

Vorwort

Künstliche Intelligenz ist nicht mehr nur ein Thema für Informatik, Forschung oder Technologieunternehmen. Sie ist Teil des Alltags geworden: in Suchmaschinen, Lernplattformen, Übersetzungsdiensten, Bildgeneratoren, Empfehlungssystemen, Textwerkzeugen und digitalen Assistenten. Schülerinnen und Schüler begegnen KI nicht erst, wenn im Unterricht ausdrücklich über sie gesprochen wird. Sie begegnen ihr längst beim Lernen, Recherchieren, Kommunizieren, Gestalten und Entscheiden.

Die ersten beiden Bände dieser Reihe haben diese Entwicklung aus unterschiedlichen Blickrichtungen betrachtet. Band 1 rückte KI im unmittelbaren Alltag von Jugendlichen in den Mittelpunkt: Schule, Lernen, Mediennutzung, Quellenarbeit, Datenschutz, Feedback und Eigenleistung. Band 2 weitete den Blick auf Gesellschaft, Politik und Zukunft: Demokratie, Überwachung, Deepfakes, Arbeit, Macht, Teilhabe und Verantwortung. Beide Bände führten immer wieder zu derselben Einsicht: Wer KI nur als Werkzeug betrachtet, versteht ihre Wirkung zu kurz. Wer nur über Folgen spricht, ohne die Funktionsweise zu verstehen, bleibt in der Kritik leicht allgemein.

Dieser dritte Band setzt deshalb an der technischen Grundlage an. Er orientiert sich an aktuellen internationalen Forschungskonsensen zur KI-Literacy, die grundlegendes KI-Wissen und kritische Perspektiven nicht gegeneinanderstellen, sondern aufeinander beziehen. Der Band fragt nicht zuerst, ob KI gut oder schlecht ist, sondern wie KI-Systeme arbeiten: wie sie aus Daten lernen, wie sie Muster erkennen, wie sie Texte erzeugen, wie sie klassifizieren, prognostizieren, empfehlen und neue Medien generieren. Zugleich fragt er, wo dabei Fehler entstehen, warum Verzerrungen nicht zufällig sind, weshalb plausible Ausgaben nicht automatisch wahr sind und warum Verantwortung nicht an ein System abgegeben werden kann.

Daraus ergibt sich eine zentrale Grundannahme: Kritisches Denken über KI braucht inhaltliches Wissen über KI. Schülerinnen und Schüler sollen nicht zu Entwicklerinnen und Entwicklern komplexer Modelle werden. Aber sie brauchen ein tragfähiges Grundverständnis davon, was KI-Systeme leisten, woran sie scheitern und warum ihre Ergebnisse immer im Zusammenhang von Daten, Zielgrößen, Training, Einsatzkontext und menschlichen Entscheidungen stehen.

Dieses Verständnis ist keine technische Zusatzqualifikation für wenige. Es ist Teil allgemeiner Bildung. Wer KI-Ausgaben beurteilen soll, muss wissen, warum ein Sprachmodell nicht in einer Datenbank geprüfter Fakten nachschlägt. Wer über Bias sprechen soll, muss verstehen, dass Verzerrungen in Daten, Labels, Zielgrößen, Proxies, Schwellenwerten und Rückkopplungen entstehen können. Wer KI verantwortungsvoll nutzen soll, muss unterscheiden können zwischen einer Ausgabe, einer Entscheidung und der Verantwortung für deren Folgen.

Daraus ergeben sich die Grundprinzipien dieses Bandes. KI-Systeme verstehen ihren Output nicht wie Menschen. Daten zeigen nur Ausschnitte der Wirklichkeit. Systeme optimieren messbare Ziele, nicht automatisch das eigentlich Gemeinte. KI-Ausgaben müssen geprüft werden. Verantwortung bleibt bei Menschen und Institutionen. Diese Prinzipien sind nicht als abstrakte Merksätze gedacht, sondern als roter Faden durch konkrete Unterrichtssituationen.

Der Band behandelt deshalb zunächst grundlegende Fragen: Was meint der Begriff „KI“? Was sind Daten in diesem Kontext? Was bedeutet Lernen als Optimierung? Was unterscheidet Training und Anwendung? Danach richtet sich der Blick auf Sprachmodelle: Tokens, Kontext, Answererzeugung, Halluzinationen, Sycophancy, Prompting und Quellenprüfung. Anschließend werden andere KI-Arten als die momentan fast ausschließlich im Rampenlicht stehenden Sprachmodelle betrachtet: Klassifikation, Prognose, Empfehlung und generative KI. Der Schlussblock fragt, wie man KI-Systeme prüft, wo Bias entsteht und wer verantwortlich bleibt.

Wichtig ist dabei die Balance. KI wird in diesem Band weder mystifiziert noch abgewertet. Sie erscheint nicht als denkendes Gegenüber, aber auch nicht als gewöhnliches Werkzeug ohne eigene Logik. KI-Systeme können nützlich sein, weil sie Muster verarbeiten, Vorschläge machen, Varianten erzeugen und große Datenmengen strukturieren können. Sie bleiben aber begrenzt, weil ihre Ausgaben von Daten, Modellarchitektur, Zielgrößen, Wahrscheinlichkeiten und Einsatzbedingungen abhängen.

Gerade für Schule ist diese Balance entscheidend. Unterricht darf sich nicht darauf beschränken, Schülerinnen und Schülern zu zeigen, wie sie bessere Prompts schreiben. Ebenso wenig genügt es, nur vor Risiken zu warnen. Zwischen Bedienung und Warnung liegt der eigentliche Bildungsauftrag: Lernende sollen verstehen, warum ein System so überzeugend irren kann, warum eine Zahl wie „90 Prozent Genauigkeit“ nicht automatisch genügt, warum gleiche Berechnung nicht automatisch fair ist und warum menschliche Kontrolle mehr bedeutet als ein kurzer Klick auf „OK“.

Dieser Band versteht KI-Bildung deshalb als Verbindung von Grundwissen und kritischer Perspektive. Grundwissen ohne Kritik bleibt technisch verkürzt. Kritik ohne Grundwissen bleibt leicht ungenau. Erst beides zusammen ermöglicht eine sachliche Haltung: KI kann genutzt werden, aber nicht blind; sie kann unterstützen, aber nicht ersetzen; sie kann Hinweise geben, aber keine Verantwortung übernehmen.

Wenn Schülerinnen und Schüler nach der Arbeit mit diesem Band nicht nur fragen „Was sagt die KI?“, sondern „Mit welchen Daten wurde gearbeitet?“, „Welche Zielgröße steckt dahinter?“, „Welche Fehler sind möglich?“, „Wer wird übersehen?“, „Wie wurde geprüft?“ und „Wer entscheidet am Ende?“, dann ist der zentrale Schritt erreicht. Sie bedienen KI nicht nur. Sie beginnen, KI zu verstehen.

Die Kapitel sind für Lerngruppen ab Klasse 10 konzipiert, besonders für die Oberstufe. Je nach Schulart, Fach, Vorwissen und Diskussionskultur der Lerngruppe sollten Umfang, sprachliche Dichte und Aufgabenformate angepasst werden.

Auch Band 3 ist als Materialband angelegt. Die Kapitel sind so konzipiert, dass sie grundsätzlich auch ohne Vorlaufkenntnisse aus anderen Kapiteln einsetzbar sind. Jedes Kapitel behandelt einen eigenständigen Aspekt der KI-Funktionsweise. Es enthält ein Arbeitsblatt für Schülerinnen und Schüler sowie ein Lehrerhandbuch mit Hintergrundwissen, didaktischer Einordnung, Lösungshinweisen, Zeitplanung und Differenzierung. Wer mehrere Kapitel kombiniert, profitiert von der inhaltlichen Progression; wer ein Kapitel herauslöst, findet alle nötigen Erklärungen im jeweiligen Lehrerhandbuch.

Aufbau der Kapitel

Die Arbeitsblätter folgen einem wiederkehrenden Grundmuster. Am Anfang steht meist ein kurzer Warm-up-Impuls. Er aktiviert Vorwissen, Alltagserfahrungen oder erste Einschätzungen. Dieser Einstieg sollte nicht nur als schnelle Einstiegsfrage behandelt, sondern am Ende möglichst noch einmal aufgegriffen werden. Gerade bei technisch-kritischen KI-Themen ist es didaktisch ergiebig, wenn Schülerinnen und Schüler merken, ob und wodurch sich ihre anfängliche Einschätzung verändert hat.

Danach führt eine kurze Einleitung in das Thema ein. Sie soll keine vollständige Sachanalyse liefern, sondern den Problemraum öffnen. Im Zentrum steht in der Regel eine Fallgeschichte, ein Dialog oder ein fiktives Szenario. Diese Texte sind bewusst alltagsnah formuliert. Sie ermöglichen eine Diskussion über KI in Schule, Medizin, Verwaltung, Stadtplanung, Empfehlungssystemen oder Medienproduktion, ohne dass Schülerinnen und Schüler persönliche Erfahrungen offenlegen müssen.

An die Fallgeschichte schließen Analysefragen an. Sie sichern zunächst das Textverständnis und führen dann zur eigentlichen Urteilsbildung. Die fachlichen Infoboxen stehen deshalb meist nicht vor der ersten Analyse. Die Schülerinnen und Schüler sollen zunächst selbst herausarbeiten, worin das Problem liegt. Erst danach werden Begriffe, Mechanismen und Hintergründe systematisiert.

Der Reflexionsteil führt über den Einzelfall hinaus. Hier geht es um Transfer, eigene Positionierung, gesellschaftliche Folgen und verantwortliches Handeln. Den Abschluss bildet je nach Kapitel eine Sicherungsaufgabe: ein Lückentext, eine Zuordnung, eine Fallbewertung, eine Regelaufgabe oder ein Mini-Fazit. Diese Aufgaben dienen nicht nur der Wiederholung, sondern sollen zentrale Begriffe für spätere Diskussionen verfügbar machen.

Das Lehrerhandbuch liefert die fachliche Absicherung. Die Lösungshinweise sind bewusst nicht als starre Musterlösungen formuliert. Bei technischen KI-Fragen geht es häufig um begründete Einordnungen, nicht um eindeutige Antworten. Deshalb zeigen die Hinweise, worauf tragfähige Antworten achten sollten, welche Differenzierungen wichtig sind und wo typische Vereinfachungen liegen.

Einsatzformen im Unterricht

Die Materialien sind grundsätzlich für 90-minütige Doppelstunden gedacht. Diese Zeit reicht in vielen Fällen für Einstieg, Fallgeschichte, Analyse, zentrale Infoboxen, Reflexion und Sicherung. Einzelne Kapitel lassen sich aber auch gut auf zwei bis drei Unterrichtsstunden verteilen. Das gilt besonders für Kapitel mit rechnerischen Anteilen, etwa zur Klassifikation, Präzision und Trefferquote oder zu Empfehlungsmodellen.

Die im Lehrerhandbuch genannten Zeitangaben sind Orientierungen. Sie sollen bei der Planung helfen, aber keine minutiöse Durchführung erzwingen. Lesezeit, Gesprächsbedarf und begriffliche Klärung werden leicht unterschätzt. Wenn eine Lerngruppe lebhaft diskutiert, ist es meist sinnvoller, einzelne Fragen gründlich zu bearbeiten, statt das Arbeitsblatt vollständig abzuarbeiten.

Für eine 45-Minuten-Fassung empfiehlt sich eine klare Reduktion: Warm-up, Fallgeschichte, zwei bis drei zentrale Analysefragen, eine Infobox und eine kurze Sicherung. Reflexionsfragen, Bonusaufgaben oder vertiefende Quellenarbeit können dann als Hausaufgabe, Gruppenauftrag oder Anschlussstunde genutzt werden.

Die Kapitel eignen sich auch für Projekttag, Methodentage oder Stationenarbeit. Dabei können verschiedene Gruppen unterschiedliche KI-Systemarten oder -aspekte bearbeiten und ihre Ergebnisse anschließend zusammenführen. Eine Gruppe untersucht beispielsweise Klassifikationsfehler, eine andere Empfehlungslogik, eine dritte generative KI, eine vierte Bias-Entstehung, eine fünfte Verantwortungsketten. Wichtig ist dann eine gemeinsame Auswertung, in der die Gruppen nicht nur Inhalte vorstellen, sondern ihre Kriterien, Abwägungen und offenen Fragen vergleichen.

Die Kapitel im Überblick

Kapitel	Thema	Schwerpunkt
1	Was ist KI eigentlich?	Grenzfälle, Abgrenzung, Leitfrage „Was tut das System genau?“
2	Daten zeigen nur Ausschnitte	Datenqualität, Repräsentativität, Label, Proxy
3	Wie Maschinen lernen	Optimierung, Fehlersignal, Mini-Modell als Brücke
4	Vom Lernen zum Anwenden	Training, Inferenz, Modellgrenzen
5	Woraus Sprachmodelle Antworten	Tokens, Kontext, Wahrscheinlichkeit, Answererzeugung
6	Wenn KI überzeugend irrt	Halluzinationen, Prüfprotokoll, Fehlertypen
7	Wenn Zustimmung belohnt wird	Sycophancy, Bestätigungsdynamik, Promptformulierung
8	Gute Fragen an KI	Promptstruktur, Prüfauftrag, Grenzen setzen
9	Chatbot oder Quelle?	Quellenprüfung, Belegfunktion, Sprachmodell vs.
10	KI als Klassifikator	Klassen, Scores, Unsicherheit, Fehlerarten
11	KI als Prognosemaschine	Wahrscheinlichkeit, Zielgröße, Rückkopplung
12	Die Logik von Empfehlungen	Ranking, Signale, Zielkonflikte, Rückkopplungsschleifen
13	KI, die Neues erzeugt	Generierung, Muster, Prüfung, Kennzeichnung
14	KI kritisch prüfen	Testdaten, Fehlerarten, Präzision, Fairnessprüfung
15	Wo Bias entsteht	Bias-Kette, Proxies, Teilgruppenprüfung, Maßnahmen
16	Wer verantwortlich bleibt	Verantwortungskette, menschliche Kontrolle, Scheinkontrolle

Die Tabelle dient der schnellen Auswahl. Die Kapitel sind nicht als Schwierigkeitsprogression gemeint, auch wenn der Band inhaltlich von Grundbegriffen zu komplexeren Systemfragen fortschreitet. Für einen niedrigschwelligen Einstieg eignen sich besonders die Kapitel zu Daten, Sprachmodellen und Halluzinationen. Für Lerngruppen mit mathematischem Interesse bieten sich Klassifikation, Prognose und Evaluation an. Für stärker gesellschaftlich ausgerichtete Diskussionen eignen sich Bias, Empfehlungssysteme und Verantwortung. Das Kapitel zu generativer KI kann gut fächerübergreifend genutzt werden, etwa mit Deutsch, Kunst, Musik oder Medienbildung.

Arbeitsblätter anpassen

Die Arbeitsblätter liegen als editierbare Dateien zum Download bereit, Sie finden den dazu notwendigen Webcode am Anfang jedes Lehrerhandbuchs. Sie dürfen und sollen an die jeweilige Lerngruppe angepasst werden. Sinnvolle Anpassungen sind zum Beispiel:

- Fallgeschichten kürzen,
- Analysefragen reduzieren,
- Infoboxen sprachlich vereinfachen,
- Lückentexte durch andere Sicherungsformen ersetzen,
- aktuelle Beispiele ergänzen,
- schulische Regeln oder landesspezifische Vorgaben einarbeiten,
- Aufgaben stärker fachbezogen ausrichten,
- zusätzliche Rechen- oder Transferaufgaben für leistungsstarke Gruppen ergänzen.

Gerade in Band 3 kann Kürzung sinnvoll sein. Einige Kapitel haben einen rechnerischen oder analytischen Kern, der je nach Vorkenntnissen und verfügbarer Zeit unterschiedlich tief bearbeitet werden kann. Es ist nicht sinnvoll, alle Teilschritte in einer Stunde vollständig abhandeln zu wollen. Besser ist eine klare Schwerpunktsetzung: ein Szenario, *ein* zentraler Begriff, *eine* Analysefrage, *eine* begründete Einordnung.

Der didaktische Kern sollte dabei erhalten bleiben. Ein Kapitel zur Klassifikation sollte nicht auf „KI ordnet Dinge ein“ reduziert werden, sondern Unsicherheit, Fehlerarten und Kontextabhängigkeit einbeziehen. Ein Kapitel zur Prognose sollte nicht nur fragen, ob KI richtig liegt, sondern zwischen Wahrscheinlichkeit, Entscheidung und Verantwortung unterscheiden. Ein Kapitel zu Empfehlungen sollte nicht nur zeigen, wie Algorithmen funktionieren, sondern Zielkonflikte, Rückkopplungen und Interessenlagen benennen. Ein Kapitel zu Bias sollte nicht bei der Feststellung „Daten können schief sein“ stehenbleiben, sondern die Entstehungsstellen entlang der Verarbeitungskette unterscheiden.

Umgang mit Quellen

Die Quellen auf den Arbeitsblättern sind knapp gehalten. Im Lehrerhandbuch werden sie ausführlicher angegeben und fachlich eingeordnet. Sie dienen der Transparenz und ermöglichen Vertiefung, müssen aber nicht in jeder Lerngruppe aktiv bearbeitet werden.

In Band 3 stammen die Quellen überwiegend aus technisch-wissenschaftlichen Publikationen, KI-Risikodokumenten internationaler Institutionen und Bildungsressourcen zu KI-Literacy. Solche Quellen sind verlässlicher als populärwissenschaftliche Darstellungen, aber nicht voraussetzungslos. Lehrkräfte können die Materialien deshalb auch nutzen, um Quellenbewusstsein einzuüben: Wer veröffentlicht? Mit welchem Zweck? Was wird gemessen, und was bleibt außen vor?

Wenn Arbeitsblätter verändert werden, sollten auch die Quellen geprüft werden. Entfernte Inhalte brauchen keine Quellen mehr. Neu eingefügte Behauptungen benötigen gegebenenfalls neue Belege. Bei rechtlichen oder datenschutzbezogenen Fragen ersetzt das Material keine Rechtsberatung und keine schulinterne Dienstanweisung.

Hinweise zu besonderen Unterrichtssituationen

Band 3 behandelt überwiegend technische und methodische Fragen. Direkt persönlich berührende Themen wie Krankheit, Arbeitsplatzverlust oder politische Verfolgung stehen nicht im Mittelpunkt. Dennoch können einzelne Kapitel Reaktionen auslösen, die didaktische Aufmerksamkeit brauchen.

Kapitel zu Bias, Empfehlungssystemen und Verantwortung können eigene Erfahrungen mit algorithmischen Entscheidungen ansprechen: Wer hat schon einmal das Gefühl gehabt, von einem System nicht richtig gesehen zu werden? Solche Erfahrungen können Diskussionen vertiefen, dürfen aber nicht eingefordert werden. Die Fallgeschichten im Buch bieten bewusst eine Schutzdistanz. Man kann über Figuren und Szenarien aus dem Material sprechen, ohne sofort über einzelne Personen in der Klasse zu sprechen.

Kapitel zu Halluzinationen und Sycophancy können bei Schülerinnen und Schülern, die KI-Systeme intensiv nutzen, Verunsicherung auslösen. Diese Verunsicherung ist produktiv und gehört zum Lernziel. Der Unterricht sollte aber nicht bei der Botschaft „KI lügt immer“ stehenbleiben, sondern zu einer differenzierten Haltung führen: Prüfen ist Kompetenz, nicht Misstrauen.

Bei stark emotionalen Diskussionen empfiehlt es sich, zur Sachebene zurückzukehren: Was ist im Fall beschrieben? Welche Informationen fehlen? Welche Akteure kommen vor? Welche Entscheidung steht zur Diskussion? So bleibt der Unterricht urteilsorientiert und verliert sich nicht in Technikeuphorie oder Technikverdrossenheit.

Kombinationen von Kapiteln

Nicht jede Lerngruppe muss alle Kapitel bearbeiten. Besonders naheliegende Kombinationen sind:

Sprachmodell-Schwächen (Kapitel 6 + 7): Halluzinationen und Sycophancy zeigen gemeinsam, warum sprachliche Überzeugungskraft kein Qualitätsmerkmal ist. Die beiden Kapitel ergänzen sich, weil sie dieselbe Grundfrage aus zwei Richtungen stellen: Kapitel 6 vom Fehler her, Kapitel 7 von der Dynamik der Bestätigung her.

Technischer Grundblock (Kapitel 1–4): Diese vier Kapitel bilden zusammen eine kompakte Einführung in Funktionsweise und Grenzen von KI – von Begriff über Daten und Optimierung bis zur Anwendung. Sie eignen sich als Einstiegsreihe für Lerngruppen ohne KI-Vorkenntnisse.

Systemische Effekte (Kapitel 11 + 12): Prognose und Empfehlung zeigen beide, wie Rückkopplungen entstehen und warum Zielgrößen die Ausgabe formen. In Kombination wird sichtbar, dass derselbe technische Mechanismus in unterschiedlichen Kontexten unterschiedliche gesellschaftliche Konsequenzen hat.

Daten und Bias (Kapitel 2 + 15): Was Daten nicht zeigen und warum das zu systematischen Verzerrungen führt – dieser Zusammenhang kann in Kapitel 2 eingeführt und in Kapitel 15 vertieft werden.

Fehlerarten quer (Kapitel 6 + 10 + 14): Halluzinationen bei Sprachmodellen, Fehlklassifikationen und Fairnessprüfung zeigen denselben Grundgedanken in drei verschiedenen Systemtypen: Wie entsteht ein Fehler, welche Folgen hat er, und wie prüft man ihn?

Prüfen und verantworten (Kapitel 14 + 15 + 16): Der Schlussblock führt Evaluation, Bias und Verantwortung als zusammenhängendes Kompetenzfeld zusammen. Diese Kombination eignet sich besonders für einen reflektierenden Abschluss einer KI-Einheit oder eines Projekttags.

Solche Kombinationen eignen sich besonders für Unterrichtsreihen, Projektstage oder fächerübergreifende Arbeit. Sie zeigen, dass sich dieselben Grundfragen – Woran erkenne ich Qualität? Wer wird durch ein System übersehen? Wer verantwortet die Folgen? – in unterschiedlichen KI-Systemtypen wiederholen, ohne identisch zu sein.

Digitale und analoge Nutzung

Die Arbeitsblätter können ausgedruckt, als PDF verteilt oder digital in Lernplattformen genutzt werden. Für Einzelarbeit, Partnerarbeit und Gruppenarbeit eignen sie sich gleichermaßen, wenn die Sozialform klar vorgegeben wird. Bei längeren Reflexionsfragen ist es oft sinnvoller, nicht alle Antworten direkt auf dem Arbeitsblatt einzutragen, sondern im Heft, in einem digitalen Dokument oder auf einem separaten Blatt zu arbeiten.

Bei Kapiteln mit Tabellen und Berechnungen ist analoge Bearbeitung oft vorzuziehen, weil das Auszählen, Einzeichnen und Vergleichen kleinschrittiger erfolgen kann. Bei Kapiteln mit Diskussionsschwerpunkt kann digitale Kollaboration – etwa gemeinsame Dokumentenbearbeitung oder Pinnwände – den Austausch erleichtern.

Worum es in diesem Band geht

Dieser Band behandelt die Funktionsweise von KI-Systemen: wie sie lernen, wie sie Antworten erzeugen, wie sie klassifizieren, vorhersagen, empfehlen und bewerten – und wo dabei Fehler, Verzerrungen und Verantwortungslücken entstehen können.

Das Ziel ist kein technisches Expertenwissen. Es geht um Grundverständnis: Was passiert, wenn ein System mit Daten trainiert wird? Warum erzeugt ein Sprachmodell manchmal plausible, aber falsche Antworten? Warum führt „gleiche Berechnung für alle“ nicht automatisch zu fairem Ergebnis? Wer ist verantwortlich, wenn ein KI-System einen Fehler macht?

Wer diese Fragen beantworten kann, kann KI-Systeme nicht nur nutzen, sondern einordnen, hinterfragen und begründet beurteilen.

Aufbau des Bandes

Der Band ist in vier inhaltliche Blöcke gegliedert.

Block 1 – Begriff, Daten und Optimierung (Kapitel 1–4) Was ist KI? Was sind Daten, und was zeigen sie nicht? Wie lernen Systeme durch Optimierung? Was ist der Unterschied zwischen Training und Anwendung?

Block 2 – Sprachmodelle (Kapitel 5–9) Wie erzeugen Sprachmodelle Antworten? Warum können sie überzeugend irren? Was ist Sycophancy? Wie formuliert man gute Aufträge, und wie unterscheidet man Sprachmodell-Ausgaben von Quellen?

Block 3 – KI jenseits des Chatbots (Kapitel 10–13) Wie klassifizieren, prognostizieren, empfehlen und erzeugen KI-Systeme? Was sind Fehlerarten, Zielgrößen, Rückkopplungen und Proxies?

Block 4 – Prüfen und verantworten (Kapitel 14–16) Wie prüft man KI-Systeme methodisch? Wo entsteht Bias? Wer trägt Verantwortung – und woran erkennt man echte menschliche Kontrolle?

Die Grundprinzipien des Bandes

Die folgenden fünf Prinzipien tragen alle 16 Kapitel. Sie werden nicht in jedem Kapitel vollständig erklärt, aber jedes Kapitel baut auf ihnen auf. Für Lehrkräfte, die den Band als Ganzes einsetzen, genügt es, diese Prinzipien einmal zu kennen. Für Lehrkräfte, die einzelne Kapitel herausnehmen, bieten die jeweiligen Lehrerhandbücher die nötige kapitelspezifische Einordnung.

1. KI-Systeme verstehen ihren Output nicht. Ein Sprachmodell erzeugt Texte als statistisch plausible Fortsetzungen. Es berechnet keinen Sachverhalt, überprüft keine Fakten und hat keine Absichten. Das gilt auch dann, wenn die Ausgabe kompetent, höflich oder überzeugend klingt. Dieses Prinzip schützt vor Vermenschlichung und vor unkritischer Übernahme von KI-Ausgaben.

2. Daten bilden nur Ausschnitte der Wirklichkeit ab. Trainingsdaten sind nie vollständig, nie neutral und nie ohne Entstehungsgeschichte. Sie zeigen, was in bestimmten Kontexten sichtbar war – nicht, was tatsächlich vorhanden ist oder was hätte sichtbar sein sollen. Wer Daten als objektive Grundlage behandelt, übersieht systematisch, was sie nicht enthalten.

3. Systeme optimieren das Messbare – nicht das Gemeinte. Jedes KI-System wird auf eine Zielgröße hin trainiert oder bewertet. Diese Zielgröße ist häufig ein Proxy: ein messbares Signal, das für etwas steht, das man eigentlich will. Klicks stehen für Interesse. Teilnahme steht für Motivation. Bearbeitungszeit steht für Lernen. Solange der Proxy nicht mit dem Gemeinten verwechselt wird, ist er nützlich. Wird er verwechselt, entstehen systematische Verzerrungen.

4. KI-Ausgaben müssen geprüft werden. Keine KI-Ausgabe darf blind übernommen werden – unabhängig davon, wie sicher oder kompetent sie wirkt. Prüfen bedeutet: Quelle suchen, Kontext beachten, Fehlerfolgen einschätzen, bei Entscheidungen mit Tragweite Fachkompetenz hinzuziehen. Prüfen ist keine Schwäche gegenüber der Technik, sondern Grundbedingung verantwortlicher Nutzung.

5. Verantwortung bleibt bei Menschen und Institutionen. Ein KI-System kann Empfehlungen, Prognosen, Klassifikationen und Texte erzeugen. Es kann keine Verantwortung übernehmen. Verantwortlich sind die Menschen und Organisationen, die ein System entwickeln, auswählen, einsetzen, kontrollieren und korrigieren. Automatisierung verteilt Verantwortung, sie hebt sie nicht auf.

Für den Einzeleinsatz

Alle Arbeitsblätter dieses Bandes sind so angelegt, dass sie auch ohne Vorlaufkenntnisse aus anderen Kapiteln einsetzbar sind. Fachbegriffe werden eingeführt, bevor sie verwendet werden. Fallbeispiele tragen die nötige Sachlogik in sich.

Empfehlenswert für die Unterrichtsvorbereitung: die fünf Grundprinzipien oben einmal lesen. Sie helfen einzuschätzen, welcher Grundgedanke hinter dem jeweiligen Kapitel steht – ohne dass das gesamte Material bekannt sein muss.

Wenn Schülerinnen und Schüler im Verlauf eines Kapitels auf ein Prinzip stoßen, das in einem anderen Kapitel vertieft wird, ist das ein Gewinn, aber keine Voraussetzung. Die Kapitel sind als eigenständige Doppelstunden konzipiert.

Zur Materialreihe

Die Materialreihe „KI und Gesellschaft“ besteht aus drei Bänden.

Band 1 behandelt KI im Schulalltag und im direkten Lebensumfeld von Jugendlichen: Sprachmodelle als Werkzeug, Datenschutz, Feedback und Quellenprüfung.

Band 2 behandelt gesellschaftliche und politische Dimensionen: Überwachung, Demokratie, Deepfakes, Empfehlungssysteme und Zukunft der Arbeit.

Band 3 (dieser Band) behandelt die technischen Grundlagen: Wie KI-Systeme funktionieren, wo Fehler entstehen und wer Verantwortung trägt.

Die Bände sind unabhängig voneinander einsetzbar. Band 3 kann auch dann genutzt werden, wenn Band 1 und Band 2 nicht bekannt sind.