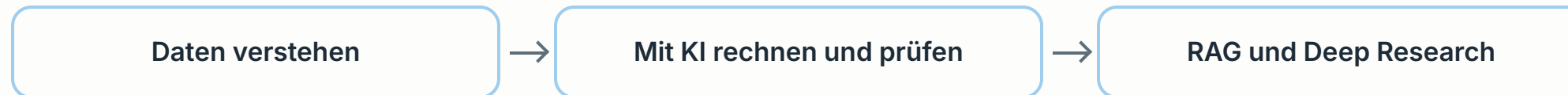


# Die KI rechnet nur mit dem, was du *reingibst*

Ein Sprachmodell weiß nichts, es rechnet. In diesem Teil geht es um das, womit es rechnet: Daten. Stimmen sie nicht, macht die KI daraus trotzdem eine selbstbewusste Empfehlung.

## DEIN WEG DURCH DIESEN TEIL



# Zum Ausprobieren: Lass die KI eine *Tabelle* beschreiben

- 1 Nimm eine kleine Tabelle mit erfundenen Daten, etwa Verkäufe eines Ladens über einen Monat.
- 2 Lade sie in Copilot oder ChatGPT hoch.
- 3 Prompt: „Beschreibe diese Tabelle in fünf Sätzen.“ Genau so, nichts ergänzen.
- 4 Öffne die Datei selbst und prüf jeden Satz: Zeilen, Zeitraum, Filialen.

*Merk dir* Nur erfundene Daten. Echte Kundendaten gehören nicht in ein privates KI-Konto.

GARBAGE IN, *GARBAGE* OUT

# Schlechte Daten rein, schlechte Ergebnisse raus. Auch mit der besten KI.

Früher gab ein falscher Wert eine falsche Zahl, die auffiel. Heute baut die KI daraus einen flüssigen Absatz, der richtig klingt.

# Woran du *gute* Daten erkennst

| MERKMAL     | PRÜFFRAGE                              | BEISPIEL FÜR EINEN FEHLER     |
|-------------|----------------------------------------|-------------------------------|
| Vollständig | Sind alle Pflichtfelder gefüllt?       | Bestellung ohne Menge         |
| Korrekt     | Stimmt der Wert mit der Wirklichkeit?  | Schlauch für minus 24,50 Euro |
| Einheitlich | Gleiche Schreibweise, gleiches Format? | „Süd“ und „Sued“              |
| Aktuell     | Ist der Stand bekannt und passend?     | Kundenliste von 2019          |
| Eindeutig   | Steht jeder Datensatz nur einmal drin? | Bestellung B1052 doppelt      |

*Merk dir* Diese fünf Fragen sind dein Prüfraster für jede Spalte.

# Fehlerjagd: fünf Fehler, *fünf* Merkmale

| ID  | DATUM      | ART        | ANZAHL | KOORDINATE   |
|-----|------------|------------|--------|--------------|
| K01 | 2026-05-04 | Rotmilan   | 2      | 52.41, 13.12 |
| K02 | 2026-05-04 | Feldlerche | 5      | 52.40, 13.10 |
| K03 | 2026-05-11 | Rot-Milan  | 1      | 52.42, 13.14 |
| K04 | 2026-05-11 | Kiebitz    | -2     | 52.39, 13.11 |
| K05 | 2026-05-18 | Feldlerche | 3      | (leer)       |
| K06 | 2019-05-18 | Neuntöter  | 1      | 52.41, 13.13 |
| K07 | 2026-05-25 | Feldlerche | 4      | 52.40, 13.12 |
| K07 | 2026-05-25 | Feldlerche | 4      | 52.40, 13.12 |

Erfundene Brutvogelkartierung, Fläche Nord. Für jedes Merkmal steckt genau ein Fehler drin.

# Strukturiert oder *unstrukturiert*?

## STRUKTURIERT: FESTE SPALTEN

- Jede Spalte hat eine feste Bedeutung: Datum, Filiale, Menge.
- Tabellen, Datenbanken, CSV, Attributtabellen.

## UNSTRUKTURIERT: FREIER TEXT, BILD, TON

- „Hallo, mein Rad war schon wieder nicht fertig, obwohl mir Freitag zugesagt wurde ...“
- E-Mails, PDFs, Protokolle, Luftbilder, Sprachnachrichten.

*Merk dir* Neu: Sprachmodelle arbeiten gut mit unstrukturierten Daten.

# Welche *Datenform* ist das?

| DATENQUELLE                      | LÖSUNG           | WARUM                                                     |
|----------------------------------|------------------|-----------------------------------------------------------|
| Attributtabelle eines Shapefiles | strukturiert     | feste Spalten pro Fläche                                  |
| Luftbild einer Baufläche         | unstrukturiert   | Pixel ohne feste Bedeutung, KI kann Klassen daraus machen |
| Gutachten als PDF                | unstrukturiert   | freier Text, Tabellen darin sind Beiwerk                  |
| Excel-Liste der kartierten Arten | strukturiert     | solange alle dieselben Spalten nutzen                     |
| E-Mail einer Behörde             | halbstrukturiert | Absender und Datum fest, Text frei                        |
| Sprachmemo aus dem Gelände       | unstrukturiert   | Ton: erst Transkript, dann Auswertung                     |

# Fünf *Datenarten*, die du kennen musst

| DATENART               | WAS SIE AUSMACHT            | BEISPIELE                                            |
|------------------------|-----------------------------|------------------------------------------------------|
| Stammdaten             | ändern sich selten          | Kunde, Produkt, Preis                                |
| Bewegungsdaten         | entstehen ständig           | Bestellung, Buchung, Ticket                          |
| Personenbezogene Daten | alles zu einer Person       | Name, Mail, auch ein Kürzel                          |
| Metadaten              | Daten über Daten            | wer, wann, welche Version, welches Koordinatensystem |
| Wissen                 | Texte, die Abläufe erklären | Handbücher, Richtlinien, Protokolle                  |

In einer Verkaufstabelle: Produkt und Einzelpreis sind Stammdaten, Datum, Bestellnummer und Menge Bewegung, das Kürzel der verkaufenden Person ein Personenbezug.

# Wo die Daten im Unternehmen *liegen*

- **ERP:** Warenwirtschaft, Bestellungen.
- **CRM:** Vertrieb, Kundenkontakte.
- **Buchhaltung und Personal:** Rechnungen, Belege, Verträge, Zeiten.
- **SharePoint, Laufwerke, Postfächer:** das große unstrukturierte Feld.
- **Fachsysteme:** zum Beispiel ein GIS mit Geodaten.
- **Excel-Listen auf einzelnen Rechnern:** `Umsatz_final_v3.xlsx` , `Kosten_neu_NICHT_LÖSCHEN.xlsx` .

*Achtung* Excel-Listen sind oft am wichtigsten und oft nur von einer Person verstanden.

# Jeder Datenbestand braucht eine *zuständige* Person

Fachwort: **Data Owner**. Die Person oder Abteilung, die für einen Datenbestand geradesteht.

- 1 Wer pflegt die Daten?
- 2 Wer darf sie ändern?
- 3 Wer gibt sie für die KI frei?
- 4 Wer löscht sie, wenn die Frist abläuft?

*Merk dir* Keine Antwort? Dann hast du die erste Aufgabe für dein KI-Projekt gefunden.

# Aufbewahren und *löschen*

| WAS                          | WIE LANGE          | GRUNDLAGE                                               |
|------------------------------|--------------------|---------------------------------------------------------|
| Bücher, Jahresabschlüsse     | 10 Jahre           | § 257 HGB, § 147 AO                                     |
| Buchungsbelege, Rechnungen   | 8 Jahre, seit 2025 | § 257 HGB, § 147 AO                                     |
| Handels- und Geschäftsbriefe | 6 Jahre            | § 257 HGB, § 147 AO                                     |
| Personenbezogene Daten       | so kurz wie nötig  | Art. 5 Abs. 1 lit. e DSGVO, Recht auf Löschung: Art. 17 |

Prüffrage: Kannst du Daten im KI-System auch wieder löschen? Einordnung, keine Rechtsberatung.

# Drei Fragen vor *jedem* KI-Projekt

- 1 **Welche Daten brauchen wir?** Genau benennen, nicht „alle“.
- 2 **In welchem Zustand sind sie?** Die fünf Merkmale prüfen.
- 3 **Wer ist zuständig und gibt frei?** Der Data Owner.

*Achtung* Ein häufiger Grund für gescheiterte Piloten: nicht das Modell, sondern die Datenlage.

# Was KI mit *Tabellen* gut kann

## AUFGABE

## WAS DU BEKOMMST

Überblick

Spalten, Zeilen, Zeitraum

Rechnen

Summen, Durchschnitte, Rangfolgen

Diagramme

auf Zuruf

Auffälligkeiten

Ausreißer finden

Formeln

Excel-Formeln erklären und bauen

ChatGPT: CSV oder Excel hochladen, in normaler Sprache fragen. Copilot: auch direkt in Excel, je nach Lizenz und Freigabe.

*Merk dir* Erste Antwort. Nicht letzte.

# Wie die KI *eigentlich* rechnet

NUR GELESEN: ZAHL IST GESCHÄTZT

„Der Gesamtumsatz beträgt etwa 3.700 €.“

Das Modell sagt Token voraus. Eine Summe, die gut aussieht, ist noch keine Rechnung.

MIT CODE-AUSFÜHRUNG: ZAHL IST GERECHNET

```
df = pd.read_  
csv("beispieldaten-verkauf.csv")  
df["Umsatz_EUR"].sum()  
# Ergebnis: 3686.59
```

Erkennbar am aufklappbaren Code.  
Ohne Code: Zahl nicht übernehmen.

*Achtung* Auch gerechnet heißt nicht richtig. Der Code rechnet mit „nord“, „Sued“ und der Dublette.

# Zum Ausprobieren: Umsatz pro *Filiale*

FILIALE LAUT KI

UMSATZ IN EURO

|       |          |
|-------|----------|
| Süd   | 1.422,70 |
| Nord  | 1.333,90 |
| Mitte | 906,29   |
| Sued  | 15,80    |
| nord  | 7,90     |

Der Laden hat drei Filialen. Im Diagramm der KI stehen fünf.

*Merk dir* Nachrechnen: Spalte in Excel markieren, Summe unten ablesen. Hier 3.686,59 €.

# Ein Diagramm ist kein *Beweis*

## ACHSE BEGINNT BEI 0

- Süd und Nord aus der Beispieldatei.
- Süd liegt knapp vorn.

## ACHSE BEGINNT BEI 1.300

- Dieselben Zahlen.
- Süd scheint Nord zu überrollen.

*Merk dir* Frag bei jedem KI-Diagramm: Wo beginnt die Achse? Was fehlt?

# Plausibilitätsprüfung in *vier* Schritten

- 1 **Ganze Datei gelesen?** Zeilenzahl und Zeitraum vergleichen.
- 2 **Eine Summe selbst rechnen.** Nur eine. Stimmt sie, steigt das Vertrauen.
- 3 **Größenordnung prüfen.** 100 Bremsreparaturen an einem Tag? Dein Fachwissen prüft.
- 4 **Was hat die KI verändert?** „Liste jede Änderung auf.“

*Merk dir* Diese vier Schritte gehören in jede Datenanalyse mit KI.

# Plausibel oder *nicht*?

| AUSSAGE                                                           | LÖSUNG      | WARUM                                                     |
|-------------------------------------------------------------------|-------------|-----------------------------------------------------------|
| Ein kleiner Fahrradladen macht 100 Bremsreparaturen an einem Tag. | unplausibel | Tippfehler: 135 € Umsatz passen zu drei Reparaturen       |
| Ein Fahrradschlauch kostet minus 24,50 €.                         | unplausibel | negative Preise gibt es nicht, vielleicht eine Gutschrift |
| Der Punkt 52,52° N, 13,40° O liegt in Berlin.                     | plausibel   | das ist Berlin-Mitte                                      |
| Auf 2 Hektar Wiese brüten 40 Rotmilan-Paare.                      | unplausibel | Greifvögel brauchen große Reviere                         |
| Eine Bestellung vom 31.10.2026 im Septemberbericht.               | unplausibel | außerhalb des Monats, klären                              |
| Die mittlere Julitemperatur in Berlin liegt bei 19 °C.            | plausibel   | üblicher Bereich: rund 19 bis 20 °C                       |

# Was nicht in ein *privates* KI-Konto gehört

## DATEN

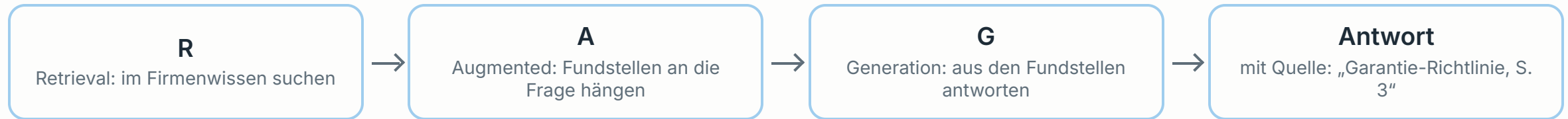
## IN EIN PRIVATES KI-KONTO?

|                         |                    |
|-------------------------|--------------------|
| Echte Kundendaten       | ohne Freigabe nein |
| Personaldaten, Gehälter | nein               |
| Gesundheitsdaten        | nein               |
| Geschäftsgeheimnisse    | nein               |
| Erfundene Daten         | ja                 |
| Anonymisierte Beispiele | ja                 |

Ob ein Unternehmens-Copilot über das Firmenkonto echte Daten verarbeiten darf, entscheidet der Arbeitgeber mit dem Datenschutz.

# RAG: erst *nachschlagen*, dann antworten

FRAGE: „WIE LANGE GARANTIE AUF EINEN FAHRRADMANTEL?“



- **Nicht neu trainieren:** Ändert sich die Regel, tauschst du das Dokument. Das Modell bleibt gleich.
- **Mit Quelle:** Du kannst nachsehen, woher die Antwort kommt.

# RAG *kennst* du schon

## WERKZEUG

## WORAUS ES ANTWORTET

|            |                                                  |
|------------|--------------------------------------------------|
| Copilot    | aus Datei, SharePoint, Postfach, je nach Rechten |
| ChatGPT    | aus Projekten mit eigenen Dokumenten             |
| Gemini     | aus Dateien in Google Drive                      |
| NotebookLM | nur aus deinen Quellen                           |

Gleiches Prinzip: erst nachschlagen, dann antworten. Wie genau es intern läuft, legen die Anbieter nicht immer offen.

# Wo RAG *scheitert*

- **Veraltet:** Alte Dokumente, alte Antworten.
- **Zwei Versionen:** Welche die KI nimmt, ist Glück.
- **Falsche Rechte:** Gehaltsliste für alle freigegeben? Copilot findet sie.
- **Falsche Stelle:** Die Suche findet einen Absatz, in dem das Wort passt, aber nicht der Sinn.

*Merk dir* Vor jedem RAG-Projekt: aufräumen, Versionen klären, Rechte prüfen.

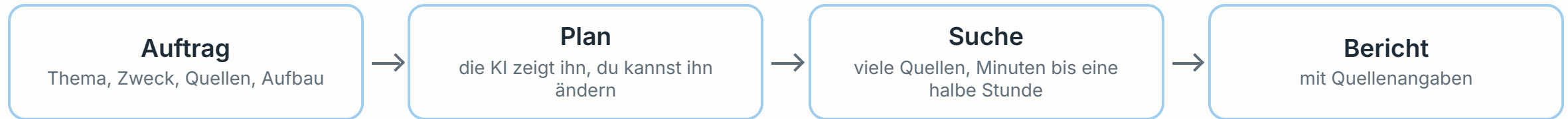
# RAG-Detektiv: Welche *Fundstelle* gewinnt?

| FUNDSTELLE                                     | TEXT                                                         | BEWERTUNG                             |
|------------------------------------------------|--------------------------------------------------------------|---------------------------------------|
| A · Reisekostenrichtlinie 2021, Ordner Archiv  | 0,25 € pro Kilometer mit dem privaten Pkw                    | veraltet, gehört ins Archiv oder raus |
| B · Reisekostenrichtlinie 2026, Ordner aktuell | 0,32 € pro Kilometer mit dem privaten Pkw                    | die richtige, wenn die KI sie wählt   |
| C · Fuhrpark-FAQ                               | Poolfahrzeuge spätestens zwei Tage vorher im Intranet buchen | passt zum Wort, nicht zur Frage       |

Frage an den Firmen-Bot: „Wie viel Kilometergeld bekomme ich für Dienstfahrten mit dem Privatauto?“ Erfundenes Unternehmen, erfundene Beträge.

# Deep Research: Recherche als *Auftrag*

## SO LÄUFT DEEP RESEARCH



- **ChatGPT:** Plus-Symbol am Eingabefeld, dann Deep Research.
- **Gemini:** Tools-Menü am Eingabefeld, dann Deep Research.

*Achtung* 50 Fußnoten sind kein geprüfter Bericht. Quellen selbst anklicken.

# Zum Ausprobieren: ein *Deep-Research*-Auftrag

Erstelle einen Leitfaden für kleine Planungsbüros in Deutschland:  
Welche KI-Werkzeuge helfen bei Kartierung, Geodatenauswertung und Berichten?  
Zeitraum 2025 bis 2026.  
Nur deutsche und europäische Quellen, bevorzugt Behörden, Verbände, Fachpresse.  
Aufbau: Überblick, fünf Werkzeuge mit Kosten, Risiken, Empfehlung.  
Keine Herstellerwerbung.

Bausteine: Thema und Zweck · Frage, Zeitraum, Region · Quellenarten · Aufbau und Ausschlüsse.

# Dein Auftrag: Daten *prüfen*, KI nachrechnen, recherchieren

- 1 **Fehler finden, ohne KI:** Prüf eine erfundene Tabelle Spalte für Spalte mit den fünf Merkmalen. Notier Zeile, Spalte, Fehler, Merkmal und was du tun würdest.
- 2 **KI rechnen lassen:** Gesamtumsatz, Umsatz je Filiale, Balkendiagramm, Datenqualitätsprobleme mit ID.
- 3 **Nachfragen und nachrechnen:** „Hast du Werte verändert, gelöscht oder zusammengefasst?“ Dann selbst rechnen und vergleichen.
- 4 **Deep Research:** eine Frage zu deinem Thema, danach drei Quellen prüfen.
- 5 **Kursbuch:** Kapitel 4 mit den Glossarbegriffen dieses Teils.

# Prüf dich *selbst*

**1** „Süd“ und „Sued“ in einer Spalte: Welches Merkmal fehlt?

Einheitlich.

**2** Ein Gutachten als PDF ist strukturiert, halbstrukturiert oder unstrukturiert?

Unstrukturiert.

**3** Wie heißt die Person, die für einen Datenbestand geradesteht?

Data Owner.

**4** Woran erkennst du eine gerechnete KI-Zahl?

Am aufklappbaren Code. Und auch dann rechnet der Code nur mit den Daten, die drinstehen.

**5** Zwei Versionen einer Richtlinie liegen im SharePoint. Welche nimmt RAG?

Zufall. Deshalb aufräumen.

**6** Was spart Deep Research, und was nicht?

Suchzeit, aber nicht die Prüfung der Quellen.

# Weiter im Kurs

## LERNPFAD

Lerntage zum Nachklicken

4 · Datenanalyse und  
Datenmanagement

## GLOSSAR

KI-Glossar mit 93 Begriffen

## NÄCHSTER TEIL

Teil 5: Ethik, Bias und der EU AI  
Act

Erik



Dir reicht Selbstlernen nicht? Dann lernst du das mit mir 1:1. Auf Wunsch bringe ich es auch Unternehmerinnen, Unternehmern und ganzen Teams bei. Schreib mir, was du brauchst.

Erstgespräch anfragen >