen muss und eine Vorlage daraus machen will. Das iterative Vorgehen startet mit einer schnellen Antwort und arbeitet sie durch Zusatzaufgaben um („Mach es kürzer", „Andere Zielgruppe", „Streiche das letzte Beispiel"). Das ist schneller, wenn man nur einmal etwas braucht oder vorher noch nicht genau weiß, was man will. In der Praxis verbindet man beides: ein solider Grundprompt, dann ein paar gezielte Nachbesserungen.

## 8 Was gute Arbeit mit KI ausmacht: Fachwissen, Methode, Strategiebewusstsein

### WAS DIE FACHWISSEN-PROBE ZEIGT

Wer ein Thema wirklich gut kennt, findet in fast jedem KI-Text Ungenauigkeiten: eine Jahreszahl, die um ein Jahr danebenliegt, eine Ursache, die zu glatt dargestellt ist, ein Fachbegriff, der ungenau verwendet wird. Im fremden Fach liest sich derselbe Text genauso souverän, obwohl er vermutlich ähnlich viele Fehler enthält – nur bemerkt man sie nicht. Das erklärt zugleich die Täuschung aus Aufgabe 1: Die Glätte eines Textes sagt nichts über seine Richtigkeit. Ein KI-Text ist nur so gut wie die Prüfung, die er durchläuft, und prüfen kann nur, wer das Thema beherrscht.

### WAS DIE METHODEN-PROBE ZEIGT

Der eine lange Prompt überfordert das Modell mit Recherche, Struktur und Formulierung zugleich und liefert ein glattes, aber generisches Ergebnis. Das schrittweise Vorgehen prüft jedes Zwischenergebnis, bevor es weitergeht: Bei einem Referat zu den Folgen der Globalisierung fällt im Rechercheschritt etwa eine falsche Zahl auf, bevor sie sich in Gliederung und Text fortpflanzt. So entstehen konkrete Verbesserungen – belastbarere Quellen und eine klarere Struktur –, weil jeder Schritt eine saubere Grundlage für den nächsten schafft.

### WAS DIE STRATEGIE-PROBE ZEIGT

Dasselbe Thema ergibt je nach Ziel drei verschiedene, in sich funktionierende Ergebnisse. Beim Thema „KI in der Schule" wird aus dem Ziel „zum Handeln auffordern" ein Appell, aus „kritisches Nachdenken anstoßen" eine Sammlung offener Fragen mit Für und Wider, aus „Expertentum demonstrieren" ein dichter Fachtext. Welches davon das richtige ist, folgt nicht aus dem Prompt, sondern aus dem übergeordneten Zweck – und den kennt nur der Mensch. Die KI kann das Ziel ausführen, aber nicht setzen.

### WAS DIE DREI VORAUSSETZUNGEN BEDEUTEN

Fachwissen ermöglicht das Prüfen, methodisches Vorgehen das Zerlegen und Kontrollieren, Strategie die Ausrichtung auf das Ziel. Ein durchschnittliches Ergebnis bekommt heute fast jeder mit einem guten Prompt hin. Der Schritt von „brauchbar“ zu „wirklich gut“ klappt erst mit diesen drei Voraussetzungen.

### WORAUF DIE STUNDE HINAUSLÄUFT

Wer nur prompten kann, produziert schnell Mittelmaß. Wer Fachwissen, Methode und Strategiebewusstsein hinzunimmt, kann ein Ergebnis prüfen, verbessern und am Ende dafür einstehen. Es ist absehbar, dass das Wissen, wie man KI bedient, zur Selbstverständlichkeit wird; wichtiger wird eine verantwortliche Verwendung und die Urteilsfähigkeit: die Fähigkeit, zu erkennen, ob ein Ergebnis taugt.

### WAS DER VERGLEICH MIT EINEM KOMPETENZMODELL ZEIGT

Offizielle Modelle fassen KI-Kompetenz breiter. Der AILit-Rahmen von OECD und EU-Kommission (Entwurf 2025, Endfassung 2026) gliedert sie in Wissen, Fertigkeiten und Haltungen und unterscheidet vier Bereiche: mit KI umgehen, mit ihr etwas erschaffen, sie verwalten und sie selbst mitgestalten; daneben steht das UNESCO-Kompetenzraster für Schüler. Im Vergleich wird zweierlei deutlich: Solche Rahmen enthalten Dimensionen, die in den drei Voraussetzungen nicht im Zentrum stehen, etwa das Verständnis, wie KI funktioniert, und der verantwortliche, ethische Umgang. Umgekehrt schärfen die drei Voraussetzungen einen Punkt, den die breiten Modelle eher allgemein behandeln: dass die Qualität eines Ergebnisses nicht an der Prompt-Technik hängt, sondern an Fachwissen, Vorgehen und Zielklarheit. Das Bedienen der KI ist nur eine Dimension unter mehreren – und die anspruchsvollen sind die menschlichen.

ZUM WEITERLESEN

# Was sonst noch im Buch steht

Die acht Stunden auf den vorigen Seiten bilden das erste von zwölf Kapiteln im *Praxisbuch KI-Kompetenz für die Schule*. Sie legen das Fundament: wie Sprachmodelle zu ihren Antworten kommen, woran sich Halluzinationen erkennen lassen, wie Verzerrungen entstehen und wie man mit KI arbeitet, ohne ihr das Urteil zu überlassen.

Darauf bauen die übrigen elf Kapitel auf. Sie führen vom Schreiben, Lernen und Recherchieren mit KI über Einblicke in eine künstliche Öffentlichkeit und die psychologische Seite der Mensch-Maschine-Beziehung bis zu Recht, Arbeitswelt und Beispielen für den Fachunterricht mit KI sowie Anregungen zum Vibe Coding (Programmieren mit KI) mit den Klassen. Ein Anhang versammelt neben sämtlichen Musterlösungen auch noch sechs Orientierungstexte zu rechtlichen Aspekten und ein Glossar.

Das folgende Inhaltsverzeichnis gibt den vollständigen Inhalt des Buches wider.

# Inhalt

*12 Kapitel · 72 Stundenideen · Anhang*

## 11 Fachunterricht mit KI



## 12 Vibe Coding und Apps



## Anhang für Lehrkräfte

GRUNDKURS KI-KOMPETENZ · BAND 1

# *Sie haben acht Stunden gesehen. Es gibt noch 64 weitere.*

Das vollständige *Praxisbuch* enthält 72 ausgearbeitete Unterrichtsstunden in zwölf Kapiteln. Sie führen vom Schreiben, Lernen und Recherchieren mit KI über Einblicke in eine künstliche Öffentlichkeit und die psychologische Seite der Mensch-Maschine-Beziehung bis zu Recht, Arbeitswelt und Beispielen für den Fachunterricht mit KI sowie Anregungen zum Vibe Coding (Programmieren mit KI) mit den Klassen. Alle Stunden sind am *AI Literacy Framework* von OECD und Europäischer Kommission ausgerichtet und decken dessen 19 Einzelkompetenzen ab.

Zusätzlich enthält das *Praxisbuch* einen umfangreichen Anhang mit sämtlichen Musterlösungen, sechs Orientierungstexten zu rechtlichen Aspekten sowie ein Glossar, der alle relevanten Begriffe rund um KI prägnant erklärt. Die 1. Auflage ist im August 2026 als E-Book erschienen. Der Text wird kontinuierlich aktualisiert; neue Auflagen stehen allen Käuferinnen und Käufern kostenfrei zur Verfügung.

Der Autor, Dr. Sven Lüder, begleitet als KI-Bildungsreferent Lehrkräfte und Schulen beim Umgang mit Künstlicher Intelligenz.

BESTELLOPTIONEN UND WEITERE MATERIALIEN

***ki-lehren.de/grundkurs-schule***

**ki-lehren.de**

1. AUFLAGE · 2026