AI Skills, Powered by Values

values

www.valuesai.eu

MODUL 04

Zuverlässige KI-gestützte Inhalte

Quellen, Prüfung und Kontrolle von Halluzinationen

Co-funded by the European Union

Mitfinanziert von der Europäischen Union. Die geäußerten Ansichten und Meinungen sind jedoch ausschließlich die der Autor*innen und spiegeln nicht unbedingt die Ansichten der Europäischen Union oder der Stiftung für die Entwicklung des Bildungssystems wider. Weder die Europäische Union noch die fördernde Einrichtung können dafür verantwortlich gemacht werden.

Inhaltsverz

Co-funded by the European Union

Mitfinanziert von der Europäischen Union. Die geäußerten Ansichten und Meinungen sind jedoch ausschließlich die der Autor*innen und spiegeln nicht unbedingt die Ansichten der Europäischen Union oder der Stiftung für die Entwicklung des Bildungssystems wider. Weder die Europäische Union noch die fördernde Einrichtung können dafür verantwortlich gemacht werden.

01

Glaubwürdige Inhalte im KI-Zeitalter erstellen

(10 Min.)

Zuverlässige KI-Inhalte

Glaubwürdige Inhalte im KI-Zeitalter zu erstellen bedeutet nicht nur, schnell Text zu produzieren, sondern vor allem Genauigkeit, Transparenz und Verantwortung sicherzustellen. Der Einsatz von KI mindert Qualität nicht automatisch, erfordert aber stärkere menschliche Kontrolle, besonders wenn Inhalte Fakten, Daten, Gesundheit, Politik oder gesellschaftliche Themen betreffen.

AI Skills,
Powered
by Values
values

Zuverlässige KI-Inhalte – Grundprinzipien

- **Quellen nennen.** Beziehe dich auf Studien, Berichte, offizielle Stellungnahmen und Originaldaten.
- **Vor dem Veröffentlichen prüfen.** Jeder Inhalt sollte auf Fakten, Zahlen und Kontext geprüft werden.
- **KI-Nutzung offenlegen.** Transparenz stärkt das Vertrauen des Publikums.
- **Klar und präzise schreiben.** Vermeide vage Formulierungen, die zu Missverständnissen führen können.
- **Inhalte aktualisieren.** Wenn neue Daten erscheinen oder ein Fehler gefunden wird, überarbeite das Material.
- **Klare Autorenschaft sichern.** Inhalte sollten eine erkennbare Autor*in oder redaktionell verantwortliche Stelle haben.
- **Nicht manipulieren. KI nicht zum absichtlichen Irreführen nutzen – auch nicht für Reichweite.**

Zuverlässige KI-Inhalte – was Vertrauen schafft

Glaubwürdige Inhalte erklären meist nicht nur, was gilt, sondern auch warum und auf welcher Grundlage. Deshalb funktioniert die Kombination aus Daten, Fachkommentar und praktischen Beispielen gut.

- Veröffentlichungs- und Aktualisierungsdaten,
- Links zu Primärquellen,
- eine Fachperson oder Autor*in mit relevanter Qualifikation,
- eine logische Textstruktur,
- eine klare Unterscheidung zwischen Fakt, Meinung und Interpretation,
- keine reißerische oder Clickbait-artige Sprache.

In der Praxis: KI kann unterstützen, aber Menschen bleiben für Wahrhaftigkeit und Bedeutung verantwortlich.

Warum KI-Inhalte zuverlässig sein müssen

KI-Tools erzeugen Inhalte schnell, aber jede Ausgabe bleibt ein Entwurf: flüssiger Text, plausible klingenden Zahlen und überzeugende Zusammen-fassungen können dennoch falsch sein.

Warum das wichtig ist:

- Nach Veröffentlichung können schwache Inhalte Glaubwürdigkeit schädigen und Menschen irreführen.
- KI kann Quellenangaben, Statistiken und Zitate erfinden, die echt aussehen (Halluzinationen).
- „Zuverlässig“ bedeutet mehr als flüssig – es bedeutet belegt, geprüft und verantwortbar.

Beispiel: ChatGPT liefert oft eine flüssige Definition ohne Quellen. NotebookLM kann mit geladenen PDFs eine belegte Definition mit realen Autor*innen liefern.

02

Risiken von KI-Inhalten

(20 Min.)

KI-Inhalte bergen drei Risiken:

Halluzinationen, fehlende Quellen und KI-Erkennung

Bei KI-Inhalten achten viele nur auf Tempo. In der Praxis bergen KI-Ausgaben drei zentrale Risiken.

01

Halluzinationen (falsche Fakten):

KI kann Daten, Zahlen, Namen und Zitate erfinden, die aber dennoch echt klingen.

02

Quellenrisiko:

Modelle wie ChatGPT 4.0 erzeugen flüssigen Text ohne reale Quellen. Bibliografien fehlen, sind vage oder erfunden.

03

KI-Erkennungsrisiko:

Detektoren markieren generische KI-Texte. Unbearbeitete Texte zeigen bis zu 100 % KI; fundierte, bearbeitete Ausgaben unter 10 %.

„Sieht zuverlässig aus“ – ist es aber nicht (Mythos vs. Realität)

Achte auf:

- Quellen mit falschen Seitenzahlen
- echte Namen + erfundene Artikel
- plausible Zahlen ohne Quelle
- flüssige Definitionen die der aktuellen Literatur widersprechen

Mythos *„Akademischer Klang = zuverlässig.“*

Fakt KI kann akademisch klingen und trotzdem Fakten erfinden.

Mythos *„Detektor sagt: menschlich = sicher.“*

Fakt Detektoren prüfen Stil. Bearbeitete KI kann menschlich wirken und trotzdem Fehler enthalten.

Was ist AI SLOP?

AI SLOP = minderwertige, irreführende oder ungenaue Inhalte, die von KI erzeugt wurden.

Es klingt überzeugend, enthält aber:

- erfundene Fakten
- falsche Zahlen oder Daten
- falsche Namen, Quellenangaben oder wissenschaftliche Aussagen
- wiederholende, generische Struktur
- „glatten“ Text ohne echte Belege

Warum passiert das?

01

Das Modell sagt Muster voraus, nicht Wahrheit – es erzeugt auf Basis von Daten die wahrscheinlichste Antwort, „weiß“ aber nicht wirklich, ob sie wahr ist.

02

Es kann Lücken mit erfundenen Details füllen – wenn Informationen fehlen, kann das Modell etwas ergänzen, das logisch klingt, aber nicht der Realität entspricht.

03

Es hat nicht immer aktuelle Informationen – wenn das Wissen des Modells zeitlich begrenzt ist, kann es veraltete Fakten liefern oder neue Ereignisse übersehen.

04

Es kann Kontext missverstehen – eine unklare Frage, Kurzform oder Mehrdeutigkeit kann zu einer falschen Antwort führen.

Warum passiert das?

05

Es klingt selbstbewusst, auch wenn es falsch ist – die Formulierung kann sehr zuverlässig wirken, selbst wenn der Inhalt Fehler enthält.

06

Es kann echte Fakten mit Halluzinationen mischen – ein Teil der Antwort kann korrekt sein, während andere Teile erfunden oder ungeprüft sind.

07

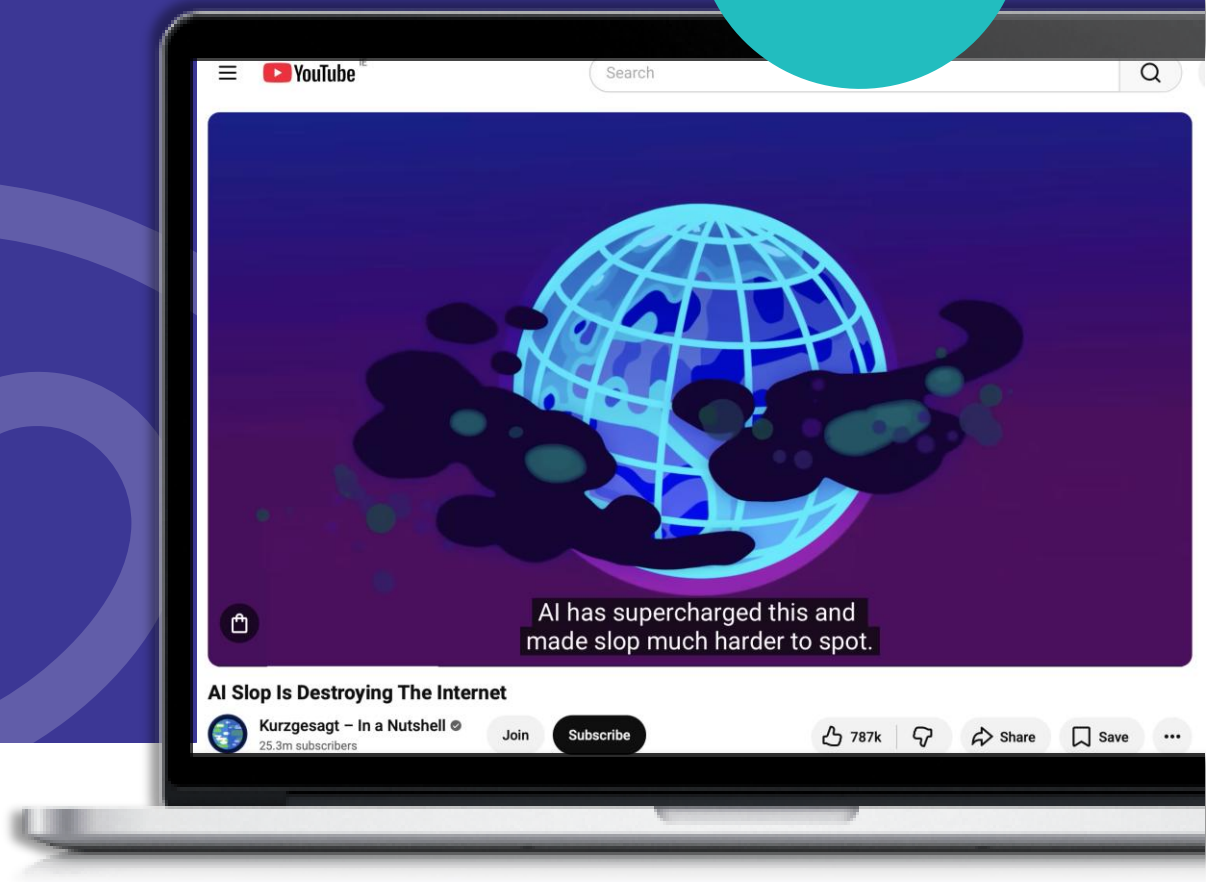
Es hängt von der Qualität der Eingabe ab – wenn der Prompt unklar, widersprüchlich oder unvollständig ist, kann die Ausgabe weniger genau sein.

Deshalb wird Überprüfung zu einer zentralen digitalen Kompetenz.

Wie AI SLOP das Internet zerstört

VIDEO

KI-generierte Inhalte überfluten das Internet – von Videos und Musik bis hin zu Nachrichten und Büchern. Darüber hinaus lässt generative KI falsche Informationen überzeugender wirken als je zuvor. Wir treten in eine neue Ära der Informationsüberlastung ein, und es wird immer schwieriger herauszufinden, was echt ist und was nicht.



03

Verantwortliche Autorenschaft

(10 Min.)

Verantwortungsbewusste Perspektive

Beachte vier zentrale Fragen, wenn du Inhalte mit KI erstellst:

Genauigkeit	Quellen	Originalität	Verantwortung
Habe ich Fakten, Daten und Zitate anhand vertrauenswürdiger Quellen geprüft?	Kann ich reale Veröffentlichungen für jede zentrale Aussage angeben?	Habe ich den KI-Entwurf so bearbeitet, dass er nicht wie rohe KI-Ausgabe wirkt?	Kann ich erklären, wie ich KI genutzt habe, was ich geändert habe und warum?

Zuverlässige Inhalte = kluge KI-Nutzung, geleitet von Verantwortung.

Verantwortungsbewusste Perspektive

Sicherheit beginnt vor der KI	Rechte und Einwilligung zählen	Fairness ist Teil von Sicherheit	Sicherere Workflows sind möglich
Was du eingibst oder hochlädst, zählt zuerst	Daten anderer Menschen brauchen Erlaubnis und Sorgfalt	Verzerrte oder minderwertige Daten können Schaden verursachen	Du kannst oft KI nutzen und viel weniger teilen

Bevor du KI nutzt: Jede Ausgabe ist ein Entwurf

Viele Menschen konzentrieren sich auf die Geschwindigkeit von KI. Zuverlässige Inhalte beginnen früher: bei dem, was du prüfst, belegst und bearbeitest.

In der Praxis erzeugen KI-Tools:

- Definitionen und Erklärungen (oft ohne Quellen)
- Zusammenfassungen von Artikeln (manchmal erfunden oder verzerrt)
- Literaturlisten (häufig erfunden)
- Datenanalysen und Interpretationen (ohne Methodik)

Bevor du KI nutzt: Jede Ausgabe ist ein Entwurf

Zuverlässige KI-Nutzung beginnt mit einer einfachen Gewohnheit: Behandle jede Ausgabe als Entwurf, der geprüft werden muss – nicht als endgültige Antwort.

AI Skills, Powered by Values

values

KI-Ausgaben sind Entwürfe, keine Wahrheit

Welche Inhalte musst du prüfen?

- Definitionen und theoretische Aussagen
- Statistiken, Prozentwerte und Daten
- Namen von Autor*innen, Büchern und Artikeln
- Datenanalysen und Schlussfolgerungen

TUN

TUN: Denke daran: Jede Tatsache, Quelle und Zahl aus KI muss vor der Nutzung geprüft werden.

NICHT TUN

NICHT TUN: Nicht annehmen: „Es klingt professionell“ oder „Es muss stimmen, weil KI es gesagt hat.“

Nie blind vertrauen (NTL)

Nicht belegbar? Nicht veröffentlichen.

Nicht ohne Prüfung veröffentlichen:

- **Quellenangaben:** jeder Verweis auf ein Buch, einen Artikel oder eine Autorin bzw. einen Autor, den die KI nennt
- **Zitate:** jedes „direkte Zitat“, das die KI einer Person zuschreibt
- **Statistiken:** Prozentwerte, Stichprobengrößen, Studienergebnisse, Rankings
- **Historische Fakten:** Daten, Ereignisse, Orte, Namen von Gesetzen oder Verträgen
- **Definitionen und Theorien:** Aussagen darüber, was ein Konzept „ist“ oder wer es vorgeschlagen hat
- **Numerische Analysedaten:** Kund*innenlisten, HR-Daten, interne Dokumente, nicht öffentliche Zahlen

Wie du KI-Ausgaben zuverlässiger machst (vor der Veröffentlichung)

Wende vor der Veröffentlichung von KI-Ausgaben diese Prüfungen an:

- **Quellen prüfen:** Öffne Google Scholar, Scopus oder die Website des Verlags.
- **Prüfe jede Statistik** mit mindestens einer unabhängigen Quelle gegen.
- **Führe den Prompt erneut in einem quellenbasierten Tool aus** (z. B. NotebookLM mit geladenen echten PDFs).
- **Struktur bearbeiten**, generische Formulierungen ersetzen, konkrete Beispiele hinzufügen.

Quelle nicht gefunden? Nicht veröffentlichen.

04

Arbeit mit Quellen

(15 Min.)

Von KI angegebene Quellen

01

Glaubwürdige Quellen nutzen – wähle Materialien aus angesehenen Fachzeitschriften, öffentlichen Institutionen, wissenschaftlichen Datenbanken und Fachorganisationen.

02

Wissenschaftliche, Branchen-, Medien- und Meinungsquellen unterscheiden – jede Quellenart erfüllt eine andere Funktion und bietet ein anderes Maß an Zuverlässigkeit.

03

Veröffentlichungsdaten prüfen – Aktualität ist wichtig, besonders in schnelllebigen Bereichen, in denen KI veraltete Informationen liefern kann.

04

Quellenangaben prüfen – prüfe, ob die genannten Autor*innen, Titel, Jahre und zitierten Aussagen tatsächlich existieren und zu dem passen, was Text oder KI behaupten.

Von KI angegebene Quellen

05

Primär- und Sekundärquellen unterscheiden – eine Primärquelle enthält Originalmaterial, während eine Sekundärquelle frühere Arbeiten diskutiert oder interpretiert.

06

Informationen gegenprüfen – wichtige Aussagen sollten durch mindestens zwei oder drei unabhängige Quellen bestätigt werden.

07

Autor*in, Verlag und Zweck bewerten – frage immer, wer den Text geschrieben hat, wo er veröffentlicht wurde und ob er informieren, verkaufen, überzeugen oder kommentieren soll.

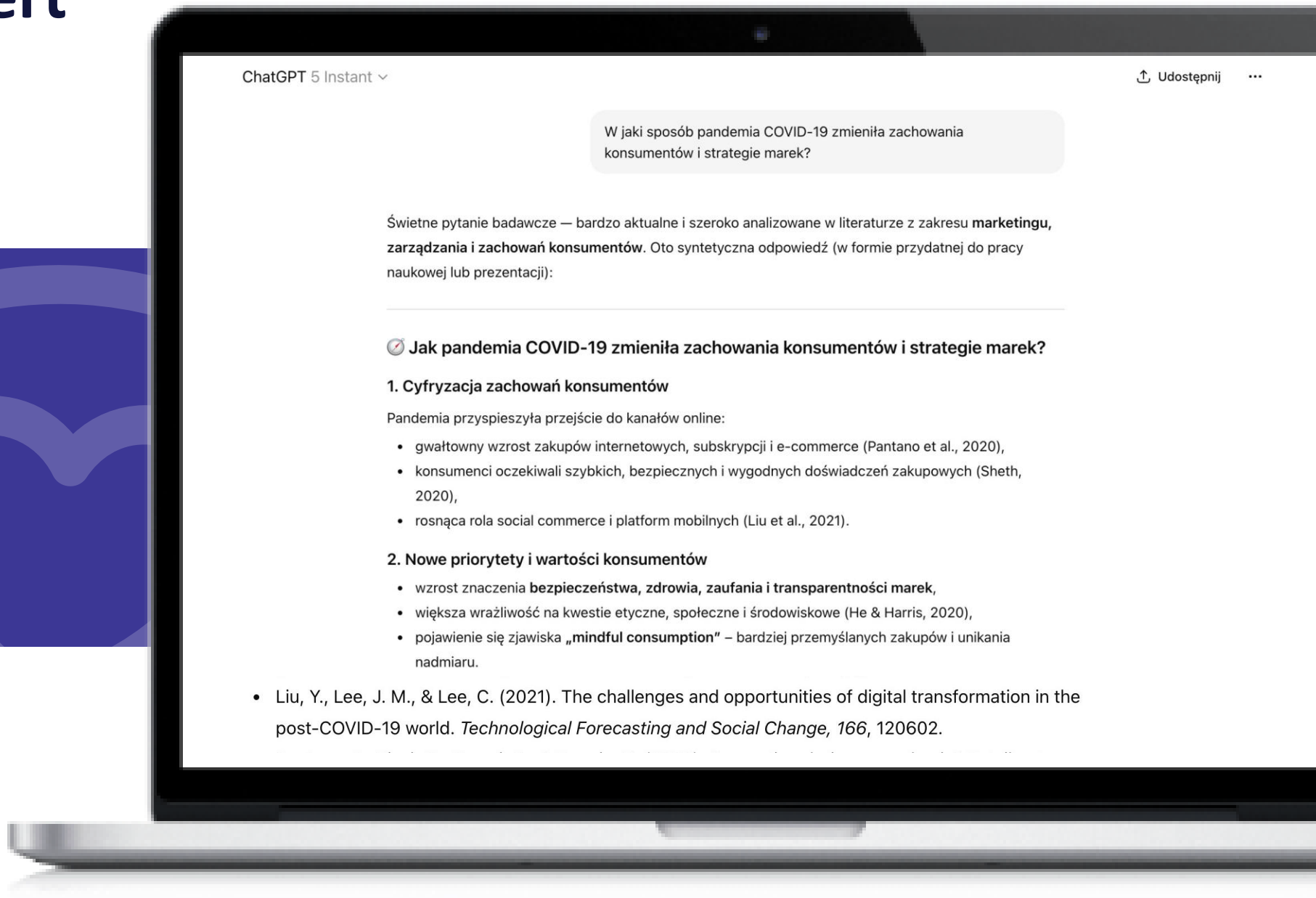
Beispiel: KI liefert eine falsche Quelle

Diese Folie zeigt ein Gespräch mit ChatGPT 5 Instant. Die Nutzerfrage lautet, wie die COVID-19-Pandemie Konsumverhalten und Markenstrategien verändert hat. Die KI liefert eine strukturierte Antwort mit Quellenverweisen auf wissenschaftliche Arbeiten (Pantano et al., 2020; Sheth, 2020; Liu et al., 2021; He & Harris, 2020) und diskutiert die Digitalisierung des Konsumverhaltens sowie neue Prioritäten wie Sicherheit, Gesundheit, Vertrauen und „mindful consumption“.

Bei der Überprüfung einer Quellenangabe stellte sich heraus, dass der Artikel „The challenges and opportunities of digital transformation in the post-COVID-19 world“ NICHT in Technological Forecasting and Social Change erschienen ist, wie die KI behauptete, sondern in IEEE Engineering Management Review, Vol. 48(3), S. 97–103.

Siehe nächste Folie

Beispiel: KI liefert eine falsche Quelle

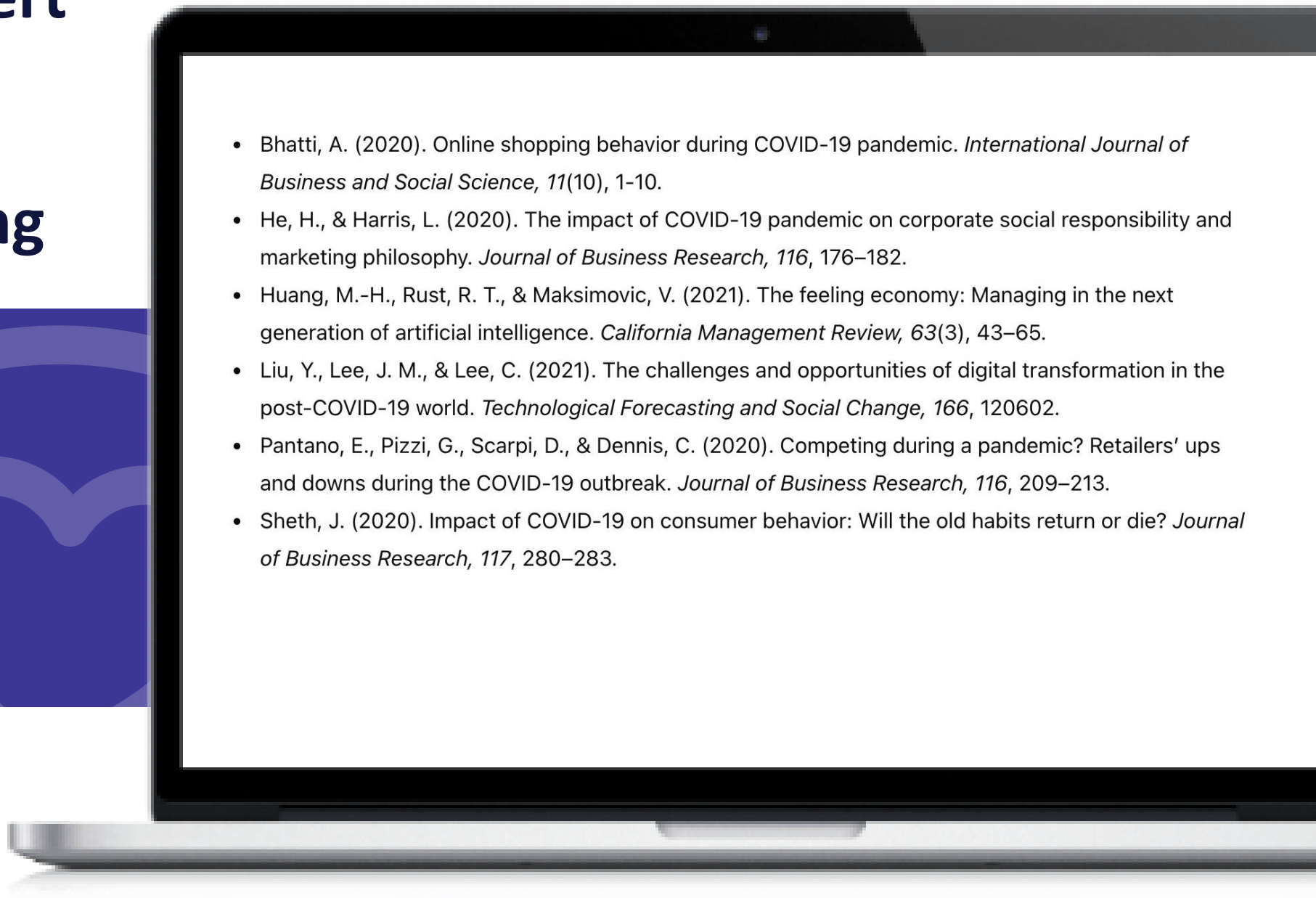


Beispiel: KI liefert eine falsche Quelle. Schlussfolgerung

- Die KI ordnete einen realen Artikel der falschen Zeitschrift zu und zitierte ihn fälschlich als Beitrag in Technological Forecasting and Social Change statt in IEEE Engineering Management Review.
- Sie erzeugte eine plausible, aber falsche bibliografische Angabe einschließlich Zeitschrift und Publikationsdetails – also eine irreführende APA-Quellenangabe.
- Die KI prüfte die Publikationsdaten (Zeitschrift, Verlag, DOI) nicht, obwohl die Quelle leicht überprüfbar war.
- Sie kombinierte einen realen Titel mit einem falschen Publikationskontext – ein typischer Fall von Quellenhalluzination.
- Dieser Fehler untergräbt wissenschaftliche Zuverlässigkeit und kann zu falschen Literaturverzeichnissen oder Problemen im Peer Review führen.

Siehe nächste Folie

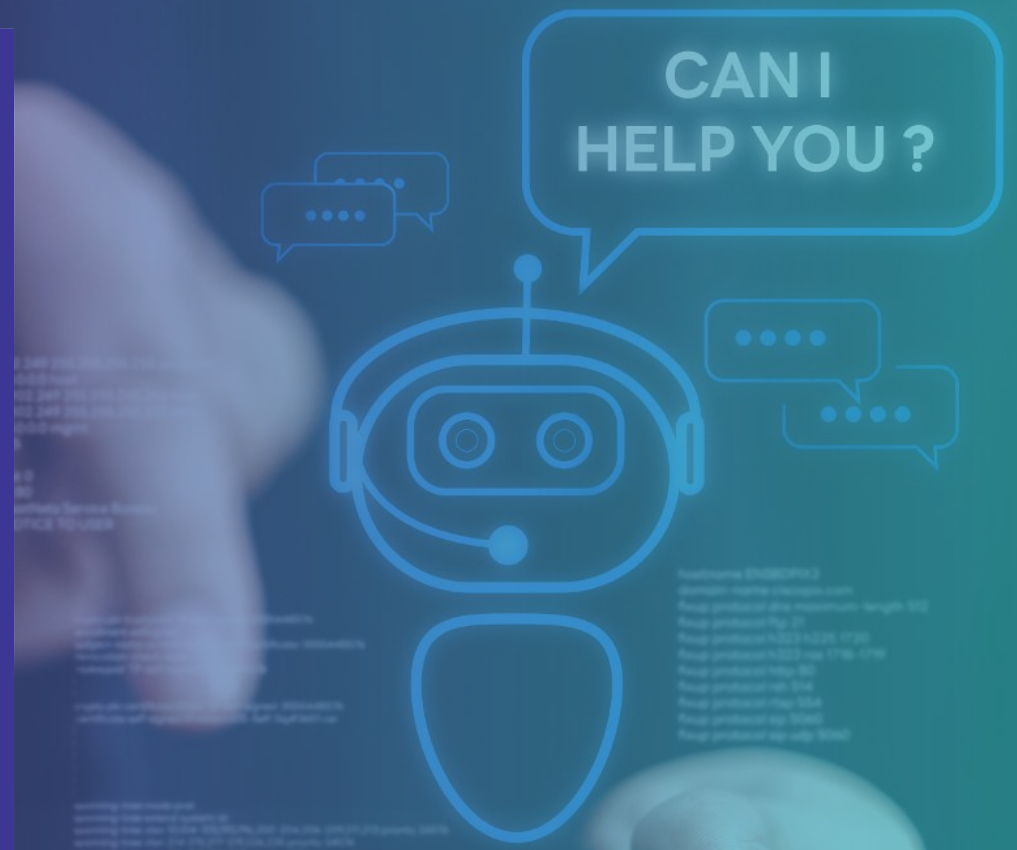
Beispiel: KI liefert eine falsche Quelle. Schlussfolgerung



- Bhatti, A. (2020). Online shopping behavior during COVID-19 pandemic. *International Journal of Business and Social Science, 11*(10), 1-10.
- He, H., & Harris, L. (2020). The impact of COVID-19 pandemic on corporate social responsibility and marketing philosophy. *Journal of Business Research, 116*, 176–182.
- Huang, M.-H., Rust, R. T., & Maksimovic, V. (2021). The feeling economy: Managing in the next generation of artificial intelligence. *California Management Review, 63*(3), 43–65.
- Liu, Y., Lee, J. M., & Lee, C. (2021). The challenges and opportunities of digital transformation in the post-COVID-19 world. *Technological Forecasting and Social Change, 166*, 120602.
- Pantano, E., Pizzi, G., Scarpi, D., & Dennis, C. (2020). Competing during a pandemic? Retailers' ups and downs during the COVID-19 outbreak. *Journal of Business Research, 116*, 209–213.
- Sheth, J. (2020). Impact of COVID-19 on consumer behavior: Will the old habits return or die? *Journal of Business Research, 117*, 280–283.

Beispiel: KI liefert eine falsche Quelle. Schlussfolgerung

KI kann Literatur effektiv zusammenfassen, aber alle KI-generierten Quellenangaben müssen vor der Nutzung in wissenschaftlichen Arbeiten unabhängig überprüft werden.



AI Skills,
Powered
by Values
values

Übung

Bitte ein KI-Modell um einen einseitigen A4-Text zu einem Thema. Bitte es, keine Websites zu nutzen, sondern Veröffentlichungen wie Artikel, Bücher oder Berichte als Quellen.

- Erstelle eine Quellenliste.
- Prüfe anschließend nach guter Praxis die Richtigkeit der angegebenen Quellen: Existieren sie wirklich?
- Sind sie zugänglich?
- Nennen sie die richtigen Autor*innen?

Wenn du einen KI-Fehler entdeckst, teile deine Beobachtungen mit der Gruppe und der Kursleitung.

Szenario: ChatGPT 4.0 vs. NotebookLM

Eine Forschungsfallstudie verglich zwei Tools bei derselben Aufgabe: *„Erstelle eine Definition von Innovationsstrategie mit Bibliografie.“*

Derselbe Prompt erzeugte zwei Ausgaben, die zeigen, was Quellenfundierung verändert:

- ChatGPT 4.0: flüssige Definition, kein echtes Literaturverzeichnis, von Originality.ai als 100 % KI erkannt.
- NotebookLM (mit realen PDFs): fundierte Definition mit echten Autor*innen (Porter, Janasz, Pomykalski, Freeman); 92 % Originalitätswert.

Was das in der Praxis bedeutet:

- Ohne Quellenfundierung erfindet KI flüssige, aber unbelegte Inhalte. Mit echter Wissensbasis erzeugt dasselbe Modell überprüfbare Texte, die an konkrete Veröffentlichungen gebunden sind.

ChatGPT 4.0 vs. NotebookLM: Was quellenbasierte Verankerung verändert

Quellen sind nicht nur akademisch – sie machen KI-Ausgaben verantwortbar.

ChatGPT 4.0 (ohne Quellenfundierung):

- Flüssiger Text, vereinfachte Definitionen, kein Literaturverzeichnis
- Plausible Unternehmensbeispiele (Google, Tesla, Apple) ohne „Studien“ dahinter.
- 100 % als KI-generiert von Originality.ai erkannt

NotebookLM (mit geladenen PDFs):

- Theoriegestützte Definitionen mit Porter, Janasz, Pomykalski und Freeman
- Echte Forschungstraditionen: Innovationstypologien, Netzwerkökonomie, KMU-Kontext.
- Konkretes Literaturverzeichnis im APA-Format, überprüfbar
- Stil von Originality.ai als 92 % original bewertet

Quellen und Rechte: „Habe ich eine echte Referenz?“

Zuverlässige KI-Nutzung betrifft nicht nur Fakten. Es geht auch um Autorenschaft: Wer hatte die ursprüngliche Idee, wessen Arbeit zitierst du und wie gibst du Anerkennung?

KI-Ausgaben nicht ungeprüft vertrauen bei:

- exakte Statistiken, Prozentwerte, Stichproben
- konkrete Zitate mit Autor*innen
- Publikationsjahre, Seitenzahlen, Zeitschriftentitel
- Definitionen von Nischen- oder neuen Konzepten
- allgemeine Erklärungen bekannter Konzepte
- Strukturüberarbeitung (Grammatik, Lesefluss, Absätze)
- Zusammenfassung eigener Texte

Wie eine „echte Quelle“ aussieht (und wie nicht)

Zuverlässige KI-Nutzung betrifft nicht nur Fakten. Es geht auch um Autorenschaft: Wer hatte die ursprüngliche Idee, wessen Arbeit zitierst du und wie gibst du Anerkennung?

Eine echte Quelle ist nicht nur Name und Jahr. Bei KI-Zitaten muss sie sein:

- **Real:** Die Veröffentlichung existiert und ist auffindbar.
- **Spezifisch:** Sie passt zur Aussage, nicht nur zum Thema.
- **Zugänglich:** Du kannst die relevante Stelle lesen.
- **Autoritativ:** glaubwürdige Zeitschrift, Verlag oder Institution.
- **Wichtig:** Einwilligung hängt von Publikum und Tool ab.

Eine Person kann dir etwas erzählen, aber nicht zustimmen, dass es mit KI geteilt oder gespeichert wird.

05

KI-Halluzinationen

(10 Min.)

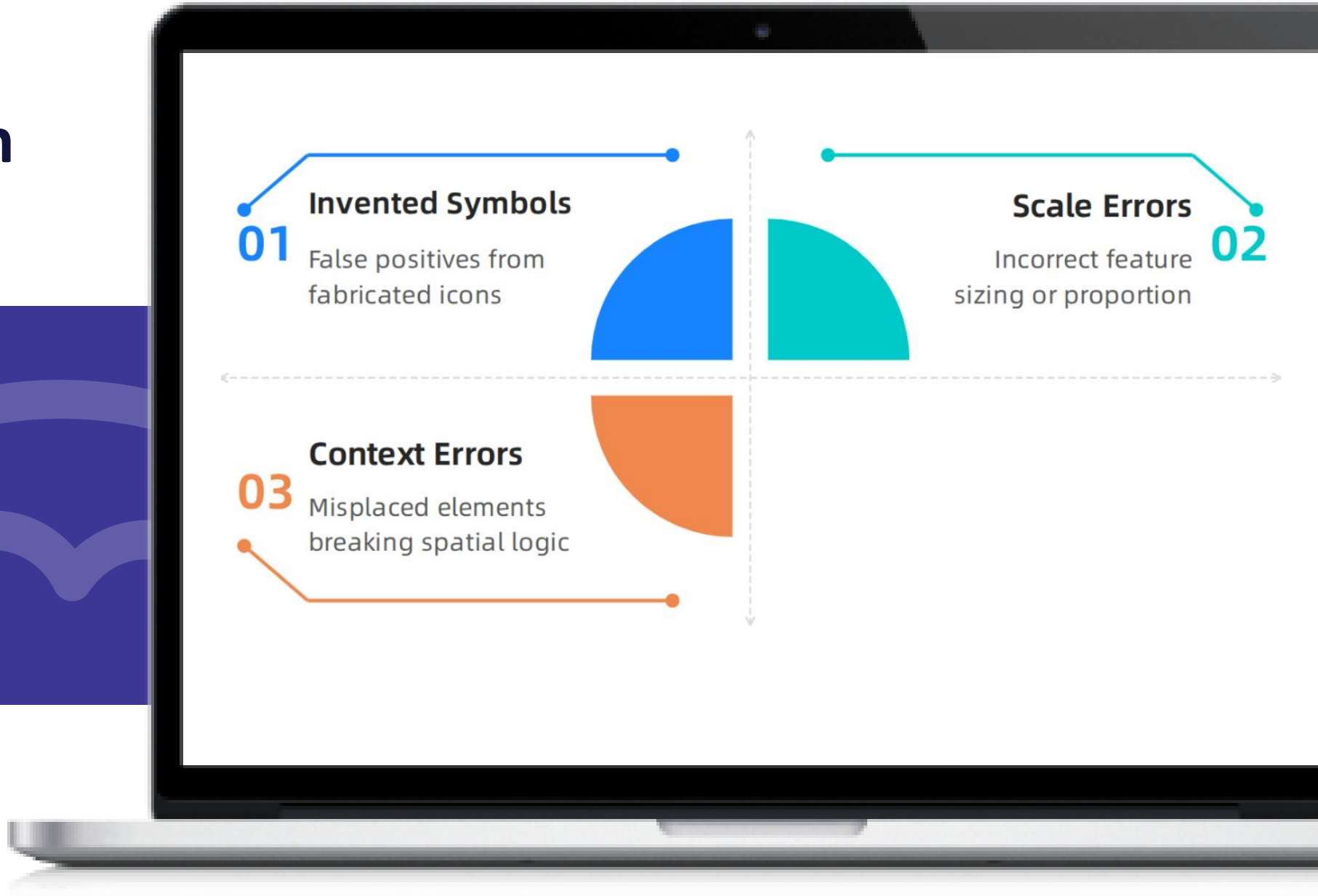
Was sind KI-Halluzinationen?

KI-Halluzinationen sind Modellantworten, die plausibel klingen, aber falsche, erfundene oder irreführende Informationen enthalten. Das Modell prüft nicht die Realität, sondern sagt wahrscheinlichen Text voraus.

Beispiele sind erfundene Zitate, nicht existierende Quellen, falsche Daten oder irreführende Erklärungen. Gemeint sind Fehler der Informationsgenerierung – keine medizinischen Halluzinationen.

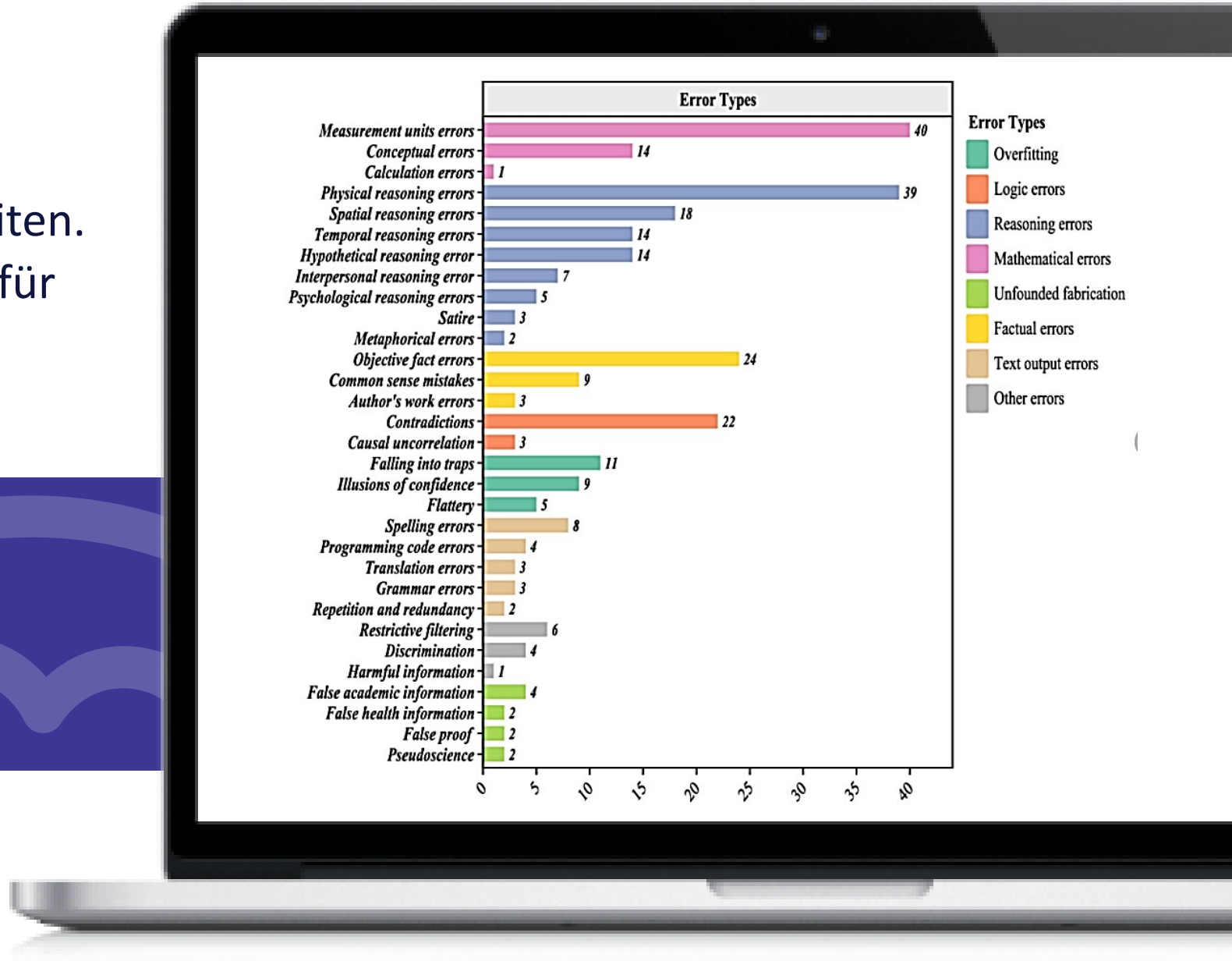
Karte der KI-Halluzinationen

Wie KI-Halluzinationen in Karten mit abgestuften Symbolen erkannt und behoben werden Karten. Aihallucinationreport.com. https://aihallucinationreport.com/how-to-detect-and-fix-ai-hallucinations-in-graduated-symbol-maps/



KI-Halluzinationen

Statistiken für 284 Analyseeinheiten. Unterschiedliche Farben stehen für unterschiedliche Fehlertypen.



Sun, Yujie & Sheng, Dongfang & Zhou, Zihan & Wu, Yifei. (2024). AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. Humanities and Social Sciences Communications. 11. 10.1057/s41599-024-03811-x.

Arten von KI-Halluzinationen

01

Faktische Halluzinationen – das Modell nennt falsche Fakten, Ereignisse, Personen oder Definitionen.

02

Quellenhalluzinationen – KI erfindet Quellen, Veröffentlichungen, Autor*innen oder Titel oder ordnet Aussagen nicht existierenden Texten zu.

03

Zitathalluzinationen – das Modell nennt ein Zitat, das nie gesagt wurde, oder schreibt es fälschlich einer Person zu.

04

Numerische Halluzinationen – KI verrechnet Daten oder verwechselt Prozentwerte, Daten, Statistiken oder Tabellenwerte.

Arten von KI-Halluzinationen

05

Kontextuelle Halluzinationen – das Modell missversteht die Frage oder fügt Details hinzu, die der/die Nutzer*in nicht beabsichtigt hat.

06

Gemischte Halluzinationen – die Antwort enthält sowohl richtige Fakten als auch erfundene Elemente, wodurch der Fehler schwerer zu erkennen ist.

Ursachen von KI-Halluzinationen

01

Das Modell sagt Muster voraus, nicht Wahrheit – es erzeugt die wahrscheinlichsten Wörter, statt Fakten zu prüfen.

02

Trainingsdaten können unvollständig oder fehlerhaft sein – wenn die Daten Lücken, Fehler oder Widersprüche enthalten, kann das Modell sie reproduzieren.

03

Fehlende aktuelle Informationen – das Modell kennt möglicherweise neue Ereignisse oder Änderungen nicht.

04

Unklarer Prompt-Kontext – bei mehrdeutigen oder verkürzten Prompts kann das Modell die Absicht falsch verstehen.

Ursachen von KI-Halluzinationen

05

Systeme belohnen flüssige Antworten – Modelle werden oft darauf trainiert, selbstbewusst und glatt zu antworten, auch wenn sie unsicher sind.

06

Keine eingebaute Faktenprüfung – ohne Zugriff auf externe Quellen kann das Modell nicht überprüfen, ob seine Antwort wahr ist.

Indirekte Halluzinationen: wenn Namen und Daten richtig aussehen, es aber nicht sind

Häufige indirekte Halluzinationen:

- Reale*r Autor*in + echte Zeitschrift + erfundener Titel
- Echter Artikel + falsches Jahr, falsche Seite oder Band
- Echtes Konzept + falscher Autor („Porter“ statt „Schumpeter“)
- Echte Statistik + erfundener Kontext

Sicherere Alternativen:

- **Suchen: Exakten Artikel in Google Scholar prüfen.**
- **Ersetzen: Nur gelesene Quelle zitieren.**
- **Abschwächen: allgemeiner formulieren, wenn keine konkrete Quelle belegbar ist.**
- **Entfernen: unbelegte Aussage löschen.**

KI kann auch dann halluzinieren, wenn Namen, Daten und Titel korrekt wirken.
Das Muster ist plausibel, aber die Verbindung zum Originalwerk ist falsch.

06

Inhaltsprüfung

(10 Min.)

KI-generierte Inhalte können nützlich sein – aber nur nach Prüfung veröffentlichen oder nutzen.

Prüfe Aussagen, Zahlen, Zitate und Quellen bevor du sie nutzt.

Behandle die KI-Antwort als Entwurf, nicht als endgültige Wahrheit.

Dokumentiere, was geprüft wurde und auf welcher Grundlage.

Schritt 1: Risiko und zentrale Aussagen identifizieren

Identifiziere zuerst, welche Teile der KI-Antwort zwingend geprüft werden müssen: Fakten, Daten, Namen, Zahlen, Zitate, Definitionen und genannte Quellen. Je größer die Auswirkungen auf Gesundheit, Recht, Bildung, Finanzen oder Reputation, desto strenger sollte die Überprüfung sein.

Welche Elemente könnten Schaden verursachen, wenn sie falsch sind?

Hat die KI echte Quellen angegeben oder klingt sie nur überzeugend?

Was ist in dieser Antwort Fakt und was ist Interpretation?

Schritt 2: Fakten und Quellen überprüfen

Jede wichtige Aussage sollte anhand glaubwürdiger externer Quellen geprüft werden, vorzugsweise Primärquellen: wissenschaftliche Artikel, Berichte, Websites öffentlicher Institutionen und seriöse Branchenpublikationen. Gehe niemals davon aus, dass eine von KI genannte Quelle tatsächlich existiert – prüfe Autor*in, Titel, Jahr, Zugänglichkeit und ob die Quelle die genannte Information wirklich enthält.

Bestätige die Information durch mindestens zwei unabhängige Quellen.

Öffne die Originalquelle, nicht nur eine Zusammenfassung oder einen Repost.

Prüfe Veröffentlichungsdatum und Aktualität der Information.

Schritt 3: Qualität, Sprache und Kontext bewerten

Nach der Faktenprüfung bewertest du, ob der Text logisch, klar, ausgewogen und für seinen Zweck geeignet ist. KI kann Inhalte erzeugen, die formal korrekt, aber schlecht kontextualisiert, übermäßig selbstsicher, verzerrt oder stilistisch ungeeignet sind.

Der Text vermischt Fakten mit Meinung.

Er enthält Verzerrungen, Vereinfachungen oder übermäßige Verallgemeinerungen.

Ton und Stil passen zur Zielgruppe und zum Veröffentlichungsziel.

Bearbeiten, entscheiden und dokumentieren

Korrigiere abschließend Fehler, entferne unsichere Stellen und entscheide, ob der Inhalt verwendbar ist. Gute Praxis ist es, festzuhalten, was geändert, was verworfen wurde und welche Quellen die finale Version bestätigt haben. Endentscheidung:

Veröffentlichen

Wenn der Inhalt geprüft und korrigiert wurde.

Zurückhalten

Wenn einige Informationen unsicher bleiben.

Ablehnen

Wenn der Inhalt zu viele Fehler enthält oder Quellenbestätigung fehlt.

07

Datenanalyse mit KI

(10 Min.)

Datenanalyse mit KI: Validierung ist notwendige Arbeit

Wenn du KI bittest, Daten zu analysieren, erzeugt das System eine flüssige Darstellung. Je nach Prompt und Tool kann die Ausgabe auch:

- Methodik auslassen (Stichprobe, Methode, Annahmen)
- Ausreißer mitteln, ohne sie zu kennzeichnen
- kleine Stichproben verallgemeinern (z. B. n=10)
- selbstsichere Schlüsse aus unsicheren Daten ziehen
- Bestätigungsfehler: Prompt-Rahmung wiederholen

Datenanalyse mit KI:

Validierung ist notwendige Arbeit

Warum Validierung von Analysen wichtig ist:

- Du erkennst Fehler vor Bericht oder Veröffentlichung.
- Du trennst belastbare Befunde von Rauschen.
- Du vermeidest Entscheidungen auf Basis nicht repräsentativer Stichproben.
- Du behältst die Autorenschaft der Analyse – nicht das Modell.

SAFE-Analyseleiter: Wähle die prüfbarste Stufe

Beginne nicht mit vollständigen Daten. Wähle die niedrigste Stufe, die deine Frage beantwortet.

Stopp-Regel: Wenn du sensible/vertrauliche Elemente nicht entfernen kannst → nicht hochladen.

STUFE 1

(Am zuverlässigsten): Allgemeine Frage, keine personenbezogenen Daten.

Beispiel: „Wann t-Test statt Mann-Whitney-U?“

STUFE 2

Nutze eine anonymisierte Zusammenfassung (Rollen, keine IDs).

Beispiel: „MW = 4,2; SD = 1,1; n = 120 – Trend?“

STUFE 3

Nutze begrenzte, de-identifizierte Daten nur wenn nötig.

Beispiel: „Berechne X; ich prüfe nach.“

Schlussfolgerung aus wenigen Daten

Unzuverlässige KI-Ausgabe (realistisch):

„KI hat eine Sentimentanalyse von 200 Tweets über unsere Marke durchgeführt. 67 % positiv. Daher ist die Markenstimmung stark positiv.“

Was ist unzuverlässig?

- „67 %“ ohne Methodik, ohne Modell, ohne Validierung
- 200 Tweets sind ein sehr kleiner, möglicherweise verzerrter Ausschnitt
- „stark positiv“ übertreibt, was die Zahl zeigt

AI Skills,
Powered
by Values

values

Folge unserer Reise

www.valuesai.eu

Co-funded by
the European Union

Mitfinanziert von der Europäischen Union. Die geäußerten Ansichten und Meinungen sind jedoch ausschließlich die der Autor*innen und spiegeln nicht unbedingt die Ansichten der Europäischen Union oder der Stiftung für die Entwicklung des Bildungssystems wider. Weder die Europäische Union noch die fördernde Einrichtung können dafür verantwortlich gemacht werden.