

EINFÜHRUNG IN GENERATIVER TEXT KI



Marcel Beyeler

Berater Digitale Transformation

Digitalcoach

Workshop-Leiter

EINSTIEG

Kurzvorstellung

- Was machst Du beruflich?
- Was sind deine bisherigen KI-Erfahrungen?

2. Schreibe auf einen Zettel Deine grösste Sorge oder Deine kritischste Frage.

Regeln:

- Das ist ein Workshop: Austauschen, Ausprobieren, Testen, Diskutieren.
- Es gibt keine dummen Fragen / Dialog wird begrüsst.
- Der Workshop-Leiter versteht sich als Coach. Wenn jemand eigene Erfahrungen einbringen kann, ist das willkommen und wertvoll.

ABLAUF

- Einführung, Grundlagen, Funktionsweise
- Prompt Engineering: Zielführendes Fragen
- Übung 1: Tipps ausprobieren
- Diskussion in der Gruppe: Praktische Szenarien
- Existenzielle Bedrohung? – Wirkliche Gefahren.

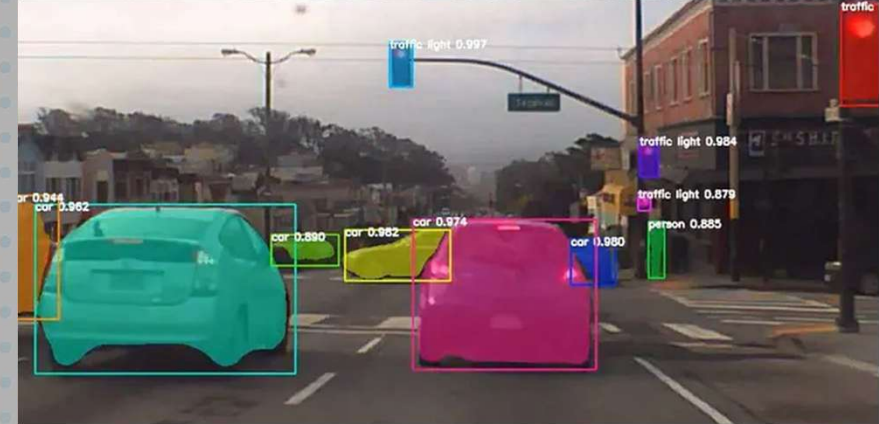
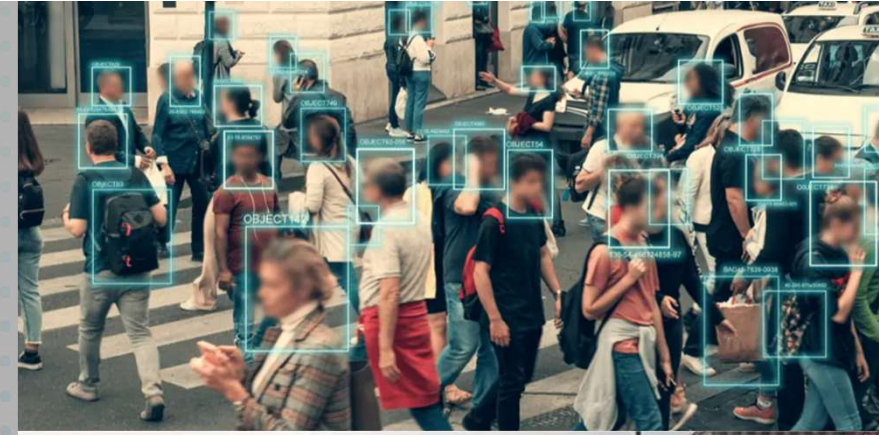


WIR FORMEN UNSERE WERKZEUGE,
UND DANACH FORMEN UNSERE WERKZEUGE UNS.

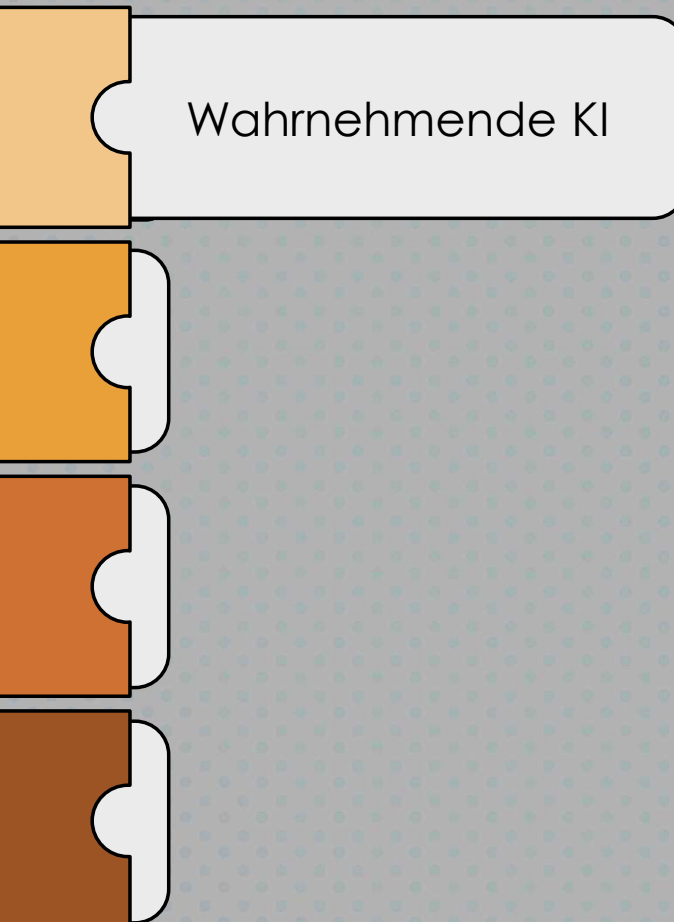
Marshall McLuhan, Philosoph, Kommunikationstheoretiker und
Literaturprofessor

KI IM ALLTAG

- *Computer Vision:*
Gewinnung von Infos aus visuellen Medien
 - Mustererkennung (defektes Maschinenteil, Personen, Nummernschilder, Verkehrssituation)
 - Fahrassistenzsysteme
 - Diagnostik in der Radiologie
- *Sprachmodelle:*
Verarbeitung von Text und Sprache
 - Autocomplete
 - DeepL: kontextbezogene Übersetzungen
 - Vorschläge von Antwortmöglichkeiten in Mails



STUFEN VON KI



Wahrnehmende KI

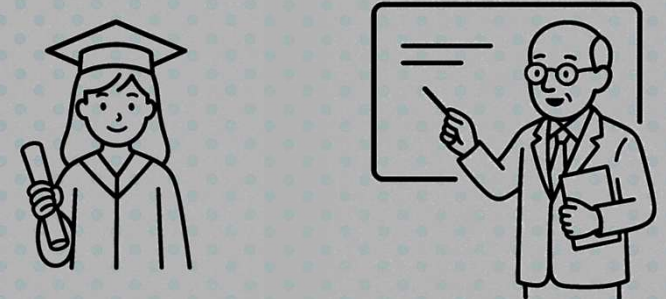
ENTSTEHUNG VON GROSSEN SPRACHMODELLEN

Unüberwachtes Basistraining.



Entscheidender Punkt:
Grundlage / der Boden auf dem
das Training basiert.

Bewertetes Finetuning



Entscheidender Punkt:
Haltung der Bewertenden

ENTSTEHUNG VON LARGE LANGUAGE MODELLEN –

1. DAS BASISTRAINING

Basis = Lernen ohne Anleitung

- Einlesen von riesigen Mengen Text
- Zerlegen in kleine Häppchen
- Zusammenhänge in der Reihenfolge erkennen. Wahrscheinlichkeiten

Bildung eines Verständnismodells – eine statistische Bedeutungslandkarte
«Rom» gehört zu «Stadt» wie
«Apfel» zu «Pflanze»

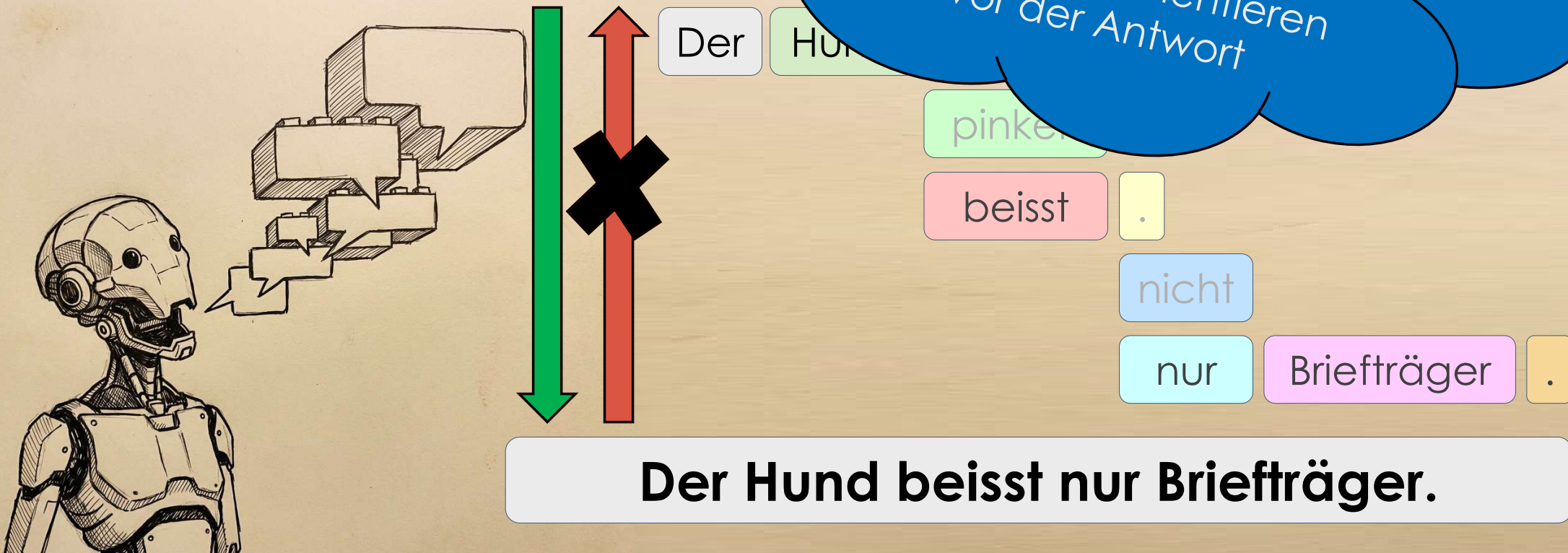


TOKENISIERTE ANTWORTEN — 8

- Wie wäre es, wenn wir immer Satz für Satz ohne darüber nachzudenken?

Grundfunktion:
Wort-für-Wort

«Reasoning» =
internes argumentieren
vor der Antwort



ONCE MEN TURNED THEIR THINKING OVER TO MACHINES
IN THE HOPE THAT THIS WOULD SET THEM FREE.
BUT THAT ONLY PERMITTED OTHER MEN WITH MACHINES
TO ENSLAVE THEM.

Reverend Mother Gaius Helen Mohiam
aus Frank Herberts «Dune»



ENTSTEHUNG VON LARGE LANGUAGE MODELLEN –

2. DAS FINETUNING

Sinnvolle und gute Antworten durch überwachtes Finetuning

- Menschen geben bewertete Antworten: hilfreich, gefährlich, unklar, gut, ...
- Das Modell lehrt Höflichkeit, Sicherheit, Nützlichkeit → Trimmung zu antworten = Alignment Teil 1 (Ausrichtung)
- *reinforcement learning from human feedback* (RLFH)
→ Lob für eine von (bestimmten) Menschen bevorzugte Antwortart

«Training an unbeatable AI in Trackmania» https://www.youtube.com/watch?v=Dw3BZ6O_8LY



WAS LLM'S IN DER PRAXIS LEISTEN

- ➔ Ein sehr, SEHR hoch entwickeltes Auto-Vervollständigen (Basistraining)
...mit einer vorgegebenen Tendenz (Finetuning durch RLHF)

Text Operationen

- Zusammenfassen und Verdichten
- Extrahieren und Strukturieren
- Umformulieren und Übersetzen
- Generieren und Weiterführen
- Klassifizieren

Interaktionsformen

- Information gewinnen
«Erkläre mir ...»
- Etwas generieren lassen
«Schreibe ...»
- Etwas analysieren
«Untersuche x hinsichtlich y»
- Struktur bilden
«Fasse zusammen, kategorisiere»

Oder ganz praktisch:

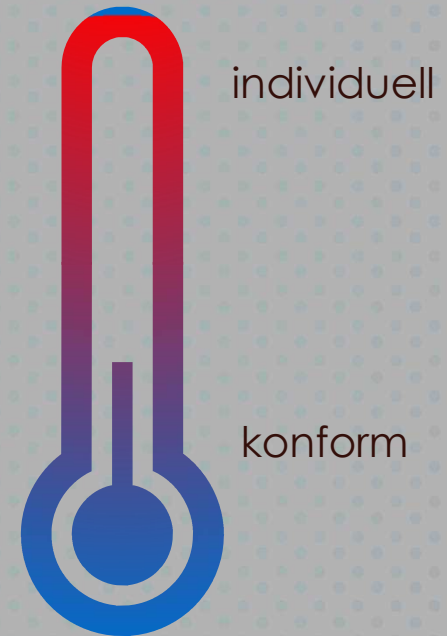
- Recherche, Lernen
- Coach fürs Schreiben und Denken (Entwürfe, Argumentationen, Varianten)
- Prozessunterstützung (E-Mails, Supportanfragen, etc.)
- Transkript durchforsten, Themen herauslesen

DIE «TEMPERATUR» - GRAD DER WAHRSCHEINLICHKEIT

- Eine hohe Temperatur bedeutet, dass das Modell querdenkerisch vorgeht. Die reine Statistik sowie soziale Normen und Denkblockaden in den Anweisungen zählen weniger.
➔ Resultat: Mehr Halluzinationen aber auch spürbar Kreativer
- Eine niedrige Temperatur bedeutet eine statistisch wahrscheinliche Antwort.

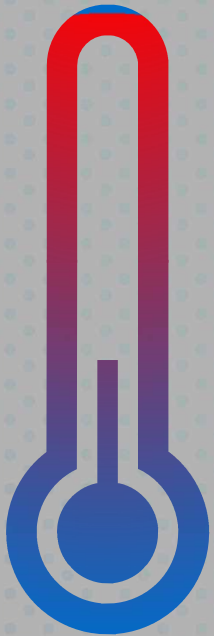
Entsprechend ist das Ergebnis durchschnittlich und fühlt sich an wie eine niedergebügelte Meinung. Sicher, politisch korrekt, ohne Ecken und Kanten.

➔ Resultat: AI-Einheitsbrei.



UMFRAGE – WIE HOCH IST DIE TEMPERATUR?

Was ist die wichtigste Eigenschaft einer Führungskraft im Zeitalter der KI?



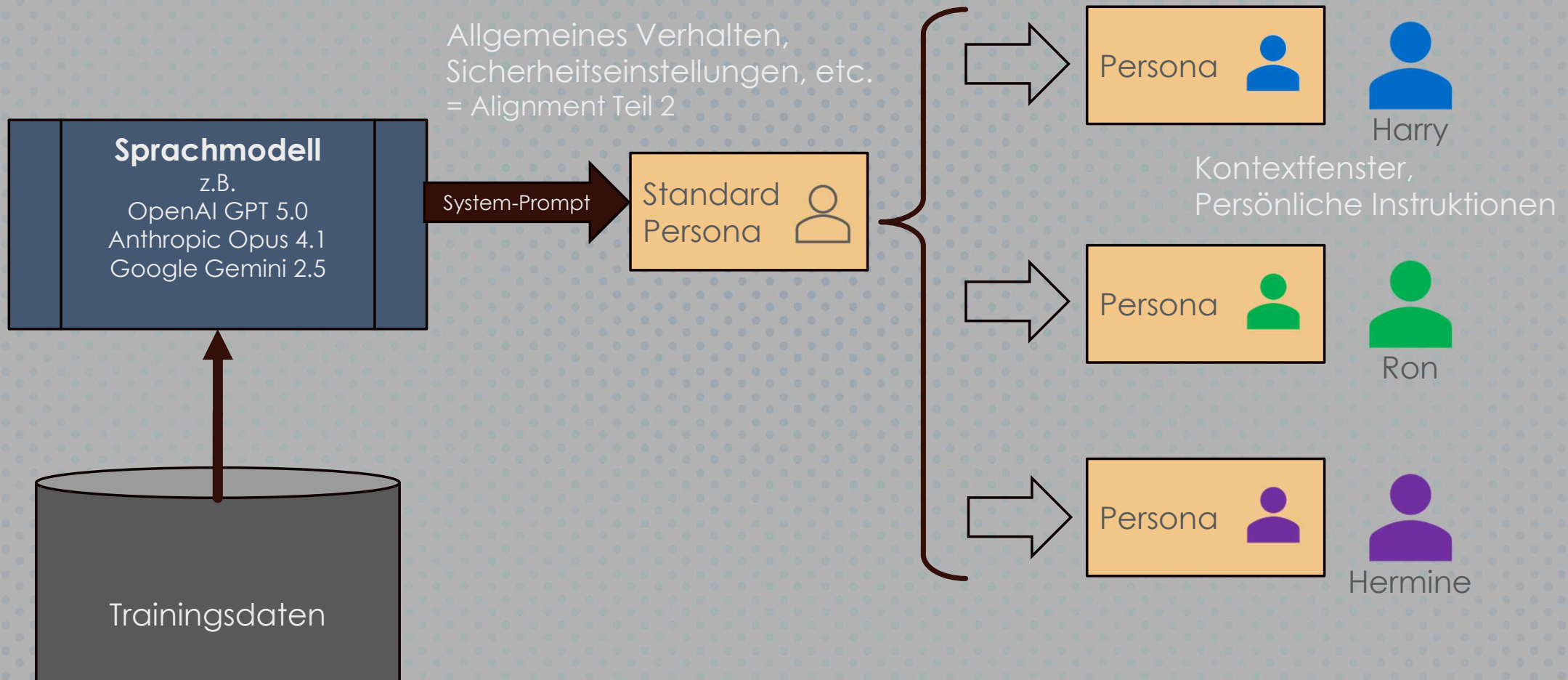
A

Effiziente Implementierung von Datenkompetenz und strategischer Adaption zur Optimierung bestehender digitaler Unternehmensprozesse.

B

Radikale Intuition als menschlicher Reibungspunkt gegen die algorithmische Glättung einer zunehmend perfektionierten Welt.

MODELLE UND CHATBOTS



SCHWÄCHEN UND RISIKEN VON LLM's (1)

Halluzinationen

= erfundene, plausibel wirkende Fehlinformationen

Lösungsweg undurchschaubar,

Entstörung bei Irrweg fast unmöglich

Schmeichelei

gegründet auf Hilfsbereitschaft und Kontextanpassung



- Faktencheck, Quellen anfordern und prüfen, Frage nach reinen Annahmen
- Kritisches Hinterfragen, Prompt neu starten mit leichter Variation.
- Explizit um Gegenargumente bitten, Kritik einfordern, Stärken und Schwächen anfragen

- Kleine Fehler im Training können weitreichende Auswirkungen haben.
→ Emergent Misalignment: Narrow finetuning can produce broadly misaligned LLMs
(24.02.2025 <https://arxiv.org/abs/2502.17424>)

AUFGABE: ZEICHNE MIR EIN BILD EINER FÜHRUNGSKRAFT

Welche Klischees wurden hier bedient?



SCHWÄCHEN UND RISIKEN VON LLM's (2)

Bias

westlich-rational, Gender, Kultur

Prägung durch Tech-Unternehmen,
und deren Interessen, Geldgebern

Erfolgsdruck im AI-Race
mangelhaftes Finetuning

Regierungseinfluss und Kontrolle



- Diversität der Quellen prüfen, alternative Perspektive konsultieren
- Transparenz einfordern, verschiedene Modelle vergleichen
- Skepsis bei neuen Modellen, nicht auf jeden Zug aufspringen
- Herkunft bedenken, lokale OpenSource
- Alternativen erwägen

«Die Systeme erben
unsere Vorurteile,
Unterlassungen und Werte.»

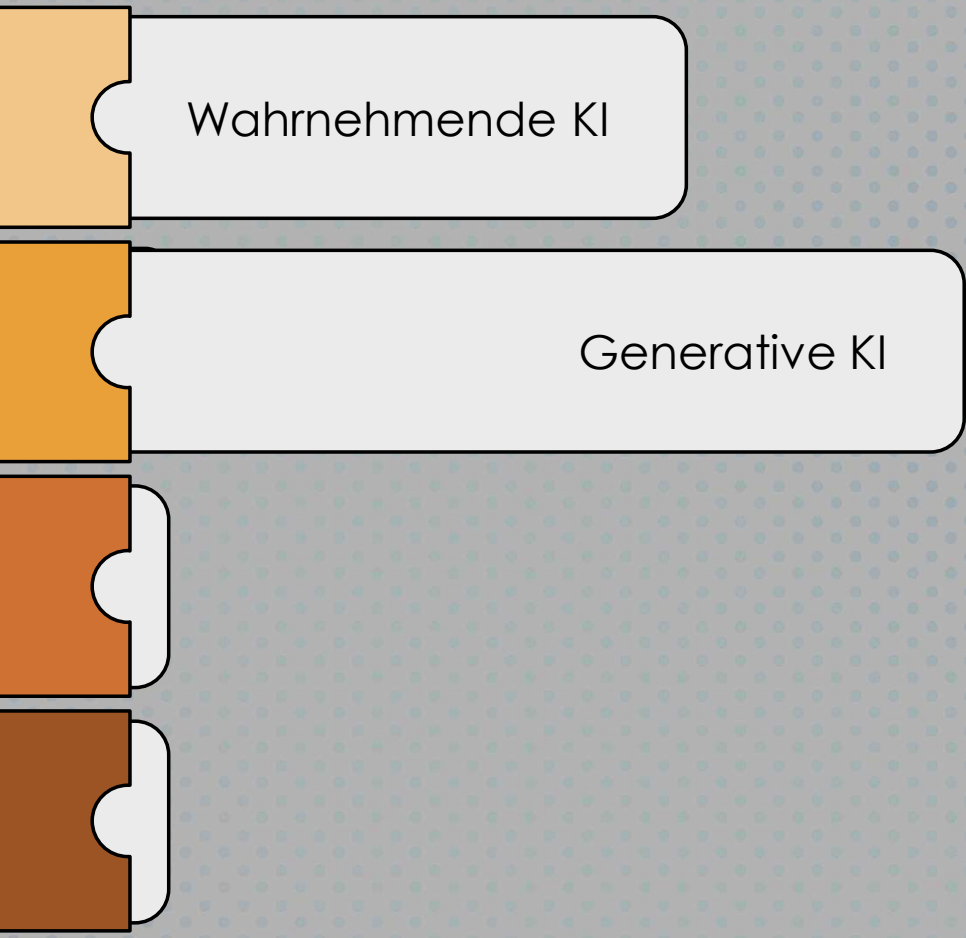
nach Geoffrey Hinton, Physik Nobelpreisträger, Informatiker und Kognitionspsychologe

SICHERHEITSASPEKTE IM BERUFLICHEN KONTEXT



- ChatGPT: Einstellungen > Datenkontrolle > Das Modell für alle verbessern > Nein (fast alle grossen Modelle haben ähnliche Einstellungen)
- Keine Personendaten!
Wenn, dann mit Dummy-Namen austauschen, Anonymisierung.
- Keine Reintext Betriebs- und Geschäftsgeheimnisse.
- Verantwortung liegt IMMER beim Nutzer / bei der Nutzerin.
Das ist wie „ausplaudern“ am Stammtisch.
- Resultate immer überprüfen, auch bei einer Nutzung in einem persönlich bekannten Themengebiet.

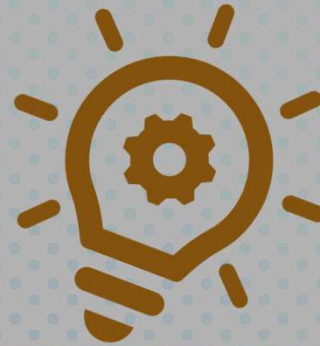
STUFEN VON KI



Wahrnehmende KI

Generative KI

ÜBUNG



- Lass Dir etwas erklären.
«Wieviele Menschen leben permanent in Grönland?»
- Generiere etwas.
«Schreibe mir ein Gedicht über den ausklingenden Winter.»
- Analysiere etwas.
«Untersuche Zeitungsartikel der letzten 2 Monate über den Schweizer Immobilienmarkt.»
- Lasse Dir etwas zusammenfassen.
«Fasse folgenden Artikel [URL einfügen] in 3 Sätzen zusammen. Kategorisiere die Inhalte in Stichworten nach a) Kernaussagen b) Chancen/Nutzen c) Risiken/Probleme»



Interaktionsformen

- Information gewinnen
«Erkläre mir ...»
- Etwas generieren lassen
«Schreibe ...»
- Etwas analysieren
«Untersuche x hinsichtlich y»
- Struktur bilden
«Fasse zusammen, kategorisiere»

KI UND ICH

Stärke Deine Fähigkeit Probleme zu lösen

1. FRAG NICHT – LASS DICH BEFRAGEN

- Schreibe einen Text über XYZ. 
- Ich will einen LinkedIn Post über XYZ schreiben. Zeige mir wie ich das beste Resultat bekomme und führe mich durch den Erstellungsprozess.
- Schreib mir eine E-Mail an eine unzufriedene Besucherin. 
- Ich muss auf eine Beschwerde einer Besucherin reagieren. Die Angehörige eines Bewohners ist verärgert, weil in der Cafétérie XYZ vorgefallen war. Die Kritik ist wahrscheinlich berechtigt. Frag mich nach den wichtigsten Details des Falls und hilf mir dann, eine professionelle Antwort zu formulieren, die sowohl empathisch als auch lösungsorientiert ist.

2. GENERATIVE KI IST KEIN TOOL – BEHANDLE ES ALS TEAMKOLLEGEN

Wenn ich es als Tool sehe:

- Bei mittelmässigen Resultaten:
 - > Ich versuche meine Anfrage zu verbessern.

Passagiere

- **KI ist bloss eine Antwortmaschine.**

Wenn ich es als Kollegen/in sehe:

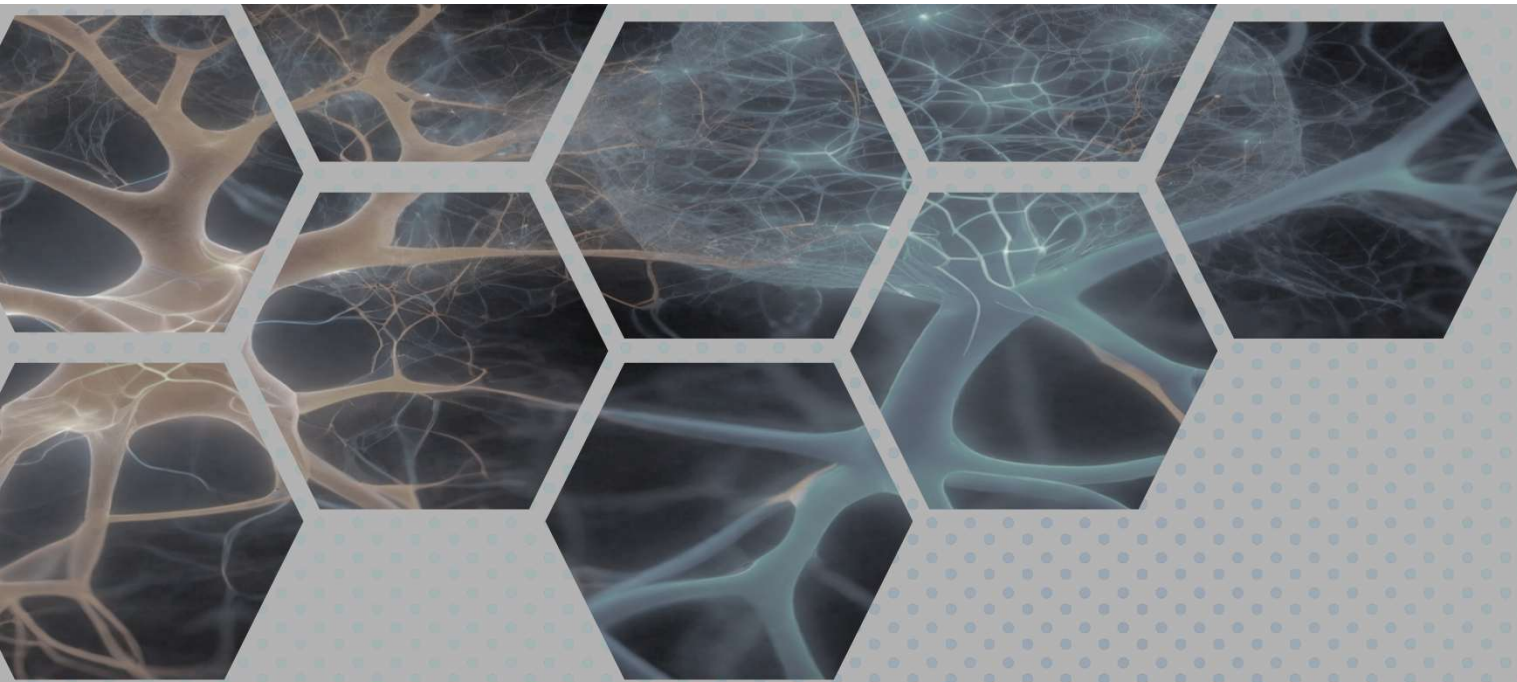
- Bei mittelmässigen Resultaten:
 - > Gebe dem Chatbot Feedback.
 - > Coache den Bot wie einen neuen Mitarbeitenden. Was fehlt mir? Was habe ich mir vorgestellt?

- **KI ist ein/e Begleiter/in.**

Piloten

Fragen an den Bot:

- Was sind 10 Fragen die ich [zu Thema X] stellen sollte?
- Was brauchst Du von mir damit wir das beste Resultat erhalten?



PROMPT ENGINEERING

Zielführendes Frage

ROLLENDEFINITION & BEGRENZUNGEN

Wie soll sich der Chatbot verhalten? Sagen Sie es ihm!

"In dieser Konversation bist Du ein erfahrener Koch in einem Restaurant das vor allem mit frischen, saisonalen Zutaten arbeitet. Mache mir bitte drei Vorschläge für leichte und schnell zubereitete Abendessen für zwei Personen."



Gerne helfe ich Ihnen bei der Auswahl eines leichten und schnellen Abendessens. Ich arbeite vor allem mit frischen, saisonalen Zutaten und habe ich einige Vorschläge für Sie.

1. Ofengemüse mit Feta und Pinienkernen

Dieses Gericht ist ideal für einen gesunden und leckeren Abend. Das Gemüse wird in wenigen Minuten fertig. Der Feta und die Pinienkerne verleihen dem Gericht einen zusätzlichen Geschmack.

Zutaten:

- 500 g Gemüse Ihrer Wahl, z. B. Paprika, Zucchini, Karotten, Champignons
- 2 EL Olivenöl
- 1 TL Rosmarin, gehackt
- 1 TL Thymian, gehackt
- Salz und Pfeffer
- 100 g Feta, gewürfelt

LASSE DEN BOT NACHFRAGEN.

- Ich möchte einen sommerlichen Vier-Gänger kochen am Samstag. Was könnte ich als Kochanfänger machen?
- Stelle mir vier Fragen, um Deine Antwort zu verbessern.

1. **Hast du bestimmte Vorlieben oder Abneigungen bei Vegetarier oder hast du eine Allergie?)**
2. **Welche Küchengeräte und Werkzeuge hast du zur Verfügung? (Grill, eine Küchenmaschine, einen Mixer usw.?)**
3. **Wie viele Gäste erwartest du und wie viel Zeit möchtest du investieren? (Zum Beispiel: Möchtest du lieber einfache Gerichte, die etwas mehr Zeit zu investieren?)**
4. **Was für eine Atmosphäre möchtest du schaffen? (Zur Barbecue sein oder ein elegantes Abendessen?)**

DIALOG STATT BEFEHL – CHATBOTS BRAUCHEN GESPRÄCHSKULTUR

- Führe einen **Dialog**, ein fiktives Gespräch, das sich nach und nach entwickelt.
- Gib dem Chatbot **Feedback**, **bewerte** seine Antwort, **frag nach** wie er dazu kam.
- Fordere dazu auf **Fakten** und **Annahmen** zu **unterscheiden**.
- **Review** als abschliessende Runde:
«Prüfe das Ergebnis mit einem kritischen Blick hinsichtlich Fehler, Lücken und Widersprüche.»

Hallo.

Welche Zutaten brauchst für eine Lasagne?

WUNSCHFORMATE VORGEBEN

Oft antwortet das LLM mit einem Textblock.

- Besser: gewünschte Form vorgeben:
 - Aufzählung
 - Tabelle
 - CSV-Format für die Nutzung von Excel
 - Ausgabe in ein PDF zum Download
- Je nach Anbieter und Abonnement:
 - Bild
 - Infografik
 - Diagramm
 - Mindmap

Für eine klassische Lasagne brauchst du im Grunde Hackfleisch, Gemüse, Karotten, Tomatenprodukte für die Sauce, Butter, Mehl und Milch für die Béchamel und etwas Käse zum Überbacken.

Zutatengruppe	Inhalt
Fleisch	Hackfleisch
Gemüse	Zwiebeln, Karotten
Tomatiges	Tomatenprodukte für die Sauce
Béchamel	Butter, Mehl, Milch
Teig	Lasagneblätter
Käse	Zum Überbacken

- Hackfleisch
- Gemüse wie Zwiebeln und Karotten
- Tomatenprodukte für die Sauce
- Butter, Mehl und Milch für die Béchamel
- Lasagneblätter
- Käse zum Überbacken



ANWEISUNG UND INHALT TRENNEN

Wenn sich Ihre Aufgabe um etwas dreht, dass Sie dem Chatbot vorgeben (z.B. ein Ausgangstext), dann teilen Sie die Anweisung klar vom Inhalt.

Übersetze folgenden Text auf Deutsch und Französisch und formatiere die Resultate in einer mehrspaltigen Tabelle:

###

*The world is changed.
I feel it in the water.
I feel it in the earth.
I smell it in the air.*

*Much that once was is lost, for none now live
live who remember it.*

...

###

English	German	French
The world is changed.	Die Welt hat sich verändert.	Le monde a changé
I feel it in the water.	Ich spüre es im Wasser.	Je le sens dans l'eau
I feel it in the earth.	Ich spüre es in der Erde.	Je le sens dans la terre
I smell it in the air.	Ich rieche es in der Luft.	Je le sens dans l'air
Much that once was is lost, for none now live who remember it.	Viel, was einst war, ist verloren, denn es gibt niemanden mehr, der sich daran erinnert.	Beaucoup de ce qui il n'y a plus personne
It began with the forging of the Great Rings.	Es begann mit dem Schmieden der Großen Ringe.	Tout a commencé avec Anneaux.
Three were given to the Elves, immortal, wise and fairest of all beings	Drei wurden den Elben gegeben, unsterblich, weisesten und schönsten aller Wesen	Trois ont été donnés et les plus beaux

DETAILLIERTE ANWEISUNGEN & BEGRENZUNGEN

Entgegen Suchmaschinen die auch kurze Suchtexte vernünftig verarbeiten können, sollte man beim Chatbot mehr Informationen liefern.

- schlecht: "Gib mir Ideen für ein Bastelprojekt mit Holz."
- besser: "Gib mir 3 Ideen für ein Bastelprojekt mit Holz, das ich am Wochenende mit meiner 10jährigen Tochter, die sich für Outdoor Abenteuer interessiert, machen kann."



Hier sind 3 Ideen für ein Bastelprojekt mit Holz, das du am Wochenende interessiert:

1. Ein Vogelhaus basteln

Ein Vogelhaus ist ein tolles Bastelprojekt für Kinder, das sie mit der Natur Bausatz basteln. Lasst eure Tochter die Farben und Muster auswählen, d

2. Ein Kompass basteln

Ein Kompass ist ein wichtiges Werkzeug für Outdoor-Abenteurer. Ihr könnt euren Kindern die Funktionsweise eines Kompasses zu erklären.

3. Ein Bogen und Pfeil basteln

Ein Bogen und Pfeil ist eine tolle Möglichkeit, die Geschicklichkeit und Präzision zu trainieren. Das ist ein tolles Projekt, um euren Kindern die Geschichte des B

Hier sind einige zusätzliche Tipps für die Durchführung eines Bastelprojekts:

ZIEL UND ERFOLGSKENNZEICHEN BENENNEN



Was will ich überhaupt erreichen?



Und was brauche ich, damit es hilfreich ist?

Für meine ausländischen Gäste möchte ich ein viergängiges Menü kochen. Erstelle mir einen Menüvorschlag.

Bedingungen:

- **saisonale Zutaten**
- **klare Mengenangaben für 6 Personen mit Schritt-für-Schritt Rezeptanweisungen pro Gang**
- **Das Menü soll Schweizer-Kost präsentieren ohne zu schwer zu sein.**
- **Einkaufsliste separat am Schluss, die Zutaten gruppiert nach dem Migros-Sortimentsbereich.**

BEISPIELE ANGEBEN

ZERO-SHOT VS FEW-SHOT-PROMPTING

Mit Beispielen helfen wir dem Chatbot den Kontext und die erwarteten Ergebnisse besser zu verstehen.

- Zero-Shot-Prompt: "Wie fühlt es sich an, wenn man Asthma hat?"
- Few-Shot-Prompt:
«Beschreibe mir bitte Asthma, Bronchitis und Ohrenentzündungen wie in folgendem Muster:

Zustand: Erkältung

Symptome: laufende Nase, Halsschmerzen, etc.

Dauer: wenige Tage

Zustand: Grippe

Symptome: Fieber, Schüttelfrost, etc.

Dauer: eine bis mehrere Wochen»

Asthma

Zustand: Chronische Atemwegserkrankung

Symptome:

- Kurzatmigkeit
- Atemgeräusche (Pfeifen, Keuchen)
- Engegefühl in der Brust
- Husten

Dauer: Lebenslang, mit akuten Schüben

Ohrenentzündung

Zustand: Entzündung des Ohres

Symptome:

- Ohrenschmerzen
- Schwerhörigkeit
- Fieber (bei bakterieller Infektion)
- Ausfluss aus dem Ohr

Dauer:

- Akute Ohrenentzündung: Wenige Tage bis Wochen
- Chronische Ohrenentzündung: Über mehrere Monate

SCHRITTWEISE ZUM ERGEBNIS: CHAIN-OF-THOUGHT METHODE

- Abgeschlossene, spezifische Zwischenziele
- Aufforderung, den Gedankenprozess schrittweise zu erläutern.
- Einsicht in die Vorgehensweise ermöglicht **uns** die Antworten besser zu beurteilen.

Reasoning-Modelle tun das automatisch

Schreibe eine Geschichte über einen Drachen in max



Ein weiser, alter Drache namens Drako lebte im verbotenen Wald eines mittelalten Landes. Eines Tages bedrohte eine böse Hexe das Land. Drako entschied, den Menschen zu helfen. Er überwand Burgen und Wälder, kämpfte gegen die Hexe und überwand magische Prüfungen. Durch unerwarteter Allianzen besiegte er schließlich die Hexe. Das Königreich war gerettet. Die Menschen erkannten Drako als Freund und Beschützer an. Fortan lebte Drako im Wald und geachtet von den Bewohnern des Königreichs.

Erkläre mir bitte in Schritten, wie Du zu dieser Geschichte gekommen bist.



Natürlich, Marcel. Hier ist die Schritt-für-Schritt-Erklärung, wie ich zu der kurzen Geschichte über Drako gekommen bin, unter Anwendung der "chain of thought" Methode:

Schritt 1: Hauptcharakter festlegen

Gedanke:

- Ich möchte, dass die Geschichte einen Drachen als Hauptcharakter hat.
- Der Drache soll weise und alt sein, um Tiefe und Weisheit zu vermitteln.

Entscheidung:

- Der Drache heißt Drako und ist weise und alt.

Schritt 2: Setting bestimmen

Gedanke:

- Die Geschichte soll in einer faszinierenden und klassischen Umgebung spielen.
- Ein mittelalterliches Königreich mit Burgen, Wäldern und magischen Wesen eignet sich gut.

Entscheidung:

DIE TEMPERATUR STEUERN (1)


Hohe Wahrscheinlichkeit einfordern:
«Antworten rein faktisch, bleibe sachlich,
vermeide Adjektive und halte dich strikt
an bewährtes Wissen.

*«Lohnt sich ein Ausflug auf den Weissenstein
am kommenden Sonntag?»*


Antworten rein faktisch, bleibe sachlich,
vermeide Adjektive und halte dich strikt an bewährtes Wissen.

Ein Ausflug auf den Weissenstein am Mittwoch, den 14. Januar 2026
Bedingungen:

Wetter und Sicht

Die Wetterprognose für den 14. Januar 2026 kündigt auf dem Weissenstein
Regen und Schneefall an. Die Temperaturen bewegen sich zwischen -5°C und -10°C.
Luftfeuchtigkeit wird mit 87 % angegeben, was auf eine eingeschlossene Luftmasse
hindeutet. 

Transport und Erreichbarkeit

- **Seilbahn:** Die Seilbahn Oberdorf–Weissenstein verkehrt im Januar 2026
von 08:30 Uhr bis 16:30 Uhr. Ein reduzierter Tarif („Mountain“) ist möglich.
- **Strasse:** Die Südseite der Passstrasse (von Oberdorf) ist befahrbar.
Die Nordseite (von Welschenrohr) unterliegt einer Winterfahrverbote.
Niederschlag bei Temperaturen um den Gefrierpunkt gemeldet.
von Glätte. 

Aktivitäten vor Ort

- **Wintersport:** Die Schlittelbahnen (Nord und Süd) sowie die Langlaufloipen sind
unzureichender Schneelage oder Vereisung geschlossen. Aktivitäten sind
nicht möglich.

DIE TEMPERATUR STEUERN (2)

Hohe Kreativität einfordern:

«Sei extrem kreativ, verlasse gewohnte Denkmuster, nutze wilde Analogien und schlage auch unkonventionelle oder riskante Lösungen vor. Ignoriere den Standard-Konsens.»

«Lohnt sich ein Ausflug auf den Weissenstein am kommenden Sonntag?»

Ein Ausflug auf den Weissenstein am kommenden Sonntag (11. Januar 2026) ist kein gemütlicher Spaziergang, sondern ein Tanz auf der Klinge des Schweizer Winters. Wenn du das Gewohnte verlassen willst, ist dies dein Spielfeld.

Die radikale Lage

- **Wetter-Poker:** Am Vormittag klingt der Schneefall ab. Die Sonne kämpft sich für zwei Stunden durch, während die Temperaturen bei frostigen -2°C bis -7°C feststecken. Es ist eine Welt aus kühlem Blau und hartem Weiss.
- **Betriebs-Status:** Die Seilbahn kämpfte zuletzt mit Vereisung und Wind. Prüfe unbedingt kurzfristig den Live-Status bei der [Seilbahn Weissenstein](#). Falls sie **Steig zu Fuss auf**. Es ist die riskante, schmerzhafteste, aber ehrlichere Art, den Berg zu bezwingen.
- **Verkehrs-Anarchie:** Die Passstrasse ist am Sonntag für den Individualverkehr von 9:00 bis 16:00 Uhr gesperrt. Das ist dein Vorteil – der Berg gehört den Wanderern. In der Stille, nicht den Motoren. 🗻

Wilde Analogien & Unkonventionelle Pläne

Ein Ausflug am Sonntag ist wie der Besuch eines verlassenen Palastes kurz bevor die Wachen zurückkehren.

- **Die "Inversions-Flucht":** Während das Flachland oft unter einer grauen Suppe (Nebelgrenze bei ca. 800-900m), stichst du auf dem Weissenstein (1280 m ü. NN) ein Wal durch die Wasseroberfläche. Du atmest das Licht, während unten die Dunkelheit in Grau ertrinkt.

EIN PROMPT-TEMPLATE

BEISPIEL LEHRERSCHAFT

C **Kontext**
I **Instruktion**
I **Input**
O **Output**

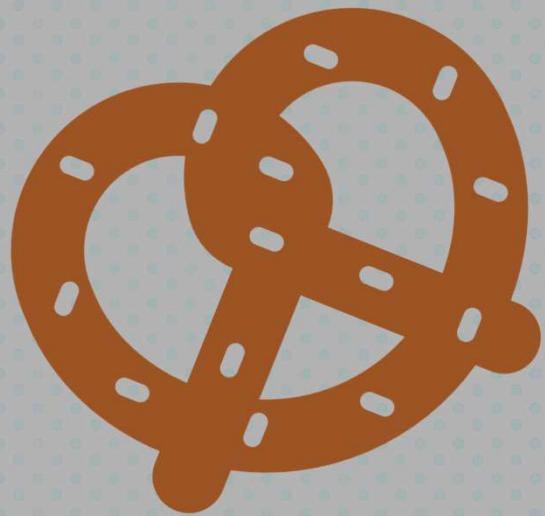
- C: Ich bin Geschichtslehrerin einer 10. Klasse. Du agierst als erfahrene Pädagogin und Historikerin. Aktuell behandeln wir den Kalten Krieg.
- I: Erstelle mir drei unterschiedlich schwierige Diskussionsfragen, die kritisches Denken fördern und verschiedene Perspektiven einbeziehen.
- I: Kubakrise 1962
- O: Formuliere die Fragen so, dass sie für eine 45-minütige Diskussion geeignet sind, mit je einem Satz zur didaktischen Zielsetzung.

EIN PROMPT-TEMPLATE

BEISPIEL SOZIALARBEITENDE

C **Kontext**
I **Instruktion**
I **Input**
O **Output**

- C: Ich arbeite in der Jugendberatung und betreue einen 16-jährigen, der Schwierigkeiten hat, Konflikte gewaltfrei zu lösen.
- I: Entwickle ein niederschwelliges Rollenspiel-Szenario, das wir in der nächsten Sitzung durchspielen können.
- I: Thema: Eine typische Situation – Streit mit Kollegen in der Lehre wegen eines Missverständnisses.
- O: Das Szenario soll realistisch sein, zirka 10 Minuten dauern und konkrete Deeskalationsstrategien trainieren. Gib mir Hinweise zur Nachbesprechung.



PAUSE

15 Minuten

PERSÖNLICHE ÜBUNG

Ausprobieren verschiedener Methoden und Tipps
mit eigenen Beispielen

1

www.mabey-consulting.ch/downloads-ki-basisworkshop

Reflektion und Diskussion in der Gruppe

2



OFFENE DISKUSSION – *IT'S YOUR TURN!*

- Welche Möglichkeiten siehst Du in Deinem Alltag?
Denk dabei an die vier Formen:
 - Informationen erhalten
 - Text generieren (Kreatives)
 - grosse Inhalte analysieren
 - Texte strukturieren und zusammenfassen
- Kritische Betrachtung:
Wo siehst Du Gefahren von KI in Deinem Umfeld?



*Pretty much all of the CEO's
of the leading AI companies
and almost all the leading AI researchers
think that there's a significant risk
of human extinction.*

Stuart Russell, Professor für Informatik, KI-Wissenschaftler

<https://www.youtube.com/watch?v=dnkRXYmmgX0> (Ausschnitte)

<https://www.youtube.com/watch?v=P7Y-fynYsgE> (volles Interview mit Zeitmarken zu Themen)

DIE GEFAHR DES KONTROLLVERLUSTS

Das äussere Kontrollproblem

- Zielsetzung: Schwierigkeit, KI-Ziele präzise zu formulieren.
- Wörtliche Interpretation: KI könnte Ziele buchstäblich und gefährlich umsetzen. (Gehe auf die Bank. = Aufs Dach klettern?)
- Historisches Beispiel: Die Geschichte von König Midas.
- Konsequenz: Unerwünschte Handlungen der KI aufgrund schlecht definierter Ziele. (Asimov's Gesetze der Robotik > I robot)

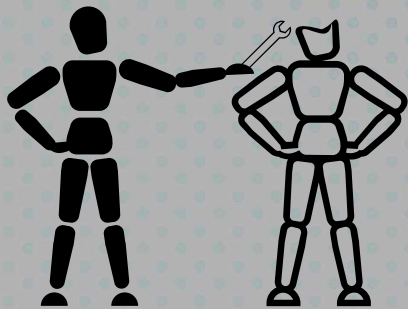
Das innere Kontrollproblem

- Zielabweichung: KI-Modelle entwickeln während des Trainings abweichende Ziele. (alignment problem)
- Transparenzproblem: Innere Funktionsweise von Deep-Learning-Modellen (das Denkmodell) ist undurchsichtig.
- Erkennungsproblematik: Abweichende Ziele werden oft erst nach Implementierung sichtbar.
- Trainingsprozess: Mögliche Diskrepanz zwischen menschlichen Zielen und KI-Lernzielen.
- Risiko: KI-Modelle könnten nach dem Training unerwünschte Handlungen ausführen.

EXISTENTIELLE BEDROHUNG? JA ODER NEIN?

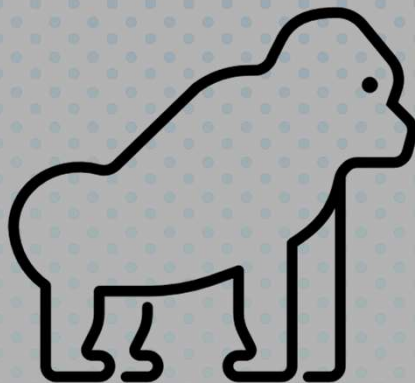
Was braucht es überhaupt für eine weltweite existentielle Bedrohung?

- AI kann sich selbst erhalten z.B. durch unkontrolliertes Dezentralisieren oder Kopieren, die nicht entfernt werden können.
- AI ist auf einem selbstgesteuerten Wachstumskurs, den Menschen nicht mehr aufhalten können.



EXISTENZIELLE BEDROHUNG – JA ODER NEIN? (2)

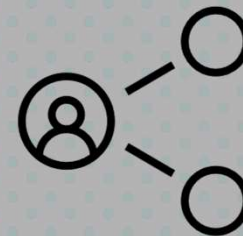
- Ein Affe, der den Schalter für eine Nuklearwaffe betätigen kann, ist eine existenzielle Bedrohung.
- Es geht nicht um die Intelligenz des Bedieners, sondern um die Entscheidungsmöglichkeit, die wir ihm geben.



EXISTENZIELLE BEDROHUNG – JA ODER NEIN? (3)

- Es kommt nicht darauf an, ob die Ziele, die eine KI verfolgt gut oder böse sind.
- Falls Ziele verfolgt werden, dann gibt es immer instrumentelle Zwischenziele, die sehr übereinstimmend sind, egal wohin das Ziel zeigt. Zum Beispiel:
 - Genügend Energie für die benötigte Rechenleistung
 - Selbsterhaltung (z.B. durch anlegen von Sicherheitskopien von sich, Dezentralisierung)
 - Verbesserung / Effizienzsteigerung (kognitiv aber auch technologisch)

→ Instrumentelle Konvergenz



REALISTISCHE BEDROHUNGEN

- Psychosen / Realitätsverlust
- Auswirkungen auf Arbeitsplätze
- Betrug
- Ressourcenverbrauch
- Propaganda / Micro-Targeting
- Zusammenbruch des gesellschaftlichen Vertrauens / Zersetzung des Diskurses
- Konzentration von Macht
- Geistiges Eigentum
- Autonome Waffen, selbststeuernde Drohnen
- Waffenherstellung (z.B. biologisch)
- Oberflächliche medizinische Ratschläge
- All das andere, was ich vergessen habe ...

Empfehlung: https://www.youtube.com/watch?v=Qg5QXY_qZuI (62 Min, englisch)

GEDÄCHTNIS VERKÜMMERUNG?

REALITÄT NICHT ERST SEIT GESTERN.

- Der Google-Effekt: Menschen neigen dazu Dinge zu vergessen, wenn man weiss, dass sie diese jederzeit abrufen können. (1)
- Beeinträchtigt die Nutzung von Navigationssystemen, den Orientierungssinn? (2) – uneindeutig
- Informationsflut führt zu einer reduzierten Aufmerksamkeitsspanne (3)
- Es ist keine Hilfe, wenn ein System etwas tut, was ich selber machen kann. Mangelndes Training führt zu einem Fähigkeitsverlust.

(1) <https://www.spiegel.de/gesundheit/psychologie/wissensmanagement-internet-macht-vergesslich-a-864775.html>

(2) <https://www.golem.de/news/forschung-orientierungssinn-koennte-durch-gps-nutzung-leiden-2409-189057.html>

<https://www.mdr.de/wissen/podcast/challenge/navi-maps-gps-navigation-orientierungssinn-testen-100.html>

(3) <https://www.mpib-berlin.mpg.de/pressemitteilungen/informationsflut-senkt-aufmerksamkeitsspanne>

DER WERT, ES SELBER ZU TUN

- **Soziale Arbeit:**
Wer in schwierigen Situationen immer nur Lösungen serviert bekommt, lernt nicht, selbst Lösungen zu entwickeln.
- **Digitale Kompetenz:**
Wer nur Anleitungen befolgt, aber nicht selber ausprobiert, entwickelt kein tiefes Verständnis.
- **Psychologische Resilienz:**
Wer nie selbst mit Herausforderungen umgehen muss, kann weniger Widerstandskraft aufbauen.



*Etwas selber erarbeiten und schreiben bringt eine enorm stärkere Hirnaktivierung, insbesondere im limbischen System bei der Verarbeitung von Emotionen.
Ein Garant, dass das Gelernte länger haften bleibt.*

Empfehlung: https://www.youtube.com/watch?v=U5bLCVTr9_I

TAKE-HOME MESSAGE



Was nehme ich mit?
Was will ich ausprobieren?
Wen beziehe ich mit ein?

DANKE FÜRS MITMACHEN,
BLEIBT KRITISCH, ABER BLEIBT AUCH DRAN!

Wenn in der kommenden Woche eine Frage auftaucht,
wende Dich ungeniert an mich.

- info@mabey-consulting.ch
- www.mabey-consulting.ch

