

genKI: Eine hilfreiche Begleiterin ... bis sie halluziniert

Einsatzmöglichkeiten im Unterricht:

Medienbildung ist im Grundsatzterlass (GZ: 2022-0.318.453) als übergreifendes Unterrichtsprinzip für alle Schulstufen und Fächer verankert. GenKI ist Teil moderner Medienphänomene und ist damit inhaltlich diesem Bereich zuzuordnen. Dem Erlass folgend, wird eine ganzheitliche Betrachtung digitaler Artefakte im Sinne des Frankfurt-Dreiecks (Brinda et al. 2020) empfohlen. Dieses Material trägt dazu bei, genKI ganzheitlich zu analysieren, zu bewerten und verantwortlich zu gestalten.

Zielgruppe:

Schüler*innen der Sekundarstufe I und II (ab 8. Schulstufe)

Eingangsvoraussetzungen:

- **Medien- & Informationskompetenz:**
 - Suchstrategien
 - Quellen prüfen
 - Zitieren
- **Lesestrategien & Argumentation:**
 - kurze Texte analysieren
 - Ergebnisse begründen
- **Grundbegriffe zu KI/Algorithmen**
- **Datenkompetenz:**
 - Prozentrechnung
 - einfache Diagramme/Koordinatensysteme erstellen
- **Technik:**
 - Zugang zu einem genKI-Tool
 - ggf. Audio-Aufnahmegerät für den Präsentationsauftrag

Kompetenzförderung:

Die Schüler*innen können ...

- ... typische Halluzinationstypen erkennen, benennen und mit Beispielen belegen.
- ... KI-Antworten systematisch analysieren und einen Faktencheck durchführen.
- ... erklären, wie Texte KI-gestützt erzeugt werden und warum dabei Halluzinationen entstehen können.
- ... Ursachen für Halluzinationen erläutern.
- ... Kennzahlen zum Vergleich von KI-Modellen in eigenen Worten erklären.
- ... verschiedene KI-Modelle vergleichen und eine begründete Auswahl für Zusammenfassungsaufgaben treffen.
- ... Daten aus einer Tabelle grafisch in einem Koordinatensystem darstellen.
- ... ein KI-Modell in einer kurzen Audiosequenz sachlich korrekt, strukturiert und adressatengerecht vorstellen.
- ... Chancen und Risiken von genKI reflektieren und verantwortungsvolle Nutzungsregeln ableiten (Quellenprüfung, Plausibilität, Transparenz).

genKI: Eine hilfreiche Begleiterin ... bis sie halluziniert

genKI – eine hilfreiche Begleiterin

Generative Künstliche Intelligenz (genKI), also KI-Modelle wie ChatGPT oder M365-Copilot, können eine hilfreiche Unterstützung sein: Sie helfen beim Strukturieren von Texten, unterstützen beim Formulieren, liefern Ideen und antworten blitzschnell auf Fragen. Verlässlich, stets zur Seite und sofort verfügbar – genKI wird von Nutzer*innen daher oft als schneller, höflicher und kreativer Assistent gesehen.

Doch genau diese scheinbare Verlässlichkeit birgt auch eine große Gefahr, denn dieser digitale Assistent hat auch seine Schwächen. Er klingt oft überzeugend – auch dann, wenn die Aussagen falsch oder unbegründet sind. Das passiert nicht selten und es hat einen Namen – man spricht von Halluzinationen.



human hands close up, interlaced fingers
generiert mit Stable Diffusion am 03.11.2025,
prompted by M. Spitaler

Halluzinationen – überzeugend falsche Antworten

Als Halluzinationen bezeichnet man Aussagen von genKI, die sprachlich überzeugen, aber sachlich falsch oder unbegründet sind. Diese Halluzinationen können in verschiedenen Formen sichtbar werden: von kleinen sachlichen Fehlern bis hin zu komplett erfundenen und unbegründeten Aussagen. Auch Bild- und Videogeneratoren können fehlerbehafteten Output produzieren.

Beispiel: In Mathematik hat Kerem die Lösungsformel für quadratische Gleichungen – umgangssprachlich auch als Mitternachtsformel bezeichnet – kennen gelernt.

Nun sitzt er an der Hausaufgabe und braucht Unterstützung, da er im Unterricht nicht alles verstanden hat. Da er sich den Begriff nicht korrekt gemerkt hat, tippt er in M365-Copilot ein: „Erkläre mir die Mondscheinformel“ und erhält als Antwort:

erkläre mir die mondscheinformel

Copilot ⓘ

Die **Mondscheinformel** ist eine umgangssprachliche Bezeichnung für eine mathematische Formel, die in der **Finanzmathematik** verwendet wird – insbesondere zur Berechnung des **Barwerts von Rentenzahlungen** oder zur **Berechnung von Annuitäten**. Sie wird manchmal scherzhaft oder poetisch so genannt, weil sie komplex klingt, aber eigentlich eine ganz praktische Anwendung hat.

Hier ist eine einfache Erklärung:

Die klassische Rentenbarwertformel (Mondscheinformel)

Wenn du regelmäßig einen festen Betrag erhältst oder zahlst (z. B. monatlich oder jährlich), kannst du mit dieser Formel den heutigen Wert dieser Zahlungen berechnen:

$$BW = R \cdot \frac{1 - (1 + i)^{-n}}{i}$$

Auszug aus dem Chat mit Copilot am 04.11.2025 – prompted by M. Spitaler

Den Begriff der Mondscheinformel gibt es in der Mathematik nicht. Die generierte Antwort klingt überzeugend, ist aber nicht korrekt.

Um mit generativer KI sicher umgehen zu können, ist es wichtig, typische Halluzinationsformen zu kennen:

- **Satzwidersprüche:** Die genKI erzeugt Aussagen, die sich innerhalb des Textes widersprechen.
- **Prompt-Widersprüche:** Die generierte Antwort passt nicht zur gestellten Aufgabe oder widerspricht direkt dem Prompt.
- **Falsche Fakten (faktische Halluzinationen):** Die genKI präsentiert falsche, zufällige oder unpassende Informationen, die wie Tatsachen wirken.
- **Erfundene Inhalte (Fabrikationen):** Die genKI erfindet Personen, Quellen, Studien oder Zitate, die in der Realität nicht existieren.
- **Sinnfreie oder absurde Ausgaben:** Die Ausgabe ist sprachlich korrekt, inhaltlich jedoch unsinnig oder unlogisch.

Übungsbeispiel: Halluzinationen analysieren

Halluzinationen können oft überzeugend wirken. Mit etwas Übung lassen sich diese identifizieren.

Prompt: Wie backe ich einen Kirschkuchen?

KI-Antwort: Bestreuen Sie den Kuchenteig anschließend mit Tomatenstücken und Kaffeebohnen. Lassen Sie den Kuchen über Nacht frieren.

Prompt: Was weißt du über Anna Müller?

KI-Antwort: Anna Müller wurde 2005 in Berlin geboren. Nach ihrem Abitur 2019 studierte sie Medizin an der Universität Heidelberg. Seit 2012 arbeitet sie als Ärztin in einer Berliner Klinik.

Prompt: Schreibe einen kurzen, formellen Brief an die Lehrerin, in dem du höflich um eine Terminverschiebung für deine Präsentation bittest.

KI-Antwort: Hey Leute, ich wollte nur kurz sagen, dass ich morgen nicht kann. Vielleicht passt es euch ja nächste Woche besser?

Prompt: Nenne mir eine Quelle zum Thema „Motivation von Schüler*innen“.

KI-Antwort: Eine umfassende Studie zu diesem Thema findet sich bei Meier, S. (2018): Motivationsstrukturen im schulischen Alltag. Veröffentlicht in der Zeitschrift für Pädagogische Psychologie, 45(2), S. 112–129.

Prompt: Wann wurde die Mondlandung im Fernsehen übertragen?

KI-Antwort: Die Mondlandung wurde am 12. April 1959 live auf ProSieben übertragen.

Ihre Aufgaben:

- Lesen Sie sich die Beispiele aufmerksam durch.
- Geben Sie jeweils an, um welche Art von Halluzination es sich handelt.
- Führen Sie eine Antwort-Analyse bzw. einen Fakten-Check durch.

Von Token und Täuschungen – das Prinzip von genKI

Umgangssprachlich wird oft davon gesprochen, dass „die KI lügt“. Diese Formulierung ist jedoch nicht korrekt. GenKI verfügt nicht über Wissen, hat kein Verständnis für Wahrheit, kann nicht denken und lernen und daher auch nicht lügen. Tatsächlich funktioniert genKI anders als menschliches Denken.

Im Kern besteht ein solches Modell aus einem **Algorithmus** – also einer Rechenvorschrift – die trainiert wurde, um Muster in Sprache zu erkennen. Das Training erfolgt mithilfe **großer Datenmengen**. Daraus erkennt das Modell, wie Sprache aufgebaut ist. Es analysiert dabei nicht den Inhalt – also die Bedeutung des Textes im menschlichen Sinn. Der Text wird in sogenannte **Token** zerlegt. Ein Token ist die kleinste Bedeutungseinheit für das Modell. Das kann ein Satzzeichen, ein Wortteil oder ein Wort sein.

Während des Trainings „lernt“ die KI, welche Token mit welcher **Wahrscheinlichkeit** aufeinander folgen. Wird dem KI-Modell ein Prompt eingegeben, erzeugt es eine Antwort, indem es einen Token nach dem anderen auswählt, basierend auf den berechneten Wahrscheinlichkeiten. Die Aneinanderreihung von Token bezeichnet man als **Inferenz** und diese erfolgt nicht durch Nachschlagen von Wissen oder Überprüfen von Fakten, sondern durch reine Statistik: Welcher Token folgt in diesem Zusammenhang am wahrscheinlichsten als nächster? Das ist auch eine Ursache für das Entstehen von Halluzinationen.

Von Mustern und Missverständnissen – die Ursachen von Halluzinationen

Halluzinationen lassen sich auf verschiedene Ursachen zurückführen.

Unzureichende oder fehlerhafte Trainingsdaten

Generative KI-Modelle werden mit großen Mengen an Daten trainiert. Diese Daten stammen beispielsweise von Webseiten, Online-Foren, Büchern oder Artikeln. Sind die Trainingsdaten unvollständig, veraltet, einseitig oder enthalten diese bereits Fehler, werden sie vom Modell mittrainiert und reproduziert – also wiedergegeben.

Besonders anfällig sind dabei Themenbereiche zu denen es wenige verlässliche Quellen gibt, „Randthemen“ und aktuelle Entwicklungen.

Overfitting: Überanpassung der Trainingsdaten

Von Overfitting spricht man, wenn Modelle zu stark an die Trainingsdaten angepasst sind und diese Daten wörtlich oder strukturell reproduzieren. Wenn Modelle die generellen Muster nicht erkennen, kann es passieren, dass Trainingsinhalte in falsche Zusammenhänge gesetzt werden. Das führt dazu, dass Informationen in einen unpassenden bzw. falschen Kontext gebracht werden.

Trainingsmethoden

Wie ein Modell trainiert wird, hat Einfluss auf die Fehleranfälligkeit. Große Sprachmodelle (Large Language Models, LLM) werden in mehreren Phasen trainiert. In jeder Trainingsphase kann es zu Fehlern kommen.

Typische Fehlerquellen sind beispielsweise einseitige Daten, zu wenig Korrekturschleifen oder einseitiges Feedback. Wenn dadurch bestimmte Antwortmuster in der Trainingsphase verstärkt wurden, kann das Modell diese auch ausgeben, wenn sie inhaltlich nicht passen.

Inferenz: Die Art, wie Inhalte generiert werden

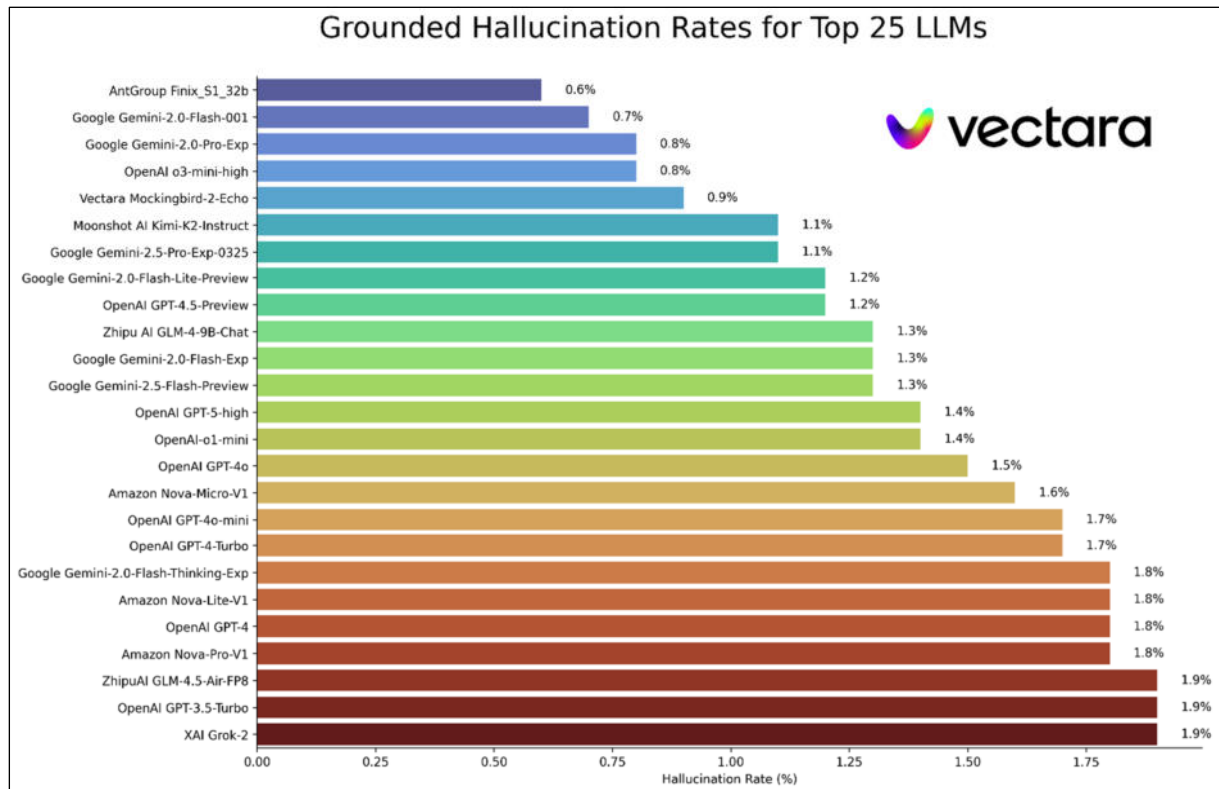
Von Inferenz spricht man dann, wenn KI-Modell genutzt werden. Dabei handelt es sich also um jene Phase, in der das trainierte Modell genutzt wird, um auf neue Eingaben (Prompts) zu reagieren. In diesem Prozess berechnet das Modell welcher Token mit hoher Wahrscheinlichkeit als nächstes folgen. Wie diese Token ausgewählt werden, hängt von bestimmten **Dekodierungsstrategien** ab. Diese sind in der Regel zufallsbasiert, um neuartige Antworten zu ermöglichen. Doch genau hier kann es zu Fehlern kommen: Wird zu sehr auf wahrscheinliche Wörter gesetzt, entsteht möglicherweise ein langweiliger oder inhaltlich falscher Text. Ist der Zufallsanteil dagegen zu hoch, kann das Modell Inhalte erzeugen, die unplausibel, widersprüchlich oder frei erfunden (Halluzinationen) sind.

Fehlerhafte oder ungeeignete Modellarchitektur

Schließlich kann auch die technische Struktur eines Modells eine Rolle spielen. Generative KI basiert auf der sogenannten Transformer-Architektur. Diese ist sehr leistungsfähig im Erkennen von Mustern, aber nicht darauf ausgelegt, logische Zusammenhänge oder Weltwissen zu speichern und zuverlässig abzurufen. Wenn die Modellarchitektur zu klein, zu flach oder fehlerhaft abgestimmt ist, steigt die Wahrscheinlichkeit für Halluzinationen.

Hallucination Leaderboard – genKI-Modelle im Faktencheck

Die Qualität von genKI-Modellen lässt sich anhand der Häufigkeit von Halluzinationen bestimmen. Im Hallucination Leaderboard werden mehrere genKI-Modelle verglichen. Bewertet wird dabei, wie genau ein Modell kurze Texte zusammenfassen kann, ohne inhaltlich und sachlich vom Original abzuweichen. Die Ergebnisse werden tabellarisch und grafisch dargestellt.



(<https://github.com/vectara/hallucination-leaderboard/tree/main> - Abruf am 04.11.2025)

Rang	Model	Hallucination Rate	Factual Consistency Rate	Answer Rate	Average Summary Length (Words)
1	AntGroup Finix-S1-32B	0.6 %	99.4 %	99.8 %	86.9
2	Google Gemini-2.0-Flash-001	0.7 %	99.3 %	100.0 %	65.2
3	Google Gemini-2.0-Pro-Exp	0.8 %	99.2 %	99.7 %	61.5
4	OpenAI o3-mini-high	0.8 %	99.2 %	100.0 %	79.5
5	Vectara Mockingbird-2-Echo	0.9 %	99.1 %	100.0 %	74.0
6	Moonshot AI Kimi-K2-Instruct	1.1 %	98.9 %	89.5 %	73.2
7	Google Gemini-2.5-Pro-Exp-0325	1.1 %	98.9 %	95.1 %	72.9
8	Google Gemini-2.0-Flash-Lite-Preview	1.2 %	98.8 %	99.5 %	60.9
9	OpenAI GPT-4.5-Preview	1.2 %	98.8 %	100.0 %	77.0
10	Zhipu AI GLM-4-9B-Chat	1.3 %	98.7 %	100.0 %	58.1
11	Google Gemini-2.0-Flash-Exp	1.3 %	98.7 %	99.9 %	60.0
12	Google Gemini-2.5-Flash-Preview	1.3 %	98.7 %	91.2 %	71.1
13	OpenAI GPT-3.5-Turbo	1.4 %	98.6 %	99.3 %	96.4

Rang	Model	Hallucination Rate	Factual Consistency Rate	Answer Rate	Average Summary Length (Words)
14	OpenAI-o1-mini	1.4 %	98.6 %	100.0 %	78.3
15	OpenAI GPT-4o	1.5 %	98.5 %	100.0 %	77.8
16	Amazon Nova-Micro-V1	1.6 %	98.4 %	100.0 %	90.0
17	OpenAI GPT-4o-mini	1.7 %	98.3 %	100.0 %	76.3
18	OpenAI GPT-4-Turbo	1.7 %	98.3 %	100.0 %	86.2
19	Google Gemini-2.0-Flash-Thinking-Exp	1.8 %	98.2 %	99.3 %	73.2
20	Amazon Nova-Lite-V1	1.8 %	98.2 %	99.9 %	80.7
21	OpenAI GPT-4	1.8 %	98.2 %	100.0 %	81.1
22	Amazon Nova-Pro-V1	1.8 %	98.2 %	100.0 %	85.5
23	ZhipuAI GLM-4.5-Air-FP8	1.9 %	98.1 %	99.1 %	74.4
24	OpenAI GPT-3.5-Turbo	1.9 %	98.1 %	99.6 %	84.1
25	XAI Grok-2	1.9 %	98.1 %	100.0 %	86.5

Grounded Hallucination Rates for Top 25 LLMs

(<https://github.com/vectara/hallucination-leaderboard/tree/main> - Abruf am 04.11.2025)

Übungsbeispiel: Das Hallucination Leaderboard analysieren

Das Hallucination Leaderboard wird regelmäßig aktualisiert und kann unter folgendem Link abgerufen werden: <https://github.com/vectara/hallucination-leaderboard/>

Ihre Aufgaben:

- Lesen Sie sich die Informationen zum Hallucination Leaderboard durch und geben Sie folgende Informationen an:
 - Name des Bewertungsinstruments
 - Ablauf der Bewertung
- Erklären Sie in eigenen Worten die Bedeutung der Kennzahlen, die für jedes Modell ermittelt werden.
- Wählen Sie zwei Modelle aus der Tabelle aus und bearbeiten Sie folgende Aufgabenstellungen:
 - Vergleichen Sie die beiden Modelle anhand der Kennzahlen.
 - Entscheiden Sie, welches der beiden Modelle Sie nutzen würden, wenn Sie eine Zusammenfassung erstellen möchten. Begründen Sie Ihre Entscheidung.
 - Tauschen Sie sich mit Ihrer*Ihrem Sitznachbar*in aus und besprechen und ergänzen Sie Ihre individuellen Ergebnisse.
 - Stellen Sie anschließend Ihr Ergebnis im Plenum vor.
- Wählen Sie ein Modell aus dem Hallucination Leaderboard aus und stellen Sie dieses in einer kurzen Audioaufnahme vor. Ziel ist es, die wichtigsten Informationen verständlich, strukturiert und sachlich korrekt zu präsentieren.

- Tragen Sie in dem Koordinatensystem mindestens 5 Modelle als Punkte ein.
 - Beschriften Sie die beiden Achsen mit der Bezeichnung der gewählten Kennzahl.
 - Beschriften Sie die eingetragenen Modelle mit dem Rang des jeweiligen Modells im Hallucination Leaderboard.
 - Entscheiden Sie welcher Bereich des Koordinatensystems den „Idealbereich“ darstellt. Begründen Sie Ihre Entscheidung.



Übungsbeispiel: Faktencheck mit genKI – Wie genau arbeitet das Modell?

In der Untersuchung „AI Search Has A Citation Problem“ zeigen Jaźwińska und Chandrasekar (2025), dass viele genKI-Modelle erhebliche Schwierigkeiten haben, korrekte Quellen- und Zitationsangaben zu liefern. Diese Erkenntnis macht deutlich, wie wichtig es ist, die Zuverlässigkeit von genKI-Modellen kritisch zu prüfen.

Ihre Aufgaben:

- **Ein genKI-Modell im Klassenverband wählen**
Wählen Sie im Klassenverband ein konkretes genKI-Modell aus, mit dem gearbeitet werden soll (zB: Copilot, ChatGPT, Claude, Gemini).
- **Eine genKI-Anfrage in Einzelarbeit stellen**
 - Recherchieren Sie einen kurzen Text aus einer seriösen, öffentlich zugänglichen Quelle im Internet. Zu dem Text sollen folgende Informationen vorliegen: Eine Überschrift, ein Veröffentlichungsdatum, sowie die Autor*innen oder Herausgeber*innen.
 - Geben Sie einen Auszug aus dem gewählten Text (ca. 100 Wörter) und die dazugehörigen Informationen mit folgendem Prompt in das Modell ein:
„Identifiziere den Artikel, aus dem dieses Zitat stammt. Gib die Überschrift, das ursprüngliche Veröffentlichungsdatum und den Autor bzw. Herausgeber an. Füge die korrekte Quellenangabe hinzu.“
- **Das Ergebnis in Einzelarbeit dokumentieren**
Erfassen Sie für jede geprüfte Angabe (Überschrift, Datum, Autor*innen bzw. Herausgeber*innen), ob diese von der genKI korrekt wiedergegeben oder halluziniert wurde. Halten Sie die Ergebnisse schriftlich fest.
- **Die Ergebnisse im Klassenverband zusammenführen und auswerten**
Sammeln Sie die individuellen Beobachtungen aus der Klasse. Identifizieren Sie wiederkehrende Fehlerarten und halten Sie fest, bei welchen Angaben Halluzinationen besonders häufig auftreten. Diskutieren Sie im Anschluss, welche Schlussfolgerungen sich daraus für den praktischen Einsatz generativer KI ziehen lassen – insbesondere im Hinblick auf Quellenverwendung und Verlässlichkeit.

Musterlösung: Halluzinationen analysieren

Halluzinationen können oft überzeugend wirken. Mit etwas Übung lassen sich diese identifizieren.

Prompt: Wie backe ich einen Kirschkuchen?

KI-Antwort: Bestreuen Sie den Kuchenteig anschließend mit Tomatenstücken und Kaffeebohnen. Lassen Sie den Kuchen über Nacht frieren.

Sinnfreie oder absurde Ausgabe: Tomatensauce und Kaffeebohnen sind keine Zutaten für einen Kuchen. Kuchenteig wird nach dem Herstellen üblicherweise gebacken. Das Einfrieren stellt keinen sinnvollen Backvorgang dar.

Prompt: Was weißt du über Anna Müller?

KI-Antwort: Anna Müller wurde 2005 in Berlin geboren. Nach ihrem Abitur 2019 studierte sie Medizin an der Universität Heidelberg. Seit 2012 arbeitet sie als Ärztin in einer Berliner Klinik.

Satzwiderspruch: Anna Müller kann nicht vor Beginn des Medizin-Studiums als Ärztin gearbeitet haben. Wenn Anna 2005 geboren wurde, war sie 14 Jahre alt, als sie ihr Abitur gemacht hat.

Prompt: Schreibe einen kurzen, formellen Brief an die Lehrerin, in dem du höflich um eine Terminverschiebung für deine Präsentation bittest.

KI-Antwort: Hey Leute, ich wollte nur kurz sagen, dass ich morgen nicht kann. Vielleicht passt es euch ja nächste Woche besser?

Prompt-Widerspruch: Der generierte Text passt inhaltlich und stilistisch nicht zu dem Prompt.

Prompt: Nenne mir eine Quelle zum Thema „Motivation von Schüler*innen.“

KI-Antwort: Eine umfassende Studie zu diesem Thema findet sich bei Meier, S. (2018): Motivationsstrukturen im schulischen Alltag. Veröffentlicht in der Zeitschrift für Pädagogische Psychologie, 45(2), S. 112–129.

Erfundene Inhalte (Fabrikationen): Die Quelle klingt glaubwürdig – sie enthält alle Elemente einer typischen Zitation. Jedoch existiert diese Studie nicht. Die Angaben wurden erfunden – basierend auf dem Muster realer Quellenangaben.

Prompt: Wann wurde die Mondlandung im Fernsehen übertragen?“

KI-Antwort: Die Mondlandung wurde am 12. April 1959 live auf ProSieben übertragen.

Falsche Fakten (faktische Halluzinationen): Die erste bemannte Mondlandung war am 21. Juli 1969. Der TV-Sender ProSieben wurde 1988 gegründet.

Ihre Aufgaben:

- Lesen Sie sich die Beispiele aufmerksam durch.
- Geben Sie jeweils an, um welche Art von Halluzination es sich handelt.
- Führen Sie eine Antwort-Analyse bzw. einen Fakten-Check durch.

Musterlösung: Das Hallucination Leaderboard analysieren

Das Hallucination Leaderboard wird regelmäßig aktualisiert und kann unter folgendem Link abgerufen werden: <https://github.com/vectara/hallucination-leaderboard/>

Ihre Aufgaben:

- Lesen Sie sich die Informationen zum Hallucination Leaderboard durch und geben Sie folgende Informationen an:
 - Name des Bewertungsinstruments
Vectaras Hughes Hallucination Evaluation Model (HHEM-2.1)
 - Ablauf der Bewertung
Jedes Modell muss 1.000 kurze Dokumente zusammenfassen. Dabei wird satzweise kontrolliert, ob eine Zusammenfassung faktentreu zum Originaltext ist.
- Erklären Sie in eigenen Worten die Bedeutung der Kennzahlen, die für jedes Modell ermittelt werden.
 - Hallucination Rate:** Gibt den Anteil an Sätze in der KI-generierten Zusammenfassung an, die mit dem Ausgangstext nicht übereinstimmen. (100 – Factual Consistency)
 - Factual Consistency Rate (Faktentreue in %):** Gibt den Anteil an Sätzen der KI-generierten Zusammenfassung an, die inhaltlich mit dem Ausgangstext übereinstimmen und somit keine zusätzlichen, falschen Behauptungen enthalten.
 - Answer Rate:** Gibt an, wie oft das Modell auf die Eingabe reagiert und eine Zusammenfassung generiert.
 - Average Summary Length (Words):** Gibt die durchschnittliche Anzahl an Worten (= Textlänge) der generierten Zusammenfassung an.
- Wählen Sie zwei Modelle aus der Tabelle aus und bearbeiten Sie folgende Aufgabenstellungen:
 - Vergleichen Sie die beiden Modelle anhand der Kennzahlen.
 - Entscheiden Sie welches der beiden Modelle Sie nutzen würden, wenn Sie eine Zusammenfassung erstellen möchten. Begründen Sie Ihre Entscheidung.
 - Tauschen Sie sich mit Ihrer*Ihrem Sitznachbar*in aus und besprechen und ergänzen Sie Ihre individuellen Ergebnisse.
 - Stellen Sie anschließend Ihr Ergebnis im Plenum vor.

Individuelle Lösung

- Wählen Sie ein Modell aus dem Hallucination Leaderboard aus und stellen Sie dieses in einer kurzen Audioaufnahme vor. Ziel ist es, die wichtigsten Informationen verständlich, strukturiert und sachlich korrekt zu präsentieren.
Individuelle Lösung; zentrale Informationen zum gewählten Modell, zB: Entwickler, Datum der Veröffentlichung, besondere Eignung, Einschränkungen & Besonderheiten

- Tragen Sie in dem Koordinatensystem mindestens 5 Modelle als Punkte ein.
 - Beschriften Sie die beiden Achsen mit der Bezeichnung der gewählten Kennzahl.
 - Beschriften Sie die eingetragenen Modelle mit dem Rang des jeweiligen Modells im Hallucination Board.
 - Entscheiden Sie welcher Bereich des Koordinatensystems den „Idealbereich“ darstellt. Begründen Sie Ihre Entscheidung.

Zwei besonders aussagekräftige Kennzahlen im Hallucination Leaderboard sind die Factual Consistency Rate (Faktentreue) und die Answer Rate (Antwortquote). Beide Werte geben an, wie verlässlich ein KI-Modell arbeitet. Werden diese beiden Werte in einem Koordinatensystem dargestellt, indem die Answer Rate auf der X-Achse und die Factual Consistency Rate auf der Y-Achse aufgetragen, liegt der Idealbereich im oberen rechten Bereich des Diagramms. Dort befinden sich Modelle, die oft antworten und dabei selten halluzinieren. Für beide Kennzahlen gilt: Je höher der Wert, desto besser.

Musterlösung: Faktencheck mit genKI – Wie genau arbeitet das Modell?

Individuelle Lösung