

*„Dürfen wir das für die
Hausaufgabe benutzen?“*

*„Bessere Frage: Was lernt ihr dabei –
und was nicht?“*

• • •

GRUNDKURS KI-KOMPETENZ · BAND 1

Praxisbuch KI-Kompetenz für die Schule

Anregungen und Aufgaben für Lehrkräfte.

71 Unterrichtsstunden zum Erlernen einer neuen Kulturtechnik.

IMPRESSUM

TITEL

Praxisbuch KI-Kompetenz für die Schule.
Anregungen und Aufgaben für Lehrkräfte. 71 Unterrichtsstunden zum
Erlernen einer neuen Kulturtechnik.
Grundkurs KI-Kompetenz, Band 1.

AUTOR UND HERAUSGEBER

Dr. Sven Lüder
Forstweg 19
15738 Zeuthen
kontakt@ki-lehren.de · ki-lehren.de

ERSTE AUFLAGE · JULI 2026

Im Eigenverlag erschienen. Alle Rechte vorbehalten.

GESTALTUNG UND SATZ

Sven Lüder

HINWEIS

Die hier vorgestellten Aufgaben wurden für den Unterricht entwickelt und sind didaktisch durchdacht. Sie ersetzen keine schulinterne Regelung zum KI-Einsatz. Vor der ersten Anwendung empfehle ich, die geltende DSGVO-Lage Ihrer Schule und die im Anhang aufgeführten Quellen zu prüfen.

SCHRIFTEN

DM Serif Display, Lora, IBM Plex Sans

Inhalt

12 Kapitel · 76 Stundenideen · Bonus · Anhang

V

Vorwort

Eine neue Kulturtechnik

1

KI verstehen und einsetzen

8 STUNDEN

- 1 Wie ein Sprachmodell zu Antworten kommt
- 2 Warum die KI so höflich ist – die drei Stufen ihres Trainings
- 3 Bei welchen Aufgaben ist die KI zuverlässig?
- 4 Halluzinationen erkennen und vorbeugen
- 5 Wenn die KI Vorurteile reproduziert – Bias am konkreten Beispiel
- 6 Warum funktioniert KI auf Englisch besser als auf Türkisch?
- 7 Wie man mit KI arbeitet: Kontext, Aufgabe, Format und Material
- 8 Was gute Arbeit mit KI ausmacht: Fachwissen, Methode, Strategie

2

Lernen mit und ohne KI

6 STUNDEN

- 9 KI als Werkzeug oder Abkürzung?
- 10 Die KI als geduldiger Privatlehrer
- 11 Hausaufgaben mit KI: Was bringt mich weiter, was nicht?
- 12 Allgemeinbildung trotz KI
- 13 Erst denken, dann prüfen – KI als Sparringpartner für meine Argumente
- 14 Texte lesen oder zusammenfassen lassen?

3

Schreiben oder schreiben lassen?

5 STUNDEN

- 15 Kantige und glatte Texte. Was ist meine eigene Stimme beim Schreiben wert?
- 16 KI-Antworten analysieren – was macht ihren typischen Ton aus?
- 17 KI als Schreibtutor – Feedback einholen, Überarbeitung verstehen
- 18 Die eigene Text-Stimme simulieren – kann eine KI nach mir klingen?
- 19 KI-assistierte Schreiben und Ghostwriting – was genießt mehr Ansehen?

4

Irgendwas mit KI-Medien

7 STUNDEN

- 20 KI-Bilder erzeugen
- 21 Bildmanipulation in historischer Perspektive
- 22 Deepfakes erkennen – Indizien und Grenzen
- 23 Wenn aus einem Foto ein Nacktbild wird – Nudify-Apps und ihre Folgen
- 24 Stimmen klonen – Anwendungen und Missbrauch
- 25 KI-Podcasts: aus Quellen ein Hörgespräch erzeugen
- 26 KI-Musik: eigene Songs erzeugen

5

Recherchieren mit KI

5 STUNDEN

- 27 Mit KI recherchieren – Treffer als Anfangshypothese
- 28 KI-Antworten als Spiegel der Mehrheitsmeinung
- 29 Wahre, plausible und konsensfähige Aussagen unterscheiden
- 30 Wikipedia, klassische Suche, KI-Suche – wann was nutzen
- 31 KI in der Wissenschaft – Chance und Krise

6

Künstliche Öffentlichkeit

5 STUNDEN

- 32 KI in der Politik – eine internationale Fallstudie
- 33 Was hilft gegen Deepfakes im politischen Kontext?
- 34 Microtargeting und algorithmische Öffentlichkeit
- 35 KI-generierte Influencer und Astroturfing
- 36 AI-Slop – warum unser Feed plötzlich anders aussieht

7

Mensch und Maschine – die psychologische Seite

4 STUNDEN

- 37 Mein Chatbot, mein Freund? – Companion-KI im Selbsttest
- 38 Die KI gibt mir immer recht – ein Problem?
- 39 Was würdet ihr einer Maschine sagen, was ihr keinem Menschen anvertraut?
- 40 Eine Woche KI-Tagebuch – schätze ich meine Nutzung richtig ein?

8

Recht, Datenschutz, Verantwortung

4 STUNDEN

- 41 Was passiert eigentlich mit meinen Daten?
- 42 Urheberrecht bei KI-Inhalten – darf ich das nutzen?
- 43 Ist Datenschutz zurecht langweilig?
- 44 Wer haftet bei Fehlern? Ein Fall durchgespielt

9

KI und Arbeitswelt

6 STUNDEN

- 45 Berufsbilder im Wandel – was lässt sich automatisieren?
- 46 „Das kann KI doch niemals!“ (oder?)
- 47 Bewerbungen schreiben mit KI – wann hilft sie wirklich?
- 48 Eigene Kompetenzen sichtbar machen, wenn alle KI nutzen
- 49 Wer macht was? Eine Aufgabe bewusst zwischen mir und der KI aufteilen
- 50 KI-Agenten am Arbeitsplatz – was sie sind, was sie verändern werden

10

Wirtschaftliche & politische Strukturen hinter der KI

5 STUNDEN

- 51 Wem gehört die KI? Marktmacht und Abhängigkeit
- 52 Energie- und Wasserkosten generativer KI
- 53 KI im militärischen Einsatz
- 54 Clickwork und Data Labeling – wer trainiert die Modelle?
- 55 Was ändert sich, wenn alles mitgelesen wird?

11

Fachschwerpunkte · Sprachen, Gesellschaft, Kunst

8 STUNDEN

- 64 Deutsch: KI-Texte analysieren
- 65 Deutsch: Materialgestützt argumentieren – die eigene Position gegen die KI behaupten
- 66 Deutsch: KI gegen Lyrik – wer kann besser dichten?
- 67 Fremdsprachen: KI als Konversationspartner – im Selbsttest
- 68 Geschichte: KI prüft historische Mythen – Faktencheck mit Quellen
- 69 Ethik/PoWi: KI als Dilemma-Sparring
- 70 Kunst: Generative Werkzeuge im Atelier
- 71 Musik: Songtext entwerfen und performen

12

Fachschwerpunkte · MINT, Sport

5 STUNDEN

- 72 Mathematik: KI-Lösungswege nachvollziehen und Fehler finden
- 73 Naturwissenschaften: Versuchsprotokoll mit KI sprachlich überarbeiten
- 74 Informatik: Code mit KI lesen, debuggen, verstehen
- 75 Naturwissenschaften: Modellbildung mit KI gegenprüfen
- 76 Sport: Trainingsplanung und Bewegungsanalyse

B

Bonus · Vibe Coding und Apps

7 STUNDEN

- B 1 Vibe Coding – einfach mal loslegen
- B 2 Eine kleine Datenstudie zu einer Schulfrage
- B 3 Ein Browser-Spiel mit KI bauen
- B 4 Eine Klassen- oder AG-Webseite generieren
- B 5 Eine kleine Alltagsautomatisierung für den Schultag
- B 6 Eine eigene App-Idee prototypisieren
- B 7 Individuelle Präsentationen mit KI erstellen

A

Anhang für Lehrkräfte

LÖSUNGEN · 9 TEXTE · GLOSSAR

Lösungen

ZU ALLEN AUFGABEN · SEITE 29

Praxis und Recht für Lehrkräfte

NEUN ORIENTIERUNGSTEXTE · SEITE [S]

- A 1 Prüfungsformate fürs KI-Zeitalter – Prävention, Adaption, Innovation
- A 2 KI-Erkennungssoftware – Möglichkeiten und Grenzen
- A 3 Verdachtsfälle didaktisch wenden statt sanktionieren
- A 4 KI-resistente Aufgabenformate gestalten
- A 5 Nudify-Vorfälle in der Schule – rechtlich und pädagogisch
- A 6 DSGVO und Schüler-Accounts – datenschutzkonforme Nutzungsmodelle
- A 7 Urheberrecht in der Schulpraxis
- A 8 Schulinterne KI-Vereinbarung partizipativ entwickeln
- A 9 Stand des KI-Schulrechts mit Quellenliste

Glossar

ZENTRALE BEGRIFFE · SEITE [S]

VORWORT

Eine neue Kulturtechnik

Die KI-Transformation ist allgegenwärtig und steht zugleich, während ich dieses Vorwort schreibe, noch ganz am Anfang. Im Jahr 2026 nutzt die große Mehrheit der Schülerinnen und Schüler im deutschsprachigen Raum längst Anwendungen, die auf generativer künstlicher Intelligenz (im Folgenden kurz: KI), insbesondere auf großen Sprachmodellen (Large Language Models, LLMs) beruhen. Viele Kinder und Jugendliche gehen mit der neuen Technologie vertrauter um als die mit ihrer Erziehung und Bildung betrauten Erwachsenen. Dabei dürften sich nur die wenigsten ihr Wissen systematisch angeeignet haben, gar mit didaktisch sinnvoller Progression. Und noch seltener wird man sie dort, wo sie vom technisch Machbaren erfahren haben, zu der ethischen Frage angeregt haben, welche Grenzen bei der KI-Nutzung gelten sollten (Stichwort: Deepfakes).

Wenn Schule ihrem Anspruch gerecht werden will, junge Menschen aufs Leben vorzubereiten, muss sie zum zentralen Ort der Auseinandersetzung mit KI und zur Vermittlung der entsprechenden Kompetenzen werden. Der Gegenstand ist geradezu prädestiniert für ein Unterrichtsgespräch auf Augenhöhe: Im Zweifel verstehen die Schülerinnen und Schüler einzelne Aspekte der Technik und ihrer Anwendung besser, wovon ihre Lehrkräfte ruhig profitieren dürfen. Diese wiederum müssen mit ihrer pädagogischen Kompetenz und Erfahrung dafür sorgen, dass die Beschäftigung mit KI im angemessenen Alter geschieht und sowohl an sinnvoller Produktivität als auch an ethischen Werten orientiert bleibt. Für genau diesen konstruktiven Unterricht ist dieses Praxisbuch gemacht. Es richtet sich zunächst an Lehrkräfte, denen es Überblick, Grundlagen und Auswahl bietet; in seinem Kern aber spricht es die Schülerinnen und Schüler selbst unmittelbar an, indem es einzelne Aspekte von KI in Form von Aufgaben für sie aufbereitet.

Mir ist wichtig, hervorzuheben, dass es in diesem Praxisbuch nie um bloßes Tool-Training geht. KI ist kein weiteres Programm, dessen Funktionen sich in einer Schulung oder einem YouTube-Tutorial wirksam erklären lassen. Stattdessen stellt ihre Nutzung eine neue Kulturtechnik dar und steht damit auf einer Ebene mit Schreiben, Lesen und Rechnen – als Basisfähigkeit, die auf diesen dreien aufbaut. Wer nicht lesen, schreiben und rechnen kann, kann auch mit KI nicht produktiv arbeiten. Und wer souverän mit ihr umgehen will, braucht sogar noch mehr: strukturiertes und kritisches Denken zumal und Fachwissen in den Bereichen, in denen man mehr als durchschnittliche Leistungen erbringen möchte.

Wie lernt man eine Kulturtechnik? Sicherlich nicht durch Ratgeber oder theoretische Erklärungen. Auch um lesen zu können, reicht es nicht, die graphische Form und Verlautung der Buchstaben zu kennen. Man braucht Praxiswissen über unzählige Situationen: Wer nicht weiß, dass ein Gedicht anders zu lesen ist als eine Gebrauchsanweisung, wird sich in unserer Kultur schertun. Hinzu kommt noch ein selbst angeeignetes Erfahrungswissen darüber, wann sich eine Technik sinnvoll einsetzen lässt und wann nicht und worauf dabei zu achten ist. Knapp zehn Jahre Lehre an verschiedenen Hochschulen, zuletzt an der University of Oxford, haben mir immer wieder gezeigt: Neue Themen erschließen sich Lernende am besten durch eigene Praxis – durch ein Ausprobieren und die Chance, dabei Fehler zu machen. Deshalb habe ich die vielen relevanten Aspekte von KI in Form konkreter Aufgaben aufbereitet.

Dieses Praxisbuch enthält 76 Stundenideen in zwölf Basiskapiteln sowie ein Bonuskapitel zu Vibe Coding (also dem Programmieren mit KI). Zusätzlich gibt es einen Anhang für Lehrkräfte, der allgemeine Orientierungstexte zum Komplex KI und Schule, Musterlösungen zu allen Aufgaben sowie ein Glossar enthält. Im Hauptteil ist jede Stunde auf einer Seite ausgearbeitet: mit Lernzielen, einer kurzen Kontext- oder Situationsbeschreibung, die den Aspekt in der Lebenswelt der Schülerinnen und Schüler verortet, sowie fünf bis sechs aufeinander aufbauende Aufgaben, die dazu einladen, sich einen Aspekt praktisch zu erarbeiten. Die Progression beginnt in der Regel mit Brainstorming oder Recherche und führt über gezielte Anwendungen zur Reflexion und Diskussionen mit Transferanteil. Damit die Aufgaben in möglichst vielen Klassenzimmern funktionieren, habe ich zu jeder Stunde Differenzierungsmöglichkeiten nach unten und oben angegeben, nenne geeignete Tools und für die wirklich Interessierten einen Text sowie eine passende Podcast-Folge als vertiefende Quellen. Insbesondere die Podcasts, hier in der Regel von öffentlich-rechtlichen Sendern, sind eine perfekte Quelle, um sich niedrigschwellig über KI zu informieren und bei diesem Thema, wo sich Technologie und Diskurs so rasant entwickeln, auf dem Laufenden zu bleiben.

Ein solches Praxisbuch kann immer nur eine Momentaufnahme sein. Ich werde mich nach meinen Möglichkeiten bemühen, es aktuell zu halten und Ihnen, sofern Sie ein Exemplar erworben haben, den digitalen Zugang zu neueren Auflagen kostenfrei zu ermöglichen. Dabei setze ich auf eine Methode, die in Reiseführern üblich ist: Schreiben Sie mir bitte eine E-Mail an kontakt@ki-lehren.de, wenn Ihnen ein Aspekt auffällt, der veraltet ist. Ich werde es Ihnen danken!

Inhaltlich gibt es neben der Nutzung von KI als Kulturtechnik noch einen zweiten Schwerpunkt, nämlich die Reflexion über Themen und Fragen, die KI aufwirft. Die Stunden vollziehen darum eine doppelte Bewegung. Sie führen die Schülerinnen und Schüler einerseits an konkrete Anwendungen heran – Texte überarbeiten, Bilder erzeugen, Quellen prüfen, KI zur Recherche nutzen – und öffnen andererseits den Blick auf dahinterliegende Fragen: Was geschieht, wenn Maschinen scheinbar fehlerfreie Texte produzieren? Welche KI-Nutzung ist förderlich, welche schädlich? Unter welchen Bedingungen werden die Modelle trainiert? Was wird aus den Berufen, in denen die Kinder und Jugendlichen einmal arbeiten werden? Fachwissen bleibt auch im KI-Zeitalter wichtig. Das gilt auch für KI-Kompetenz selbst, die sich darum neben Anwendungswissen auch durch thematische Expertise auszeichnen muss.

Bewusst habe ich auch schwierige Themen aufgenommen, etwa Nudify-Apps, die Lage der Click-Worker oder KI und psychische Gesundheit. Sie gehören in den Unterricht, weil sie die Welt der Schülerinnen und Schüler schon heute prägen. Die dazugehörigen Aufgaben sollen vor allem zur Diskussion einladen. Die fachbezogenen Kapitel am Ende – Sprachen, Gesellschaft und Kunst sowie MINT und Sport – zeigen beispielhaft, wie sich KI in einem Fach verankern lässt, ohne dessen Kern zu verdrängen. Der Anhang für Lehrkräfte bündelt neben den Lösungen auch informative Texte zu einigen Themen, die eng mit KI zusammenhängen, über den Unterricht im engeren Sinne aber hinausgehen, etwa zum Umgang mit Prüfungsformaten und Verdachtsfällen oder was zum Thema Datenschutz und Schulrecht relevant ist.

Die folgenden Seiten verstehen sich als Anregung, KI im Unterricht zu behandeln. Die Kapitel und die Abfolge der Aufgaben habe ich didaktisch durchdacht; was Sie davon für Ihre Klassen übernehmen, bleibt Ihnen überlassen. Wichtig ist mir vor allem, dass die Auseinandersetzung mit KI in der Schule stattfindet. Sie ist der einzige Ort, an dem sich verlässlich sicherstellen lässt, dass Kinder und Jugendliche dem Thema chancengerecht und dabei didaktisch begleitet begegnen. Wir allen müssen, so meine ich, unser Möglichstes tun, damit sich Kinder und Jugendliche und junge Erwachsene KI als Kulturtechnik so erschließen, dass sie ihnen selbst – wie auch den Menschen, mit denen sie verbunden sind – am Ende mehr nützen möge als schaden.

Zeuthen, im Frühjahr 2026

Sven Lüder

ZUM AUFBAU

Wie Sie mit diesem Praxisbuch arbeiten

Dieses Praxisbuch behandelt in Kapitel 1 die Grundlagen kompetenter KI-Nutzung und ist damit geeignet, bei ihren Klassen ein Fundament zu legen, um die neue Kulturtechnik zu erlernen. In seinen weiteren zwölf Kapiteln, deren Stunden frei kombinierbar sind, soll es dazu anregen, KI-Themen, die Ihnen für Ihre Klassen spannend und wichtig erscheinen, im Unterricht zu behandeln. Die Namen der Kapitel und Stunden sind so gewählt, dass sie eine klare Idee vom Inhalt und der Facheignung vermitteln. Zugleich sollen sie neugierig machen, einzelne Stunden herauszugreifen und Aufgaben frei zu übernehmen oder anzupassen.

Alle Stundenseiten folgen demselben Aufbau, den das Schema zeigt.

TITEL UND STUNDENNUMMER
LERNZIELE „Ich kann ...“
KONTEXT ODER SITUATION Verortung in der Lebenswelt
AUFGABEN 1–6 mit aufsteigender Progression
DIFFERENZIERUNG Leichter · Schwerer
GEEIGNETE TOOLS
VERTIEFUNG In der Regel: Text + Podcastfolge
→ LÖSUNG IM ANHANG

LERNZIELE

Die Ziele geben der Lehrkraft den schnellsten Überblick zum Stundeninhalt. Außerdem sollen sie der Selbsteinschätzung dienen und sind darum aus Perspektive der Schülerinnen und Schüler formuliert.

AUFGABEN

Die Aufgaben sprechen die Klasse direkt an; Sie können sie projizieren oder kopieren. Die Reihenfolge führt vom Sammeln und Recherchieren über ein Anwenden und Ausprobieren zur Reflexion. Passende Sozialformen und die Zeitplanung ergeben sich meistens aus der Aufgabenstellung. Zu jeder Aufgabe steht eine Musterlösung im Anhang.

DIFFERENZIERUNG

„Leichter“ nennt eine reduzierte Variante, die oft ohne Schüler-Endgeräte auskommt. „Schwerer“ erweitert die Stunde für schnelle Lerngruppen oder Oberstufenkurse.

Diese verschiedenen Elemente liefern die didaktisch aufbereitete Substanz einer Unterrichtsstunde, ohne Ihnen einen Lehrstil vorzuschreiben. Wie sich aus einer Stundenseite ein vollständiger Verlauf bauen lässt, zeigen zwei exemplarische Stundenentwürfe im ersten Kapitel (nach den Stunden 1 und 7).

KAPITEL 1 · 8 STUNDEN

1

KI verstehen, prüfen, einsetzen

Das erste Kapitel schafft die Grundlagen. In acht Stunden bauen die Schülerinnen und Schüler ein Verständnis für die allgemeine Funktionsweise von Generativer KI auf und erhalten eine strukturierte Anleitung, wie sie die Technologie reflektiert nutzen können. Konzepte wie Halluzinationen, Bias und das Prompt-Schema werden spätere Kapitel immer wieder aufgreifen, ohne sie neu zu erklären.

STUNDEN IN DIESEM KAPITEL

- 1 Wie ein Sprachmodell zu Antworten kommt
- 2 Warum die KI so höflich ist – die drei Stufen ihres Trainings
- 3 Bei welchen Aufgaben ist die KI zuverlässig?
- 4 Halluzinationen erkennen und vorbeugen
- 5 Wenn die KI Vorurteile reproduziert – Bias am konkreten Beispiel
- 6 Warum funktioniert KI auf Englisch besser als auf Türkisch?
- 7 Wie man mit KI arbeitet: Kontext, Aufgabe, Format und Material
- 8 Was gute Arbeit mit KI ausmacht: Fachwissen, Methode, Strategiebewusstsein

1 Wie ein Sprachmodell zu Antworten kommt

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 – und 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.

EXEMPLARISCHER STUNDENENTWURF

Stunde 1 im Verlauf – ein Beispiel für 45 Minuten

Dieser Entwurf zeigt, wie sich aus der Stundenseite ein vollständiger Verlauf entwickeln lässt. Er ist ein Vorschlag – Zeiten und Sozialformen richten sich nach Ihrer Lerngruppe.

VERLAUFSPLAN

ZEIT	PHASE	UNTERRICHTSGESCHEHEN	SOZIALFORM · MEDIEN
5'	Einstieg	An der Tafel stehen zwei Satzanfänge. Die Schüler schreiben kurz drei mögliche Fortsetzungen. Ihre Vorschläge werden als Strichliste gezählt; die Häufung weniger Favoriten wird sichtbar (Aufgabe 1).	Plenum · Tafel
5'	Überleitung	Die Lehrkraft löst auf: Genau so arbeitet ein Sprachmodell – nur mit Milliarden von Beispielsätzen. Die Lernziele werden projiziert.	Plenum · Beamer
10'	Erarbeitung I	Tandems lassen die KI für zwei Satzanfänge die fünf wahrscheinlichsten Fortsetzungen mit Prozentwerten schätzen und vergleichen flache und spitze Verteilungen (Aufgabe 2).	Tandem · KI-Zugang
10'	Erarbeitung II	Tokenizer-Experiment: Ein deutscher Satz wird zerlegt, der Unterschied zwischen Wort und Token geklärt (Aufgabe 3).	Tandem · Tokenizer
8'	Sicherung	Die Faustregel „spitze Verteilung = sichere KI“ wird formuliert und mit dem Begriff Halluzination verbunden (Aufgabe 4).	Einzelarbeit, dann Plenum
7'	Reflexion	Diskussion: Ist die Wortauswahl der KI mit menschlichem Sprechen vergleichbar? (Aufgabe 5)	Plenum

MÖGLICHE SCHWIERIGKEITEN

- › Es ist möglich, dass die KI Prozentschätzungen verweigert oder bei jedem Versuch andere Werte liefert. Das ist kein Fehler, sondern ein Anlass, über den Zufallsanteil der Ausgabe zu sprechen.
- › Der Tokenizer zerlegt deutsche Wörter manchmal unerwartet kleinteilig; diese Irritation bereitet die Erkenntnis von Stunde 6 (Sprachenungleichheit) vor.
- › Ohne Schülergeräte wird Aufgabe 2 zur Demonstration am Beamer (siehe Differenzierung „Leichter“).

„Der Hund läuft über die ...“ – flache Verteilung

Straße	41 %
Wiese	22 %
Brücke	12 %
Kreuzung	8 %
Ziellinie	5 %

„Zwei plus zwei ergibt ...“ – spitze Verteilung

vier	97 %
4	2 %
fünf	< 1 %

Ist die Verteilung spitz, „weiß“ das Modell die Fortsetzung.

Ist sie flach, sind viele Fortsetzungen plausibel – hier beginnt das Risiko der Halluzination.

Abb.: Zwei Wahrscheinlichkeitsverteilungen für das nächste Wort – Tafelbild zur Sicherung. Werte schematisch.

2 Warum die KI so höflich ist – die drei Stufen ihres Trainings

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. Widerspricht 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 schlaueste KI der Welt?“, 24.09.2024.

3 Bei welchen Aufgaben ist die KI zuverlässig?

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): „Was kann GPT-5 (und was nicht)?“, 12.08.2025.

4 Halluzinationen erkennen und vorbeugen

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 (Zusammenfassung im OpenAI-Blog).
- › 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.

5 Wenn die KI Vorurteile reproduziert – Bias am konkreten Beispiel

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. Wo künstliche Intelligenz sich irrt“, Heyne, 2019.
- › KI verstehen (Deutschlandfunk): „KI und Sexismus – Wie KI Rollenbilder zementiert“, 18.09.2025.

6 Warum funktioniert KI auf Englisch besser als auf Türkisch?

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.

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

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“, docs.claude.com.
- › Lüder, S.: „Prompt-Schema“, ki-lehren.de.
- › Der KI-Podcast (ARD): „Prompten wir alle falsch?“, 17.02.2026.

EXEMPLARISCHER STUNDENENTWURF

Stunde 7 im Verlauf – ein Beispiel für 45 Minuten

Der zweite Beispielentwurf zeigt eine stärker gelenkte Erarbeitung: Die Klasse baut denselben Prompt in drei Schritten aus und vergleicht nach jedem Schritt den Output.

VERLAUFSPLAN

ZEIT	PHASE	UNTERRICHTSGESCHEHEN	SOZIALFORM · MEDIEN
5'	Einstieg	Der Minimalprompt „Schreib mir etwas über Globalisierung“ läuft live am Beamer; die Klasse bewertet das Ergebnis (Aufgabe 1).	Plenum · Beamer
7'	Problematisierung	Was fehlt dem Text? Kriterien werden gesammelt; die Briefing-Analogie (Zeitarbeiter ohne Kontext) wird eingeführt.	Plenum · Tafel
18'	Erarbeitung	Tandems bauen den Prompt in drei Schritten aus: erst Kontext, dann Aufgabe und Format, zuletzt eigenes Material. Nach jedem Schritt wird der Output kurz verglichen (Aufgaben 2–4).	Tandem · KI-Zugang
8'	Sicherung	Das Prompt-Schema wird als Tafelbild festgehalten (vgl. Abbildung); Anschlussfrage: Was gehört in den System-Prompt? (Aufgabe 5)	Plenum · Tafel
7'	Transfer	Jedes Tandem legt eine wiederverwendbare Prompt-Vorlage für ein eigenes Vorhaben an und sichert sie (Aufgabe 6).	Tandem

MÖGLICHE SCHWIERIGKEITEN

- › Die Verbesserung von Schritt zu Schritt ist nicht immer spektakulär – dann hilft der direkte Vergleich zweier Outputs nebeneinander am Beamer.
- › Vor dem Materialanhang (Aufgabe 4) das Urheberrecht ansprechen, bevor Lehrbuchseiten hochgeladen werden (vgl. Anhang A7).

KONTEXT

Zweck, Zielgruppe, eigene Rolle

AUFGABE

Präzise Handlung

FORMAT

Dateiformat, Länge, Sprachregister

MATERIAL

Eigene Beispiele, Quellen, Daten



Je mehr Bausteine ein Prompt enthält, desto weniger muss das Modell raten – und desto näher liegt der Output am eigenen Bedarf.

Abb.: Das Prompt-Schema als Tafelbild zur Sicherung.

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

LERNZIELE

- › erkennen, dass es nicht ausreicht, prompts 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 Prompts 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 eindrucklichsten 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](https://www.ki-lehren.de).
- › Mollick, E.: „Co-Intelligenz. Leben und Arbeiten mit künstlicher Intelligenz“, Plassen, 2025.
- › KMK: „Handlungsempfehlungen zum Umgang mit KI in Schule und Unterricht“, [kmk.org](https://www.kmk.org), 2024.

KAPITEL 2 · 6 STUNDEN

2

Lernen mit und ohne KI

Mit den Werkzeugen aus Kapitel 1 in der Hand lässt sich untersuchen, wie sich das Lernen verändert, wenn man jede Frage einfach durch die KI beantworten lassen könnte. Sieben Stunden zielen auf die wesentlichen Unterschiede, ob man die KI als Werkzeug oder Abkürzung einsetzt, zur konstruktiven Hilfe oder aus Bequemlichkeit. Dabei brauchen die Stunden selbst nur wenige Tools. In erster Linie geht es hier darum, Reflexionsangebote zu machen. Eine Stunde geht weiter: Sie stellt mit einem mehrwöchigen Klassenexperiment eine bekannte Studie zum Lernen mit und ohne KI im Kleinen nach.

STUNDEN IN DIESEM KAPITEL

- | | |
|--|--|
| 9 KI als Werkzeug oder Abkürzung? | 12 Allgemeinbildung trotz KI |
| 10 Die KI als geduldiger Privatlehrer | 13 Erst denken, dann prüfen – KI als Sparringpartner für meine Argumente |
| 11 Hausaufgaben mit KI: Was bringt mich weiter, was nicht? | 14 Texte lesen oder zusammenfassen lassen? |

9 KI als Werkzeug oder Abkürzung?

LERNZIELE

- › die Funktion von KI im eigenen Lernprozess reflektieren
- › unterscheiden, wann KI das Lernen unterstützt und wann sie es ersetzt
- › eine eigene Position zur Nutzung im Schulalltag formulieren

SITUATION

Ein Schüler lässt sich ein halbes Jahr lang jede Englisch-Hausaufgabe von der KI formulieren. Die Aufgaben sind tadellos, die Noten darauf gut. In der Klassenarbeit, ganz ohne KI, fehlen ihm dann die Wörter und Satzmuster, die er nie selbst geübt hat. Hat ihm, was ihm kurzfristig geholfen hat, langfristig geschadet?

AUFGABEN

- 1 Besprecht im Plenum: Für welche produktiven Zwecke habt ihr in der letzten Woche KI genutzt? Sammelt die zehn häufigsten Anwendungen an der Tafel.
- 2 Sortiert im Tandem die Liste in zwei Spalten: „unterstützt mein Lernen“ und „ersetzt mein Lernen“. Wo seid ihr unsicher?
- 3 Greift drei Fälle heraus, bei denen ihr unsicher wart, und klärt, worauf es ankommt. Nehmt anschließend einen davon und überlegt konkret: Wie müsstet ihr die KI einsetzen, damit sie euer Lernen unterstützt statt es zu ersetzen?
- 4 Recherchiert die Begriffe „Cognitive Offloading“ und „Skill Atrophy“. Passt das Bild „KI als Abkürzung beim Denken“ zu dem, was diese Begriffe beschreiben? Wo trifft es, wo greift es zu kurz?
- 5 Formuliert für euch selbst drei Kriterien (z.B. „Habe ich es zuerst selbst versucht?“), die sicherstellen, dass KI euer Lernen nicht ersetzt, und stellt sie anschließend im Plenum vor. Gibt es Unterschiede?
- 6 Diskutiert: Wer sollte entscheiden, welche Nutzung sinnvoll ist – ihr selbst, die Lehrkraft, die Schule, die Eltern? Ab welchem Alter kann man das für sich selbst entscheiden?

DIFFERENZIERUNG

LEICHTER

Aufgaben 1 bis 3. Das Bild der Abkürzung aus Aufgabe 4 wird im Plenum besprochen, ohne dass die Begriffe recherchiert oder eigene Kriterien formuliert werden müssen.

SCHWERER

Informiert euch über ein historisches Beispiel für eine Technik, die Fähigkeiten hat verkümmern lassen (z.B. Taschenrechner, Navigationsgerät, Rechtschreibprüfung). Was ging verloren, was wurde gewonnen, und was folgt daraus für den Umgang mit KI?

GEEIGNETE TOOLS

Diese Stunde braucht kein KI-Tool. Optional eine Wortwolke aus eurer Liste (etwa über [mentimeter.com](https://www.mentimeter.com)).

VERTIEFUNG

- › Mollick, E.: „Co-Intelligenz. Leben und Arbeiten mit künstlicher Intelligenz“, Plassen, 2025.
- › Der KI-Podcast (ARD): „Macht KI uns dümmer?“, 11.03.2025.

10 Die KI als geduldiger Privatlehrer

LERNZIELE

- › KI als Erklärinstrument einsetzen
- › typische Stärken und Grenzen der KI-Tutoring-Funktion erkennen
- › einschätzen, in welchen Fächern die KI ein guter Tutor ist (und in welchen weniger)

SITUATION

Eine Schülerin hat in Mathe seit Wochen ein Thema nicht verstanden. Im Unterricht traut sie sich nicht zu fragen, zu Hause läuft das Gespräch in eine Sackgasse. Sie öffnet ChatGPT und fragt einfach drauflos – ohne Hemmung, mit allen Lücken. Was zurückkommt, überrascht sie.

AUFGABEN

- 1 Wählt im Tandem ein Thema, das ihr im Unterricht nur halb verstanden habt, und lasst es euch von einer KI erklären. Arbeitet mit gezielten Rückfragen: Bittet um ein konkretes Beispiel, um eine Analogie und darum, die noch unklare Stelle anders zu erklären. Hört erst auf, wenn ihr das Thema in eigenen Worten wiedergeben könnt.
- 2 Vergleicht im Tandem: Was hat die KI besser erklärt als Schulbuch oder Lehrkraft, was schlechter? Welche eurer Rückfragen haben am meisten gebracht?
- 3 Aktiviert den Lernmodus eurer KI und vergleicht ihn mit dem normalen Chat. Das geht in Gemini über das Plus-Menü „Lernhilfe“, in ChatGPT mit „/“, dann „Studieren und Lernen“ im Menü, in Claude über Plus → Stil → „Lernen“ (sucht im Zweifel kurz nach den Funktionen). Lasst euch dasselbe Thema einmal normal und einmal im Lernmodus erklären. Was macht der Lernmodus anders?
- 4 Dreht die Rollen um: Bittet die KI, euch zum Thema abzufragen, statt es euch zu erklären – sie stellt Fragen, ihr antwortet, sie gibt Rückmeldung. Was bringt dieses Abfragen, das reines Erklären-Lassen nicht leistet?
- 5 Diskutiert: Was ändert sich am Lernen, wenn jeder Schüler einen kostenlosen Privatlehrer hat?

DIFFERENZIERUNG

LEICHTER

Aufgaben 1-4 mit einem von der Lehrkraft vorgegebenen Thema. Die Schüler dokumentieren, an welcher Stelle der Erklärung sie gedanklich ausgestiegen sind.

SCHWERER

Vergleicht die KI-Erklärung mit einem Erklärvideo (Khan Academy, Lehrerschmidt, simpleclub). Wo erklärt die KI besser, wo das Video? Was passiert, wenn ihr die KI bittet, das Video zu kommentieren?

GEEIGNETE TOOLS

ChatGPT, Claude oder Gemini in der freien Webversion. Für den Vergleich Khan Academy (de.khanacademy.org) oder ein Erklärvideo aus dem aktuellen Stoffbereich.

VERTIEFUNG

- › Mollick, E.: „Co-Intelligenz. Leben und Arbeiten mit künstlicher Intelligenz“, Plassen, 2025.
- › Tech, KI & Schmetterlinge (Sascha Lobo): „KI und Bildung: Dreamteam oder Höllen-Duo? (mit Bob Blume)“, 08.10.2025.

11 Hausaufgaben mit KI: Was bringt mich weiter, was nicht?

LERNZIELE

- › ein Experiment zum KI-Einsatz bei Hausaufgaben durchführen und gemeinsam auswerten
- › unterstützte Leistung und tatsächliches Lernen auseinanderhalten
- › die eigenen Ergebnisse mit einer Studie vergleichen und daraus eine eigene Linie ableiten

KONTEXT

Die KI erledigt die Hausaufgaben in fünf Minuten, kurzfristig ist das super – und langfristig?

AUFGABEN

- 1 Die Klasse teilt sich für die nächsten drei Wochen in drei Gruppen, die ihre Hausaufgaben im Fach unterschiedlich bearbeiten. Gruppe A lässt die KI die Aufgaben komplett erledigen und schaut sich die Lösungen sorgfältig an. Gruppe B arbeitet mit KI: erst selbst versuchen, dann nachfragen, sich Schwierigkeiten erklären lassen und die eigene Lösung prüfen, ohne sich die fertige Lösung geben zu lassen. Gruppe C arbeitet ohne KI.
- 2 Jedes Gruppenmitglied führt ein kurzes Protokoll: pro Hausaufgabe den Zeitaufwand, eine Selbsteinschätzung „Wie kompetent fühle ich mich beim Stoff?“ (Skala 1 bis 5) und Notizen.
- 3 Zwischenfazit nach etwa anderthalb Wochen: Vergleicht die Protokolle bisher. Wo unterscheiden sich die Gruppen in Zeit und Kompetenzgefühl? Wie wohl fühlt ihr euch mit eurem Vorgehen?
- 4 Am Ende schreibt die ganze Klasse einen kurzen Test zum Stoff – ohne KI. Wertet gemeinsam aus: Wie schneiden die Teams ab, und wie passt das zu Zeitaufwand und Kompetenzgefühl aus den Protokollen?
- 5 Vergleicht euer Ergebnis mit der Studie, die dieses Experiment mit fast 1000 Schülern durchgeführt hat (Bastani u. a., 2024): Wo deckt sich euer Ergebnis, wo weicht es ab, und woran könnte das liegen?
- 6 Leitet aus eurem Experiment für euch persönlich drei Regeln für die KI-Nutzung bei Hausaufgaben ab.

DIFFERENZIERUNG

LEICHTER

Nehmt am Experiment teil und nutzt für das Protokoll eine von der Lehrkraft vorbereitete Vorlage. Auswertung (Aufgabe 4) und Studienvergleich (Aufgabe 5) erfolgen gemeinsam im Plenum.

SCHWERER

Falls es Regelungen zum Umgang mit KI an eurer Schule gibt – habt ihr nach eurem Experiment Verbesserungsvorschläge? Wenn ja, formuliert sie. Wenn nein, schreibt einen ersten Entwurf. Bittet die Schulleitung um ein Gespräch, um euren Text zu besprechen.

GEEIGNETE TOOLS

ChatGPT, Claude oder Gemini in der freien Webversion. Eine reale Hausaufgabe aus dem aktuellen Unterricht oder eine gleichwertige Übungsaufgabe.

VERTIEFUNG

- › Bastani, H. u. a.: „Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics“, PNAS 122 (26), 2025.
- › KI verstehen (Deutschlandfunk): „KI auf dem Campus – Wie Hochschullehrer KI kreativ nutzen“, 22.05.2025.

LÖSUNGSTEIL · ZU ALLEN AUFGABEN

Lösungen

Dieser Teil sammelt die Lösungen zu allen Aufgaben des Hefts. 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, benenne ich hier den Spielraum und die Kriterien einer guten Antwort statt einer einzigen Musterlösung. 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. Genau 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. Wichtig: Die genannten Prozentwerte sind keine echten internen Werte des Modells, sondern eine Illustration. Genau diese Form, mal spitz, mal flach, ist der Kern dessen, wie ein Sprachmodell zu Antworten kommt: Es wählt aus einer Verteilung. Ist sie flach, kann dieselbe Frage in einem neuen Chat ein anderes Wort liefern.

› 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 Warum die KI so höflich ist – die drei Stufen ihres Trainings

› 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 sind kein Ausdruck einer Meinung, sondern erlerntes Verhalten, also 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 UNTERSCHIEDEN

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 bei seiner richtigen 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**

Man sollte die Heuristik nicht als Endergebnis behandeln, sondern als Arbeitsdokument. 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 einreichen 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.