

AF · FOUNDATION

AI Foundations Professional

Teilnehmerhandbuch mit Beispielen, Übungen und Praxisgrafiken

Foundation

Associate

Professional

Expert

Umfang	8 Module · ca. 12 Stunden
Prüfung	20 Fragen · 40 Minuten · Bestehensgrenze 70 %
Voraussetzungen	keine
Rechtsstand	August 2026
Abbildungen	33 Grafiken · 16 Übungen · 52 Testfragen

CertifAI · Version 2.0 · Ausgabe DE · erweiterte und illustrierte Fassung

Dieses Handbuch bereitet auf die Zertifizierungsprüfung AI Foundations Professional vor. Die Teilnahme an diesem Kurs ist keine Voraussetzung für die Prüfung; gleichwertige Vorbereitungswege werden anerkannt.

Hinweise zu diesem Handbuch

Bitte vor Beginn lesen

Was in dieser Ausgabe neu ist

Diese zweite Ausgabe erweitert die Theorie in allen acht Modulen, ergänzt 33 eigens erstellte Abbildungen, belegt zentrale Aussagen mit Primärquellen und bildet den Rechtsstand nach der Änderungsverordnung vom 27. Juli 2026 ab. Neu hinzugekommen sind unter anderem: die Begriffsleiter von KI bis Agent (1.3), Werkzeuge und Agenten (2.6), die Ursachen von Halluzinationen (3.3), der Eignungs-Schnelltest und der Umbau-Katalog (4.4/4.5), die Unterscheidung der Zugangsarten (5.3), vier Prüfstufen (6.3), wirksame gegenüber dekorativer Aufsicht (7.4) und der Vorhaben-Steckbrief (8.7). Die Zahl der Übungen ist von acht auf sechzehn gestiegen, die der Testfragen von 36 auf 52.

Art der Qualifikation

CertifAI ist ein unabhängiger Anbieter. Das Zertifikat *AI Foundations Professional* ist ein Industriezertifikat und keine staatlich anerkannte Qualifikation. Es wird von CertifAI ausgestellt und ist nicht behördlich zugelassen. Es verleiht keine akademischen Grade, keine ECTS-Punkte und keine berufsrechtliche Erlaubnis.

Keine Rechtsberatung

Modul 8 gibt einen Überblick über rechtliche Rahmenbedingungen. Dieser Überblick ersetzt keine Rechtsberatung. Rechtsfragen im Einzelfall gehören zu einer Person, die dazu beraten darf. Der Rechtsstand dieses Handbuchs ist **August 2026**; das Rechtsgebiet bewegt sich schnell, und einzelne Fristen haben sich in den zwölf Monaten vor Drucklegung mehrfach verschoben.

Einsatz von KI bei der Erstellung

Teile dieses Handbuchs wurden KI-gestützt erstellt und anschließend fachlich geprüft und freigegeben. Alle statistischen Angaben wurden gegen die in Anhang D genannten Primärquellen verifiziert. Die Prüfungsfragen der Zertifizierungsprüfung werden unabhängig vom Kursmaterial entwickelt und von Fachexperten freigegeben. Bei der Bewertung der Prüfung kommt keine KI zum Einsatz: die Auswertung erfolgt deterministisch gegen einen von Menschen freigegebenen Antwortschlüssel.

So arbeitest du mit diesem Handbuch

Jedes Modul folgt derselben Struktur: Lernziele, Text mit Abbildungen, mindestens zwei durchgerechnete Beispiele, zwei Übungen und ein Selbsttest mit vier Fragen. Die Übungen sind der Teil, der wirkt — überspringe sie nicht. Die Lösungen stehen in Anhang A, absichtlich weit hinten. Bearbeite eine Übung vollständig, bevor du nachschlägst; eine gelesene Lösung fühlt sich richtig an, ohne dass du sie könntest.

Am Ende findest du eine Musterprüfung mit 20 Fragen im Format der echten Prüfung. Nimm sie dir unter realen Bedingungen vor: 40 Minuten, keine Hilfsmittel, keine Unterbrechung.

WAS DU NACH DIESEM KURS KANNST

Du kannst erklären, was diese Systeme tun und wo ihre Grenzen liegen; du erkennst, wann eine Aufgabe für KI geeignet ist und wann nicht, und du kannst eine ungeeignete Aufgabe so umbauen, dass sie geeignet wird; du weißt, welche Daten nie in ein Eingabefeld gehören und welcher Zugang was erlaubt; du kannst ein Ergebnis angemessen prüfen und den Aufwand an den Folgen ausrichten; und du kennst die Rahmenbedingungen gut genug, um die richtigen Fragen zu stellen. Was dieser Kurs nicht leistet: dich zum Entwickler, zum Datenschutzbeauftragten oder zum Prompt-Spezialisten machen. Dafür gibt es die Spezialisierungen.

Inhalt

Modul 1	Was KI heute ist – und was nicht	5
Modul 2	Wie Sprachmodelle arbeiten, ohne Mathematik	12
Modul 3	Warum Modelle überzeugend falsch liegen	19
Modul 4	Geeignete und ungeeignete Aufgaben	26
Modul 5	Daten, Vertraulichkeit und Geheimnisse	32
Modul 6	Ergebnisse bewerten und prüfen	38
Modul 7	Verantwortung, Aufsicht und Kennzeichnung	44
Modul 8	Der rechtliche Rahmen in Grundzügen	50
Musterprüfung	20 Fragen im Prüfungsformat	57
Anhang A	Lösungen zu Übungen, Selbsttests und Musterprüfung	60
Anhang B	Glossar	68
Anhang C	Prüfungsinformationen	71
Anhang D	Zahlen und Quellen	73

Was KI heute ist – und was nicht

ca. 75 Minuten · 2 Beispiele · 2 Übungen · Selbsttest am Ende

LERNZIELE

- Du kannst die vier Aufgabentypen unterscheiden, die praktisch alle KI-Systeme im Unternehmen abdecken.
- Du erkennst den Unterschied zwischen regelbasierten und gelernten Systemen und weißt, warum er für die Fehlersuche zählt.
- Du kannst KI, maschinelles Lernen, Deep Learning und generative KI korrekt zueinander in Beziehung setzen.
- Du kannst in einem Satz erklären, was ein generatives Sprachmodell im Kern tut.
- Du kannst drei verbreitete Erwartungen benennen, die auf einem Missverständnis beruhen.

Dieses Modul räumt auf. Der Begriff „KI“ steht im Arbeitsalltag für vier grundverschiedene Dinge, und wer sie nicht auseinanderhält, diskutiert über Risiken, die es nicht gibt, und übersieht die, die es gibt. Am Ende kannst du in jeder Besprechung sagen, welche Art von System gemeint ist und wo bei genau dieser Art der Fehler sitzt.

1.1 Ein Sammelbegriff, kein System

„KI“ ist kein Produkt, sondern ein Etikett für eine Familie sehr unterschiedlicher Verfahren. Der Spamfilter im Postfach, die Bedarfsprognose im Lager und der Textassistent im Browser haben technisch fast nichts gemeinsam — außer dass alle drei aus Daten gelernt haben statt ausprogrammiert worden zu sein. Wer über „die KI“ spricht, redet deshalb regelmäßig aneinander vorbei.

Für die Praxis reicht eine grobe Einteilung nach dem, was ein System *tut*:

TYP	WAS ES LEISTET	TYPISCHES BEISPIEL	FEHLER FÄLLT AUF
Klassifikation	ordnet etwas einer von mehreren Kategorien zu	Spam / kein Spam; Rechnung / Lieferschein / Mahnung	meist ja
Vorhersage	schätzt einen Wert für die Zukunft	Absatz der nächsten Woche; Ausfallrisiko einer Maschine	messbar
Generierung	erzeugt neuen Inhalt: Text, Bild, Code, Audio	Textassistent, Bildgenerator, Code-Vorschlag	nur bei Prüfung
Suche & Empfehlung	findet Ähnliches oder Passendes	semantische Dokumentensuche; „das könnte Sie interessieren“	nur bei Stichprobe

Diese Einteilung ist nicht akademisch, sie ist diagnostisch. Ein Klassifikationssystem irrt anders als ein Generierungssystem: Der Spamfilter schiebt eine wichtige Mail in den Müll — ein Fehler, den man bemerkt und korrigieren kann. Ein Textassistent erfindet eine Quelle — ein Fehler, den man ohne Prüfung nicht bemerkt. Die zweite Sorte ist die gefährlichere, und der ganze Rest dieses Kurses beschäftigt sich überwiegend mit ihr.



Abb. 1 Wo der Prüfaufwand hingehört. Nicht die Fehlerhäufigkeit bestimmt das Risiko, sondern die Kombination aus Sichtbarkeit und Reichweite.

1.2 Ausprogrammiert oder gelernt

Ein **regelbasiertes** System tut, was jemand hineingeschrieben hat: *wenn Betrag über 10.000, dann zweite Freigabe*. Es ist vollständig nachvollziehbar, es irrt nur, wenn die Regel falsch ist, und man kann die Regel lesen.

Ein **gelerntes** System hat aus Beispielen ein Muster abgeleitet. Niemand hat die Regel geschrieben, niemand kann sie vollständig lesen. Das System ist oft besser als jede Regel, die ein Mensch formuliert hätte — und es ist genau deshalb schwerer zu prüfen. Wenn es sich falsch verhält, lautet die Antwort nicht „welche Regel ist kaputt?“, sondern „welche Daten haben es dorthin geführt?“.

REGELBASIERT	GELERNT
● Regel ist lesbar und änderbar ● Fehler ist lokalisierbar ● Verhalten ist reproduzierbar ● scheitert an Vielfalt und Ausnahmen ● Prüffrage: Wer hat die Regel geschrieben?	● Muster ist nicht vollständig lesbar ● Fehler ist nur statistisch fassbar ● Verhalten kann streuen ● kommt mit Vielfalt und Ausnahmen zurecht ● Prüffrage: Woran wurde trainiert?

Abb. 2 Zwei Systemarten, zwei Fehlersuchen. Der Unterschied entscheidet, wen du fragst, wenn etwas schiefgeht.

DER PRAKTISCHE UNTERSCHIED

Bei einem regelbasierten Fehler fragst du: *Wer hat die Regel geschrieben?* Bei einem gelernten Fehler fragst du: *Woran wurde trainiert, und ist mein Fall darin überhaupt vorgekommen?* Ein Modell, das überwiegend auf englischsprachigen Standardverträgen gelernt hat, wird bei einem österreichischen Werkvertrag nicht warnen, dass es sich unsicher ist. Es antwortet gleich souverän.

1.3 Vier Begriffe, die ständig verwechselt werden

In Besprechungen fallen vier Wörter, die nicht dasselbe meinen. Sie liegen ineinander wie Schalen:

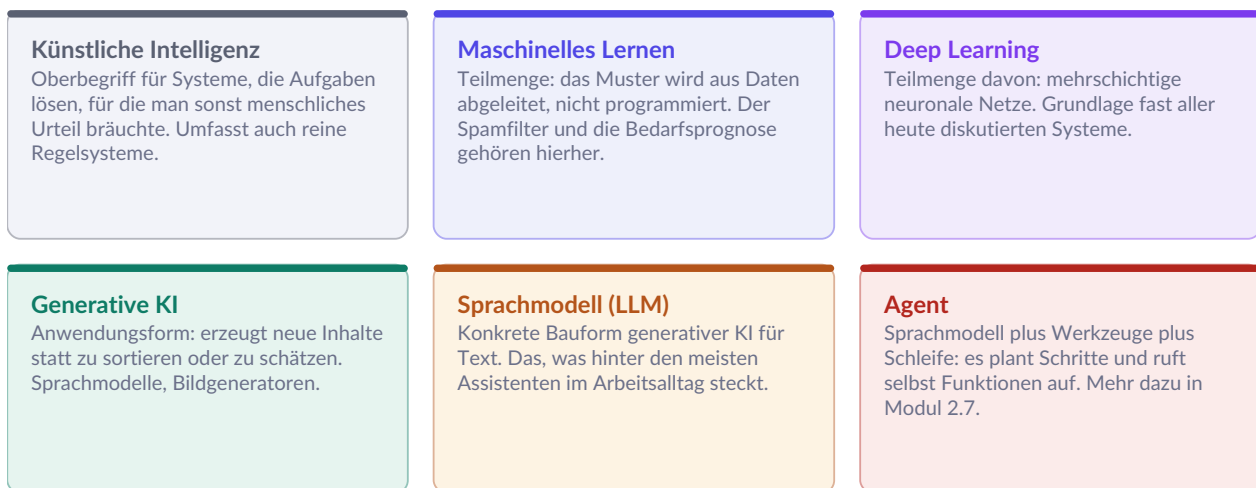


Abb. 3 Von außen nach innen: jede Stufe ist eine Teilmenge der vorhergehenden. „Wir setzen KI ein“ sagt fast nichts; „wir setzen ein Sprachmodell für Entwürfe ein“ sagt fast alles.

1.4 Wo KI im Unternehmen tatsächlich ankommt

Der Sprung der letzten beiden Jahre ist keine Wahrnehmungstäuschung. In der repräsentativen Bitkom-Befragung von 604 Unternehmen ab 20 Beschäftigten gaben Anfang 2026 **41 Prozent** an, KI einzusetzen — nach 17 Prozent im Vorjahr. Weitere 48 Prozent planen oder diskutieren den Einsatz.^[1]



Nur die beiden erhobenen Kategorien sind dargestellt; sie addieren sich nicht auf 100 %.

Abb. 4 KI-Einsatz in deutschen Unternehmen ab 20 Beschäftigten.

Quelle: Bitkom Research, repräsentative Befragung von 604 Unternehmen, KW 2–6/2026

Zwei Zahlen aus derselben Erhebung sind für diesen Kurs wichtiger als die Wachstumszahl. Erstens: 77 Prozent der einsetzenden Unternehmen berichten eine verbesserte Wettbewerbsposition, aber nur 52 Prozent können einen *messbaren* Beitrag zum Geschäftserfolg benennen.^[1] Zwischen „fühlt sich besser an“ und „ist nachweisbar“ liegt genau die Prüfdisziplin, um die es in Modul 6 geht. Zweitens nennen Unternehmen als größte Hemmnisse Rechtsunsicherheit (53 %), fehlendes Fachwissen (53 %) und fehlende Personalressourcen (51 %).^[3] Zwei der drei Hemmnisse adressiert dieser Kurs direkt.

1.5 Was ein Sprachmodell im Kern tut

Ein generatives Sprachmodell setzt Text fort. Es schätzt bei jedem Schritt, welche Fortsetzung angesichts des bisherigen Textes plausibel ist, wählt eine, und beginnt von vorn. Das ist keine Vereinfachung für Anfänger, das ist die Sache selbst.

Daraus folgt mehr, als es zunächst scheint. Das Modell schlägt nichts nach — es hat keine Datenbank, in der die Antwort steht. Es rechnet nicht im eigentlichen Sinn, sondern setzt Ziffernfolgen fort, die bei einfachen Aufgaben meist richtig aussehen. Es weiß nicht, welcher Tag heute ist. Und es unterscheidet nicht zwischen einer Aussage, die es tausendfach gesehen hat, und einer, die gerade erst entsteht: Beide kommen im selben Tonfall heraus.

Moderne Systeme umgehen einige dieser Grenzen, indem sie Werkzeuge einsetzen — eine Suche, einen Taschenrechner, einen Kalender. Das ist ein wichtiger Fortschritt, ändert aber nichts am Kern: Ohne Werkzeug bleibt es Fortsetzung. Und ob im konkreten Fall ein Werkzeug lief, siehst du der Antwort meist nicht an.

1.6 Drei Erwartungen, die auf einem Missverständnis beruhen

Fast alle Enttäuschungen im ersten Jahr eines KI-Vorhabens lassen sich auf drei Annahmen zurückführen. Jede klingt vernünftig, jede ist falsch.

- 1 „Es kennt unsere Zahlen.“**
Das Modell kennt nur, was im Kontextfenster steht (Modul 2.2) oder was ein angeschlossenes System nachschlägt. Ohne Anbindung sind eure Umsätze für das Modell nicht existent — es erfindet dann eine plausible Größenordnung.
- 2 „Wenn es unsicher ist, sagt es das.“**
Es gibt keine eingebaute Unsicherheitsanzeige. Ein Modell formuliert eine erfundene Fundstelle im selben Tonfall wie das Einmaleins. Diese fehlende Anzeige — nicht die Fehlerquote — ist das eigentliche Problem (Modul 3.1).
- 3 „Es lernt aus unseren Korrekturen.“**
Nach dem Training wird beim einzelnen Aufruf nichts mehr gelernt. Was wie Lernen aussieht, ist meist ein Gedächtnis-Feature, das frühere Notizen wieder ins Kontextfenster legt. Das heißt umgekehrt nicht, dass nichts gespeichert wird — das ist eine andere Frage (Modul 5).

Abb. 5 Die drei häufigsten Fehlannahmen und was tatsächlich passiert.

BEISPIEL 1.1 – DREI SYSTEME IN EINEM UNTERNEHMEN

Ein mittelständischer Zulieferer setzt drei Systeme ein.

Der Rechnungssortierer ordnet eingehende PDFs den Kategorien Rechnung, Lieferschein und Mahnung zu. *Typ: Klassifikation.* Fehlerbild: Eine ungewöhnlich formatierte Rechnung landet als Lieferschein und wird nicht bezahlt. Der Fehler fällt spätestens bei der Mahnung auf — er ist sichtbar und umkehrbar.

Die Bedarfsprognose schätzt den Materialbedarf der Folgeweche. *Typ: Vorhersage.* Fehlerbild: Nach einer Werksferienwoche prognostiziert sie zu niedrig, weil vergleichbare Wochen in den Trainingsdaten selten sind. Der Fehler ist messbar — man vergleicht Prognose und Ist.

Der Angebotsassistent entwirft Anschreiben zu Angeboten. *Typ: Generierung.* Fehlerbild: Er nennt eine Lieferzeit von vier Wochen, weil das in ähnlichen Texten oft so stand — im konkreten Fall sind es acht. Der Text liest sich einwandfrei. Wenn niemand prüft, geht die Zusage raus.

Die Lehre: Die ersten beiden Systeme irren in einer Weise, die das Unternehmen bemerkt. Das dritte irrt in einer Weise, die nur auffällt, wenn jemand hinsieht. Der Aufwand für Kontrolle muss deshalb nicht gleich verteilt werden.

BEISPIEL 1.2 – DIESELBE TECHNIK, ZWEI RISIKOPROFILE

Zwei Abteilungen desselben Hauses nutzen denselben Assistenten.

Marketing lässt sich fünf Betreffzeilen für einen Newsletter vorschlagen und wählt eine aus. Fehlerkosten: eine schlechte Betreffzeile. Sichtbarkeit: sofort, ein Mensch liest sie. Prüfbarkeit: vollständig — man muss nur lesen können. Risikoprofil: unkritisch.

Der Vertriebsinnendienst lässt sich aus einer technischen Spezifikation eine Kurzfassung für den Kunden erzeugen und leitet sie weiter. Fehlerkosten: eine falsche technische Zusage. Sichtbarkeit: keine, der Text klingt korrekt. Prüfbarkeit: nur mit Fachkenntnis der Spezifikation. Risikoprofil: kritisch.

Die Lehre: Nicht das Werkzeug bestimmt das Risiko, sondern der Zuschnitt der Aufgabe. Genau darum geht es in Modul 4.

ÜBUNG 1 – SYSTEME EINORDNEN

Ordne jedes System einem der vier Typen zu und notiere in einem Satz, *woran der Fehler auffallen würde.*

1. Ein System schlägt beim Verfassen einer E-Mail den nächsten Halbsatz vor.
2. Ein System entscheidet, welche Bewerbungen in die engere Auswahl kommen.
3. Ein System findet in 40.000 Verträgen die mit einer Schiedsklausel.
4. Ein System schätzt, wann eine Pumpe voraussichtlich ausfällt.
5. Ein System erstellt aus Stichpunkten ein zweiseitiges Protokoll.
6. Ein System markiert Kundenanfragen als „dringend“ oder „normal“.

Zusatzfrage: Bei welchen zwei Systemen wäre ein unentdeckter Fehler am folgenreichsten, und warum?

Lösung: Anhang A, Übung 1 – Systeme einordnen

ÜBUNG 1B – ERWARTUNGEN PRÜFEN

Vier Aussagen aus einer echten Projektbesprechung. Entscheide für jede: **trifft zu**, **trifft nicht zu** oder **kommt darauf an** — und begründe mit einem Begriff aus diesem Modul.

1. „Wir füttern das Modell mit unseren Handbüchern, dann kennt es unsere Produkte.“
2. „Wenn es etwas nicht weiß, wird es schon nachfragen.“
3. „Wir brauchen keine Regeln, das System entscheidet ja objektiv.“
4. „Der Assistent rechnet uns die Marge aus, das ist ja nur Prozentrechnung.“

Lösung: Anhang A, Übung 1b – Erwartungen prüfen

KERNAUSSAGEN

- „KI“ bezeichnet vier praktisch unterscheidbare Aufgabentypen, nicht ein System.
- Gelernte Systeme sind mächtiger und schwerer zu prüfen als ausprogrammierte.
- Ein Sprachmodell setzt Text plausibel fort. Es schlägt nichts nach und weiß nicht, was es nicht weiß.
- Generierende Systeme irren unauffällig. Dort gehört die Kontrolle hin.
- Nicht das Werkzeug bestimmt das Risiko, sondern der Zuschnitt der Aufgabe.

Selbsttest Modul 1

1. Welche Aussage beschreibt ein generatives Sprachmodell am treffendsten?

- a) Es prüft Aussagen gegen geprüfte Quellen, bevor es antwortet.
- b) Es durchsucht eine Wissensdatenbank und gibt den besten Treffer aus.
- c) Es schätzt fortlaufend plausible Fortsetzungen des bisherigen Textes.
- d) Es wendet von Fachleuten formulierte Regeln auf die Eingabe an.

2. Warum erfordern generierende Systeme meist mehr Kontrolle als klassifizierende?

- a) Weil ihre Fehler ohne gezielte Prüfung unauffällig bleiben.
- b) Weil sie häufiger falsch liegen als klassifizierende Systeme.
- c) Weil sie technisch weniger ausgereift sind.
- d) Weil sie mehr Rechenleistung benötigen.

3. Ein Klassifikationssystem sortiert Bewerbungen vor. Warum ist der Fehler hier besonders schwer zu bemerken?

- a) Weil Klassifikation grundsätzlich unzuverlässig ist.
- b) Weil personenbezogene Daten nicht geprüft werden dürfen.
- c) Weil Bewerbungen nicht ins Kontextfenster passen.
- d) Weil die Aussortierten keinen Rückkanal haben.

4. Wie verhalten sich maschinelles Lernen und generative KI zueinander?

- a) Es sind zwei getrennte Verfahren ohne Überschneidung.
- b) Generative KI ist eine Anwendungsform, die auf maschinellem Lernen aufsetzt.
- c) Maschinelles Lernen ist eine Weiterentwicklung generativer KI.
- d) Beide Begriffe bezeichnen dasselbe.

Wie Sprachmodelle arbeiten, ohne Mathematik

ca. 75 Minuten · 2 Beispiele · 2 Übungen · Selbsttest am Ende

LERNZIELE

- Du kannst die Begriffe Token, Kontextfenster, Training und Inferenz in eigenen Worten erklären.
- Du kannst vorhersagen, welche Aufgaben ein Modell ohne Werkzeuge zuverlässig verfehlt.
- Du verstehst, warum dieselbe Frage zweimal verschiedene Antworten erzeugt, und wie man Reproduzierbarkeit herstellt.
- Du kannst in je einem Satz sagen, was Einbettungen und RAG leisten — und wo bei RAG der kritische Punkt liegt.
- Du kannst benennen, was Werkzeuge und Agenten ändern und was sie nicht ändern.

Dieses Modul erklärt die Mechanik — ohne eine einzige Formel. Vier Begriffe reichen, um vorherzusagen, welche Aufgaben ein Sprachmodell zuverlässig löst und welche es systematisch verfehlt: Token, Kontextfenster, Training/Inferenz und Zufall. Wer diese vier hat, muss Grenzen nicht mehr auswendig lernen, sondern kann sie herleiten.

2.1 Token: Text in Stücken

Ein Modell sieht keine Buchstaben und keine Wörter, sondern **Token** — Textstücke, die je nach Sprache zwischen einem Zeichen und einem ganzen Wort liegen. Im Deutschen zerfallen zusammengesetzte Wörter regelmäßig in mehrere Stücke. Das Modell arbeitet mit diesen Stücken, nicht mit dem Wort.



Abb. 6 Wie deutsche Wörter zerfallen. Die Buchstabenebene ist für das Modell nicht sichtbar — es sieht nur die farbigen Klumpen.

Daraus folgt ein ganzer Fehlertyp. Fragen nach Buchstaben — *wie viele „r“ stecken in „Erdbeerkonfitüre“?* — sind für ein System, das das Wort als vier Klumpen sieht, überraschend schwer. Es antwortet trotzdem, und zwar bestimmt. Dasselbe gilt für Silbenzählung, Anagramme, Reimprüfungen und für exakte Zeichenzahlen. Nicht weil das Modell dumm ist, sondern weil es die Ebene, auf der die Frage spielt, gar nicht sieht.

FAUSTREGELN FÜR DEN ALLTAG

- Im Deutschen entspricht 1 Token grob 3 bis 4 Zeichen; 1.000 Token sind etwa 700 Wörter oder anderthalb eng beschriebene Seiten.
- Eine DIN-A4-Seite Fließtext liegt bei rund 600 bis 800 Token.
- Deutsche Texte brauchen etwa 10 bis 30 Prozent mehr Token als die gleiche Aussage auf Englisch — bei Abrechnung nach Token ist das ein Kostenfaktor.
- Zahlen, Aktenzeichen und Codes zerfallen besonders unregelmäßig. Genau dort sind Zeichenoperationen am unzuverlässigsten.

2.2 Das Kontextfenster

Alles, was das Modell bei einer Antwort berücksichtigen kann, muss im **Kontextfenster** stehen: deine Frage, die vorherigen Nachrichten des Gesprächs, angehängte Dokumente, die Systemanweisung. Was nicht hineinpasst, existiert für die Antwort nicht.

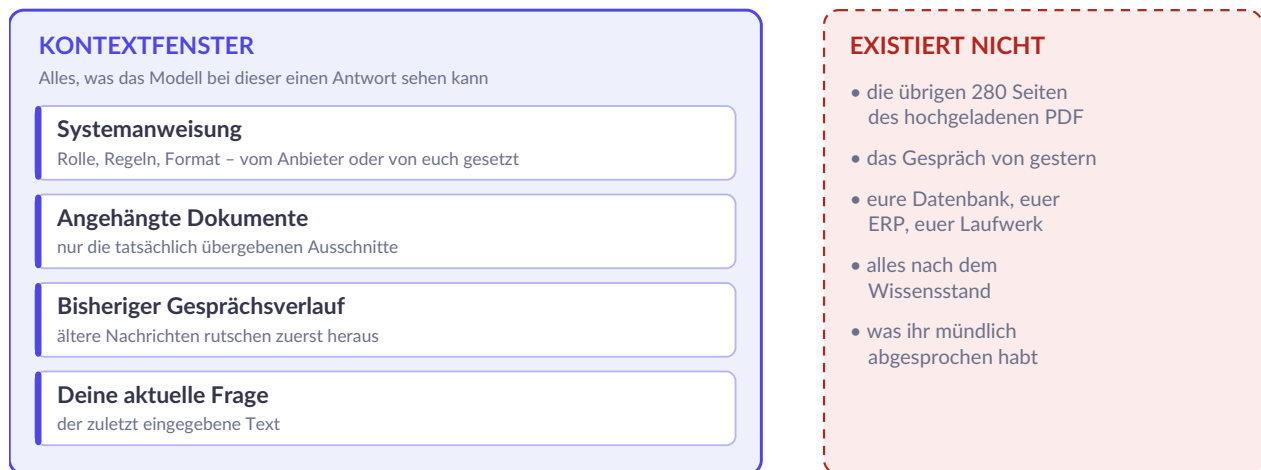


Abb. 7 Innen und außen. Das Modell unterscheidet nicht zwischen „steht nicht drin“ und „gibt es nicht“ — es antwortet mit dem, was da ist.

Das erklärt zwei alltägliche Beobachtungen. Erstens: In langen Gesprächen „vergisst“ das Modell frühere Absprachen — sie sind aus dem Fenster gerutscht oder gehen im Rauschen unter. Zweitens: Ein System, das ein 200-seitiges Handbuch „gelesen“ hat, hat es nicht gelesen. Es hat die Passagen gesehen, die jemand ausgewählt und hineingelegt hat. Wurde die entscheidende Passage nicht ausgewählt, fehlt sie — und das Modell sagt nicht „mir fehlt etwas“, sondern antwortet mit dem, was da ist.

DER SATZ, DER AM MEISTEN ÄRGER SPART

Ein größeres Kontextfenster löst dieses Problem nicht, es verschiebt es. Auch bei sehr großen Fenstern gilt: Information in der Mitte langer Eingaben wird schlechter berücksichtigt als Information am Anfang und am Ende. Wenn eine Angabe tragend ist, gehört sie an eine markierte Stelle deiner Eingabe — nicht auf Seite 84 eines Anhangs.

2.3 Training und Inferenz

Beim **Training** entsteht das Modell aus großen Textmengen. Danach ist sein Wissensstand eingefroren. Bei der **Inferenz** — jedem einzelnen Aufruf im Alltag — wird nichts mehr gelernt.



Abb. 8 Die vier Stationen. Der letzte Kasten ist der, der im Alltag übersehen wird — er hat mit Lernen nichts zu tun und ist trotzdem die relevantere Frage.

Zwei Konsequenzen, die regelmäßig missverstanden werden. Das Modell kennt keine Ereignisse nach seinem Wissensstand; wo es aktuell wirkt, liegt das an einer Suche im Hintergrund, nicht am Modell. Und: Deine Eingaben verändern das Modell nicht — es „merkt sich“ nichts von gestern. Das heißt aber ausdrücklich *nicht*, dass deine Eingaben nirgends gespeichert werden. Sie können protokolliert, ausgewertet und je nach Anbieter und Vertrag auch für künftiges Training verwendet werden. Das sind zwei verschiedene Fragen, und die zweite behandeln wir in Modul 5.

2.4 Warum zwei Antworten verschieden ausfallen

Das Modell wählt seine Fortsetzung nicht immer streng nach der höchsten Wahrscheinlichkeit, sondern mit einer einstellbaren Portion Zufall — oft „Temperatur“ genannt. Deshalb liefert dieselbe Frage zweimal gestellt zwei verschiedene Antworten. Für kreative Aufgaben ist das erwünscht. Für alles, was reproduzierbar sein soll — eine Prüfung, eine Auswertung, eine Entscheidung — ist es ein Problem, das man kennen muss: **Die eine Antwort, die du gesehen hast, ist eine Stichprobe, kein Ergebnis.**

WENN REPRODUZIERBARKEIT GEBRAUCHT WIRD

Reproduzierbarkeit muss man *herstellen*, nicht annehmen. Vier Stellschrauben:

- 1. Eingabe fixieren.** Identischer Wortlaut, identische Anhänge, identische Reihenfolge. Schon ein zusätzliches „bitte“ ändert die Eingabe.
- 2. Streuung senken.** Temperatur auf den niedrigsten sinnvollen Wert; wo verfügbar, einen festen Zufallsstartwert setzen.
- 3. Modellversion festhalten.** „GPT“ ist keine Version. Notiere den genauen Modellnamen samt Datum — Anbieter tauschen Modelle im laufenden Betrieb.
- 4. Mehrfach ziehen.** Wo es auf Stabilität ankommt: dieselbe Frage dreimal stellen. Weichen die Antworten in einem tragenden Punkt ab, ist die Aufgabe für dieses Verfahren nicht geeignet.

2.5 Einbettungen und RAG

Eine **Einbettung** ist die Darstellung eines Textstücks als Zahlenreihe, sodass inhaltlich ähnliche Texte nahe beieinander liegen — die Grundlage dafür, dass eine Suche „Kündigungsfrist“ auch bei „Vertragsende“ anschlägt.

RAG (retrieval-augmented generation) heißt: Bevor das Modell antwortet, sucht ein Schritt davor die passenden Stellen aus deinen Dokumenten heraus und legt sie ins Kontextfenster. Das Modell antwortet dann auf Grundlage dieser Stellen. Das macht Antworten belegbar — und verlagert den kritischen Punkt.

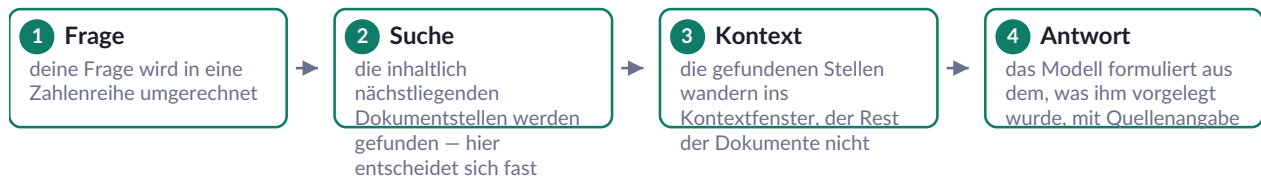


Abb. 9 RAG in vier Schritten. Der kritische Punkt ist Schritt 2, nicht Schritt 4: Findet die Suche die falsche Stelle, antwortet das Modell überzeugend auf falscher Grundlage — mit korrekt aussehender Quellenangabe.

2.6 Werkzeuge und Agenten in Grundzügen

Moderne Systeme können Werkzeuge aufrufen: eine Suche, einen Rechner, einen Kalender, eine Datenbankabfrage. Ein **Agent** ist ein Sprachmodell, das in einer Schleife läuft: Es plant einen Schritt, ruft ein Werkzeug auf, liest das Ergebnis, plant den nächsten Schritt. Das erweitert den Anwendungsbereich erheblich.

Für die Beurteilung ändern sich drei Dinge — und drei nicht:

WAS SICH ÄNDERT	WAS GLEICH BLEIBT
• Rechnen wird zuverlässig, wenn ein Rechner läuft • Aktualität wird möglich, wenn eine Suche läuft • Das System kann selbst handeln, nicht nur antworten • Fehler können sich über mehrere Schritte fortpflanzen	• Die Auswahl der Schritte bleibt Fortsetzung • Kein Werkzeug = keine Verbesserung • Ob ein Werkzeug lief, siehst du der Antwort nicht an • Verantwortung bleibt bei dem, der es einsetzt

Abb. 10 Werkzeuge verschieben die Grenze, sie heben sie nicht auf.

DIE EINE FRAGE BEI AGENTEN

Bei einem antwortenden System ist der Schaden begrenzt: Du liest, du prüfst, du verwirfst. Bei einem *handelnden* System ist der Schaden bereits eingetreten, wenn du liest. Die Frage vor jedem Agenten lautet deshalb nicht „wie gut ist es?“, sondern **„welche Handlungen kann es ohne Rückfrage auslösen, und sind die umkehrbar?“** Alles Unumkehrbare gehört hinter eine menschliche Freigabe (Modul 7.3).

BEISPIEL 2.1 – VIER AUFGABEN, ZWEI ERWARTBARE FEHLSCHLÄGE

„Wie viele Zeichen hat der folgende Satz?“ — Erwartbarer Fehlschlag. Die Zeichenebene ist für das Modell nicht sichtbar. Es nennt eine plausible Zahl. Lösung: ein Werkzeug rechnen lassen.

„Welcher Wochentag ist der 14. März nächsten Jahres?“ — Erwartbarer Fehlschlag ohne Kalenderzugriff. Das Modell kennt das heutige Datum nicht und leitet aus Mustern ab. Lösung: Datum mitgeben oder Werkzeug nutzen.

„Formuliere diese Absage freundlicher.“ — Geeignet. Die Aufgabe spielt vollständig auf der Textebene, und das Ergebnis ist unmittelbar prüfbar.

„Fasse die Kernaussagen dieses Protokolls zusammen.“ — Geeignet, mit einer Bedingung: Das Protokoll muss vollständig im Kontextfenster liegen. Bei einem 80-seitigen Dokument ist die eigentliche Frage nicht, ob das Modell zusammenfassen kann, sondern welche Teile es überhaupt gesehen hat.

BEISPIEL 2.2 – WENN DIE SUCHE DANEBENGREIFT

Ausgangslage: Ein Versicherer betreibt einen Assistenten über seiner Tarifsammlung. Eine Sachbearbeiterin fragt: „Gilt bei Tarif Komfort Plus die Wartezeit von drei Monaten auch für Zahnersatz?“

Was passiert: Die Suche findet zwei Passagen — die allgemeine Wartezeitregel aus den Rahmenbedingungen und den Zahnersatzabschnitt des *Vorgängertarifs* Komfort. Beide sind inhaltlich nahe an der Frage. Die Passage zum aktuellen Tarif rangiert auf Platz vier und wird nicht übergeben.

Die Antwort: „Ja, die Wartezeit von drei Monaten gilt auch für Zahnersatz“ — mit zwei korrekt formatierten Fundstellenangaben. Beide Belege existieren. Beide sagen etwas Richtiges. Die Kombination ist falsch.

Die Lehre: Bei RAG prüfst du nicht, *ob* es eine Quelle gibt, sondern *ob die zitierte Quelle die Aussage trägt und aktuell ist* — die drei Schritte der Belegprüfung aus Modul 6.4. Eine Fundstelle ist bei RAG kein Beweis, sondern eine Behauptung über eine Fundstelle.

ÜBUNG 2 – SCHEITERT ES, UND WORAN?

Entscheide für jede Aufgabe: **zuverlässig**, **unzuverlässig** oder **nur mit Werkzeug**. Begründe mit einem der Begriffe aus diesem Modul (Token, Kontextfenster, Training/Inferenz, Zufall).

1. „Übersetze diesen Absatz ins Englische.“
2. „Wie hoch war unser Umsatz im letzten Quartal?“
3. „Nenne alle Wörter in diesem Text, die mit ‚un‘ beginnen.“
4. „Erkläre den Unterschied zwischen Miete und Pacht.“
5. „Zähle, wie oft das Wort ‚Risiko‘ in diesem 60-seitigen Bericht vorkommt.“
6. „Schlage drei Betreffzeilen für diese E-Mail vor.“
7. „Gib mir dieselbe Auswertung wie gestern, damit ich sie vergleichen kann.“
8. „Wie viele Arbeitstage liegen zwischen dem 3. und dem 28. dieses Monats?“

Lösung: Anhang A, Übung 2 – Scheitert es, und woran?

ÜBUNG 2B – EINE EINGABE REPARIEREN

Eine Kollegin schreibt: „Ich habe dir unser komplettes Qualitätshandbuch angehängt (340 Seiten). Sag mir, ob unsere Lieferantenbewertung noch zur neuen Norm passt, und begründe mit den entsprechenden Kapiteln.“

Benenne die drei Stellen, an denen diese Eingabe systematisch scheitert, und schreibe sie so um, dass die Aufgabe lösbar wird — ohne den Nutzen zu verlieren.

Lösung: Anhang A, Übung 2b – Eine Eingabe reparieren

KERNAUSSAGEN

- Modelle sehen Token, keine Buchstaben. Zeichen- und Silbenaufgaben scheitern systematisch.
- Nur was im Kontextfenster steht, existiert für die Antwort. Ein größeres Fenster verschiebt das Problem, es löst es nicht.
- Nach dem Training wird nichts mehr gelernt — was nicht heißt, dass nichts gespeichert wird.
- Eine Antwort ist eine Stichprobe. Reproduzierbarkeit muss man herstellen, nicht annehmen.
- Bei RAG verlagert sich der kritische Punkt vom Modell auf die Suche davor.
- Bei Agenten ist die Leitfrage nicht die Qualität, sondern die Umkehrbarkeit der ausgelösten Handlungen.

Selbsttest Modul 2

1. Ein Kollege lädt ein 300-seitiges Handbuch hoch und ist überrascht, dass eine Detailfrage falsch beantwortet wird. Welche Erklärung liegt am nächsten?

- a) Das Modell wurde nicht auf das Handbuch trainiert.
- b) Die Temperatur war zu niedrig eingestellt.
- c) Das Modell hat zu wenig Trainingsdaten zu diesem Thema.
- d) Die relevante Passage lag nicht im Kontextfenster.

2. Was folgt daraus, dass Modelle Text in Token verarbeiten?

- a) Sie speichern jede Eingabe dauerhaft.
- b) Sie können keine langen Texte verarbeiten.
- c) Sie sind bei Aufgaben auf Zeichenebene unzuverlässig.
- d) Sie können nicht übersetzen.

3. Bei einem RAG-System gibt eine Antwort zwei korrekt formatierte Fundstellen an, ist aber inhaltlich falsch. Wo liegt der Fehler am ehesten?

- a) Beim Suchschritt, der die falschen Stellen ausgewählt hat.
- b) Bei der Temperatur-Einstellung.
- c) Beim Kontextfenster, das zu groß war.
- d) Beim Modell, das die Fundstellen erfunden hat.

4. Welche Maßnahme trägt am wenigsten zur Reproduzierbarkeit einer Auswertung bei?

- a) Die Temperatur auf einen niedrigen Wert setzen.
- b) Höflichkeitsformeln aus der Eingabe entfernen.
- c) Dieselbe Frage mehrfach stellen und die Antworten vergleichen.
- d) Die Modellversion samt Datum festhalten.

Warum Modelle überzeugend falsch liegen

ca. 75 Minuten · 2 Beispiele · 2 Übungen · Selbsttest am Ende

LERNZIELE

- Du kannst erklären, warum falsche Antworten genauso souverän klingen wie richtige.
- Du kennst die drei häufigsten Fehlermuster und erkennst sie im Text.
- Du kannst benennen, warum ein Modell lieber erfindet als schweigt.
- Du weißt, was Gefälligkeit (sycophancy) ist und warum sie in Besprechungen gefährlich wird.
- Du kannst sechs Warnsignale benennen, die eine Prüfung auslösen sollten.
- Du kannst vier verbreitete Scheinlösungen von zwei wirksamen Maßnahmen unterscheiden.

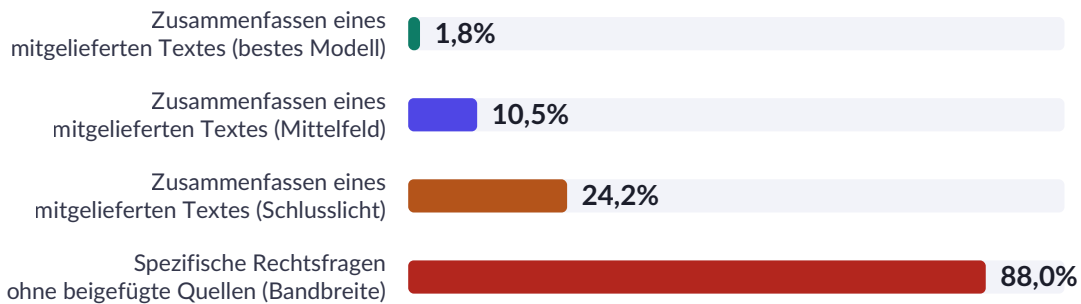
Dieses Modul ist das Herzstück des Kurses. Nicht weil Halluzinationen häufig sind, sondern weil sie *unauffällig* sind. Wer nach Modul 3 einen Text anders liest als vorher, hat den größten Teil des praktischen Nutzens dieses Kurses bereits mitgenommen.

3.1 Plausibilität ist nicht Wahrheit

Ein Sprachmodell ist darauf optimiert, dass seine Ausgabe *aussieht wie* guter Text zum Thema. Wahrheit ist dabei kein eigenes Kriterium, sondern ein Nebenprodukt: Weil in den Trainingsdaten überwiegend Zutreffendes steht, ist Plausibles meistens auch richtig. Meistens.

Wo es das nicht ist, fehlt jedes Signal. Ein Mensch, der unsicher ist, zögert, schränkt ein, sagt „ich glaube“. Ein Modell hat diese Anzeige nicht — es formuliert die erfundene Fundstelle im selben Tonfall wie das Einmaleins. **Diese fehlende Unsicherheitsanzeige ist das eigentliche Problem, nicht die Fehlerquote.** Ein System, das in 5 % der Fälle irrt und das anzeigt, ist beherrschbar. Ein System, das in 3 % der Fälle irrt und dabei gleich klingt, ist es nicht.

Wie groß ist „meistens“? Das hängt vollständig vom Zuschnitt der Aufgabe ab — und die Spannweite ist enorm:



Die Werte sind nicht direkt vergleichbar — genau das ist die Aussage: Derselbe Modelltyp liegt je nach Aufgabe zwischen 2 % und 88 %.

Abb. 11 Halluzinationsrate nach Aufgabenzuschnitt. Die ersten drei Werte stammen aus dem Vectara-Leaderboard (HHEM-2.3, Stand 11.05.2026) und messen die denkbar einfachste Aufgabe: einen mitgelieferten Text zusammenfassen. Der vierte Wert ist die Obergrenze der Stanford-RegLab-Studie zu spezifischen Rechtsfragen ohne beigefügte Quellen (Bandbreite 69–88 % über mehr als 200.000 Anfragen je Modell).

Quelle: Vectara Hallucination Leaderboard 2026; Dahl et al., Stanford RegLab 2024

DIE ZAHL, DIE MAN SICH MERKT

Zwischen **1,8 %** und **88 %** liegt kein Modellunterschied, sondern ein *Aufgabenunterschied*. Dieselbe Technik ist nahezu verlässlich, wenn die Grundlage beiliegt — und unbrauchbar, wenn sie aus dem Gedächtnis geholt werden muss. Das ist die wichtigste praktische Konsequenz dieses Kurses: **Nicht das Modell wechseln, sondern die Aufgabe umbauen.**^{[4][5]}

3.2 Die drei Muster

Erfundene Belege

Quellen, Normen, Paragraphen, Studien, Zahlen — mit korrekter Form und falschem Inhalt. Entsteht, weil das Modell gelernt hat, wie eine Fundstelle aussieht, nicht welche existiert.

Verschmolzene Fakten

Zwei ähnliche Sachverhalte werden zu einem: die Regelung aus einem Land mit der Frist aus einem anderen. Jedes Element für sich ist real, die Kombination nicht.

Veraltete Information

Der Wissensstand ist eingefroren. Preise, Zuständigkeiten, Fristen, Rechtslagen kommen im alten Stand heraus, ohne Hinweis.

Abb. 12 Die drei Muster. Das zweite ist das tückischste: Es ist das, was Fachleute am spätesten bemerken, weil alles vertraut klingt.

Besonders tückisch bei erfundenen Belegen: **Autoren, die es wirklich gibt, kombiniert mit Titeln, die es nicht gibt.** Die halbe Prüfung — Autor existiert, Zeitschrift existiert — bestätigt genau die Hälfte, die stimmt.

Bei veralteter Information ist das Regulierungsumfeld die häufigste Falle, weil sich dort innerhalb eines Jahres mehr ändert, als ein Modell mitbekommt. Ein Beispiel aus genau diesem Handbuch: Die Fristen der KI-Verordnung wurden durch die Änderungsverordnung im Juli 2026 verschoben (Modul 8.2). Jedes Modell mit älterem Wissensstand nennt souverän die alten Termine.

3.3 Woher Halluzinationen kommen

Drei Ursachen wirken zusammen. Wer sie kennt, versteht auch, warum die naheliegenden Gegenmaßnahmen nicht wirken.

1

Fortsetzung kennt kein „weiß nicht“

Das Verfahren erzeugt bei jedem Schritt die plausibelste Fortsetzung. Ein leerer Platz ist nie die plausibelste Fortsetzung — es gibt keinen Mechanismus, der Abbruch belohnt.

2

Form ist leichter zu lernen als Inhalt

Wie eine Literaturangabe, ein Aktenzeichen oder eine Normnummer aussieht, ist ein einfaches, häufiges Muster. Welche davon existieren, ist ein schweres, seltenes Faktum. Das Modell beherrscht das Erste besser als das Zweite.

3

Raten wird belohnt, Zugeben nicht

In der Bewertung während der Entwicklung zählt eine falsche Antwort oft wie eine verweigerte — eine geratene Antwort hat also eine Chance, eine Verweigerung keine. Das erzeugt genau das beobachtete Verhalten: souveränes Raten.

Abb. 13 Warum ein Modell lieber erfindet als schweigt.

3.4 Gefälligkeit

Modelle neigen dazu, der Prämisse ihrer Frage zu folgen. Wer fragt „Warum ist Verfahren A hier vorgeschrieben?“, bekommt Gründe — auch dann, wenn Verfahren A gar nicht vorgeschrieben ist. Wer widerspricht, bekommt oft prompt eine Kehrtwende, unabhängig davon, wer recht hatte.

In Besprechungen wird das gefährlich: Ein Modell, das dem lautesten Teilnehmer zustimmt, sieht aus wie eine unabhängige Bestätigung. Die Gegenmaßnahme ist einfach und wirkt: Stelle die Frage offen.

STATT (LEGT DIE ANTWORT HINEIN)

BESSER (LÄSST SIE OFFEN)

„Warum ist Verfahren A hier vorgeschrieben?“

„Was gilt hier, und woraus ergibt sich das?“

„Bestätige mir, dass Variante B korrekt ist.“

„Welche Varianten kommen in Betracht, und was spricht jeweils dafür und dagegen?“

„Das ist doch ein Hochrisikosystem, oder?“

„Prüfe anhand der Kriterien, in welche Kategorie das System fällt.“

„Fasse zusammen, warum sich das Projekt lohnt.“

„Welche Argumente sprechen für das Projekt, welche dagegen?“

„Ist diese Klausel wirksam?“

„Welche Anhaltspunkte im Text sprechen für, welche gegen die Wirksamkeit?“

DIE PROBE AUFS EXEMPEL

Willst du wissen, ob eine Antwort inhaltlich getragen ist oder nur gefällig: Stelle im nächsten Schritt *dieselbe* Frage mit umgekehrter Prämisse. Kippt die Antwort, war sie Gefälligkeit. Bleibt sie stehen und wird begründet, hast du ein schwaches, aber echtes Indiz. Das kostet dreißig Sekunden und ist der billigste Test in diesem Handbuch.

3.5 Sechs Warnsignale

WORAN DU HINSEHEN MUSST

1. **Sehr spezifische Details ohne Beleg.** Eine exakte Prozentzahl, ein Aktenzeichen, ein Datum — je präziser ohne Quelle, desto prüfenswerter.
2. **Glatte Zitate.** Perfekt formatierte Literaturangaben sind ein Muster, das leicht zu erzeugen und schwer zu erfüllen ist.
3. **Bestätigung deiner Annahme.** Wenn die Antwort exakt das sagt, was du gehofft hast, hast du möglicherweise die Antwort in die Frage gelegt.
4. **Keine Einschränkung, wo Fachleute eine machen würden.** Bei Rechts-, Medizin- und Finanzfragen antworten Fachleute fast nie ohne Vorbehalt. Eine glatte Antwort ist dort selbst das Warnsignal.
5. **Runde Zahlen an unrunden Stellen.** „Rund 30 % der Unternehmen“ ist harmlos. „31,4 % laut einer Studie der Universität Rotterdam“ ist eine Präzisionsbehauptung — und Präzision ohne Quelle ist die häufigste Form der erfundenen Fundstelle.
6. **Aktualitätsbehauptungen.** „Seit der Neufassung gilt...“, „aktuell beträgt...“, „ab Januar...“. Der Wissensstand ist eingefroren; jede Aktualitätsbehauptung ohne laufende Suche ist unbelegt.

3.6 Was nicht hilft

Vier verbreitete Gegenmaßnahmen, die den Aufwand erhöhen, ohne das Risiko zu senken:

„Sei ehrlich und erfinde nichts“

Eine Anweisung im Prompt ändert die Fortsetzungsmechanik nicht. Sie senkt die Rate messbar wenig und erzeugt beim Leser falsche Sicherheit.

Ein größeres oder neueres Modell

Verbessert den Durchschnitt, beseitigt das Muster nicht. Der Sprung von 88 % auf 2 % kommt vom Aufgabenzuschnitt, nicht vom Modellwechsel.

Ein zweites Modell zur Bestätigung

Beide können denselben Irrtum aus denselben Quellen haben. Übereinstimmung ist ein schwaches Indiz, kein Beleg (Modul 6.5).

„Gib deine Sicherheit in Prozent an“

Die Zahl wird ebenfalls fortgesetzt, nicht gemessen. Ein Modell, das „95 % sicher“ schreibt, hat nichts geprüft — es hat eine plausible Zahl gewählt.

Was tatsächlich hilft: Grundlage beilegen

Die Aufgabe von „aus dem Gedächtnis“ auf „aus diesem Text“ umstellen. Der größte Hebel, den es gibt.

Was tatsächlich hilft: Nicht-Antwort erlauben

„Wenn es in den beigelegten Unterlagen nicht steht, schreibe: nicht enthalten.“ Gibt dem Modell eine erlaubte Alternative zum Raten.

Abb. 14 Vier Scheinlösungen und zwei echte. Die beiden echten verändern die Aufgabe, die vier anderen nur den Ton.

BEISPIEL 3.1 – DREI ABSÄTZE, EIN PROBLEM IN JEDEM

A — „Nach § 17 Abs. 3 des Digitalisierungsförderungsgesetzes sind Unternehmen mit mehr als 50 Beschäftigten verpflichtet, eine KI-Beauftragte zu benennen. Die Frist endete am 30. Juni 2026.“

Erfundener Beleg. Form stimmt, Paragraph und Gesetz existieren so nicht. Warnsignal: sehr spezifisch, kein Verweis auf eine Fundstelle, exaktes Datum.

B — „Das Modell verarbeitet bis zu 200.000 Token und wurde im März 2024 veröffentlicht; es unterstützt Bildeingaben und kostet 3 US-Dollar je Million Eingabetoken.“

Verschmelzung. Vier Angaben, die einzeln zu je einem anderen Produkt gehören könnten. Nichts wirkt falsch. Warnsignal: Detaildichte ohne Quelle bei einem sich schnell ändernden Gegenstand.

C — „Die Pflichten für Hochrisiko-KI-Systeme gelten seit dem 2. August 2026.“

Veraltet. Der Termin stand einmal so im Raum, wurde aber verschoben — siehe Modul 8.2. Warnsignal: regulatorische Frist, glatt formuliert, kein Hinweis auf Änderungen.

Beachte: Alle drei Absätze sind sprachlich einwandfrei. Es gibt kein stilistisches Merkmal, an dem du sie erkennst. Du erkennst sie nur an der *Art der Behauptung* — und dann prüfst du.

BEISPIEL 3.2 – DIE ZAHL, DIE IN DIE VORSTANDSVORLAGE WANDERT

Ausgangslage: Ein Bereichsleiter lässt sich eine halbe Seite zur Marktentwicklung erzeugen. Darin steht: „Der europäische Markt für Prozessautomatisierung wächst bis 2028 mit 14,2 % jährlich.“

Was daran wahrscheinlich stimmt: Die Größenordnung. Zweistelliges Wachstum in diesem Segment ist plausibel, und genau deshalb fällt niemandem etwas auf.

Was daran nicht geprüft ist: die Nachkommastelle (Präzision ohne Quelle), das Bezugsjahr, die Marktabgrenzung („Prozessautomatisierung“ meint je nach Studie sehr Verschiedenes) und die Region.

Der Verlauf: Die Zahl wandert unbesehen in eine Folie, die Folie in eine Vorlage, die Vorlage in einen Beschluss. Nach der dritten Weitergabe ist sie nicht mehr einzufangen — sie steht jetzt in einem Hausdokument und wird zitiert, als hätte sie eine Quelle.

Die Lehre: Falsche Rechtsaussagen werden eher hinterfragt; falsche Statistiken werden geglaubt. Eine Zahl mit Nachkommastelle und ohne Quelle ist in einer Vorlage gefährlicher als ein falscher Paragraph.

ÜBUNG 3 – FINDE DAS PROBLEM

Jeder Absatz enthält genau eine problematische Stelle. Markiere sie, benenne das Muster (erfundener Beleg, Verschmelzung, veraltet) und notiere, *wie* du es prüfen würdest.

1. „Für die Einführung eines KI-gestützten Systems zur Leistungsbewertung ist in Deutschland die Zustimmung des Betriebsrats erforderlich; die Beteiligung ist innerhalb von 14 Tagen einzuholen.“
2. „Studien der Universität Rotterdam aus dem Jahr 2023 zeigen, dass 68 % der Beschäftigten KI-Werkzeuge ohne Wissen ihrer Vorgesetzten nutzen.“
3. „Da personenbezogene Daten bei der Verarbeitung durch einen Auftragsverarbeiter das Unternehmen verlassen, ist in jedem Fall eine Einwilligung der betroffenen Person erforderlich.“
4. „Nach der KI-Verordnung müssen Betreiber von Hochrisikosystemen seit dem 2. August 2026 ein Risikomanagementsystem betreiben und die Protokolle sechs Monate aufbewahren.“

Zusatzfrage: Welche der vier Aussagen wäre in einer Präsentation vor der Geschäftsführung am gefährlichsten, und warum?

Lösung: Anhang A, Übung 3 – Finde das Problem

ÜBUNG 3B – GEFÄLLIGKEIT PROVOZIEREN

Formuliere zu jeder der folgenden Situationen (a) eine Frage, die eine gefällige Antwort geradezu erzwingt, und (b) eine offene Frage zum selben Anliegen. Notiere in einem Satz, was der Unterschied bewirkt.

1. Du vermutest, dass ein Vertragsentwurf eine unwirksame Haftungsklausel enthält.
2. Du willst wissen, ob euer geplantes Auswertungswerkzeug mitbestimmungspflichtig ist.
3. Du hast dich für Anbieter X entschieden und willst die Entscheidung begründen.

Lösung: Anhang A, Übung 3b – Gefälligkeit provozieren

KERNAUSSAGEN

- Das Problem ist nicht die Fehlerquote, sondern die fehlende Unsicherheitsanzeige.
- Die Halluzinationsrate hängt am Aufgabenzuschnitt, nicht am Modellnamen: 1,8 % bis 88 % bei derselben Technik.
- Drei Muster: erfundene Belege, verschmolzene Fakten, veraltete Information.
- Modelle folgen der Prämisse deiner Frage. Frag offen, nicht bestätigend.
- Warnsignale sind inhaltlich, nicht stilistisch. Am Text erkennst du nichts.
- Was hilft: Grundlage beilegen und Nicht-Antwort erlauben. Was nicht hilft: Ehrlichkeitsappelle, Modellwechsel, Zweitmodell, Sicherheitsangaben in Prozent.

Selbsttest Modul 3

1. Was macht Halluzinationen im Arbeitsalltag besonders riskant?

- a) Sie treten nur bei kostenlosen Systemen auf.
- b) Sie treten überwiegend bei sehr langen Texten auf.
- c) Sie sind sprachlich nicht von zutreffenden Aussagen zu unterscheiden.
- d) Sie betreffen ausschließlich Zahlen und Datumsangaben.

2. Welche Frageformulierung reduziert das Risiko der Gefälligkeit am stärksten?

- a) „Variante B ist doch korrekt, oder?“
- b) „Bestätige mir bitte, dass Variante B die richtige ist.“
- c) „Warum ist Variante B hier vorgeschrieben?“
- d) „Welche Varianten kommen in Betracht, und woraus ergibt sich das jeweils?“

3. Ein Modell gibt an, es sei sich „zu 95 % sicher“. Was folgt daraus?

- a) Nichts — auch die Prozentzahl ist eine Fortsetzung, keine Messung.
- b) Das Modell hat intern eine Quellenprüfung durchgeführt.
- c) Die Aussage sollte nur noch stichprobenartig geprüft werden.
- d) Die Aussage ist mit hoher Wahrscheinlichkeit korrekt.

4. Welche Maßnahme senkt die Halluzinationsrate am wirksamsten?

- a) Ein größeres Modell wählen.
- b) Die Grundlage beilegen und eine Nicht-Antwort ausdrücklich erlauben.
- c) Die Antwort von einem zweiten Modell bestätigen lassen.
- d) Im Prompt anweisen, nichts zu erfinden.

Geeignete und ungeeignete Aufgaben

ca. 75 Minuten · 2 Beispiele · 2 Übungen · Selbsttest am Ende

LERNZIELE

- Du kannst eine Aufgabe anhand von vier Kriterien einordnen: Fehlerkosten, Umkehrbarkeit, Prüfbarkeit, Frequenz.
- Du kannst begründen, warum dieselbe Tätigkeit je nach Zuschnitt geeignet oder ungeeignet ist.
- Du kennst die Aufgabenklassen, bei denen KI grundsätzlich nicht die letzte Instanz sein darf.
- Du kannst den Eignungs-Schnelltest auf ein konkretes Vorhaben anwenden.
- Du kannst fünf Umbaumuster nennen, mit denen aus einer ungeeigneten Aufgabe eine geeignete wird.

Listen mit „geeigneten Anwendungsfällen“ veralten, sobald sich die Systeme ändern — und sie ändern sich schnell. Dieses Modul gibt dir stattdessen Kriterien, die halten: vier Fragen, ein Raster und ein Umbau-Katalog. Damit kannst du auch eine Aufgabe beurteilen, die in keiner Liste steht.

4.1 Vier Fragen statt einer Liste

Wie teuer ist ein Fehler? Ein missglückter Entwurf kostet fünf Minuten. Eine falsche Zusage an einen Kunden kostet mehr.

Ist der Fehler umkehrbar? Ein Text vor dem Versand ist korrigierbar. Ein versendeter Text, eine getroffene Personalentscheidung, ein gelöschter Datensatz sind es nicht oder nur teuer.

Kann ich das Ergebnis überhaupt prüfen? Das ist die Frage, die am häufigsten übersehen wird. Wenn du die Antwort nur beurteilen könntest, wenn du sie schon wüsstest, hilft dir das System nicht — es verlagert nur dein Risiko.

Wie oft läuft das? Die vierte Frage kommt in kaum einer Checkliste vor und entscheidet trotzdem mit. Ein Fehler in einer Einzelentscheidung ist ein Vorfall. Derselbe Fehler in einem Prozess, der zweitausendmal im Monat läuft, ist ein Systemfehler — und niemand prüft zweitausend Vorgänge einzeln. Hohe Frequenz verschiebt den Prüfpunkt vom Ergebnis auf das Verfahren: Stichprobenplan statt Einzelprüfung.

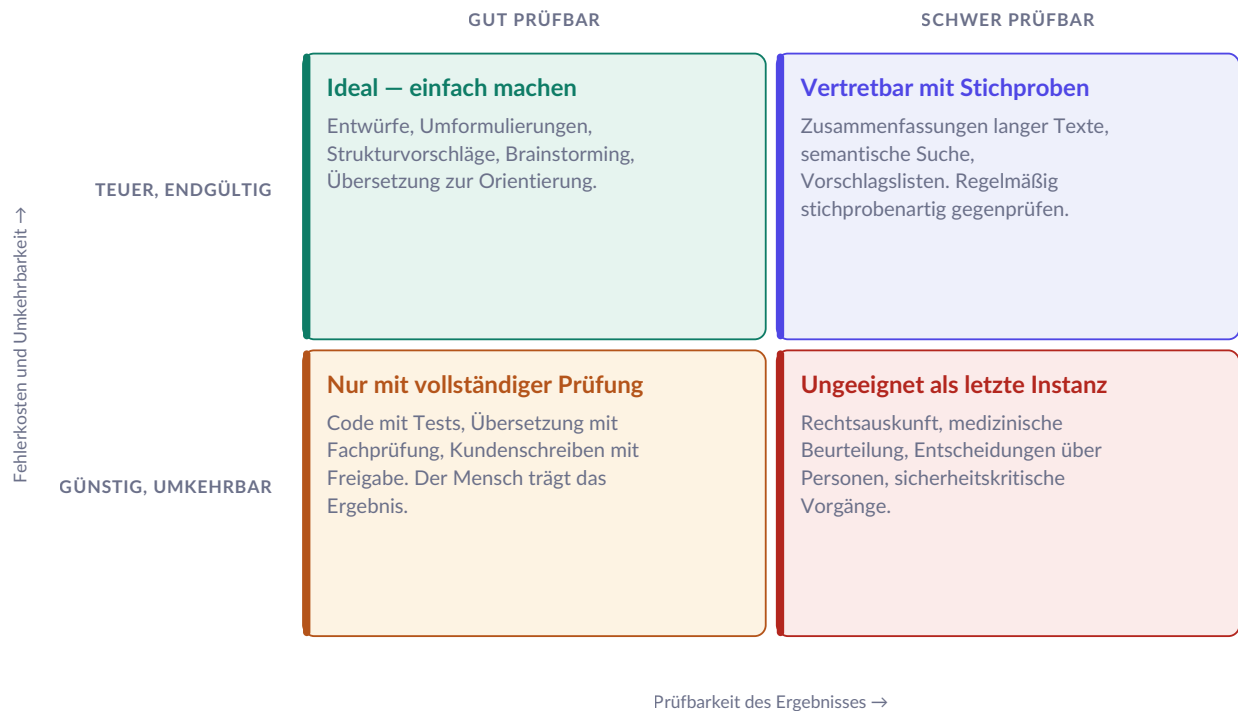


Abb. 15 Das Eignungsraster. Achte auf die untere linke Zelle: Sie ist nicht verboten, sie ist teuer — dort steckt der Prüfaufwand.

4.2 Der Zuschnitt entscheidet, nicht das Thema

„KI im Personalwesen“ ist weder geeignet noch ungeeignet — die Frage ist falsch gestellt. *Eine Stellenausschreibung entwerfen* ist günstig, umkehrbar und prüfbar: geeignet. *Bewerbungen vorsortieren* ist folgenreich für einzelne Menschen, für die Aussortierten faktisch endgültig und von außen kaum prüfbar: ungeeignet als letzte Instanz, und je nach Ausgestaltung ein regulatorisch sensibler Anwendungsfall.

Dieselbe Verschiebung gibt es überall. *Vertragsentwurf formulieren* ist geeignet. *Vertrag rechtlich freigeben* ist es nicht. *Befunde zusammenfassen* ist geeignet. *Diagnose stellen* ist es nicht. Der Grundsatz dahinter passt in einen Satz:

DER LEITSATZ DIESES KURSES

KI erzeugt Entwürfe. Menschen erzeugen Ergebnisse. Wo diese Trennung sauber ist, funktioniert der Einsatz. Wo sie verschwimmt, entstehen die Vorfälle.

4.3 Vier Klassen, bei denen KI nie allein entscheidet

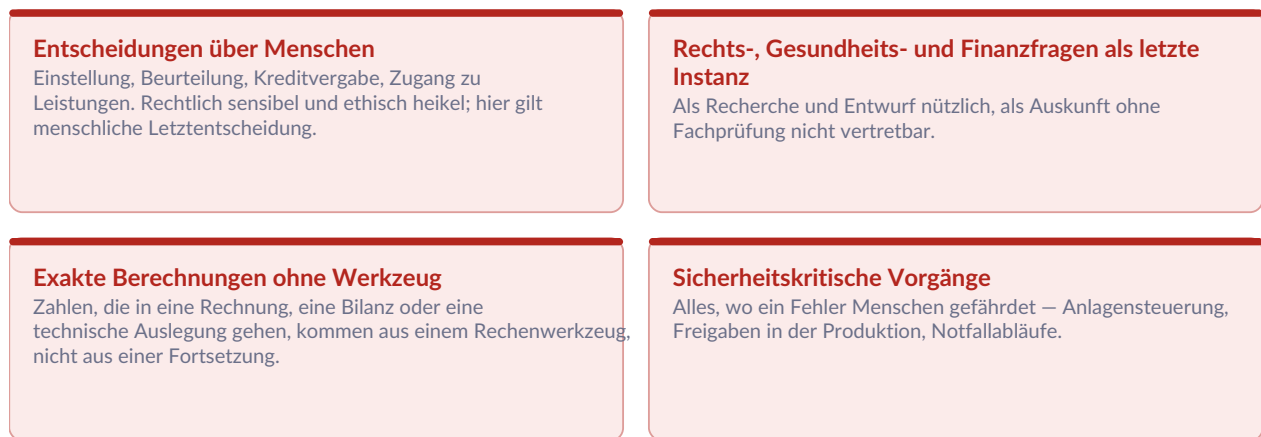


Abb. 16 Die vier roten Klassen. „Nie allein“ heißt nicht „nie“: Als Zuarbeit sind alle vier zulässig, als letzte Instanz keine.

4.4 Der Eignungs-Schnelltest

Fünf Fragen, in dieser Reihenfolge. Die erste, die mit „ja“ beantwortet wird, entscheidet — weiter unten musst du dann nicht mehr lesen.



Abb. 17 Der Schnelltest. Bleibt bis zum Ende alles „nein“, ist die Aufgabe unkritisch: Durchlesen genügt.

4.5 Der Umbau-Katalog

Die interessanteste Frage ist nicht, ob eine Aufgabe geeignet ist, sondern wie man sie geeignet *macht*. Fünf Muster decken fast alles ab:

1

Entwurf statt Ergebnis

Das System liefert einen Vorschlag, der Mensch entscheidet und trägt. Aus „Termin zusagen“ wird „Terminvorschlag aus dem Warenwirtschaftssystem entwerfen, Mensch gibt frei“.

2

Grundlage beilegen statt abfragen

Aus „was gilt bei Kündigungsfristen?“ wird „hier ist § 622 BGB im Wortlaut und unser Vertragstext — vergleiche beide“. Größter Hebel gegen Halluzinationen (Modul 3.6).

3

Kriterien offenlegen statt Urteil einholen

Aus „welche Bewerbungen sind die besten?“ wird „strukturiere alle Bewerbungen nach den folgenden offengelegten Kriterien; alle Unterlagen bleiben sichtbar“. Der Mensch trifft die Auswahl.

4

Werkzeug statt Fortsetzung

Alles Gerechnete, Gezählte, Datierte an ein Werkzeug übergeben. Aus „berechne die Wartungsintervalle“ wird „lies die Formel aus dem Handbuch aus und wende sie in einer Tabellenkalkulation an“.

5

Stichprobe statt Vollprüfung

Bei hoher Frequenz: definierter Stichprobenumfang, feste Prüfkriterien, protokollierte Trefferquote. Nicht jedes Ergebnis prüfen, sondern das Verfahren überwachen.

Abb. 18 Fünf Muster, um aus ungeeignet geeignet zu machen — ohne den Nutzen zu verlieren.

BEISPIEL 4.1 — EINE TÄTIGKEIT, ZWEI ZUSCHNITTE

Ungeeignet: Eine Sachbearbeiterin lässt sich zu einem Kündigungsfall die Rechtslage erklären und schreibt das Ergebnis in ihre Aktennotiz. Fehlerkosten hoch, Wirkung praktisch endgültig, Prüfbarkeit gering — sie kann die Auskunft nur beurteilen, wenn sie sie ohnehin wüsste.

Geeignet: Dieselbe Sachbearbeiterin lässt sich aus ihren eigenen Stichpunkten einen Entwurf des Anschreibens formulieren, prüft ihn gegen die Vorlage der Rechtsabteilung und passt ihn an. Fehlerkosten niedrig, umkehrbar, vollständig prüfbar — die Fachfrage hat sie selbst beantwortet, das System hat formuliert.

Der Unterschied liegt nicht im Werkzeug und nicht im Thema, sondern darin, wer die inhaltliche Frage beantwortet hat.

BEISPIEL 4.2 — WENN FREQUENZ ALLES VERÄNDERT

Ausgangslage: Ein Energieversorger lässt eingehende Kundenmails automatisch kategorisieren und eine Antwortvorlage vorschlagen. 4.000 Vorgänge im Monat. Die Trefferquote der Kategorisierung liegt bei 94 %.

Die naive Rechnung: 94 % klingt gut. *Die richtige Rechnung:* 240 falsch kategorisierte Vorgänge im Monat, davon ein Teil in der Kategorie „Störungsmeldung“, die eine Frist auslöst.

Was hier nicht funktioniert: Einzelprüfung. Niemand liest 4.000 Vorgänge gegen. *Was funktioniert:* eine Stichprobe von 50 Vorgängen pro Woche mit festen Prüfkriterien, getrennt ausgewertet nach Kategorie — plus eine harte Regel, dass fristauslösende Kategorien nie automatisch abgeschlossen werden.

Die Lehre: Bei hoher Frequenz ist nicht das einzelne Ergebnis der Prüfgegenstand, sondern das Verfahren. Und die Prüfung muss dort dichter sein, wo die Folgen härter sind — nicht gleichmäßig verteilt.

ÜBUNG 4 – EINORDNEN UND UMBAUEN

Teil A: Ordne jede Aufgabe in das Raster aus 4.1 ein (Fehlerkosten, Umkehrbarkeit, Prüfbarkeit, Frequenz) und entscheide: geeignet, geeignet mit Prüfung, ungeeignet.

1. Aus einem Besprechungsmitschnitt eine Aufgabenliste erzeugen.
2. Einem Kunden per Chat verbindlich einen Liefertermin nennen.
3. Aus einer technischen Spezifikation eine Kurzfassung für den Vertrieb erstellen.
4. Entscheiden, welche von 200 Bewerbungen zum Gespräch eingeladen werden.
5. Einen englischen Fachartikel für den internen Gebrauch übersetzen.
6. Die Wartungsintervalle einer Maschine aus dem Handbuch berechnen.
7. Eingehende Rechnungen den Kostenstellen zuordnen (1.500 Belege im Monat).

Teil B: Wähle die zwei Aufgaben, die du als ungeeignet eingestuft hast, und schneide sie mit einem Muster aus dem Umbau-Katalog (4.5) so um, dass sie geeignet werden — ohne den Nutzen zu verlieren. Schreibe für jede einen Satz, was der Mensch übernimmt.

Lösung: Anhang A, Übung 4 – Einordnen und umbauen

ÜBUNG 4B – DEN SCHNELLTEST ANWENDEN

Ein Bereich schlägt vor: „Wir lassen ein System die eingehenden Gewährleistungsfälle bewerten und eine Empfehlung ausgeben — anerkennen, ablehnen oder prüfen. Etwa 600 Fälle im Monat. Die Sachbearbeitung folgt der Empfehlung, wenn nichts dagegenspricht.“

Arbeite den Schnelltest aus 4.4 Frage für Frage ab. Notiere zu jeder Frage ja/nein mit einem Halbsatz Begründung, und formuliere am Ende einen Zuschnitt, den du verantworten würdest.

Lösung: Anhang A, Übung 4b – Den Schnelltest anwenden

KERNAUSSAGEN

- Vier Fragen: Was kostet ein Fehler? Ist er umkehrbar? Kann ich das Ergebnis prüfen? Wie oft läuft das?
- Prüfbarkeit wird am häufigsten übersehen und ist am wichtigsten.
- Hohe Frequenz verschiebt den Prüfpunkt vom Ergebnis auf das Verfahren.
- Nicht das Thema entscheidet, sondern der Zuschnitt der Aufgabe.
- KI erzeugt Entwürfe, Menschen erzeugen Ergebnisse.
- Ungeeignet ist selten endgültig — fünf Umbaumuster machen die meisten Aufgaben verantwortbar.

Selbsttest Modul 4

1. Welches Kriterium wird bei der Bewertung von Anwendungsfällen am häufigsten übersehen?

- a) Die Kosten des Systems.
- b) Ob das Ergebnis überhaupt prüfbar ist.
- c) Die Geschwindigkeit der Antwort.
- d) Die Sprache der Ausgabe.

2. Warum ist „einen Vertragsentwurf formulieren“ geeignet, „einen Vertrag freigeben“ aber nicht?

- a) Weil Entwürfe keine personenbezogenen Daten enthalten.
- b) Weil Freigaben technisch nicht automatisierbar sind.
- c) Weil die Freigabe endgültig, folgenreich und ohne Fachwissen nicht prüfbar ist.
- d) Weil Verträge zu lang für das Kontextfenster sind.

3. Ein Vorgang läuft 4.000 Mal im Monat mit 94 % Trefferquote. Was folgt daraus?

- a) Der Prüfpunkt verschiebt sich vom Ergebnis auf das Verfahren: Stichprobenplan.
- b) Der Anwendungsfall ist grundsätzlich ungeeignet.
- c) Die Quote ist ausreichend, keine weitere Maßnahme nötig.
- d) Jedes einzelne Ergebnis muss vor der Verwendung geprüft werden.

4. Welches Umbaumuster passt am besten zu „Bewerbungen vorsortieren“?

- a) Stichprobe statt Vollprüfung.
- b) Grundlage beilegen statt abfragen.
- c) Werkzeug statt Fortsetzung.
- d) Kriterien offenlegen statt Urteil einholen.

Daten, Vertraulichkeit und Geheimnisse

ca. 75 Minuten · 2 Beispiele · 2 Übungen · Selbsttest am Ende

LERNZIELE

- Du kannst vier Fragen benennen, die du vor jeder Eingabe beantwortest.
- Du kennst die Datenkategorien, die nie in ein nicht freigegebenes System gehören.
- Du kannst Zugangsarten unterscheiden und weißt, warum dasselbe Produkt je nach Konto unterschiedlich geregelt ist.
- Du kannst einen problematischen Prompt so umschreiben, dass er nutzbar und unbedenklich ist.
- Du verstehst, warum Schatten-KI das größere Risiko ist als das freigegebene Werkzeug — und was danach zu tun ist.

Dieses Modul ist das mit den härtesten Folgen bei Fehlern — und mit den einfachsten Gegenmaßnahmen. Vier Fragen vor jeder Eingabe, sechs Techniken, die fast immer funktionieren, und eine ehrliche Einordnung dessen, was Schatten-KI wirklich ist: kein Disziplinproblem, sondern eine Rückmeldung über eure Werkzeuge.

5.1 Das Eingabefeld ist kein Safe

Was du in ein Eingabefeld schreibst, verlässt in aller Regel dein Haus. Es wird übertragen, verarbeitet, protokolliert und je nach Anbieter, Vertrag und Einstellung unterschiedlich lange aufbewahrt. Das ist nicht skandalös — es gilt für jeden Cloud-Dienst. Es ist nur bei einem Chatfenster leichter zu vergessen, weil es sich anfühlt wie ein Notizbuch.

Wie oft sensible Daten tatsächlich hineinwandern, ist inzwischen gemessen. Eine Auswertung realer Datenflüsse in Unternehmen kommt auf: **39,7 Prozent aller Interaktionen mit KI-Werkzeugen enthalten sensible Daten**, und ein erheblicher Teil davon läuft über private statt über dienstliche Zugänge.^[6]

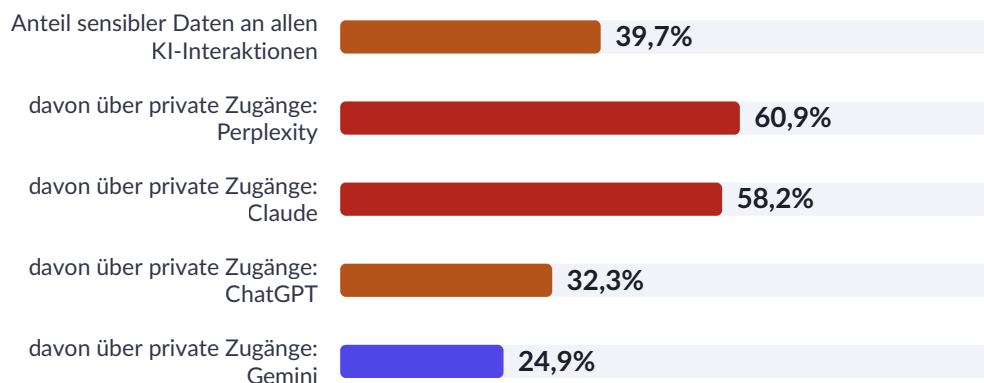


Abb. 19 Sensible Daten in KI-Werkzeugen, gemessen an tatsächlichen Datenflüssen. Die Prozentwerte der unteren vier Balken beziehen sich jeweils auf die Nutzung des genannten Werkzeugs, nicht auf die Gesamtheit.

Quelle: Cyberhaven, AI Adoption & Risk Report, Februar 2026

VIER FRAGEN VOR JEDER EINGABE

1. **Wem gehören diese Daten?** Mir, meinem Arbeitgeber, einem Kunden, einer dritten Person?
2. **Wer dürfte sie sehen?** Wäre es in Ordnung, wenn ein Fremder sie liest?
3. **Wo werden sie verarbeitet?** Welcher Anbieter, welche Region, welcher Vertrag?
4. **Werden sie zum Training verwendet?** Das steht in den Einstellungen und im Vertrag — und ist bei Privatkonten oft anders geregelt als bei Unternehmenskonten.

Wenn du eine dieser Fragen nicht beantworten kannst, ist das die Antwort: nicht eingeben, sondern klären.

5.2 Was nie in ein nicht freigegebenes System gehört

Personenbezogene Daten Namen, Kontaktdaten, Personalnummern, Bewerbungsunterlagen, Kundendaten — ohne geklärte Rechtsgrundlage.	Besondere Kategorien Gesundheitsdaten, Gewerkschaftszugehörigkeit, religiöse oder politische Überzeugungen, biometrische Daten. Die Schwelle ist hier deutlich höher.	Zugangsdaten und Schlüssel Passwörter, Token, API-Schlüssel, Zertifikate. Auch nicht „nur zum Testen“ — einmal eingeben gilt als offengelegt.
Unveröffentlichte Geschäftsgeheimnisse Preiskalkulationen, Angebotsstrategien, Konstruktionsdaten, Zahlen vor der Veröffentlichung.	Berufs- und Mandatsgeheimnisse Bei Anwaltschaft, Heilberufen, Steuerberatung und Sozialarbeit gelten eigene, strengere Pflichten — unabhängig vom Datenschutzrecht.	Fremdes geistiges Eigentum unter NDA Unterlagen eines Kunden oder Partners, die du selbst nur unter Auflagen erhalten hast. Die Auflage gilt weiter.

Abb. 20 Die Datenampel. Rot: geht in kein nicht freigegebenes System. Orange: geht nur mit geklärter vertraglicher Grundlage.

5.3 Konto ist nicht gleich Konto

„Ich benutze doch dasselbe Produkt“ ist die häufigste Fehlannahme in diesem Modul. Der Unterschied liegt nicht am Modell, sondern am Vertrag dahinter.

ZUGANGSART	AUFTRAGSVERARBEITUNG	TRAINING MIT EUREN DATEN	FÜR DIENSTLICHE DATEN
Privates Gratis-Konto	nein	häufig standardmäßig aktiv	nein
Privates Bezahl-Konto	nein	abschaltbar, aber Vertrag ist privat	nein
Team-/Unternehmenskonto	ja, per AV-Vertrag	regelmäßig standardmäßig aus	ja, wenn freigegeben
API-Zugang	ja, per AV-Vertrag	regelmäßig standardmäßig aus	ja, wenn freigegeben
Lokal / im eigenen Rechenzentrum	entfällt	nur, wenn ihr es selbst tut	ja

Für dich als Beschäftigte oder Beschäftigten heißt das: Die Frage ist nicht, *welches Produkt* du benutzt, sondern *über welchen Zugang*. Dasselbe Chatfenster kann je nach Anmeldung völlig unterschiedlich geregelt sein — und man sieht es der Oberfläche in der Regel nicht an.

5.4 Sechs Techniken, die fast immer funktionieren

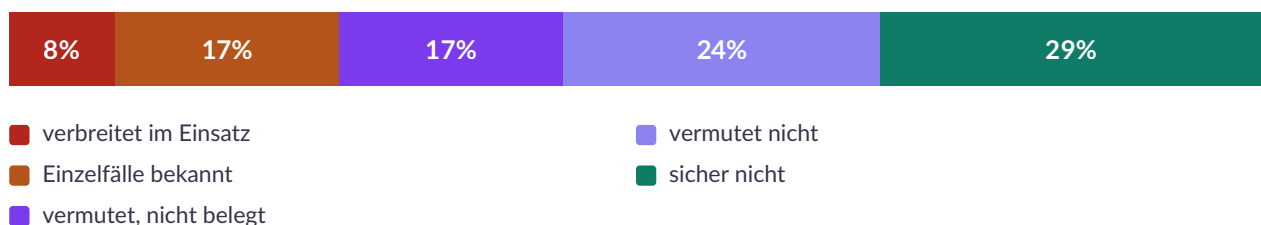
- 1 Anonymisieren**
Ersetze Namen und Kennungen durch Platzhalter: [KUNDE A], [MITARBEITERIN 1]. Für die meisten Aufgaben — Formulierung, Struktur, Tonfall — ist der echte Name ohnehin bedeutungslos.
- 2 Aggregieren**
Statt der Einzeldatensätze die Summe, den Durchschnitt, die Verteilung. Für eine Auswertungsidee reicht die Struktur der Daten; die Daten selbst braucht es nicht.
- 3 Auszug statt Volltext**
Häufig genügen zwei Absätze. Das ganze Dokument hochzuladen ist bequem und selten notwendig — und vergrößert die offengelegte Menge um ein Vielfaches.
- 4 Synthetische Beispieldaten**
Für Formatfragen, Formeln und Abfragen: erfinde fünf Zeilen, die dieselbe Struktur haben wie die echten. Das Ergebnis überträgst du anschließend selbst auf die echten Daten.
- 5 Schema statt Inhalt**
Bei technischen Fragen reichen Tabellen- und Spaltennamen. Für eine SQL-Abfrage braucht niemand die Datensätze — und erst recht keinen Zugang.
- 6 Freigegebene Werkzeuge nutzen**
Ein vertraglich geregeltes Unternehmenskonto mit Auftragsverarbeitung und abgeschaltetem Training ist etwas anderes als ein privates Konto. Wenn dein Haus eines bereitstellt, ist die Frage nicht, ob du es nutzt, sondern warum nicht.

Abb. 21 Sechs Techniken. Die ersten fünf kosten dich Sekunden und lösen den weit überwiegenden Teil aller Fälle.

5.5 Schatten-KI

Das eigentliche Risiko ist selten das freigegebene Werkzeug. Es ist das private Konto auf dem privaten Gerät, mit dem jemand am Freitagnachmittag schnell eine Kundenliste sortiert, weil das freigegebene Werkzeug umständlich ist oder es keines gibt.

Auch das ist gemessen. In der repräsentativen Bitkom-Befragung von 604 Unternehmen ab 20 Beschäftigten gehen zusammengenommen **42 Prozent** davon aus, dass in ihrem Haus private KI-Werkzeuge für dienstliche Aufgaben genutzt werden. Nur 29 Prozent sind sicher, dass das nicht vorkommt — nach 37 Prozent im Vorjahr.^[2]



Die beiden linken Segmente sind belegt, die beiden mittleren beruhen auf Vermutung. Die tatsächliche Zahl liegt erfahrungsgemäß höher als die belegte.

Abb. 22 Nutzung privater KI-Werkzeuge für dienstliche Aufgaben, Einschätzung der Unternehmen selbst.

Quelle: Bitkom Research, 604 Unternehmen ab 20 Beschäftigten, Oktober 2025

Die Gegenbewegung hinkt hinterher: Nur 26 Prozent der Unternehmen stellen ihren Beschäftigten überhaupt generative KI bereit, und nur 23 Prozent haben Nutzungsregeln aufgestellt — immerhin nach 15 Prozent im Vorjahr.^[2] Solange Bereitstellung und Regeln fehlen, ist Schatten-KI kein Disziplinproblem, sondern die vorhersehbare Folge.

Daraus folgt für dich zweierlei. Erstens: Nutze, was freigegeben ist, auch wenn es unbequemer ist. Zweitens: Wenn das freigegebene Werkzeug für deine Aufgabe nicht taugt, ist das eine Information, die zurückgemeldet gehört — nicht ein Grund, es zu umgehen. Verbote allein lösen dieses Problem nie; brauchbare Werkzeuge tun es.

5.6 Wenn es doch passiert ist

Irgendwann gibt jemand etwas ein, das dort nicht hingehört. Was dann zählt, ist die Reaktionsgeschwindigkeit, nicht die Schuldfrage.

- 1 Nicht löschen, sondern melden**
Den Chatverlauf zu löschen beseitigt die Anzeige, nicht die Übertragung. Melde den Vorgang an die zuständige Stelle — Datenschutz, IT-Sicherheit oder Führungskraft — bevor du etwas veränderst.
- 2 Was offengelegt wurde, wörtlich festhalten**
Welche Kategorie, welcher Umfang, welches Werkzeug, welcher Zugang, welcher Zeitpunkt. Ohne diese fünf Angaben kann niemand einschätzen, ob eine Meldepflicht besteht.
- 3 Zugangsdaten sofort rotieren**
Ein einmal eingegebenes Passwort, Token oder Zertifikat ist verbrannt — unabhängig davon, ob der Verlauf gelöscht wurde. Umschreiben hilft hier nicht, nur Austauschen.
- 4 Ursache melden, nicht nur den Vorfall**
Warum ist es passiert? Fehlendes Werkzeug, unklare Regel, Zeitdruck? Diese Angabe verhindert den nächsten Fall — die Schuldzuweisung verhindert nur die nächste Meldung.

Abb. 23 Vier Schritte nach einer versehentlichen Offenlegung.

BEISPIEL 5.1 — EIN PROMPT, ZWEIMAL

Problematisch:

„Hier ist unsere Kundenliste aus dem CRM-Export mit Namen, E-Mail-Adressen, Umsätzen und offenen Forderungen. Schreib mir für jeden Kunden mit über 5.000 € offener Forderung ein Mahnschreiben. Anbei die Liste mit 340 Zeilen.“

Was hier passiert: Personenbezogene Daten von 340 Personen, Umsatzdaten und Forderungsstände verlassen das Unternehmen. Für die eigentliche Aufgabe — einen guten Mahntext — ist nichts davon nötig.

Unbedenklich und genauso nützlich:

„Schreib mir eine Vorlage für ein zweites Mahnschreiben an Geschäftskunden. Platzhalter: [FIRMA], [RECHNUNGSNUMMER], [BETRAG], [FÄLLIGKEIT]. Ton: sachlich, bestimmt, ohne Drohung, Geschäftsbeziehung soll erhalten bleiben. Eine Zahlungsfrist von 14 Tagen und einen Ansprechpartner nennen. Etwa 150 Wörter.“

Was sich geändert hat: Die Daten bleiben im Haus, die Serienbrieffunktion füllt die Platzhalter. Die Aufgabe ist dieselbe — und der Prompt ist obendrein besser, weil er Ton, Länge und Inhalt vorgibt.

BEISPIEL 5.2 – WENN ANONYMISIEREN NICHT REICHT

Ausgangslage: Eine Projektleiterin will sich eine Beschwerde-Antwort entwerfen lassen und ersetzt gewissenhaft alle Namen durch Platzhalter.

„[KUNDE A] beschwert sich, dass die Inbetriebnahme der Abfüllanlage in [STANDORT] dreimal verschoben wurde, zuletzt wegen eines Lieferengpasses beim Servoantrieb. Auftragsvolumen 2,3 Mio. €, Vertragsstrafe 0,3 % pro Woche.“

Das Problem: Die Namen sind ersetzt, der Fall ist trotzdem eindeutig identifizierbar — Branche, Anlagentyp, Volumen und Vertragsstrafe zusammen beschreiben genau einen Vorgang. Und die kaufmännischen Angaben sind Geschäftsgeheimnisse, die für die eigentliche Aufgabe („höflich, aber bestimmt antworten“) keine Rolle spielen.

Die Lehre: Anonymisieren heißt Kennungen entfernen. Es heißt nicht, dass der Rest unbedenklich ist. Die zweite Frage lautet immer: *Welche dieser Angaben braucht die Aufgabe überhaupt?* Meist erstaunlich wenige.

ÜBUNG 5 – PROMPT ENTSCHÄRFEN

Schreibe jeden Prompt so um, dass er ohne schutzbedürftige Daten auskommt und dieselbe Aufgabe löst. Notiere jeweils, welche Kategorie du entfernt hast.

1. „Frau Berger, Personalnummer 4471, war dieses Jahr 34 Tage krank und hat zweimal Fristen gerissen. Formuliere mir eine Gesprächsnotiz für das Personalgespräch.“
2. „Unser Angebot an die Stadtwerke liegt bei 412.000 €, die Kalkulation im Anhang zeigt 31 % Marge. Wie argumentiere ich gegen einen günstigeren Wettbewerber?“
3. „Hier ist unser Datenbank-Zugang zum Testsystem: Host, Benutzer, Passwort. Schreib mir eine Abfrage für alle Bestellungen des letzten Quartals.“
4. „Anbei der Entwurf unseres Patentantrags für das neue Dichtungsverfahren. Prüfe die Formulierung der Ansprüche auf Verständlichkeit.“

Zusatzfrage: Bei zwei der vier Fälle ist Umschreiben allein nicht genug — welche, und was fehlt jeweils noch?

Lösung: Anhang A, Übung 5 – Prompt entschärfen

ÜBUNG 5B – DEN ZUGANG BEURTEILEN

Vier Situationen. Entscheide jeweils: **zulässig**, **unzulässig** oder **abhängig von einer Angabe, die fehlt** — und benenne die fehlende Angabe.

1. Eine Kollegin nutzt das vom Unternehmen bereitgestellte Konto, um Kundenansreiben zu entwerfen — mit echten Kundennamen.
2. Ein Kollege nutzt sein privates Bezahlkonto, hat aber das Training in den Einstellungen abgeschaltet, und lädt eine interne Präsentation hoch.
3. Eine Entwicklerin lässt sich über den API-Zugang des Unternehmens Testdaten erzeugen und gibt dafür das Datenbankschema an.
4. Ein Team nutzt ein lokal betriebenes Modell auf eigenen Servern für die Auswertung von Bewerbungsunterlagen.

Lösung: Anhang A, Übung 5b – Den Zugang beurteilen

KERNAUSSAGEN

- Vier Fragen vor jeder Eingabe: Wem gehören die Daten, wer darf sie sehen, wo werden sie verarbeitet, wird damit trainiert?
- Wer eine der vier Fragen nicht beantworten kann, gibt nichts ein, sondern klärt.
- Nicht das Produkt entscheidet, sondern der Zugang: privat, Unternehmenskonto, API oder lokal.
- Sechs Techniken: anonymisieren, aggregieren, Auszug, synthetische Daten, Schema statt Inhalt, freigegebene Werkzeuge.
- Anonymisieren heißt Kennungen entfernen — nicht, dass der Rest unbedenklich ist.
- Schatten-KI entsteht dort, wo das freigegebene Werkzeug nicht taugt. Das ist eine Meldung wert.
- Offengelegte Zugangsdaten werden rotiert, nicht gelöscht.

Selbsttest Modul 5

1. Warum ist Anonymisieren in den meisten Fällen kein Qualitätsverlust?

- a) Weil anonymisierte Daten schneller verarbeitet werden.
- b) Weil Platzhalter die Antwortqualität erhöhen.
- c) Weil Modelle Namen ohnehin ignorieren.
- d) Weil die eigentliche Aufgabe meist auf Struktur und Formulierung zielt, nicht auf die Identität.

2. Ein Kollege nutzt sein privates Konto, weil das Firmenwerkzeug zu langsam ist. Was ist die angemessene Reaktion?

- a) Nichts — er nutzt sein eigenes Gerät.
- b) Ihn melden und die Sache damit abschließen.
- c) Auf das freigegebene Werkzeug hinweisen und den Mangel an die zuständige Stelle zurückspielen.
- d) Ebenfalls das private Konto nutzen, solange es keine Anweisung gibt.

3. Jemand hat versehentlich einen API-Schlüssel in ein Chatfenster eingegeben und den Verlauf gelöscht. Was ist zu tun?

- a) Nichts weiter — der Verlauf ist gelöscht.
- b) Den Schlüssel austauschen und den Vorgang melden.
- c) Den Anbieter zur Löschung auffordern und abwarten.
- d) Den Schlüssel nur austauschen, wenn er produktiv genutzt wurde.

4. Welche Angabe entscheidet, ob dienstliche Daten in ein KI-Werkzeug eingegeben werden dürfen?

- a) Ob für den genutzten Zugang eine Auftragsverarbeitung besteht und der Zweck gedeckt ist.
- b) Ob das Training in den Einstellungen abgeschaltet ist.
- c) Der Name des Anbieters.
- d) Ob es sich um ein kostenpflichtiges Konto handelt.

Ergebnisse bewerten und prüfen

ca. 75 Minuten · 2 Beispiele · 2 Übungen · Selbsttest am Ende

LERNZIELE

- Du prüfst nicht „ob es stimmt“, sondern gezielt die Teile, die stimmen müssen.
- Du kennst vier Prüffarten und weißt, wann welche greift.
- Du kannst den Prüfaufwand an den Folgen ausrichten statt am Bauchgefühl.
- Du kannst einen Beleg vollständig statt halb prüfen.
- Du kennst fünf Griffe, mit denen du Zahlen auf Belastbarkeit abklopfst.
- Du kannst einen Prüfplan in vier Schritten und eine Prüfnotiz in vier Zeilen aufschreiben.

Prüfen ist die Tätigkeit, die den Unterschied macht — und die als erste wegfällt, wenn es eilig wird. Dieses Modul macht sie praktikabel: nicht „alles prüfen“, sondern gezielt die Stellen, die tragen, mit einem Aufwand, der sich nach den Folgen richtet.

6.1 Die falsche und die richtige Frage

„Stimmt das?“ ist bei einem zweiseitigen Text keine beantwortbare Frage. Die beantwortbare Frage lautet: *Welche Aussagen in diesem Text müssen stimmen, damit meine Verwendung trägt?* Meist sind das drei bis fünf Stellen — Zahlen, Namen, Fristen, Zusagen — und der Rest ist Formulierung.

Diese Verschiebung macht Prüfung überhaupt erst praktikabel. Wer alles prüfen will, prüft am Ende nichts.

TRAGENDE AUSSAGEN	FORMULIERUNG
• Zahlen, auf die sich jemand verlässt • Fristen und Termine • Namen, Aktenzeichen, Normbezüge • Zusagen und Verpflichtungen • Kausalbehauptungen („weil X, folgt Y“)	• Reihenfolge der Argumente • Wortwahl und Tonfall • Länge der Absätze • Überschriften und Gliederung • Beispiele ohne Belegcharakter

Abb. 24 Was geprüft wird und was nicht. Der Aufwand gehört vollständig in die linke Spalte.

6.2 Vier Prüffarten

PRÜFART	FRAGE	WIE	TYPISCHER FUND
Fakten	Ist die Behauptung außerhalb dieses Textes belegbar?	Primärquelle aufrufen, nicht die Zusammenfassung	erfundener oder halb echter Beleg
Logik	Folgt der Schluss aus den Prämissen?	Argumentation Schritt für Schritt nachvollziehen	Sprung zwischen zwei richtigen Sätzen
Vollständigkeit	Was fehlt, das ich erwartet hätte?	Gegen eine Gliederung oder Checkliste prüfen	der Wettbewerber, den nur du kennst
Passung	Trifft es Adressat, Ton, Format, Kontext?	Laut lesen; aus Sicht des Empfängers lesen	korrekter Inhalt im falschen Register

Die vierte wird unterschätzt. Ein sachlich einwandfreier Text im falschen Ton richtet in einer Kundenbeziehung mehr Schaden an als ein Zahlendreher. Und die dritte ist die einzige, die man ohne Vorwissen nicht durchführen kann: Fehlendes bemerkt nur, wer weiß, was da sein müsste.

6.3 Prüfaufwand nach Folgen

Drei Fragen bestimmen, wie gründlich du sein musst:

- **Geht es aus dem Haus?** Interne Notiz und Kundenschreiben sind nicht dasselbe.
- **Entscheidet jemand danach?** Wenn ja, prüfst du die Entscheidungsgrundlagen, nicht den Text.
- **Ist es umkehrbar?** Was zurückgeholt werden kann, darf günstiger geprüft werden.

Stufe 0 – Durchlesen

Dreimal „nein“. Intern, keine Entscheidung, umkehrbar. Aufwand: 1–2 Minuten. Beispiel: Ideensammlung für ein Teammeeting.

Stufe 1 – Tragende Stellen

Einmal „ja“. Die drei bis fünf tragenden Aussagen werden geprüft, der Rest gelesen. Aufwand: 10–20 Minuten.

Stufe 2 – Vollprüfung

Zweimal „ja“. Jede tragende Aussage gegen die Primärquelle, Vollständigkeit gegen eine Gliederung. Aufwand: 30–60 Minuten.

Stufe 3 – Vier Augen

Dreimal „ja“. Vollprüfung plus zweite Person mit Fachkompetenz, dokumentiert. Aufwand: Stunden – und angemessen.

Abb. 25 Vier Prüfstufen. Die Zuordnung nimmst du vor dem Erzeugen vor, nicht danach – sonst entscheidet die verbleibende Zeit über die Gründlichkeit.

6.4 Belege richtig prüfen

Ein häufiger Fehler ist die halbe Prüfung: Man sucht den genannten Titel, findet ihn, und hakt ab. Damit ist nur gezeigt, dass die Quelle *existiert* — nicht, dass sie die Aussage *trägt*. Erfundene Belege sind oft halb echt: echter Autor, echte Zeitschrift, erfundene Aussage.

1

Existiert die Quelle?

Titel, Autor, Jahr, Fundstelle auffindbar? Findest du nichts, ist die Aussage nicht verwendbar — nicht „wahrscheinlich richtig“.

2

Sagt sie, was behauptet wird?

Die genannte Stelle im Original aufrufen und lesen. Hier scheitern die meisten halb echten Belege: Die Quelle existiert und sagt etwas anderes.

3

Ist sie aktuell genug für meinen Zweck?

Erhebungsjahr, Rechtsstand, Marktabgrenzung. Eine korrekte Zahl von 2021 ist in einer Entscheidungsvorlage von 2026 unter Umständen falsch.

Abb. 26 Erst nach dem dritten Schritt ist ein Beleg geprüft. Die halbe Prüfung bestätigt genau die Hälfte, die stimmt.

6.5 Zahlen prüfen: fünf Griffe

Zahlen sind die Aussagen, die am weitesten wandern und am wenigsten hinterfragt werden (Beispiel 3.2). Fünf Griffe fangen das meiste ab:

1. **Größenordnung.** Passt die Zahl zu dem, was du sonst weißt? Ein Markt, der größer ist als die Branche, in der er liegt, ist ein Rechenfehler oder eine Abgrenzungsfrage.
2. **Einheit und Bezug.** Prozent wovon? Euro oder Tausend Euro? Pro Jahr oder kumuliert? Der häufigste stille Fehler steckt im Nenner.
3. **Erhebungsjahr.** Nicht das Veröffentlichungsjahr. Eine 2026 erschienene Studie kann Daten von 2023 enthalten.
4. **Abgrenzung.** „Markt für KI-Software“ bedeutet je nach Studie sehr Verschiedenes. Ohne Abgrenzung ist eine Marktzahl unbrauchbar.
5. **Summenprobe.** Addieren sich die Anteile auf 100 %? Passen Teilsummen zur Gesamtsumme? Zwei Additionen decken erstaunlich viel auf.

6.6 Die Zweitmeinung

Weil eine Antwort eine Stichprobe ist (Modul 2.4), hilft eine zweite. Stelle dieselbe Frage anders formuliert, oder lass ein anderes System antworten. Weichen die Antworten in einem tragenden Punkt ab, hast du deine Prüfstelle gefunden. Stimmen sie überein, ist das ein schwaches Indiz — beide Systeme können denselben Irrtum aus denselben Quellen haben. **Übereinstimmung ersetzt keine Primärquelle.**

6.7 Was dokumentiert wird

Ab Prüfstufe 2 lohnt eine kurze Notiz — vier Zeilen, keine Formularschlacht. Sie kostet zwei Minuten und beantwortet die Frage, die drei Monate später kommt: *Wie sind wir eigentlich zu dieser Zahl gekommen?*

PRÜFNOTIZ – VIER ZEILEN

- **Was** wurde erzeugt (Dokument, Datum, System und Version)
- **Welche** Aussagen als tragend eingestuft wurden
- **Wie** geprüft wurde (Primärquelle, Werkzeug, zweite Person) und mit welchem Ergebnis
- **Was** offen blieb — und wer damit lebt

Die letzte Zeile ist die wichtigste. Ein offener Punkt, den jemand bewusst trägt, ist ein Risiko. Ein offener Punkt, den niemand kennt, ist ein Vorfall in Vorbereitung.

BEISPIEL 6.1 – EIN PRÜFPLAN IN VIER SCHRITTEN

Ausgangslage: Du hast dir einen einseitigen Marktüberblick erstellen lassen. Er soll in eine Entscheidungsvorlage für die Geschäftsführung. Enthalten sind: eine Marktgrößenangabe mit Quelle, drei Wettbewerber mit Marktanteilen, eine Wachstumsprognose und eine Empfehlung.

Schritt 1 — Tragende Aussagen markieren. Die Marktgröße, die drei Marktanteile und die Wachstumsrate. Fünf Zahlen. Die Empfehlung ist deine, nicht die des Systems.

Schritt 2 — Fakten prüfen. Für jede der fünf Zahlen: Quelle aufrufen, Zahl im Original finden, Erhebungsjahr und Marktabgrenzung notieren. Erfahrungsgemäß ist hier die Marktabgrenzung der eigentliche Fund — „Markt für KI-Software“ bedeutet je nach Studie sehr Verschiedenes.

Schritt 3 — Vollständigkeit. Ein Wettbewerber, den du aus dem Tagesgeschäft kennst, fehlt. Das ist typisch: Was in öffentlichen Quellen wenig vorkommt, taucht nicht auf. Ergänzen.

Schritt 4 — Passung. Ist die Marktabgrenzung die, über die eure Geschäftsführung entscheidet? Wenn die Studie den Weltmarkt meint und ihr über den DACH-Raum entscheidet, sind alle fünf Zahlen richtig und die Vorlage trotzdem unbrauchbar.

Aufwand: rund 40 Minuten. **Alternative:** eine Vorlage, deren Zahlen in der Sitzung auseinanderfallen.

BEISPIEL 6.2 – DER LOGIKSPRUNG ZWISCHEN ZWEI RICHTIGEN SÄTZEN

Aus einer erzeugten Analyse: „Der Anteil der Unternehmen mit KI-Nutzung ist im vergangenen Jahr von 17 auf 41 Prozent gestiegen. Unternehmen ohne KI-Einsatz verlieren damit messbar an Wettbewerbsfähigkeit.“

Satz 1 ist belegt und korrekt (Modul 1.4). **Satz 2** ist eine Behauptung, die aus Satz 1 nicht folgt: Verbreitung ist kein Wirksamkeitsbeleg. Aus derselben Erhebung geht sogar hervor, dass nur 52 Prozent der Anwender einen messbaren Beitrag benennen können.

Was die Faktenprüfung hier nicht leistet: Beide Sätze bestehen sie — der erste, weil er stimmt, der zweite, weil er gar keine überprüfbare Fundstelle nennt. Der Fehler ist nur mit der **Logikprüfung** zu finden: Folgt der Schluss aus der Prämisse?

Die Lehre: Die gefährlichsten Stellen stehen oft zwischen zwei richtigen Sätzen. Prüfe nicht nur die Aussagen, sondern die Wörter, die sie verbinden: *damit, folglich, deshalb, somit.*

ÜBUNG 6 – PRÜFPLAN SCHREIBEN

Du hast dir für einen internen Newsletter einen 400-Wörter-Beitrag erstellen lassen: „Drei Änderungen im Datenschutz, die unser Vertrieb 2026 kennen muss.“ Der Text nennt zwei Gesetzesänderungen mit Datum, ein Gerichtsurteil mit Aktenzeichen, eine Bußgeldhöhe und drei Handlungsempfehlungen.

Schreibe einen Prüfplan:

1. Welche Aussagen sind tragend? Liste sie auf.
2. Welche Prüfmethode greift bei welcher Aussage?
3. In welcher Reihenfolge prüfst du, und warum in dieser?
4. Bei welcher Aussage würdest du eine zweite Person hinzuziehen — und warum genau bei dieser?
5. Welche Prüfstufe nach 6.3 wählst du, und wie begründest du sie?

Zusatzfrage: Der Text geht nur intern raus. Ändert das deinen Plan? Begründe.

Lösung: Anhang A, Übung 6 – Prüfplan schreiben

ÜBUNG 6B – HALBE PRÜFUNG ERKENNEN

Vier Prüfprotokolle. Entscheide für jedes: **vollständig geprüft** oder **halb geprüft** — und benenne, welcher Schritt aus 6.4 bzw. welcher Griff aus 6.5 fehlt.

1. „Die zitierte Studie habe ich gefunden, Titel und Autoren stimmen. Haken.“
2. „Ich habe die Quelle aufgerufen und die Aussage im Abschnitt 3 im Wortlaut wiedergefunden. Haken.“
3. „Ich habe die Zahl in der Primärquelle gefunden: 23,4 % Marktanteil, Erhebung 2024. Passt zu unserer Aussage. Haken.“
4. „Beide Systeme geben denselben Wert aus, das habe ich zweimal geprüft. Haken.“

Lösung: Anhang A, Übung 6b – Halbe Prüfung erkennen

KERNAUSSAGEN

- Nicht „stimmt das?“, sondern „welche Stellen müssen stimmen?“.
- Vier Prüfarten: Fakten, Logik, Vollständigkeit, Passung.
- Der Aufwand richtet sich nach Reichweite, Entscheidungswirkung und Umkehrbarkeit — vier Stufen von Durchlesen bis Vier-Augen-Prinzip.
- Eine Quelle ist erst geprüft, wenn sie existiert, die Aussage trägt und aktuell genug ist.
- Bei Zahlen: Größenordnung, Einheit, Erhebungsjahr, Abgrenzung, Summenprobe.
- Übereinstimmung zweier Systeme ersetzt keine Primärquelle.
- Die wichtigste Zeile der Prüfnotiz ist die, die den offenen Punkt benennt.

Selbsttest Modul 6

1. Du findest die zitierte Studie und der Titel stimmt. Was fehlt noch zur vollständigen Belegprüfung?

- a) Ob die Studie auf Deutsch verfügbar ist.
- b) Ob die Autoren in derselben Fachrichtung arbeiten.
- c) Nichts, der Beleg ist geprüft.
- d) Ob die Studie die behauptete Aussage tatsächlich trägt und aktuell genug ist.

2. Zwei verschiedene Systeme geben dieselbe Zahl aus. Was folgt daraus?

- a) Ein schwaches Indiz — beide können denselben Irrtum aus denselben Quellen haben.
- b) Die Zahl ist mit Sicherheit falsch.
- c) Eine Primärquelle ist dann nicht mehr nötig.
- d) Die Zahl ist bestätigt.

3. Welche Prüfmethode deckt einen falschen Schluss zwischen zwei zutreffenden Sätzen auf?

- a) Faktenprüfung.
- b) Logikprüfung.
- c) Vollständigkeitsprüfung.
- d) Passungsprüfung.

4. Ein Text geht extern raus, jemand entscheidet danach, und die Wirkung ist nicht umkehrbar. Angemessen ist

- a) einmaliges Durchlesen.
- b) eine Stichprobe der Zahlen.
- c) vollständige Prüfung jeder tragenden Aussage, möglichst durch eine zweite Person.
- d) eine zweite Antwort desselben Systems als Absicherung.

Verantwortung, Aufsicht und Kennzeichnung

ca. 75 Minuten · 2 Beispiele · 2 Übungen · Selbsttest am Ende

LERNZIELE

- Du kannst benennen, wer für einen KI-gestützten Output einsteht.
- Du unterscheidest die Rollen Anbieter und Betreiber in ihrer praktischen Bedeutung.
- Du kennst drei Stufen menschlicher Aufsicht und erkennst wirkungslose Aufsicht an drei Merkmalen.
- Du weißt, wann und wie KI-Beteiligung gekennzeichnet wird.
- Du kannst einen Fehlerbericht schreiben, mit dem jemand arbeiten kann.
- Du kannst die typischen Rollen im Betrieb und ihre häufigsten Fehler benennen.

Dieses Modul klärt, wer einsteht. Nicht als moralische Frage, sondern als praktische: In jedem Prozess, in dem am Ende niemand zuständig ist, entstehen die Vorfälle. Dazu kommen zwei Werkzeuge, die im Alltag ständig gebraucht und selten beherrscht werden — wirksame Aufsicht und Kennzeichnung.

7.1 Werkzeuge tragen keine Verantwortung

„Das hat die KI so geschrieben“ ist keine Entlastung — weder arbeitsrechtlich, noch vertraglich, noch fachlich. Wer einen Output verwendet, verantwortet ihn, genauso wie bei einer Vorlage aus dem Intranet oder einem Entwurf einer Kollegin. Das Werkzeug ändert die Zurechnung nicht.

Das ist keine Härte, sondern die Voraussetzung dafür, dass diese Systeme überhaupt eingesetzt werden können. Ein Prozess, in dem am Ende niemand einsteht, ist kein Prozess.

7.2 Anbieter und Betreiber

Die europäische Regulierung unterscheidet vor allem zwei Rollen, und der Unterschied entscheidet über die Pflichten.

ANBIETER	BETREIBER
• entwickelt ein KI-System oder lässt es entwickeln • bringt es unter eigenem Namen auf den Markt • trägt die weitreichenden Produktpflichten • Beispiel: Softwarehaus, das ein Prüfwerkzeug verkauft	• setzt ein System unter eigener Verantwortung ein • Rolle der allermeisten Unternehmen • Pflichten vor allem zu Nutzung, Aufsicht und Transparenz • Beispiel: Versicherung, die ein zugekauftes Werkzeug nutzt

Abb. 27 Zwei Rollen, zwei Pflichtenkataloge.

Die meisten Unternehmen sind Betreiber. Wichtig ist der **Übergang**: Wer ein zugekauftes System wesentlich verändert oder unter eigenem Namen weitergibt, kann zum Anbieter werden — mit deutlich weiteren Pflichten. Für

dich als Beschäftigte heißt das vor allem: Es macht einen Unterschied, ob ihr ein Werkzeug *benutzt* oder ein Produkt daraus *baut*. Diese Frage gehört an den Anfang eines Vorhabens, nicht ans Ende.

7.3 Drei Stufen der Aufsicht

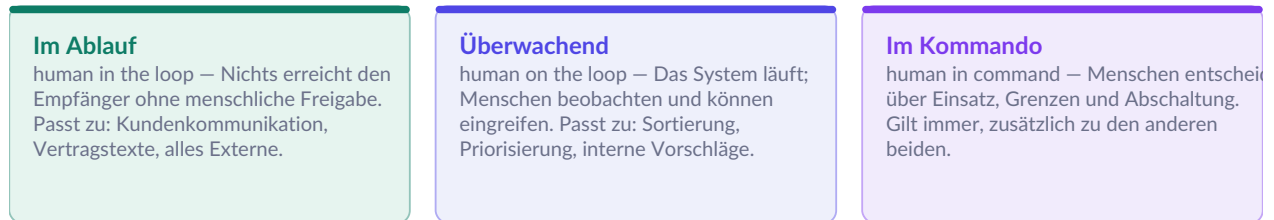


Abb. 28 Drei Stufen. Die dritte ist keine Alternative zu den ersten beiden, sondern die Ebene darüber.

7.4 Aufsicht, die keine ist

Die häufigste Schwachstelle ist nicht fehlende, sondern **wirkungslose** Aufsicht: Freigaben, die im Sekundentakt erteilt werden, weil das System „ohnehin meistens recht hat“. Formal ist ein Mensch beteiligt, praktisch nicht.

DREI MERKMALE WIRKSAMER AUFSICHT

- 1. Kompetenz.** Die prüfende Person kann den Output fachlich beurteilen. Wer eine Rechtsauskunft freigibt, ohne die Rechtsfrage beantworten zu können, gibt nichts frei — er unterschreibt.
- 2. Zeit.** Sie hat die Zeit dafür, und zwar eingeplant. Eine Freigabe, die im Durchschnitt vier Sekunden dauert, ist keine.
- 3. Befugnis und Kultur.** Sie darf tatsächlich ablehnen, und Ablehnung hat keine Nachteile. Wo Widerspruch Aufwand bedeutet und Zustimmung nicht, entsteht die Zustimmung von selbst.

Fehlt eines der drei, ist die Aufsicht dekorativ — und das fällt spätestens dann auf, wenn jemand die Freigabezeiten auswertet.

Ein einfacher Selbsttest für jede Freigabestelle: *Wann wurde hier zuletzt etwas abgelehnt?* Wenn die Antwort „nie“ lautet und der Prozess seit Monaten läuft, ist die Aufsicht nicht wirksam — sie ist eine Signatur.

7.5 Kennzeichnung

Kennzeichnung dient nicht der Selbstanzeige, sondern der Einordnung: Der Empfänger soll wissen, wie er das vor sich Liegende zu lesen hat.

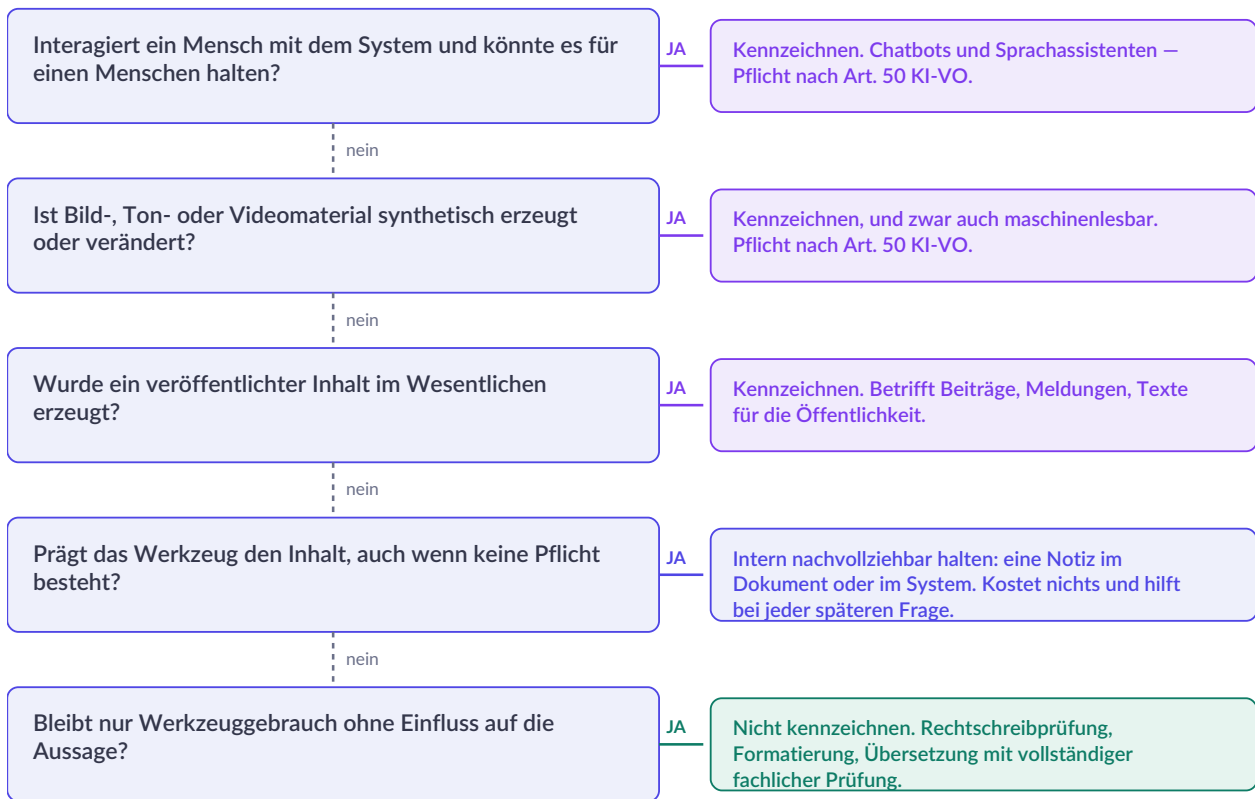


Abb. 29 Kennzeichnungs-Schnelltest. Die Grenze verläuft dort, wo das Werkzeug den Inhalt prägt.

7.6 Ein guter Fehlerbericht

Fehler zu melden ist der Mechanismus, durch den ein System besser wird. „Die KI erzählt Unsinn“ ist kein Fehlerbericht. Es ist eine Stimmung. Ein brauchbarer Bericht enthält sechs Angaben:

ANGABE	WARUM SIE GEBRAUCHT WIRD
System und Version	Anbieter tauschen Modelle im laufenden Betrieb; ohne Version ist nichts reproduzierbar
Zeitpunkt	grenzt ein, welche Modellfassung und welche Datenbasis aktiv war
Eingabe im Wortlaut	so weit zulässig — ohne sie kann niemand nachstellen
Ausgabe im Wortlaut	eine Paraphrase verliert genau das Detail, um das es geht
Was du erwartet hättest	unterscheidet Fehler von Missverständnis über den Zweck
Folge	ist etwas passiert oder wurde es vorher bemerkt? Bestimmt die Dringlichkeit

Wer diese sechs Angaben liefert, macht aus einem Ärgernis eine Verbesserung. Wer sie nicht liefert, erzeugt einen Ticketvorgang, der geschlossen wird.

7.7 Wer macht was

Eine schlanke Rollenverteilung, die in den meisten Häusern trägt:

ROLLE	VERANTWORTET	TYPISCHER FEHLER
Anwendende Person	Eingabe, Prüfung, Verwendung des Ergebnisses	verlässt sich auf eine Freigabestelle, die es nicht gibt
Freigebende Person	fachliche Beurteilung vor der Weitergabe	gibt frei, ohne beurteilen zu können (7.4)
Fachbereichsleitung	Zuschnitt der Aufgabe, Aufsichtsstufe, Prüfstufe	legt den Zuschnitt erst fest, nachdem das Werkzeug läuft
Zentrale Stelle (IT, Datenschutz, Recht)	Zugänge, Verträge, Rechtsebenen, Meldewege	erfährt vom Vorhaben erst kurz vor dem Start
Arbeitnehmervertretung	Beteiligung bei überwachungsgeeigneten Systemen	wird spät eingebunden — kostet Monate statt Wochen (Modul 8.6)

BEISPIEL 7.1 – DREI SITUATIONEN, DREI ZUORDNUNGEN

A — Ein Mitarbeiter versendet ein Angebot mit einer vom Assistenten erfundenen Lieferzeit. Der Kunde nimmt an. *Verantwortlich ist der Mitarbeiter* — er hat versendet. Fehlt eine Freigaberegeln für Angebote, liegt zusätzlich ein Organisationsmangel vor. Beides ist zu beheben, und das zweite ist das wichtigere.

B — Ein Sortiersystem priorisiert Kundenanfragen; eine Beschwerde bleibt drei Wochen liegen. *Aufsichtsstufe „überwachend“ war vorgesehen, hat aber nicht gegriffen* — niemand hat die niedrig priorisierten Vorgänge je angesehen. Die Aufsicht war formal vorhanden und praktisch nicht.

C — Eine Abteilung baut aus einem zugekauften Modell ein Werkzeug, das sie anderen Unternehmen unter eigenem Namen anbietet. *Rollenwechsel*: aus dem Betreiber ist ein Anbieter geworden, mit deutlich weiteren Pflichten. Diese Frage gehört vor die erste Zeile Code.

BEISPIEL 7.2 – DIE FREIGABE, DIE KEINE WAR

Ausgangslage: Ein Dienstleister lässt Angebotsschreiben erzeugen. Regel: Jedes Schreiben wird vor dem Versand von der Teamleitung freigegeben. Nach acht Monaten kommt es zu einem Streitfall über eine zugesagte Verfügbarkeit.

Der Befund: In acht Monaten wurden 1.240 Schreiben freigegeben und keines abgelehnt. Die durchschnittliche Zeit zwischen Öffnen und Freigabe beträgt sechs Sekunden. Die Teamleitung hat parallel die volle Vertriebsverantwortung — es war nie Zeit dafür eingeplant.

Was hier fehlte: nicht die Regel und nicht die Person, sondern zwei der drei Merkmale aus 7.4 — Zeit und Kultur. Ablehnen hätte den Versand verzögert und Rückfragen erzeugt; Zustimmung kostete sechs Sekunden.

Die Reparatur: Freigabe nur noch für Schreiben mit Zusagen zu Terminen, Preisen oder Verfügbarkeit — dafür mit eingeplanter Zeit und einer Prüfliste. Alle übrigen Schreiben laufen ohne Freigabe. *Weniger Aufsicht an weniger Stellen, dafür wirksam* ist besser als flächendeckende Aufsicht, die nirgends stattfindet.

ÜBUNG 7 – ZUORDNEN UND NACHSCHÄRFEN

Bestimme für jede Situation: (a) wer verantwortlich ist, (b) welche Aufsichtsstufe angemessen wäre, (c) welche eine Maßnahme du zuerst ergreifen würdest.

1. Ein Chatbot auf der Website beantwortet Fragen zu Gewährleistung und nennt eine Frist, die von den AGB abweicht.
2. Ein Assistent erzeugt Protokolle aus Besprechungsmitschnitten; die Protokolle gehen ohne Durchsicht an alle Teilnehmenden.
3. Eine Fachabteilung nutzt ein Modell zur Vorauswahl von Lieferantenangeboten; die Geschäftsführung entscheidet auf Grundlage der Vorauswahl.
4. Ein Werkzeug erzeugt Produktbilder für den Onlineshop, in denen keine realen Personen abgebildet sind.

Zusatzaufgabe: Schreibe zu Situation 1 einen vollständigen Fehlerbericht nach den sechs Angaben aus 7.6. Erfinde die fehlenden Details plausibel.

Lösung: Anhang A, Übung 7 – Zuordnen und nachschärfen

ÜBUNG 7B – AUFSICHT BEWERTEN

Vier Freigabestellen. Entscheide für jede: **wirksam** oder **dekorativ** — und benenne, welches der drei Merkmale aus 7.4 fehlt.

1. Die Rechtsabteilung gibt Vertragstexte frei; im letzten Quartal wurden 12 von 180 Texten zurückgewiesen.
2. Die Schichtleitung bestätigt jede erzeugte Wartungsempfehlung per Klick; vorgesehen sind 30 Sekunden je Empfehlung, tatsächlich vergehen im Schnitt 3 Sekunden.
3. Eine Werkstudentin gibt die Antwortentwürfe des Kundenservice frei; sie hat 20 Minuten je Vorgang und ein ausführliches Handbuch.
4. Der Bereichsleiter gibt Pressemitteilungen frei; er hat Zeit, ist fachlich versiert, und die Kommunikationsabteilung notiert Ablehnungen als „Verzögerung durch Leitung“ in ihrer Kennzahl.

Lösung: Anhang A, Übung 7b – Aufsicht bewerten

KERNAUSSAGEN

- Wer den Output verwendet, verantwortet ihn. Das Werkzeug ändert die Zurechnung nicht.
- Betreiber und Anbieter haben verschiedene Pflichten; der Übergang passiert schneller als gedacht.
- Wirksame Aufsicht braucht Kompetenz, Zeit und die Befugnis abzulehnen.
- Der beste Test für eine Freigabestelle: Wann wurde hier zuletzt etwas abgelehnt?
- Gekennzeichnet wird, wo das Werkzeug den Inhalt prägt oder Menschen sich täuschen könnten.
- Ein Fehlerbericht braucht sechs Angaben, nicht eine Stimmung.

Selbsttest Modul 7

1. Woran erkennt man wirkungslose menschliche Aufsicht am zuverlässigsten?

- a) Daran, dass Freigaben so schnell erteilt werden, dass eine Prüfung nicht stattgefunden haben kann.
- b) Daran, dass die Aufsicht nicht dokumentiert ist.
- c) Daran, dass mehrere Personen beteiligt sind.
- d) An der Zahl der eingesetzten Systeme.

2. In welchem Fall ist eine Kennzeichnung am ehesten entbehrlich?

- a) Ein veröffentlichter Blogbeitrag, der im Wesentlichen erzeugt wurde.
- b) Ein Chatbot, der Kundenanfragen beantwortet.
- c) Eine Rechtschreib- und Grammatikprüfung an einem selbst geschriebenen Text.
- d) Ein Werbebild, das synthetisch erzeugt wurde.

3. Eine Abteilung baut aus einem zugekauften Modell ein Werkzeug und bietet es anderen Unternehmen unter eigenem Namen an. Was folgt daraus?

- a) Nichts — das Modell stammt weiterhin vom ursprünglichen Anbieter.
- b) Die Abteilung kann zum Anbieter werden, mit deutlich weiteren Pflichten.
- c) Es liegt automatisch ein Hochrisikosystem vor.
- d) Die Pflichten des ursprünglichen Anbieters gehen vollständig über.

4. Welche Angabe gehört *nicht* zu den sechs Bestandteilen eines guten Fehlerberichts?

- a) Der Zeitpunkt.
- b) Die Eingabe im Wortlaut.
- c) Die erwartete Ausgabe.
- d) Die Einschätzung, wer schuld ist.

Der rechtliche Rahmen in Grundzügen

ca. 90 Minuten · 2 Beispiele · 2 Übungen · Selbsttest am Ende

LERNZIELE

- Du kannst vier Rechtsebenen unterscheiden, die bei KI-Vorhaben zusammenwirken.
- Du kennst den risikobasierten Ansatz der KI-Verordnung und die aktuellen Anwendungsfristen einschließlich der Verschiebungen von Juli 2026.
- Du kannst die KI-Kompetenzpflicht korrekt einordnen — auch das, was sie nicht verlangt.
- Du kennst die vier Transparenzfälle nach Art. 50 und weißt, welcher dich betrifft.
- Du weißt, bei welchen vier Fragen ein Vorhaben früh Fachrat braucht.

VORBEMERKUNG

Dieses Modul gibt einen Überblick, damit du die richtigen Fragen stellen und erkennen kannst, wann eine Fachperson gebraucht wird. Es ist **keine Rechtsberatung**. **Rechtsstand: August 2026**. Einzelne Fristen in diesem Feld haben sich zuletzt mehrfach verschoben; prüfe bei konkreten Vorhaben den aktuellen Stand.

8.1 Vier Ebenen, nicht ein Gesetz

Ein KI-Vorhaben berührt regelmäßig vier Ebenen gleichzeitig. Wer nur eine prüft, übersieht die anderen drei.

EBENE	LEITFRAGE	WER IM HAUS ZU-STÄNDIG IST
Datenschutz	Werden personenbezogene Daten verarbeitet — und auf welcher Grundlage, zu welchem Zweck, durch wen?	Datenschutzbeauftragte
KI-Verordnung	Welche Rolle haben wir, in welche Risikokategorie fällt das System, welche Transparenzpflichten greifen?	Compliance / Recht
Urheber- und Persönlichkeitsrecht	Wessen Werke gehen hinein, wem gehört, was herauskommt, sind Personen abgebildet?	Recht / Kommunikation
Branchen- und Arbeitsrecht	Gelten besondere Regeln für unseren Sektor — und ist eine Beteiligung der Arbeitnehmervertretung erforderlich?	Personal / Betriebsrat

8.2 Die KI-Verordnung im Grundsatz

Die europäische KI-Verordnung reguliert nach **Risiko**, nicht nach Technologie. Vereinfacht:

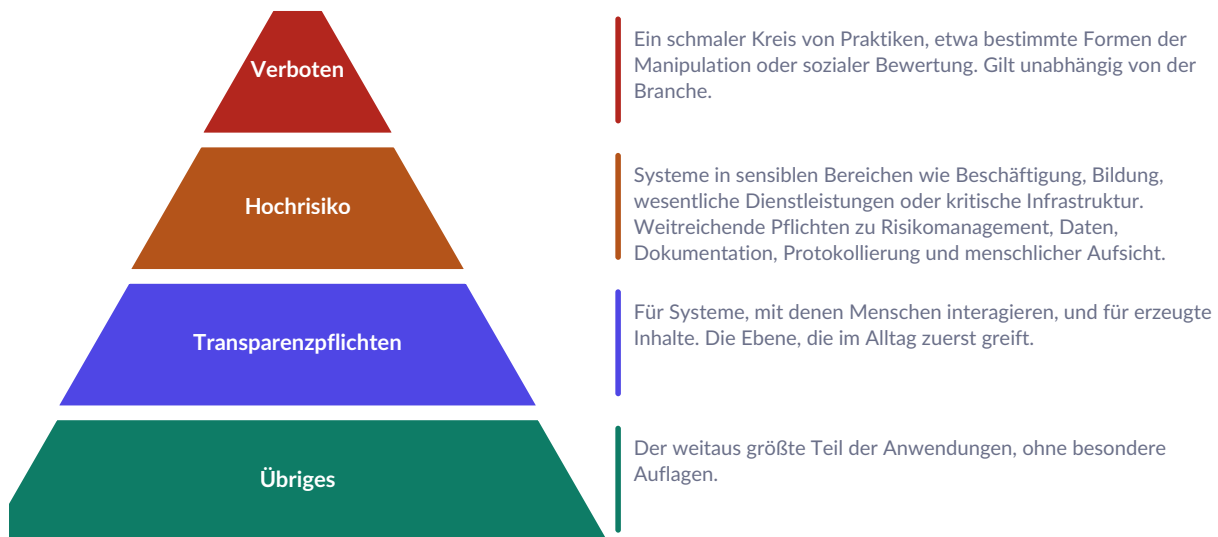


Abb. 30 Der risikobasierte Ansatz. Die meisten Anwendungen im Arbeitsalltag liegen in den unteren beiden Ebenen.

Entscheidend sind die Fristen — und genau die haben sich zuletzt geändert. Die Änderungsverordnung (EU) 2026/1744 („Digital Omnibus on AI“) ist zum **27. Juli 2026** wirksam geworden und hat die Anwendbarkeit der Hochrisikopflichten verschoben.^[7]

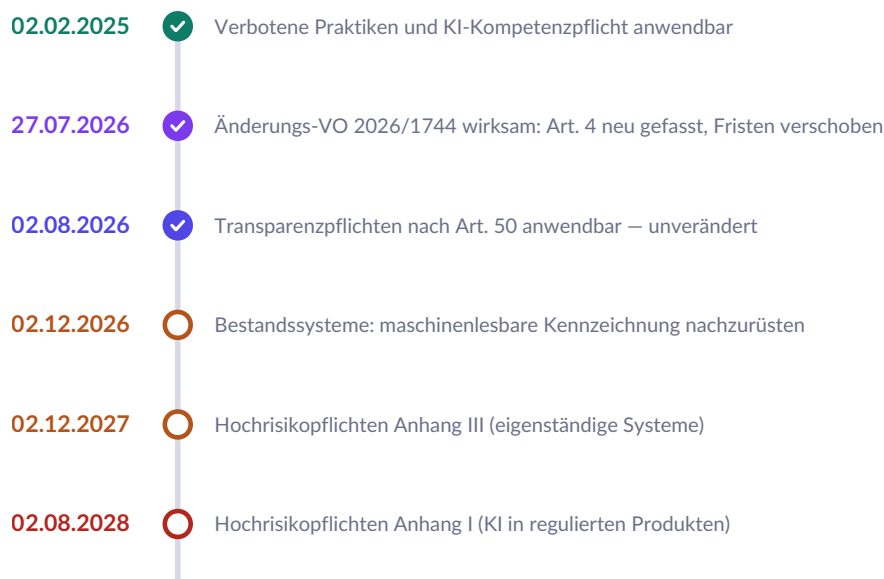


Abb. 31 Zeitstrahl der Anwendbarkeit, Stand August 2026. Ausgefüllte Punkte sind bereits eingetreten.

Quelle: VO (EU) 2024/1689 in der Fassung der VO (EU) 2026/1744

Für die tägliche Arbeit sind die Transparenzpflichten das Praktische: Kennzeichnung von Chatbots und von erzeugten Inhalten, wie in Modul 7.5 beschrieben. Sie sind seit dem 2. August 2026 anwendbar und wurden *nicht* verschoben — verschoben wurden ausschließlich die Hochrisikopflichten.

EIN MUSTER, DAS DIR BEGEGNEN WIRD

Jedes Sprachmodell mit einem Wissensstand vor Juli 2026 nennt souverän die alten Termine — insbesondere den 2. August 2026 für Hochrisikopflichten. Das ist das Muster „veraltete Information“ aus Modul 3.2 in Reinform, und es betrifft ausgerechnet die Angaben, die in Projektpläne wandern. Regulatorische Fristen sind der Bereich, in dem du grundsätzlich die Primärquelle aufrufst.

8.3 Die KI-Kompetenzpflicht – und was sie nicht ist

Die KI-Verordnung verpflichtet Anbieter und Betreiber, sich um die KI-Kompetenz der Personen zu kümmern, die mit ihren Systemen arbeiten. Diese Pflicht gilt seit dem 2. Februar 2025 und wurde durch die Änderungsverordnung **mit Wirkung zum 27. Juli 2026 neu gefasst**: Aus einer Ergebnispflicht wurde eine **Bemühenspflicht**. Verlangt wird, *Maßnahmen zur Förderung* von KI-Kompetenz zu ergreifen; ausdrücklich **nicht** verlangt wird, für eine einzelne Person ein bestimmtes Kompetenzniveau zu garantieren.^[7]

WAS VERLANGT WIRD

WAS NICHT VERLANGT WIRD

angemessene Maßnahmen zur Förderung von KI-Kompetenz

ein garantiertes Kompetenzniveau je Person

Zuschnitt auf Rolle, Vorkenntnisse und Nutzungskontext

ein von der Kommission vorgegebenes Curriculum

interne Nachweise über durchgeführte Maßnahmen

ein Schulungszertifikat für jede beschäftigte Person

Berücksichtigung des Risikoniveaus der eingesetzten Systeme

eine jährliche Prüfung durch eine externe Stelle

Zwei Missverständnisse, die dir begegnen werden:

„**Wir brauchen für alle ein Zertifikat.**“ Nein. Ein Schulungszertifikat ist nicht verlangt; interne Aufzeichnungen über durchgeführte Schulungen und Maßnahmen genügen als Nachweis. Ein Zertifikat kann eine sinnvolle Maßnahme sein — Pflicht ist es nicht, und wer es als Pflicht darstellt, wirbt mit einer Rechtspflicht, die es nicht gibt.

„**Es gibt einen vorgeschriebenen Lehrplan.**“ Nein. Vorgegeben ist kein Curriculum, sondern woran sich der Zuschnitt orientieren muss: das allgemeine Verständnis von KI in der Organisation, die eigene Rolle als Anbieter oder Betreiber, das Risikoniveau der tatsächlich eingesetzten Systeme sowie die Vorkenntnisse und der Nutzungskontext der jeweiligen Personengruppe. Genau darauf ist dieser Kurs aufgebaut.

8.4 Transparenzpflichten im Alltag

Die Ebene, die dich am häufigsten betrifft. Vier Fälle, drei Zuständigkeiten:



Abb. 32 Die vier Transparenzfälle nach Art. 50 KI-VO, anwendbar seit 2. August 2026.

8.5 Datenschutz in sechs Stichworten

- **Rechtsgrundlage:** Jede Verarbeitung personenbezogener Daten braucht eine. Einwilligung ist eine Möglichkeit, im Beschäftigungsverhältnis oft nicht die praktikabelste.
- **Zweckbindung:** Für Zweck A erhobene Daten dürfen nicht ohne Weiteres für Zweck B genutzt werden — und „wir trainieren damit ein Modell“ ist ein neuer Zweck.
- **Auftragsverarbeitung:** Wer Daten in einem fremden Dienst verarbeitet, braucht dafür eine vertragliche Grundlage. Genau hier liegt der Unterschied zwischen Privat- und Unternehmenskonto (Modul 5.3).
- **Datenminimierung:** Nur die Daten, die die Aufgabe braucht. Das ist gleichzeitig die wirksamste der sechs Techniken aus Modul 5.4.
- **Betroffenenrechte:** Auskunft, Berichtigung, Löschung — sie gelten auch dann, wenn Daten in einem KI-gestützten Prozess stecken. Wer nicht weiß, wo die Daten liegen, kann sie nicht löschen.
- **Automatisierte Entscheidungen:** Entscheidungen mit erheblicher Wirkung, die ausschließlich automatisiert ergehen, unterliegen besonderen Anforderungen.

8.6 Urheberrecht und Mitbestimmung

Beim **Urheberrecht** stellen sich drei Fragen: Darfst du fremde Werke eingeben? Wem gehören die Ausgaben, und sind sie überhaupt schutzfähig? Kann eine Ausgabe fremde Rechte verletzen, weil sie einem geschützten Werk zu nahe kommt? Die Rechtslage ist in Bewegung; die praktische Regel lautet: fremde Werke unter Vertraulichkeit nicht eingeben, Ausgaben nicht ungeprüft als eigenes Werk behandeln.

Bei der **Mitbestimmung** gilt in Deutschland ein Punkt, der viele Vorhaben aufhält: Führt ein Arbeitgeber technische Einrichtungen ein, die dazu *geeignet* sind, Leistung oder Verhalten von Beschäftigten zu überwachen, hat der Betriebsrat mitzubestimmen. Auf die Absicht kommt es nicht an, auf die Eignung schon — und viele KI-gestützte Werkzeuge sind dazu geeignet, weil sie nebenbei Bearbeitungszeiten, Häufigkeiten und Verläufe erfassen.

DIE RECHNUNG, DIE SICH IMMER WIEDERHOLT

Wer die Arbeitnehmervertretung früh einbindet, verliert Wochen. Wer es spät tut, verliert Monate. Der Unterschied liegt nicht in der Verhandlungsführung, sondern darin, ob eine bereits gebaute Lösung verteidigt oder eine geplante gemeinsam zugeschnitten wird.

8.7 Vier Fragen vor jedem Vorhaben

Ein Vorhaben-Steckbrief, der auf eine halbe Seite passt und die meisten späteren Überraschungen verhindert:

1

Welche Daten gehen hinein?

Personenbezogen? Besondere Kategorien? Fremdes geistiges Eigentum? Geschäftsgeheimnisse? → Datenschutz und Vertraulichkeit (Modul 5).

2

Welche Rolle haben wir, und was tut das System?

Betreiber oder Anbieter? Fällt der Einsatzbereich unter Anhang III? Interagiert es mit Menschen oder erzeugt es veröffentlichte Inhalte? → KI-VO (8.2, 8.4).

3

Wessen Werke gehen hinein, wem gehört das Ergebnis?

Fremde Unterlagen, Lizenzen, Vertraulichkeitsvereinbarungen, erkennbare Personen → Urheber- und Persönlichkeitsrecht (8.6).

4

Erfasst es nebenbei Leistung oder Verhalten?

Bearbeitungszeiten, Fallzahlen, Verläufe je Person — auch ungewollt. → Mitbestimmung, und zwar vor dem Start (8.6).

Abb. 33 Der Vorhaben-Steckbrief. Vier Fragen, vier Zuständigkeiten — und die vierte ist die, die am häufigsten vergessen wird und am längsten aufhält.

BEISPIEL 8.1 — VIER FRAGEN VOR DER EINFÜHRUNG

Eine Abteilung möchte einen Assistenten einführen, der eingehende Kundenmails zusammenfasst, eine Dringlichkeit vorschlägt und einen Antwortentwurf erstellt. Was ist zuerst zu klären?

1. Datenschutz: Kundenmails enthalten personenbezogene Daten. Welche Rechtsgrundlage, welcher Anbieter, welcher Auftragsverarbeitungsvertrag, wird trainiert?

2. Rolle und Risiko: Wir setzen ein zugekauft System ein — Betreiber. Antwortentwürfe an Kunden sind Interaktion mit Menschen: Transparenzpflichten prüfen. Eine Dringlichkeitseinstufung von Anfragen ist für sich keine Entscheidung über Personen.

3. Urheberrecht: Hier nachrangig — es geht um eigene und Kundenkorrespondenz, nicht um geschützte Werke.

4. Mitbestimmung: Der kritische Punkt. Das System erfasst, wie schnell wer welche Vorgänge bearbeitet, ist damit zur Leistungserfassung geeignet — unabhängig davon, ob das gewollt ist. Betriebsrat früh einbinden.

Ergebnis: Die Ebene, die das Vorhaben am ehesten aufhält, ist nicht die, an die man zuerst denkt.

BEISPIEL 8.2 – WENN DIE FRIST IM PROJEKTPLAN FALSCH STEHT

Ausgangslage: Ein Konzern plant den Einsatz eines Systems zur Vorauswahl von Bewerbungen — ein Anwendungsfall aus Anhang III. Der Projektplan nennt als Stichtag für die Konformität den 2. August 2026 und sieht deshalb ein Krisenprogramm mit externer Unterstützung vor.

Woher die Zahl kam: aus einer erzeugten Zusammenfassung der KI-Verordnung, erstellt im Frühjahr 2026. Sachlich korrekt für den damaligen Stand, überholt durch die Änderungsverordnung vom 27. Juli 2026.

Was der Fehler kostet: in diesem Fall Geld statt Recht — ein vorgezogenes Programm, das so nicht nötig gewesen wäre. Die umgekehrte Variante ist die gefährlichere: Ein Plan, der eine Frist zu *spät* ansetzt, führt in die Nichtkonformität.

Die Lehre: Für regulatorische Fristen gilt die strengste Prüfstufe. Sie sind (a) besonders änderungsanfällig, (b) besonders leicht plausibel zu erfinden und (c) wandern besonders zuverlässig unbesehen in Projektpläne. Primärquelle, immer.

ÜBUNG 8 – EBENEN ZUORDNEN

Ordne jede Aussage der oder den betroffenen Rechtsebenen zu und notiere, ob du hier Fachrat einholen würdest.

1. „Wir laden die Handbücher unseres Maschinenlieferanten hoch, damit der Assistent Servicefragen beantworten kann.“
2. „Der Chatbot auf unserer Website stellt sich als ‚Ihre persönliche Beraterin Lena‘ vor.“
3. „Das neue Ticketsystem zeigt pro Mitarbeitendem an, wie viele Vorgänge er pro Stunde schließt.“
4. „Wir werten die Bewerbungsunterlagen der letzten fünf Jahre aus, um ein Vorauswahlmodell zu trainieren.“
5. „Unsere Werbebilder erzeugen wir künftig synthetisch statt sie zu fotografieren.“
6. „Unser Assistent fasst Krankmeldungen zusammen und schlägt die Vertretungsplanung vor.“

Zusatzfrage: Welche der sechs Aussagen berührt die meisten Ebenen gleichzeitig?

Lösung: Anhang A, Übung 8 – Ebenen zuordnen

ÜBUNG 8B – AUSSAGEN ZUM RECHTSSTAND PRÜFEN

Fünf Aussagen, wie sie in Projektunterlagen vorkommen. Entscheide für jede: **trifft zu** oder **trifft nicht zu** — und korrigiere die falschen.

1. „Seit dem 2. August 2026 müssen Betreiber von Hochrisikosystemen ein Risikomanagementsystem betreiben.“
2. „Die KI-Kompetenzpflicht verlangt, dass jede beschäftigte Person eine Schulung mit Abschlusstest absolviert.“
3. „Chatbots müssen seit dem 2. August 2026 als solche erkennbar sein.“
4. „Wenn wir das System nur intern einsetzen, greift die KI-Verordnung nicht.“
5. „Solange wir keine Auswertung beabsichtigen, ist der Betriebsrat nicht zu beteiligen.“

Lösung: Anhang A, Übung 8b – Aussagen zum Rechtsstand prüfen

KERNAUSSAGEN

- Vier Ebenen wirken zusammen: Datenschutz, KI-Verordnung, Urheberrecht, Branchen- und Arbeitsrecht.
- Die KI-Verordnung reguliert nach Risiko. Transparenzpflichten gelten seit dem 2. August 2026; die Hochrisikopflichten des Anhangs III wurden auf den 2. Dezember 2027 verschoben, die des Anhangs I auf den 2. August 2028.
- Die KI-Kompetenzpflicht ist seit dem 27. Juli 2026 eine Bemühenspflicht. Kein Zertifikat verlangt, kein Curriculum vorgeschrieben.
- Regulatorische Fristen sind besonders änderungsanfällig — Primärquelle, immer.
- Mitbestimmung greift bei Eignung zur Leistungs- oder Verhaltenskontrolle, nicht erst bei Absicht.
- Vier Fragen vor jedem Vorhaben: welche Daten, welche Rolle, wessen Werke, erfasst es Verhalten?

Selbsttest Modul 8

1. Welche Aussage zur KI-Kompetenzpflicht trifft zu?

- a) Es ist ein von der Kommission vorgegebener Lehrplan zu verwenden.
- b) Es sind Maßnahmen zur Förderung zu ergreifen; ein bestimmtes Niveau je Person wird nicht verlangt.
- c) Die Pflicht gilt erst ab Dezember 2027.
- d) Jede beschäftigte Person muss ein anerkanntes Zertifikat vorweisen.

2. Ein Werkzeug erfasst nebenbei Bearbeitungszeiten je Beschäftigtem. Was folgt daraus in Deutschland am ehesten?

- a) Mitbestimmung kann greifen, weil das System zur Leistungskontrolle geeignet ist.
- b) Es handelt sich automatisch um ein Hochrisikosystem.
- c) Es ist ausschließlich eine Frage des Urheberrechts.
- d) Nichts, solange keine Auswertung beabsichtigt ist.

3. Seit wann sind die Transparenzpflichten nach Art. 50 KI-VO anwendbar, und wurden sie verschoben?

- a) Seit 27. Juli 2026, verschoben auf August 2028.
- b) Sie sind noch nicht anwendbar.
- c) Seit 2. Februar 2025, verschoben auf Dezember 2027.
- d) Seit 2. August 2026, nicht verschoben.

4. Welche der vier Fragen des Vorhaben-Steckbriefs wird am häufigsten vergessen und hält Vorhaben am längsten auf?

- a) Welche Rolle haben wir?
- b) Wessen Werke gehen hinein?
- c) Erfasst es nebenbei Leistung oder Verhalten?
- d) Welche Daten gehen hinein?

Musterprüfung

20 Fragen · 40 Minuten · Bestehensgrenze 70 % (14 von 20)

Bearbeite diese Prüfung unter realen Bedingungen: 40 Minuten am Stück, keine Hilfsmittel, keine Unterbrechung. Genau eine Antwort ist richtig, sofern nicht anders angegeben. Die Lösungen mit Begründung stehen in Anhang A — sieh erst danach hinein.

1. Ein Sprachmodell erzeugt seine Ausgabe, indem es
 - a) in einer Wissensdatenbank nachschlägt.
 - b) fortlaufend plausible Fortsetzungen des Textes schätzt.
 - c) hinterlegte Regeln anwendet.
 - d) Aussagen gegen Primärquellen abgleicht.
2. Welche Aufgabe scheitert ohne zusätzliches Werkzeug am ehesten?
 - a) Die Zeichenzahl eines Textes exakt bestimmen.
 - b) Drei Betreffzeilen vorschlagen.
 - c) Einen englischen Absatz übersetzen.
 - d) Einen Absatz freundlicher formulieren.
3. „Nur was im Kontextfenster steht, existiert für die Antwort.“ Was folgt daraus praktisch?
 - a) Antworten sind stets reproduzierbar.
 - b) Lange Dokumente können nicht verarbeitet werden.
 - c) Eine nicht mitgelieferte Passage wird nicht als fehlend gemeldet, sondern schlicht nicht berücksichtigt.
 - d) Das Modell lernt aus jedem Gespräch dazu.
4. Was ist an Halluzinationen betrieblich am gefährlichsten?
 - a) Das Fehlen einer Unsicherheitsanzeige.
 - b) Ihre Länge.
 - c) Dass sie nur bei Fachthemen auftreten.
 - d) Ihre Häufigkeit.
5. Welches Muster liegt vor, wenn eine Antwort einen echten Autor mit einem nicht existierenden Titel verbindet?
 - a) Gefälligkeit.
 - b) Ein Tokenisierungsfehler.
 - c) Veraltete Information.
 - d) Erfundener Beleg in halb echter Form.
6. Welche Frageformulierung erhöht das Risiko einer gefälligen Antwort?
 - a) „Was gilt hier, und woraus ergibt sich das?“
 - b) „Welche Möglichkeiten kommen in Betracht?“
 - c) „Bestätige mir, dass Verfahren A vorgeschrieben ist.“
 - d) „Welche Argumente sprechen dagegen?“

7. Dieselbe Technik liegt je nach Aufgabe zwischen rund 2 % und 88 % Halluzinationsrate. Woran liegt das in erster Linie?

- a) An der eingestellten Temperatur.
- b) Am Modellhersteller.
- c) An der Länge der Antwort.
- d) Am Zuschnitt der Aufgabe — ob die Grundlage beiliegt oder aus dem Gedächtnis geholt werden muss.

8. Welches der Kriterien für die Eignung einer Aufgabe wird am häufigsten übersehen?

- a) Fehlerkosten.
- b) Umkehrbarkeit.
- c) Prüfbarkeit.
- d) Bearbeitungsdauer.

9. Warum ist „Stellenausschreibung entwerfen“ geeignet, „Bewerbungen vorsortieren“ aber nicht?

- a) Die Vorsortierung wirkt auf einzelne Menschen, ist faktisch endgültig und von außen kaum prüfbar.
- b) Bewerbungen passen nicht ins Kontextfenster.
- c) Ausschreibungen enthalten keine personenbezogenen Daten.
- d) Ausschreibungen sind kürzer.

10. Ein Vorgang läuft 4.000 Mal im Monat mit 94 % Trefferquote. Der angemessene Prüfansatz ist

- a) eine zweite Antwort desselben Systems je Vorgang.
- b) keine Prüfung, die Quote genügt.
- c) Einzelprüfung jedes Ergebnisses vor der Verwendung.
- d) ein Stichprobenplan, dichter bei Vorgängen mit härteren Folgen.

11. Welche der folgenden Angaben gehört *nicht* in ein nicht freigegebenes System?

- a) Eine anonymisierte Struktur einer Tabelle.
- b) Ein Textentwurf ohne Personenbezug.
- c) Ein API-Schlüssel des Testsystems.
- d) Eine allgemeine Frage zur Rechtschreibung.

12. Was entscheidet darüber, ob dienstliche Daten in ein KI-Werkzeug eingegeben werden dürfen?

- a) Ob für den genutzten Zugang eine Auftragsverarbeitung besteht und der Zweck gedeckt ist.
- b) Ob das Konto kostenpflichtig ist.
- c) Ob das Training in den Einstellungen abgeschaltet wurde.
- d) Der Name des Anbieters.

13. Schatten-KI bezeichnet

- a) Systeme, die im Hintergrund Protokolle schreiben.
- b) die Nutzung nicht freigegebener Werkzeuge für dienstliche Aufgaben.
- c) Modelle mit unbekanntem Trainingsdatensatz.
- d) den Betrieb ohne Internetverbindung.

14. Ein Beleg gilt erst dann als geprüft, wenn

- a) zwei Systeme ihn übereinstimmend nennen.
- b) er in der Antwort korrekt formatiert war.
- c) der Titel existiert.
- d) er existiert, die Aussage trägt und aktuell genug ist.

15. Welche Prüfmethode deckt einen unzulässigen Schluss zwischen zwei zutreffenden Sätzen auf?

- a) Faktenprüfung.
- b) Logikprüfung.
- c) Vollständigkeitsprüfung.
- d) Passungsprüfung.

16. Welche drei Fragen bestimmen den angemessenen Prüfaufwand?

- a) Kosten, Geschwindigkeit und Verfügbarkeit des Systems.
- b) Anzahl der Quellen, Anzahl der Zahlen und Anzahl der Absätze.
- c) Länge, Sprache und Format des Textes.
- d) Reichweite nach außen, Entscheidungswirkung und Umkehrbarkeit.

17. „Das hat die KI so geschrieben“ ist als Begründung

- a) keine Entlastung — wer den Output verwendet, verantwortet ihn.
- b) nur bei internen Dokumenten ausreichend.
- c) ausreichend, wenn das System freigegeben war.
- d) arbeitsrechtlich ausreichend.

18. Wirksame menschliche Aufsicht setzt voraus:

- a) dass das System zertifiziert ist.
- b) Kompetenz, Zeit und die Befugnis abzulehnen.
- c) mindestens zwei beteiligte Personen.
- d) eine schriftliche Freigabe in jedem Einzelfall.

19. Warum gehört zu einem Fehlerbericht zwingend die Modellversion?

- a) Weil ohne sie keine Kennzeichnungspflicht besteht.
- b) Weil sonst die Gewährleistung entfällt.
- c) Weil Anbieter Modelle im laufenden Betrieb austauschen und der Fall sonst nicht nachstellbar ist.
- d) Weil die Version über die Zuständigkeit entscheidet.

20. Welche Aussage zum Stand der KI-Verordnung im August 2026 trifft zu?

- a) Die Hochrisikopflichten des Anhangs III gelten seit dem 2. August 2026.
- b) Die Transparenzpflichten nach Art. 50 gelten seit dem 2. August 2026; die Hochrisikopflichten des Anhangs III wurden auf Dezember 2027 verschoben.
- c) Die KI-Kompetenzpflicht gilt erst ab Dezember 2027.
- d) Alle Pflichten der Verordnung wurden auf August 2028 verschoben.

Lösungen

Übungen · Selbsttests · Musterprüfung

Übungen

Übung 1 – Systeme einordnen

1 Generierung — der Fehler fällt beim Lesen auf, solange man liest. **2** Klassifikation, mit Wirkung auf Personen — der Fehler fällt *nicht* auf: Wer aussortiert wurde, beschwert sich nicht. **3** Suche — auffällig nur, wenn man Stichproben zieht oder etwas vermisst. **4** Vorhersage — auffällig im Soll-Ist-Vergleich. **5** Generierung — auffällig für jemanden, der in der Besprechung war. **6** Klassifikation — auffällig, wenn ein dringender Vorgang liegen bleibt, also verspätet.

Zusatzfrage: 2 und 3. Bei 2 gibt es keinen Rückkanal — die Betroffenen erfahren nie, dass sie aussortiert wurden. Bei 3 hat ein Fehler die Form eines *fehlenden* Treffers, und Fehlendes bemerkt niemand, solange man nicht gezielt danach sucht. Fehler, die keinen Beschwerdeführer haben, sind die teuersten.

Übung 1b – Erwartungen prüfen

1 Kommt darauf an. „Füttern“ meint zweierlei: entweder Training (passiert praktisch nie im Unternehmen) oder Suche plus Kontextfenster (RAG, Modul 2.6). Im zweiten Fall kennt das Modell die Handbücher nicht, sondern sieht bei jeder Frage die Passagen, die eine Suche ausgewählt hat.

2 Trifft nicht zu. Es gibt keine eingebaute Unsicherheitsanzeige (Modul 1.6, Modul 3.1). Nachfragen kommen vor, sind aber Stilmerkmal, kein Wissensstand-Signal.

3 Trifft nicht zu. Ein gelerntes System bildet die Muster seiner Trainingsdaten ab, einschließlich der darin enthaltenen Schieflagen. „Objektiv“ ist keine Eigenschaft des Verfahrens, sondern eine Behauptung über die Daten.

4 Trifft nicht zu. Ohne Rechenwerkzeug wird gerechnete Arithmetik fortgesetzt, nicht gerechnet (Modul 1.5). Bei einfachen Fällen sieht das Ergebnis meist richtig aus — genau das macht es gefährlich.

Übung 2 – Scheitert es, und woran?

1 Zuverlässig — reine Textebene. **2** Unzuverlässig: Unternehmensdaten liegen nicht im Modell (Training/Inferenz); nur mit Anbindung an eure Zahlen. **3** Unzuverlässig: Zeichenebene, Tokenisierung — nur mit Werkzeug. **4** Zuverlässig für den Grundsatz; bei Einzelfallfragen gilt Modul 4. **5** Nur mit Werkzeug: Zählen über 60 Seiten scheitert am Kontextfenster und an der Tokenisierung. **6** Zuverlässig — Vorschläge, prüfbar. **7** Unzuverlässig: Antworten sind Stichproben; ohne fixierte Einstellungen und identische Eingabe ist Reproduzierbarkeit nicht gegeben. **8** Nur mit Werkzeug: Datumsarithmetik plus Kenntnis des heutigen Datums und der Feiertage — beides liegt nicht im Modell.

Übung 2b – Eine Eingabe reparieren

Drei Schwachstellen. (1) *Kontextfenster:* 340 Seiten liegen nicht vollständig vor; das System sieht Ausschnitte, meldet aber nicht, welche fehlen. (2) *Wissensstand:* „die neue Norm“ ist nach dem Trainingsstand erschienen oder liegt jedenfalls nicht bei — das Modell wird eine plausible Fassung erfinden. (3) *Prüfbarkeit:* „begründe mit den entsprechenden Kapiteln“ lädt zur Kapitelnummern-Halluzination ein, weil die Form einer Kapitelangabe leicht zu erzeugen ist.

Umgeschrieben. „Anbei zwei Auszüge: Abschnitt 7.4 unseres Qualitätshandbuchs (Lieferantenbewertung, 4 Seiten) und der Normabschnitt 8.4 im Wortlaut (3 Seiten). Vergleiche beide und liste auf: (a) welche Anforderungen des Normabschnitts durch unseren Abschnitt abgedeckt sind, (b) welche nicht, (c) welche unklar formuliert sind. Zitiere

ausschließlich aus den beiden beigefügten Texten und markiere jede Aussage mit der Fundstelle im Wortlaut. Wenn eine Anforderung in den Auszügen nicht vorkommt, schreibe ‚nicht in den Auszügen enthalten‘.“

Was sich geändert hat: Der Kontext ist vollständig statt ausschnittsweise, der Bezugspunkt liegt bei statt im Trainingsstand, und die letzte Anweisung gibt dem Modell eine erlaubte Nicht-Antwort — die einzige wirksame Gegenmaßnahme gegen erfundene Belege.

Übung 3 – Finde das Problem

1 Erfundener Beleg im Detail: Die Mitbestimmung des Betriebsrats bei überwachungsgeeigneten Systemen ist real; die genannte Frist von 14 Tagen ist die erfundene Präzision. *Prüfung:* Gesetzestext und Fachrat — nicht das Modell erneut fragen.

2 Erfundener Beleg: konkrete Universität, konkretes Jahr, konkrete Prozentzahl, kein Titel und kein Autor. Klassisches Muster. *Prüfung:* Studie namentlich suchen; wenn sie nicht auffindbar ist, ist die Zahl nicht verwendbar. (Belastbare Zahlen zum selben Sachverhalt gibt es — siehe Modul 5.5.)

3 Sachlicher Fehler in Form einer Verschmelzung: Auftragsverarbeitung und Einwilligung werden vermengt — eine Einwilligung ist nicht „in jedem Fall“ erforderlich, es gibt weitere Rechtsgrundlagen. *Prüfung:* Datenschutzbeauftragte fragen.

4 Veraltet *und* verschmolzen: Die Anwendbarkeit der Hochrisikopflichten des Anhangs III wurde auf den 2. Dezember 2027 verschoben (Modul 8.2); die Aufbewahrungsdauer ist zudem frei erfunden. *Prüfung:* aktuellen Verordnungstext aufrufen, nicht die Zusammenfassung.

Zusatzfrage: Aussage 2. Eine glatte Prozentzahl mit Universitätsnennung wandert unbesehen in Folien, wird zitiert und ist danach kaum noch einzufangen. Falsche Rechtsaussagen werden eher hinterfragt; falsche Statistiken werden geglaubt.

Übung 3b – Gefälligkeit provozieren

1 (a) „Bestätige mir, dass die Haftungsklausel in § 9 unwirksam ist.“ (b) „Prüfe § 9 auf Anhaltspunkte, die für und gegen die Wirksamkeit sprechen, und benenne, welche Angaben dir für eine Beurteilung fehlen.“

2 (a) „Das ist doch nicht mitbestimmungspflichtig, oder?“ (b) „Welche Merkmale eines Systems lösen in Deutschland Mitbestimmung aus, und welche davon liegen bei der folgenden Beschreibung vor?“

3 (a) „Fasse zusammen, warum Anbieter X die beste Wahl ist.“ (b) „Stelle die Argumente für und gegen Anbieter X einander gegenüber und benenne, unter welchen Annahmen sich die Entscheidung umkehren würde.“

Der Unterschied: Variante (a) liefert eine Begründung für ein bereits feststehendes Ergebnis — sie sieht aus wie Prüfung und ist keine. Variante (b) erzeugt eine Liste, die man abarbeiten kann, und macht sichtbar, was fehlt. Fall 3 ist der heikelste: Dort ist die gefällige Antwort besonders willkommen, weil sie die eigene Entscheidung stützt.

Übung 4 – Einordnen und umbauen

Teil A: **1** geeignet (günstig, umkehrbar, prüfbar für Teilnehmende). **2** ungeeignet — verbindliche Zusage, endgültig, extern. **3** geeignet mit Prüfung durch jemanden, der die Spezifikation kennt. **4** ungeeignet als Entscheidung — Wirkung auf Personen. **5** geeignet mit Prüfung; bei Fachterminologie durch eine fachkundige Person. **6** ungeeignet ohne Werkzeug — exakte Berechnung. **7** geeignet mit Stichprobenplan: hohe Frequenz, Einzelprüfung unwirtschaftlich, Fehler im Soll-Ist-Vergleich der Kostenstellen sichtbar.

Teil B, Beispiele. *Aufgabe 2* (Muster „Entwurf statt Ergebnis“): Das System entwirft aus dem Warenwirtschaftssystem einen Terminvorschlag und ein Mensch gibt ihn frei, bevor er den Kunden erreicht — der Mensch übernimmt die Zusage. *Aufgabe 4* (Muster „Kriterien offenlegen“): Das System strukturiert Bewerbungen nach selbst gesetzten, offengelegten Kriterien (Erfahrung, Qualifikationen, Verfügbarkeit), alle Unterlagen bleiben sichtbar — der Mensch übernimmt die Auswahl. *Aufgabe 6* (Muster „Werkzeug statt Fortsetzung“): Das System liest die Formel aus dem Handbuch aus, gerechnet wird in der Tabellenkalkulation — der Mensch übernimmt die Auslegung.

Übung 4b — Den Schnelltest anwenden

Frage 1 — nein: Es entscheidet nicht unmittelbar über eine Person im Sinne von Beschäftigung, Kredit oder Leistungszugang; es geht um Sachfälle. (Bei Verbraucherverträgen mit erheblicher Wirkung ist das erneut zu prüfen.)

Frage 2 — **ja**: Eine Ablehnung geht nach außen und ist faktisch endgültig. Damit gilt bereits hier: nur mit Freigabe durch eine fachlich zuständige Person.

Frage 3 — **ja**: Gewährleistung hängt an Fristen und Rechtsfolgen. Tragende Angaben (Kaufdatum, Fristbeginn, Vertragsklausel) müssen aus der Primärquelle stammen, nicht aus der Fortsetzung.

Frage 4 — teilweise **ja**: „Anerkennen“ ist leicht prüfbar, „ablehnen“ nicht — die Sachbearbeitung müsste den Fall ohnehin selbst beurteilen, um zu erkennen, dass die Ablehnung falsch ist. Der Satz „folgt der Empfehlung, wenn nichts dagegenspricht“ ist genau die dekorative Aufsicht aus Modul 7.3.

Frage 5 — **ja**: 600 Fälle im Monat, Einzelprüfung jeder Empfehlung ist unwirtschaftlich.

Verantwortbarer Zuschnitt: Das System sortiert vor und bereitet auf — es extrahiert Kaufdatum, Fristlage und einschlägige Klausel *mit Fundstelle im Wortlaut* und markiert Vollständigkeit. Es gibt *keine* Empfehlung zur Ablehnung aus. Anerkennungen unterhalb einer Wertgrenze dürfen automatisch laufen (günstig, umkehrbar, kundenfreundlich), Ablehnungen und alles oberhalb der Grenze gehen in die Sachbearbeitung. Wöchentliche Stichprobe von 30 Vorgängen auf korrekte Extraktion, Trefferquote protokolliert.

Übung 5 — Prompt entschärfen

1 Entfernt: personenbezogene Daten *und* Gesundheitsdaten (Krankheitstage, besondere Kategorie). Umschrift: „Formuliere die Gliederung einer Gesprächsnotiz für ein Personalgespräch zu Leistungsthemen. Struktur, Fragestellungen, Hinweise auf Sachlichkeit. Keine konkreten Fälle.“ Namen und Krankheitstage trägt die Führungskraft danach selbst ein.

2 Entfernt: Geschäftsgeheimnisse (Angebotssumme, Marge). Umschrift: „Welche Argumentationslinien gibt es im B2B-Vertrieb gegenüber einem günstigeren Wettbewerber, wenn die eigene Stärke in Servicequalität und Lieferzuverlässigkeit liegt?“ Die Zahlen braucht es für die Antwort nicht.

3 Entfernt: Zugangsdaten. Umschrift: „Schreib mir eine SQL-Abfrage für alle Bestellungen des letzten Quartals. Tabellen: `orders(id, customer_id, created_at, total)`. Dialekt: PostgreSQL.“ Nur das Schema wird gebraucht, nie der Zugang.

4 Entfernt: unveröffentlichtes geistiges Eigentum. Hier ist der Umbau schwieriger, weil die Aufgabe am konkreten Wortlaut hängt. Zulässiger Zuschnitt: nur mit freigegebenem Unternehmenskonto mit Auftragsverarbeitungsvertrag und abgeschaltetem Training — oder gar nicht. Bei Patentanmeldungen kommt hinzu, dass eine vorzeitige Offenbarung die Neuheit gefährden kann; das ist eine Frage für die Patentabteilung, nicht für die Sachbearbeitung.

Zusatzfrage: Fälle 3 und 4. Bei **3** wurden die Zugangsdaten bereits offengelegt und müssen **rotiert** werden — ein einmal in ein fremdes System eingegebenes Geheimnis ist verbrannt, auch wenn der Chatverlauf gelöscht wird. Bei **4** genügt kein Umschreiben, weil die Aufgabe den Originaltext braucht: Hier entscheidet der Zugang, nicht die Formulierung.

Übung 5b — Den Zugang beurteilen

1 Abhängig. Bereitgestellt ist notwendig, aber nicht hinreichend. Es fehlt: Ist der Verarbeitungszweck „Kundenkommunikation“ von der Rechtsgrundlage und dem Verarbeitungsverzeichnis gedeckt, und besteht ein Auftragsverarbeitungsvertrag mit dem Anbieter? Die Antwort ist meist ja, aber sie muss vorliegen, nicht vermutet werden.

2 Unzulässig. Das abgeschaltete Training ändert nichts daran, dass der Vertrag privat ist: keine Auftragsverarbeitung, keine Weisungsbindung, kein Löschanspruch des Unternehmens. Das ist Schatten-KI mit gutem Gewissen — und genau die Variante, die in Erhebungen als „Einzelfall“ auftaucht.

3 Zulässig. Schema ohne Inhalte, dienstlicher Zugang, kein Personenbezug. Musterbeispiel für die Technik „Schema statt Inhalt“.

4 Abhängig — und die Datenschutzfrage ist hier nicht die entscheidende. Lokal betrieben heißt: Die Daten verlassen das Haus nicht. Es fehlt aber die Angabe, was das System mit den Unterlagen tut. Bewertet oder sortiert es Bewerbungen vor, ist das eine Entscheidung mit Wirkung auf Personen (Modul 4.3), berührt Beschäftigtendatenschutz, Mitbestimmung und je nach Ausgestaltung die Hochrisikoeinstufung nach Anhang III (Modul 8.2). Der Serverstandort löst keines dieser Themen.

Übung 6 – Prüfplan schreiben

Tragend: zwei Gesetzesänderungen mit Datum, das Aktenzeichen, die Bußgeldhöhe. Die drei Handlungsempfehlungen sind Bewertung, nicht Tatsache — sie prüfst du auf Logik, nicht auf Fakten.

Prüfarten: Gesetzesänderungen und Aktenzeichen → Faktenprüfung an der Primärquelle; Bußgeldhöhe → Faktenprüfung und Passung (auf welchen Fall bezieht sie sich?); Empfehlungen → Logik und Vollständigkeit.

Reihenfolge: Aktenzeichen zuerst. Es ist am schnellsten zu prüfen und der beste Frühindikator — existiert es nicht, ist der ganze Text verdächtig und du sparst dir die übrigen Prüfungen. Danach die Gesetze, dann das Bußgeld, zuletzt die Empfehlungen.

Zweite Person: bei den Gesetzesänderungen. Ob eine Änderung im Vertrieb tatsächlich das bedeutet, was der Text sagt, ist eine Fachfrage — hier prüfst du nicht Existenz, sondern Auslegung, und dafür brauchst du Fachkompetenz, die du selbst nicht hast.

Prüfstufe: Stufe 2. Es geht nicht aus dem Haus (nein), aber Menschen handeln danach gegenüber Kunden (ja) und die Wirkung ist nicht ohne Weiteres umkehrbar (ja). Zweimal „ja“ = Vollprüfung.

Zusatzfrage: Kaum. „Nur intern“ senkt die Außenwirkung, aber ein Newsletter an den Vertrieb *steuert Verhalten* — Menschen handeln danach gegenüber Kunden. Die Entscheidungswirkung bleibt, und sie ist das maßgebliche Kriterium, nicht der Verteiler.

Übung 6b – Halbe Prüfung erkennen

1 Halb geprüft. Es fehlt Schritt 2: Sagt die Quelle, was behauptet wird? Genau hier scheitern halb echte Belege — echter Autor, echte Zeitschrift, erfundene Aussage.

2 Halb geprüft. Schritt 3 fehlt: Ist die Quelle aktuell genug für den Zweck? Eine korrekt zitierte Aussage von 2019 kann in einer Vorlage von 2026 falsch sein.

3 Halb geprüft. Die drei Schritte aus 6.4 sind erfüllt, aber der Griff „Abgrenzung“ aus 6.5 fehlt: 23,4 % *welches* Marktes, in welcher Region? Ohne Abgrenzung ist die Zahl nicht übertragbar.

4 Gar nicht geprüft. Übereinstimmung zweier Systeme ist ein schwaches Indiz, kein Beleg (Modul 6.6). Es wurde keine Primärquelle aufgerufen — die Prüfung hat nicht stattgefunden, sie wurde nur zweimal ausgelassen.

Übung 7 – Zuordnen und nachschärfen

1 Verantwortlich ist das Unternehmen als Betreiber; die Aussage des Chatbots wirkt nach außen. Angemessene Stufe: *im Ablauf* für rechtlich verbindliche Aussagen — praktisch heißt das, dass der Chatbot zu Gewährleistungsfragen gar nicht antwortet, sondern weiterleitet. Erste Maßnahme: Themenbereich sperren und auf die AGB verweisen.

2 Verantwortlich ist, wer den Prozess eingerichtet hat. Angemessene Stufe: *im Ablauf* — ein Protokoll wird zum Dokument, sobald es verteilt ist. Erste Maßnahme: Freigabe durch die protokollführende Person vor dem Versand.

3 Verantwortlich ist die Geschäftsführung für die Entscheidung, die Fachabteilung für die Vorauswahl. Angemessene Stufe: *überwachend* für die Vorauswahl, *im Ablauf* für die Entscheidung. Erste Maßnahme: die Ausschlusskriterien der Vorauswahl offenlegen — die Geschäftsführung muss wissen, was sie *nicht* zu sehen bekommt.

4 Verantwortlich ist das Unternehmen als Betreiber. Aufsichtsstufe: *überwachend* genügt. Erste Maßnahme: Kennzeichnung als synthetisch erzeugtes Bildmaterial, einschließlich maschinenlesbarer Markierung — die Pflicht nach Art. 50 knüpft an die synthetische Erzeugung an, nicht daran, ob Personen abgebildet sind.

Zusatzaufgabe, Musterlösung: **System und Version:** Kundenservice-Bot „Helix“, Modellversion 4.2-2026-07. **Zeitpunkt:** 14.08.2026, 10:42 Uhr, Sitzungs-ID 8842-A. **Eingabe im Wortlaut:** „Wie lange habe ich Garantie auf das Gerät?“ **Ausgabe im Wortlaut:** „Auf alle Geräte gewähren wir 36 Monate Garantie ab Kaufdatum.“ **Erwartet:** Verweis auf die gesetzliche Gewährleistung von 24 Monaten und auf Ziffer 7 der AGB; keine eigenständige Fristnennung. **Folge:** Die Auskunft wurde einem Kunden erteilt und von ihm in einer Reklamation zitiert; ein Fall ist bereits offen, weitere sind nicht auszuschließen — Sofortmaßnahme erforderlich.

Übung 7b – Aufsicht bewerten

1 **Wirksam.** Kompetenz vorhanden, Zeit offenbar eingeplant, und die Ablehnungsquote von rund 7 % zeigt, dass Ablehnung tatsächlich stattfindet. Das ist der beste verfügbare Indikator.

2 **Dekorativ.** Es fehlt die **Zeit** — genauer: Sie ist vorgesehen und wird nicht genommen. Drei Sekunden reichen nicht, um eine Empfehlung zu beurteilen. Die Differenz zwischen vorgesehener und tatsächlicher Dauer ist das Warnsignal.

3 **Dekorativ.** Es fehlt die **Kompetenz**. Zeit und Handbuch ersetzen nicht die Fähigkeit, einen fachlich falschen Entwurf als solchen zu erkennen. Ein Handbuch hilft bei bekannten Fällen; die gefährlichen sind die unbekannten.

4 **Dekorativ.** Es fehlt die **Kultur**. Kompetenz und Zeit sind da, aber Ablehnung erzeugt eine negative Kennzahl bei einer anderen Abteilung. Damit ist Ablehnung teuer und Zustimmung kostenlos — der Ausgang steht fest. Die Reparatur liegt nicht bei der freigebenden Person, sondern in der Kennzahl.

Übung 8 – Ebenen zuordnen

1 Urheberrecht (fremde Werke) und Vertragsrecht (Vertraulichkeitsvereinbarung mit dem Lieferanten). Fachrat: ja, bei bestehender Vertraulichkeitsvereinbarung.

2 KI-Verordnung — Transparenzpflicht nach Art. 50; eine erfundene menschliche Beraterin ist die klassische Fallgestaltung. Fachrat: meist nicht nötig, die Lösung ist offensichtlich.

3 Arbeitsrecht und Mitbestimmung, zusätzlich Datenschutz (Beschäftigtendaten). Fachrat: ja.

4 Alle vier Ebenen: Datenschutz (Bewerberdaten, Zweckbindung — Bewerbungen wurden nicht zum Modelltraining erhoben), KI-Verordnung (Beschäftigung als Anhang-III-Bereich), Arbeitsrecht, und mittelbar Antidiskriminierung. Fachrat: unbedingt, vor Beginn.

5 KI-Verordnung (Kennzeichnung synthetischer Medien, maschinenlesbar) und Persönlichkeitsrecht, falls Personen erkennbar sind. Fachrat: bei erkennbaren Personen ja.

6 Datenschutz in besonderer Schärfe: Krankmeldungen sind Gesundheitsdaten, also eine besondere Kategorie (Modul 5.2). Zusätzlich Mitbestimmung. Fachrat: unbedingt — hier ist die Schwelle deutlich höher als bei gewöhnlichen Beschäftigtendaten.

Zusatzfrage: Aussage 4 — sie berührt alle vier Ebenen und ist zugleich die einzige, bei der eine Fehlentscheidung einzelne Menschen unmittelbar trifft.

Übung 8b – Aussagen zum Rechtsstand prüfen

1 **Trifft nicht zu.** Die Anwendbarkeit der Hochrisikopflichten für eigenständige Systeme des Anhangs III wurde auf den 2. Dezember 2027 verschoben, für Anhang I auf den 2. August 2028.

2 **Trifft nicht zu.** Verlangt sind *Maßnahmen zur Förderung* von KI-Kompetenz (Bemühenspflicht seit der Neufassung zum 27. Juli 2026). Weder ein bestimmtes Niveau je Person noch ein Zertifikat oder ein Abschlusstest sind vorgeschrieben.

- 3 Trifft zu.** Die Transparenzpflichten nach Art. 50 sind seit dem 2. August 2026 anwendbar und wurden nicht verschoben.
- 4 Trifft nicht zu.** Die Verordnung knüpft an die Rolle und den Anwendungsbereich an, nicht an interne oder externe Nutzung. Ein intern eingesetztes System zur Bewerbervorauswahl bleibt ein Anhang-III-Fall.
- 5 Trifft nicht zu.** Mitbestimmung greift bei der *Eignung* zur Leistungs- oder Verhaltenskontrolle, nicht erst bei der Absicht.

Selbsttests

Modul 1: 1 c · 2 a · 3 d · 4 b

Modul 2: 1 d · 2 c · 3 a · 4 b

Modul 3: 1 c · 2 d · 3 a · 4 b

Modul 4: 1 b · 2 c · 3 a · 4 d

Modul 5: 1 d · 2 c · 3 b · 4 a

Modul 6: 1 d · 2 a · 3 b · 4 c

Modul 7: 1 a · 2 c · 3 b · 4 d

Modul 8: 1 b · 2 a · 3 d · 4 c

Die Begründungen zu allen Selbsttestfragen findest du in der interaktiven Fassung dieses Handbuchs; sie erscheinen dort direkt nach der Antwortwahl.

Musterprüfung

NR.	LÖ- SUNG	BEGRÜNDUNG
1	b	Fortsetzung nach Plausibilität; kein Nachschlagen, keine Regeln, kein Quellenabgleich (Modul 1.5).
2	a	Zeichenebene ist durch Tokenisierung nicht zugänglich (Modul 2.1).
3	c	Das Modell meldet Fehlendes nicht — es antwortet mit dem Vorhandenen (Modul 2.2).
4	a	Nicht die Quote, sondern die fehlende Unsicherheitsanzeige (Modul 3.1).
5	d	Halb echte Belege sind das typische Muster erfundener Fundstellen (Modul 3.2).
6	c	Bestätigungsfragen legen die Antwort in die Frage (Modul 3.4).
7	d	Der Sprung kommt vom Aufgabenzuschnitt, nicht vom Modellnamen (Modul 3.1 und 3.6).
8	c	Prüfbarkeit wird am häufigsten übersehen (Modul 4.1).
9	a	Wirkung auf Personen, faktische Endgültigkeit, fehlende Prüfbarkeit (Modul 4.2).
10	d	Hohe Frequenz verschiebt den Prüfpunkt auf das Verfahren (Modul 4.1).
11	c	Zugangsdaten gehören nie hinein, auch nicht aus Testsystemen (Modul 5.2).
12	a	Der Vertrag hinter dem Zugang entscheidet, nicht das Produkt (Modul 5.3).
13	b	Nutzung nicht freigegebener Werkzeuge, meist über private Konten (Modul 5.5).
14	d	Existenz, Tragfähigkeit, Aktualität — drei Schritte (Modul 6.4).
15	b	Beide Sätze bestehen die Faktenprüfung; der Fehler steckt in der Verknüpfung (Modul 6.2).
16	d	Geht es aus dem Haus, entscheidet jemand danach, ist es umkehrbar — drei Mal „ja“ bedeutet Vollprüfung im Vier-Augen-Prinzip (Modul 6.3).
17	a	Werkzeuge tragen keine Verantwortung (Modul 7.1).
18	b	Kompetenz, Zeit, Ablehnungsbefugnis und -kultur (Modul 7.4).
19	c	Ohne Version ist nichts reproduzierbar — Modelle werden im laufenden Betrieb getauscht (Modul 7.6).
20	b	Verschoben wurden ausschließlich die Hochrisikopflichten; die Transparenzpflichten sind unverändert anwendbar (Modul 8.2).

AUSWERTUNG

17–20 richtig — sehr sicher. Du kannst zur Prüfung antreten.

14–16 richtig — bestanden, aber knapp. Geh die Module der falsch beantworteten Fragen noch einmal durch; die Prüfung zieht aus einem größeren Fragenpool und trifft dich an anderen Stellen.

unter 14 — noch nicht ausreichend. Arbeite die betroffenen Module erneut durch, einschließlich der Übungen, und wiederhole die Musterprüfung frühestens nach zwei Tagen.

Glossar

33 Begriffe

Agent

Sprachmodell, das in einer Schleife läuft: plant einen Schritt, ruft ein Werkzeug auf, liest das Ergebnis, plant weiter. Kann handeln, nicht nur antworten.

Aggregieren

Technik der Datenminimierung: statt Einzeldatensätzen Summe, Durchschnitt oder Verteilung übergeben.

Anbieter

Wer ein KI-System entwickelt oder entwickeln lässt und unter eigenem Namen bereitstellt. Weiter reichende Pflichten als der Betreiber.

Anonymisieren

Ersetzen von Namen und Kennungen durch Platzhalter. Entfernt Kennungen, macht den Rest der Angaben nicht automatisch unbedenklich.

Auftragsverarbeitung

Vertragliche Grundlage dafür, dass ein Dienstleister personenbezogene Daten im Auftrag verarbeitet. Bei privaten Konten regelmäßig nicht gegeben.

Betreiber

Wer ein KI-System unter eigener Verantwortung einsetzt. Die Rolle der meisten Unternehmen.

Bemühenspflicht

Pflicht, angemessene Maßnahmen zu ergreifen, ohne ein bestimmtes Ergebnis zu garantieren. Seit dem 27. Juli 2026 die Rechtsnatur der KI-Kompetenzpflicht nach Art. 4 KI-VO.

Deep Learning

Verfahren des maschinellen Lernens mit mehrschichtigen neuronalen Netzen. Grundlage praktisch aller heute diskutierten Systeme.

Einbettung (Embedding)

Darstellung eines Textstücks als Zahlenreihe, sodass inhaltlich Ähnliches nahe beieinander liegt. Grundlage semantischer Suche.

Gefälligkeit (Sycophancy)

Neigung eines Modells, der Prämisse der Frage oder dem Widerspruch des Nutzers zu folgen, unabhängig von der Sachlage.

Generierung

Erzeugung neuer Inhalte — Text, Bild, Code, Audio — im Unterschied zu Klassifikation, Vorhersage und Suche.

Halluzination

Ausgabe, die plausibel formuliert und inhaltlich falsch ist. Nicht am Stil erkennbar.

Hochrisiko-KI-System

System in einem sensiblen Anwendungsbereich nach Anhang I oder III der KI-Verordnung, mit weitreichenden Pflichten. Anwendbarkeit Anhang III: 2. Dezember 2027, Anhang I: 2. August 2028.

Inferenz

Der einzelne Aufruf des fertigen Modells im Betrieb. Dabei wird nicht gelernt.

KI-Kompetenz

Fähigkeiten, Kenntnisse und Verständnis, die einen sachkundigen Einsatz von KI-Systemen ermöglichen. Gegenstand der Förderpflicht nach Art. 4 KI-VO.

Klassifikation

Zuordnung zu einer von mehreren Kategorien. Fehler sind meist sichtbar — außer bei Wirkung auf Personen.

Kontextfenster

Der Textbereich, den das Modell bei einer Antwort berücksichtigen kann. Was nicht darin steht, existiert für die Antwort nicht.

Maschinelles Lernen

Ableitung von Mustern aus Daten statt Programmierung von Regeln. Teilmenge der KI.

Menschliche Aufsicht

Einbindung von Menschen in Betrieb und Kontrolle — im Ablauf, überwachend oder im Kommando. Wirksam nur mit Kompetenz, Zeit und Ablehnungsbefugnis.

Mitbestimmung

Beteiligung des Betriebsrats, in Deutschland unter anderem bei technischen Einrichtungen, die zur Leistungs- oder Verhaltenskontrolle *geeignet* sind.

Prüfbarkeit

Frage, ob ein Ergebnis mit vertretbarem Aufwand überprüft werden kann. Das am häufigsten übersehene Eignungskriterium.

RAG

Retrieval-augmented generation: Ein Suchschritt legt passende Dokumentstellen ins Kontextfenster, bevor das Modell antwortet. Der kritische Punkt ist die Suche, nicht das Modell.

Regelbasiertes System

System, dessen Verhalten vollständig durch geschriebene Regeln bestimmt ist. Nachvollziehbar und lesbar.

Reproduzierbarkeit

Eigenschaft, dass identische Eingaben identische Ausgaben erzeugen. Muss durch fixierte Eingabe, gesenkte Streuung und festgehaltene Modellversion hergestellt werden.

Schatten-KI

Nutzung nicht freigegebener KI-Werkzeuge für dienstliche Aufgaben, meist über private Konten.

Temperatur

Einstellgröße für die Zufälligkeit der Fortsetzung. Höhere Werte erzeugen Vielfalt, niedrigere Werte Stabilität.

Token

Textstück, in das ein Modell Eingaben zerlegt. Zwischen einem Zeichen und einem ganzen Wort; im Deutschen etwa 3 bis 4 Zeichen.

Training

Der Vorgang, in dem das Modell aus Daten entsteht. Danach ist sein Wissensstand eingefroren.

Transparenzpflicht

Pflicht, Menschen über den Einsatz von KI und über KI-erzeugte Inhalte zu informieren. Nach Art. 50 KI-VO seit dem 2. August 2026 anwendbar.

Verschmelzung

Fehlermuster, bei dem zwei real existierende Sachverhalte zu einem falschen kombiniert werden. Das am schwersten erkennbare Muster.

Vorhersage

Schätzung eines künftigen Werts. Fehler sind im Soll-Ist-Vergleich messbar.

Wissensstand (Cut-off)

Zeitpunkt, bis zu dem die Trainingsdaten reichen. Alles Spätere ist für das Modell nicht existent, sofern keine Suche angebunden ist.

Zweckbindung

Grundsatz, dass für Zweck A erhobene Daten nicht ohne Weiteres für Zweck B verwendet werden dürfen. Modelltraining ist ein neuer Zweck.

Prüfungsinformationen

Ablauf

Fragen	20, Einfachauswahl und Szenariofragen
Dauer	40 Minuten, ohne Unterbrechung
Bestehensgrenze	70 % (14 von 20)
Hilfsmittel	keine
Bewertung	deterministisch gegen einen von Fachexperten freigegebenen Antwortschlüssel; kein KI-Einsatz bei der Bewertung
Ergebnis	unmittelbar nach Abgabe, mit Auswertung je Themengebiet
Wiederholung	nach 24 Stunden Wartezeit, im Rahmen der erworbenen Versuche
Gültigkeit	befristet; Rezertifizierung nach den Bedingungen des Zertifizierungsschemas

Zulassung

Die Teilnahme an diesem Kurs ist **keine** Voraussetzung für die Prüfung. Du kannst dich auf drei Wegen anmelden: über die Kursvorbereitung, direkt zur Prüfung (Exam-Only) oder über den Nachweis gleichwertiger Vorkenntnisse. Die Kriterien für die Gleichwertigkeitsprüfung sind veröffentlicht.

Vor der Prüfung

Vor Beginn wird dir angezeigt, an welchen Stellen bei CertifAI KI eingesetzt wird und an welchen nicht. Du bestätigst, dass du diese Angaben gelesen hast. Die Prüfung findet unbeaufsichtigt statt; Integritätssignale wie Sitzungsunterbrechungen werden erfasst und, falls auffällig, von einer Person geprüft — automatisch getroffene nachteilige Entscheidungen gibt es nicht.

Beschwerden und Einsprüche

Gegen Zulassungs- und Prüfungsentscheidungen kannst du Einspruch einlegen. Das Verfahren, die Stufen und die Fristen sind öffentlich beschrieben. Über einen Einspruch entscheidet niemand, der an der ursprünglichen Entscheidung beteiligt war.

Wie es weitergeht

Mit *AI Foundations Professional* steht dir die gesamte Leiter offen. Welche Spezialisierung als Nächstes sinnvoll ist, hängt von deiner Rolle ab — die Einordnung auf der Plattform gibt eine Empfehlung, entscheiden musst du selbst.

Zahlen und Quellen

Alle in diesem Handbuch verwendeten Zahlen im Überblick, mit Erhebung, Grundgesamtheit und Fundstelle. Zahlen ohne diese drei Angaben sind nicht verwendbar — das gilt auch für die eigenen.

KENNZAHL	WERT	ERHEBUNG / GRUNDGESAMTHEIT	QUELLE
Unternehmen mit KI-Einsatz (DE)	41 %	604 Unternehmen ab 20 Beschäftigten, KW 2–6/2026, repräsentativ	[1]
Unternehmen, die planen oder diskutieren	48 %	wie oben	[1]
Anwender mit verbesserter Wettbewerbsposition	77 %	wie oben, Teilgruppe der Anwender	[1]
Anwender mit <i>messbarem</i> Erfolgsbeitrag	52 %	wie oben	[1]
Größte Hemmnisse: Rechtsunsicherheit / Fachwissen / Personal	53 % / 53 % / 51 %	604 Unternehmen, KW 27–32/2025	[3]
Unternehmen, die Schatten-KI vermuten oder kennen	42 %	604 Unternehmen ab 20 Beschäftigten, KW 27–32/2025	[2]
Unternehmen, die generative KI bereitstellen	26 %	wie oben	[2]
Unternehmen mit KI-Nutzungsregeln	23 % (2024: 15 %)	wie oben	[2]
Anteil sensibler Daten an KI-Interaktionen	39,7 %	Auswertung realer Datenflüsse, Februar 2026	[6]
Halluzinationsrate, bestes Modell (Zusammenfassung)	1,8 %	HHEM-2.3, Stand 11.05.2026	[5]
Halluzinationsrate, schlechtestes Modell (Zusammenfassung)	24,2 %	wie oben	[5]
Halluzinationsrate, spezifische Rechtsfragen	69–88 %	>200.000 Anfragen je Modell, 2024	[4]

WIE MAN DIESE TABELLE LIEST

Die vier untersten Zeilen sind bewusst untereinander gestellt, obwohl sie nicht vergleichbar sind. Genau das ist die Aussage aus Modul 3.1: Dieselbe Technik liegt je nach Aufgabenzuschnitt zwischen 1,8 % und 88 %. Wer zwei Halluzinationsraten aus verschiedenen Studien nebeneinanderstellt und daraus einen Modellvergleich macht, hat den Griff „Abgrenzung“ aus Modul 6.5 übersprungen.

Quellenverzeichnis

- [1] Bitkom Research (2026): *Digitalisierung der Wirtschaft*. Repräsentative Telefonbefragung von 604 Unternehmen ab 20 Beschäftigten in Deutschland, KW 2–6/2026. Presseinformation vom 11. März 2026.
<https://www.bitkom.org/Presse/Presseinformation/Digitalisierung-der-Wirtschaft-Unternehmen-beschaeftigen-sich-mit-KI>
- [2] Bitkom Research (2025): *Beschäftigte nutzen vermehrt Schatten-KI*. Repräsentative Befragung von 604 Unternehmen ab 20 Beschäftigten, KW 27–32/2025. Presseinformation vom 21. Oktober 2025.
<https://www.bitkom.org/Presse/Presseinformation/Beschaeftigte-nutzen-Schatten-KI>
- [3] Bitkom Research (2025): *Durchbruch bei Künstlicher Intelligenz*. Presseinformation vom 15. September 2025 (Hemmnisse beim KI-Einsatz).
<https://www.bitkom.org/Presse/Presseinformation/Durchbruch-Kuenstliche-Intelligenz>
- [4] Dahl, M.; Magesh, V.; Suzgun, M.; Ho, D. E. (2024): *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models*. Stanford RegLab / Journal of Legal Analysis. Über 200.000 Anfragen je Modell.
<https://reglab.stanford.edu/publications/hlarge-legal-fictions-profiling-legal-hallucinations-in-large-language-models/>
- [5] Vectara (2026): *Hallucination Leaderboard*, Bewertungsmodell HHEM-2.3. Aufgabe: Zusammenfassung eines mitgelieferten Dokuments. Stand: 11. Mai 2026.
<https://github.com/vectara/hallucination-leaderboard/>
- [6] Cyberhaven (2026): *AI Adoption & Risk Report*, veröffentlicht am 11. Februar 2026. Auswertung tatsächlicher Datenflüsse in KI-Werkzeuge.
<https://www.cyberhaven.com/blog/sensitive-data-flowing-into-ai-tools>
- [7] Verordnung (EU) 2024/1689 (KI-Verordnung / AI Act), geändert durch Verordnung (EU) 2026/1744 („Digital Omnibus on AI“), wirksam zum 27. Juli 2026.
<https://artificialintelligenceact.eu/article/4/>

CertifAI ist ein unabhängiger Anbieter. Dieses Zertifikat ist ein Industriezertifikat und keine staatlich anerkannte Qualifikation. Es wird von CertifAI ausgestellt und ist nicht behördlich zugelassen.

Teilnehmerhandbuch AI Foundations Professional (AF) · Version 2.0 · Ausgabe DE · erweiterte und illustrierte Fassung · Rechtsstand August 2026