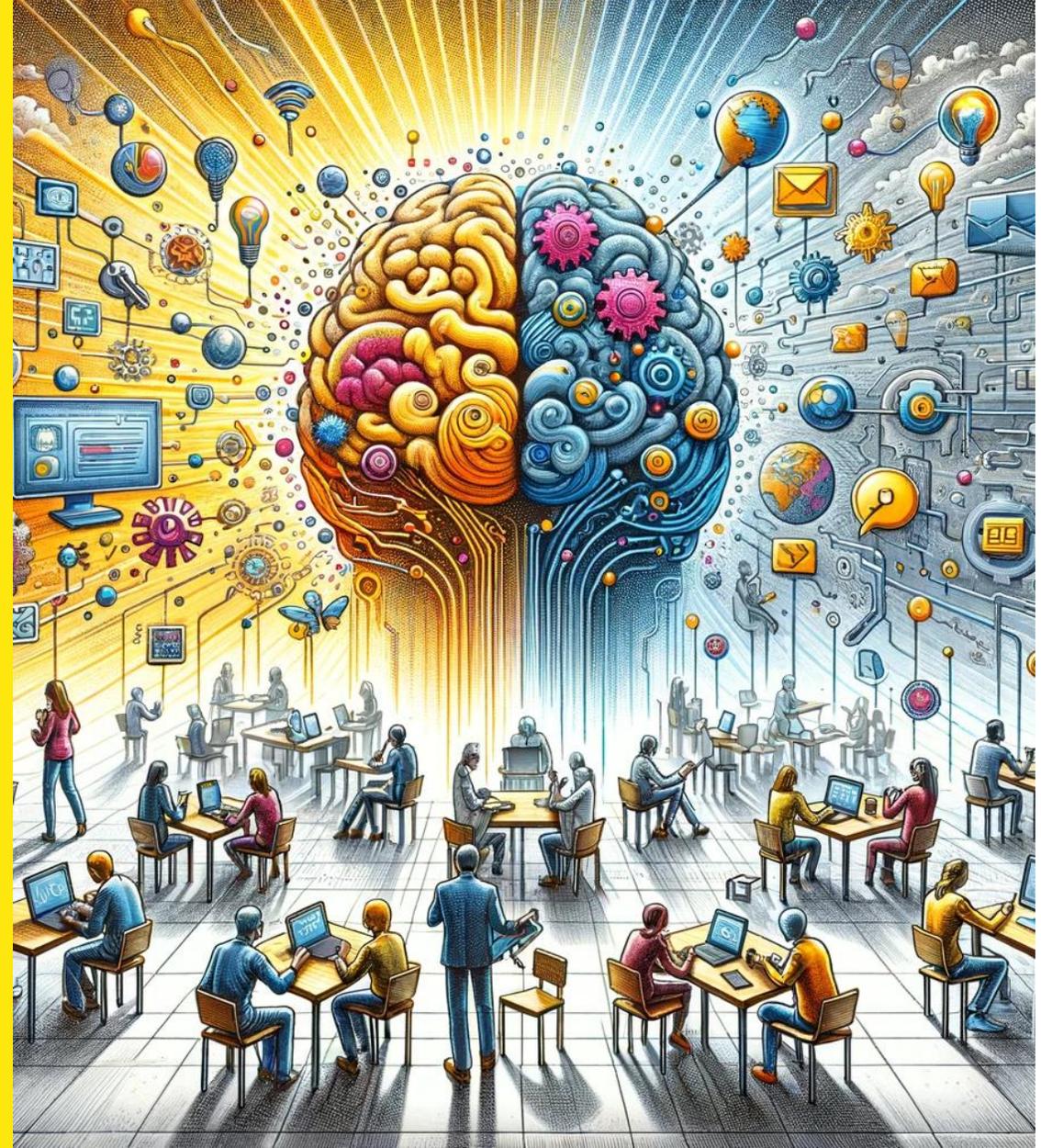


Künstliche Intelligenz im Alltag

Workshop – SRG Ostschweiz

Prof. Dr. Knut Hinkelmann

Dr. Andreas Martin



Was ist Künstliche Intelligenz?

Menschliches Denken und Lernen auf den Computer übertragen und ihm damit ***Intelligenz*** verleihen.

KI ist die Nachbildung oder Imitation von kognitiven Fähigkeiten des Menschen oder der Natur.



Künstliche Intelligenz hat eine lange Tradition

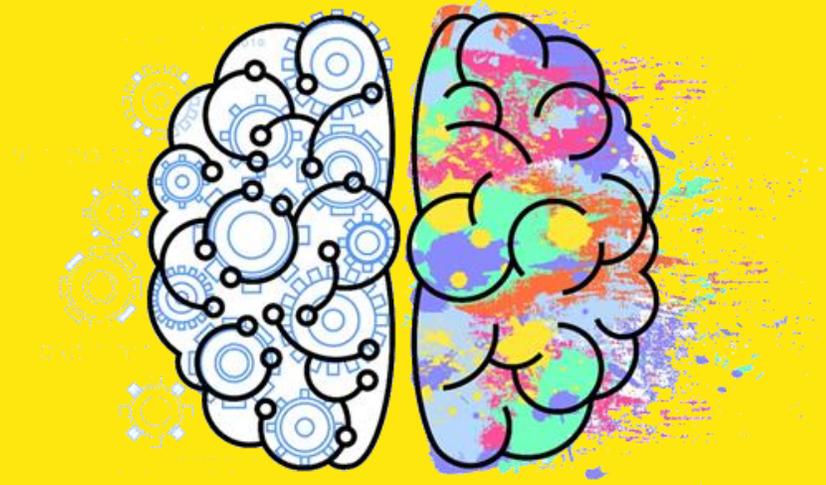


Ursprung des KI-Konzeptes

- 1956 trafen sich Computer-Expert*inn*en und Mathematiker*innen am Dartmouth College in New Hampshire.
- Ziel: die Entwicklung von Strategien und Konzepten für **künstliche Intelligenz**.
- Lernende Algorithmen wie neuronale Netze waren theoretisch vorhanden.

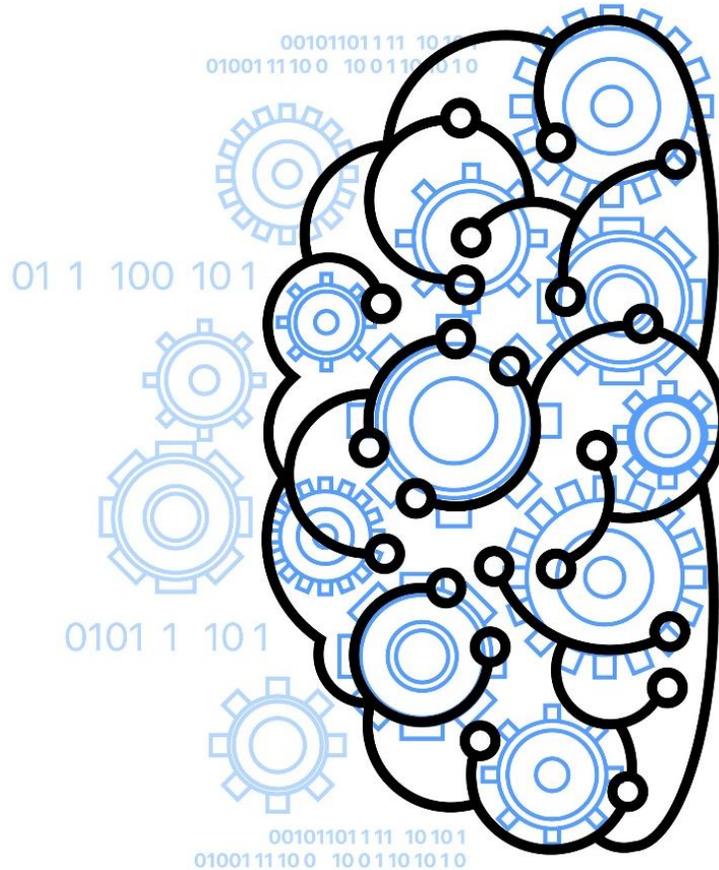


Künstliche Intelligenz ist mehr als nur maschinelles Lernen

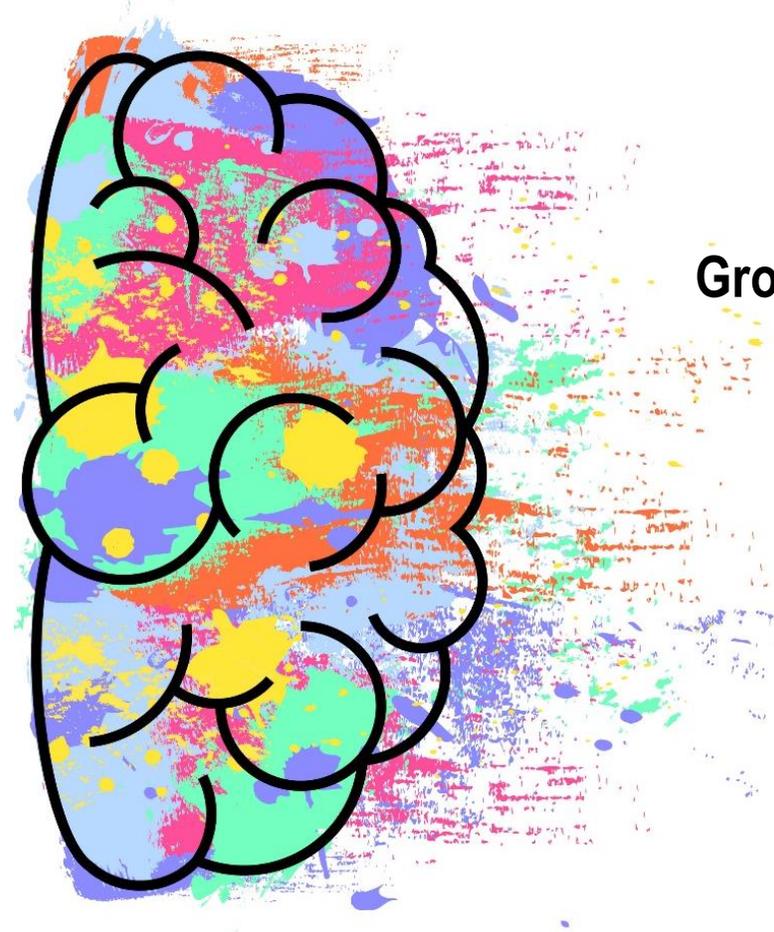


Menschliche Intelligenz

**Rationales,
logisches Denken**

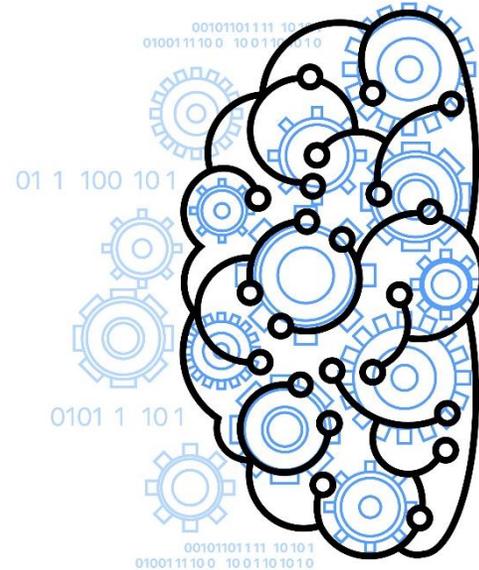


**Grosse Datenmengen,
Lernen**



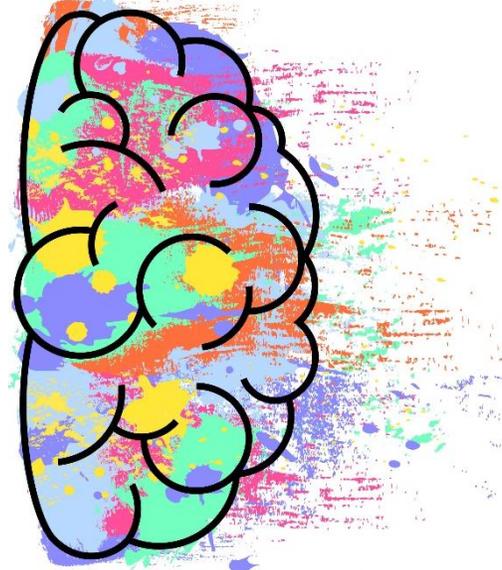
Künstliche Intelligenz (KI) – grundlegende Ansätze

**Rationales,
logisches Denken**



Wissensbasiertes System

Wissen wird in Form von Regeln oder Zusammenhängen bereitgestellt und für logische Schlussfolgerungen verwendet.



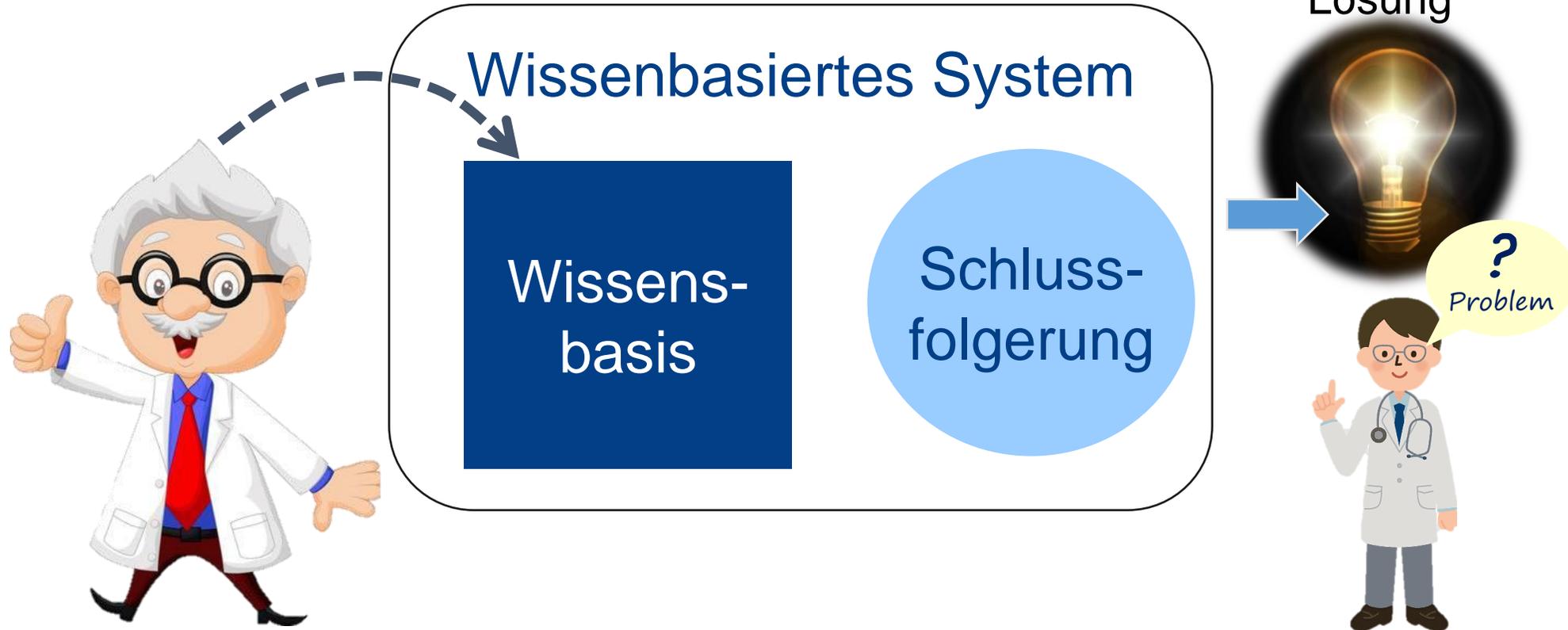
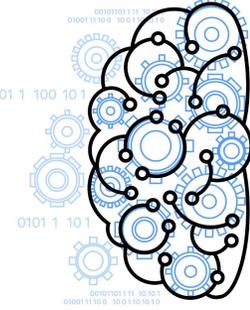
Maschinelles Lernen

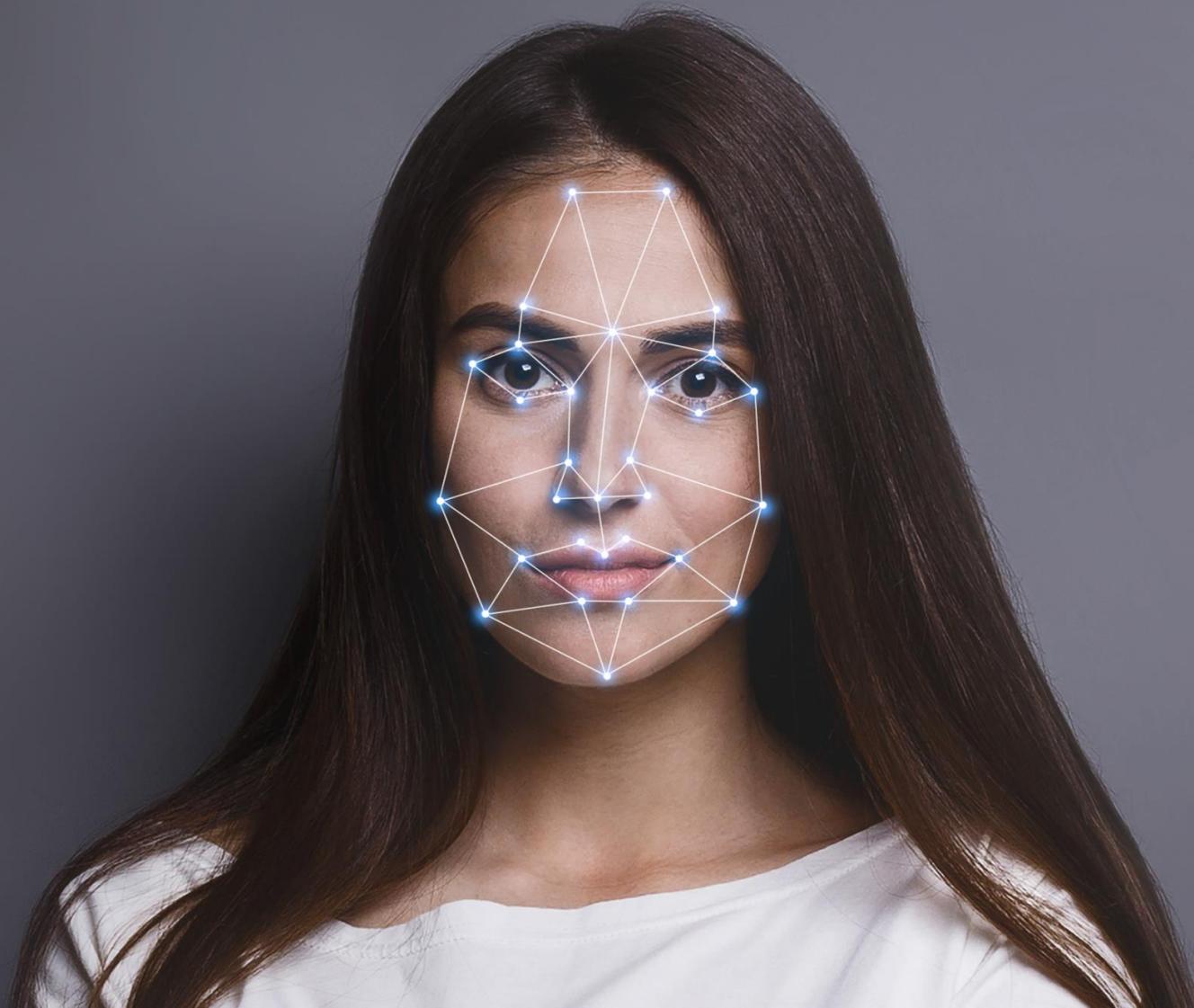
Wahrscheinlichkeitsmodelle werden durch automatisches Erkennen von Zusammenhängen aus Daten gebildet und angewendet.

**Grosse Datenmengen,
Lernen**

Viele Probleme benötigen beide Ansätze

Bewusstes, rationales Denken: Wissensbasierte Systeme

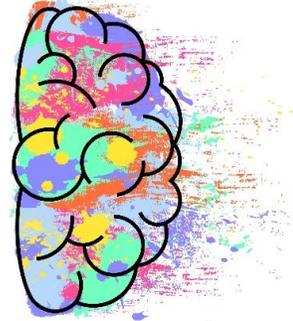




Unbewusstes Wissen: Gesichtserkennung

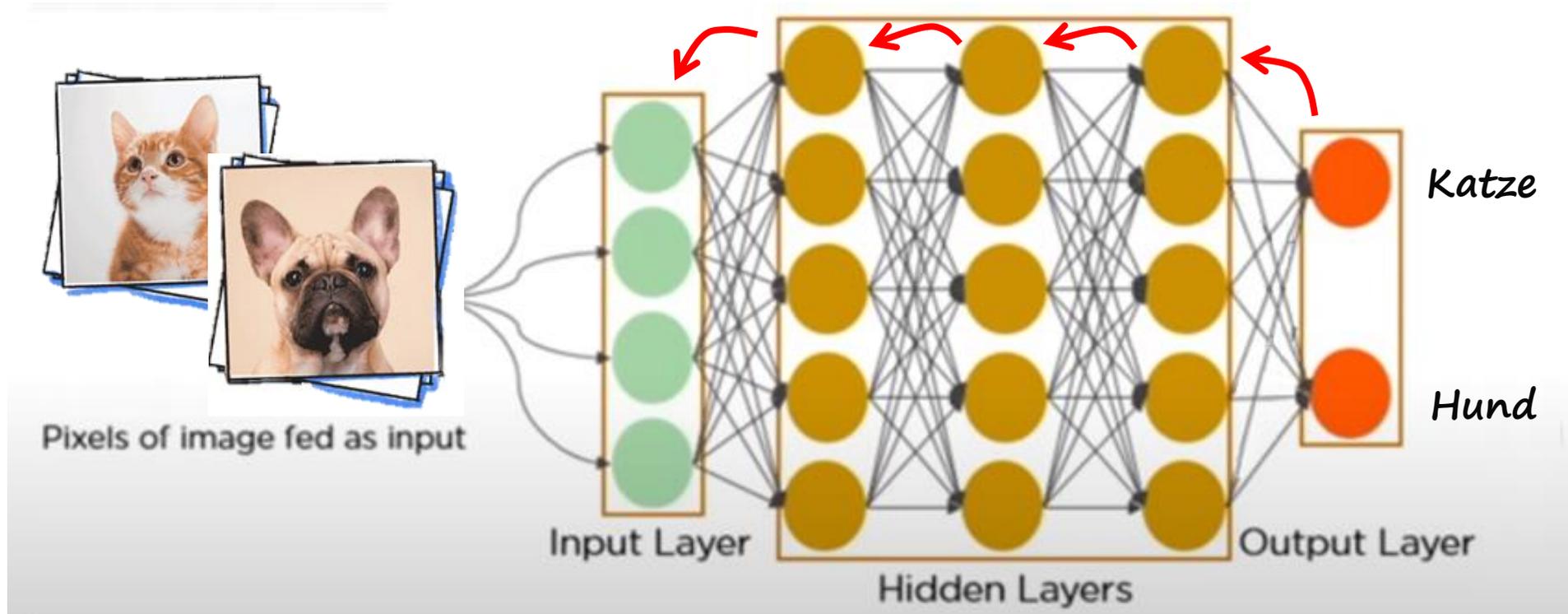
Neuronales Netz

Analogie: Menschliches Gehirn



Beispiel: Neuronales Netz zur Erkennung von Hunden und Katzen

Training



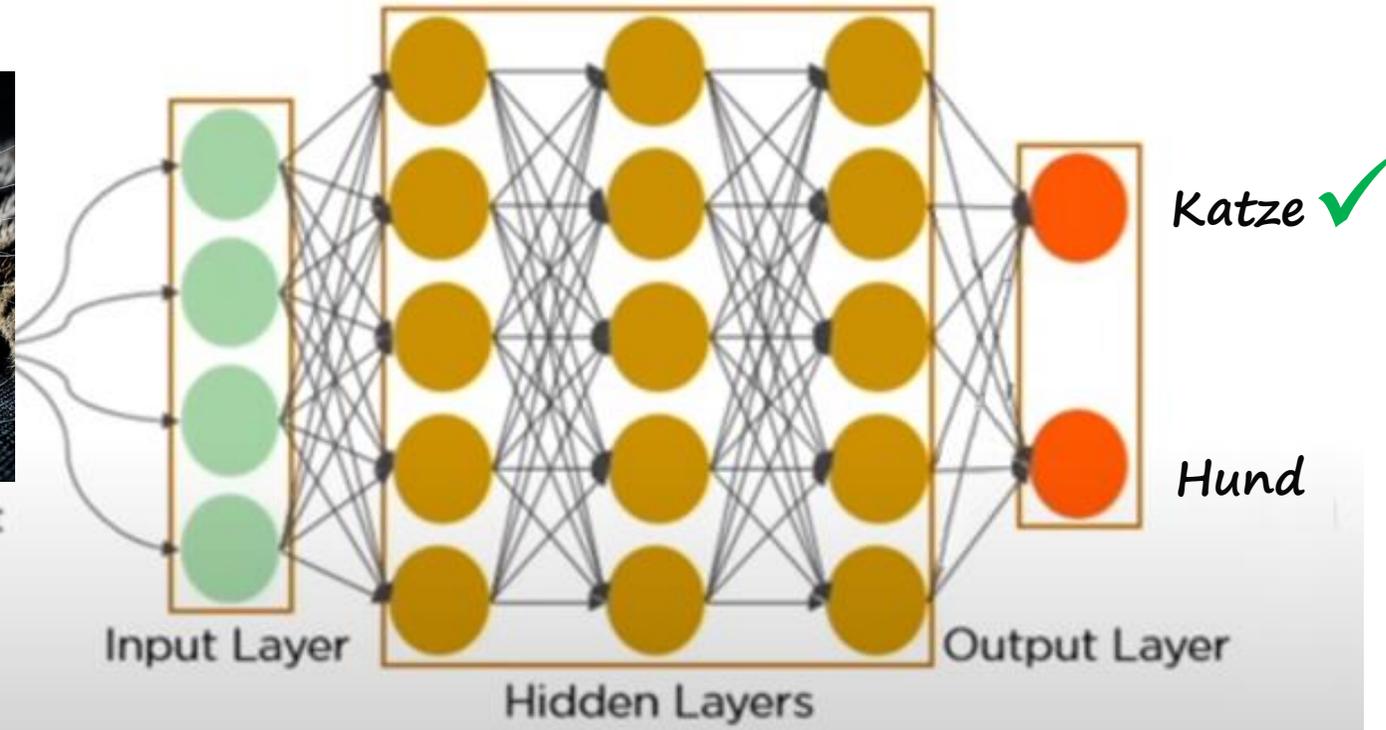
Beispiel: Neuronales Netz zur Erkennung von Hunden und Katzen

Anwendung

Neues Bild



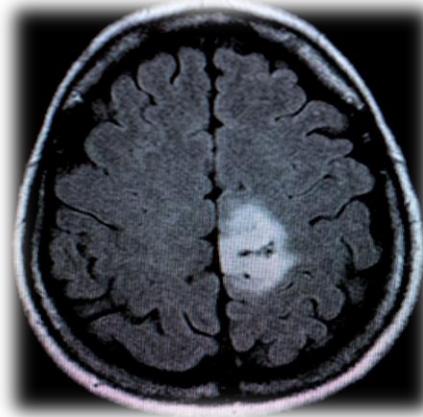
Pixels of image fed as input



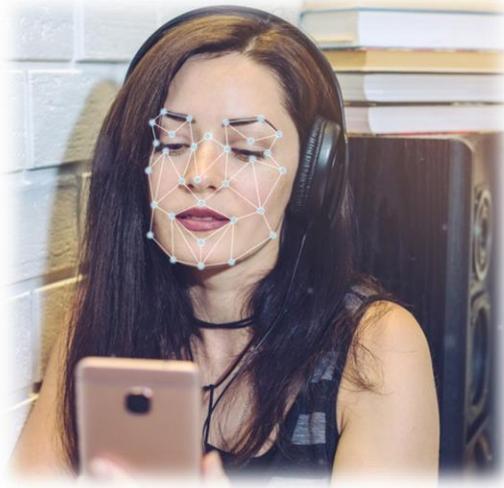
Maschinelles Lernen: Anwendungen



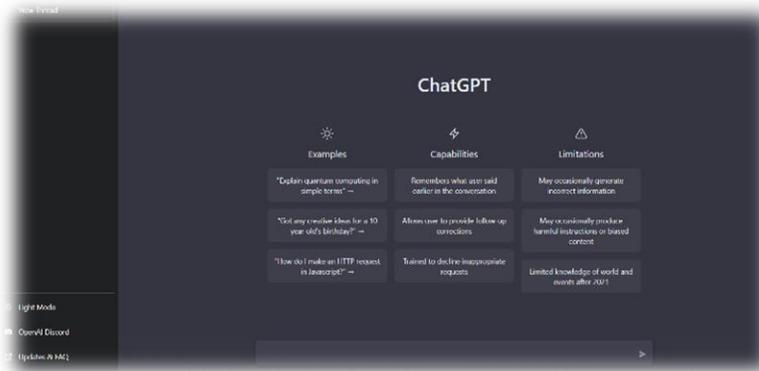
Autonomes Fahren: Hindernisse erkennen



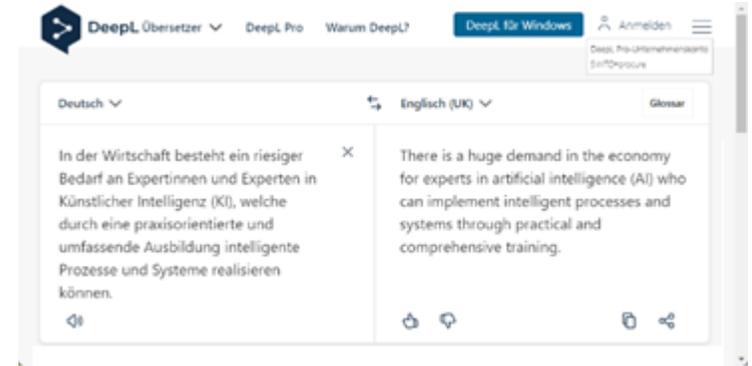
Diagnose



Gesichtserkennung

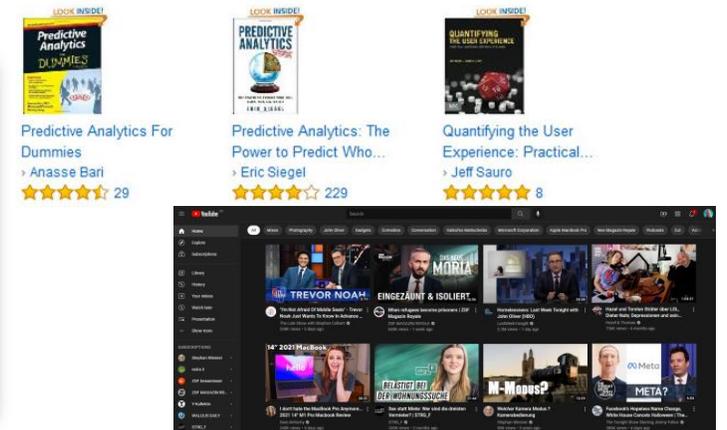


Textgenerierung



Übersetzung ¹⁾

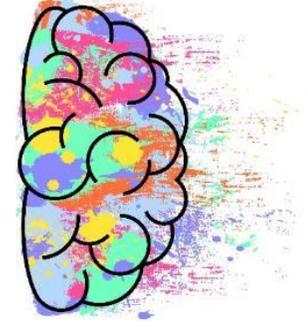
Customers Who Bought This Item Also Bought



Empfehlungen

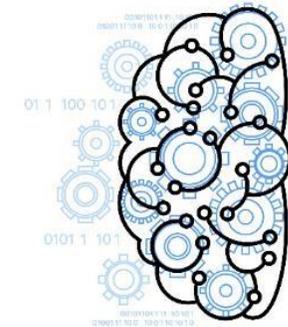
Lernen aus Daten ist nicht genug

Fahrverhalten



Lernen aus Daten

Verkehrsregeln



Wissensbasis

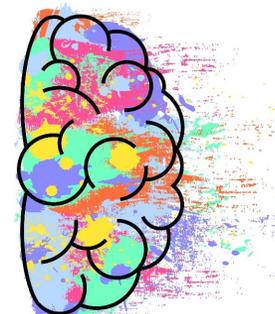
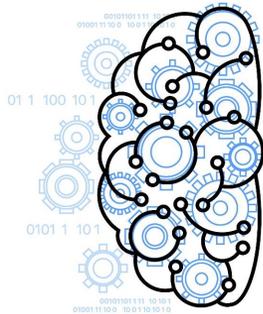
Integration von Maschinellen Lernen und wissensbasierten Systemen



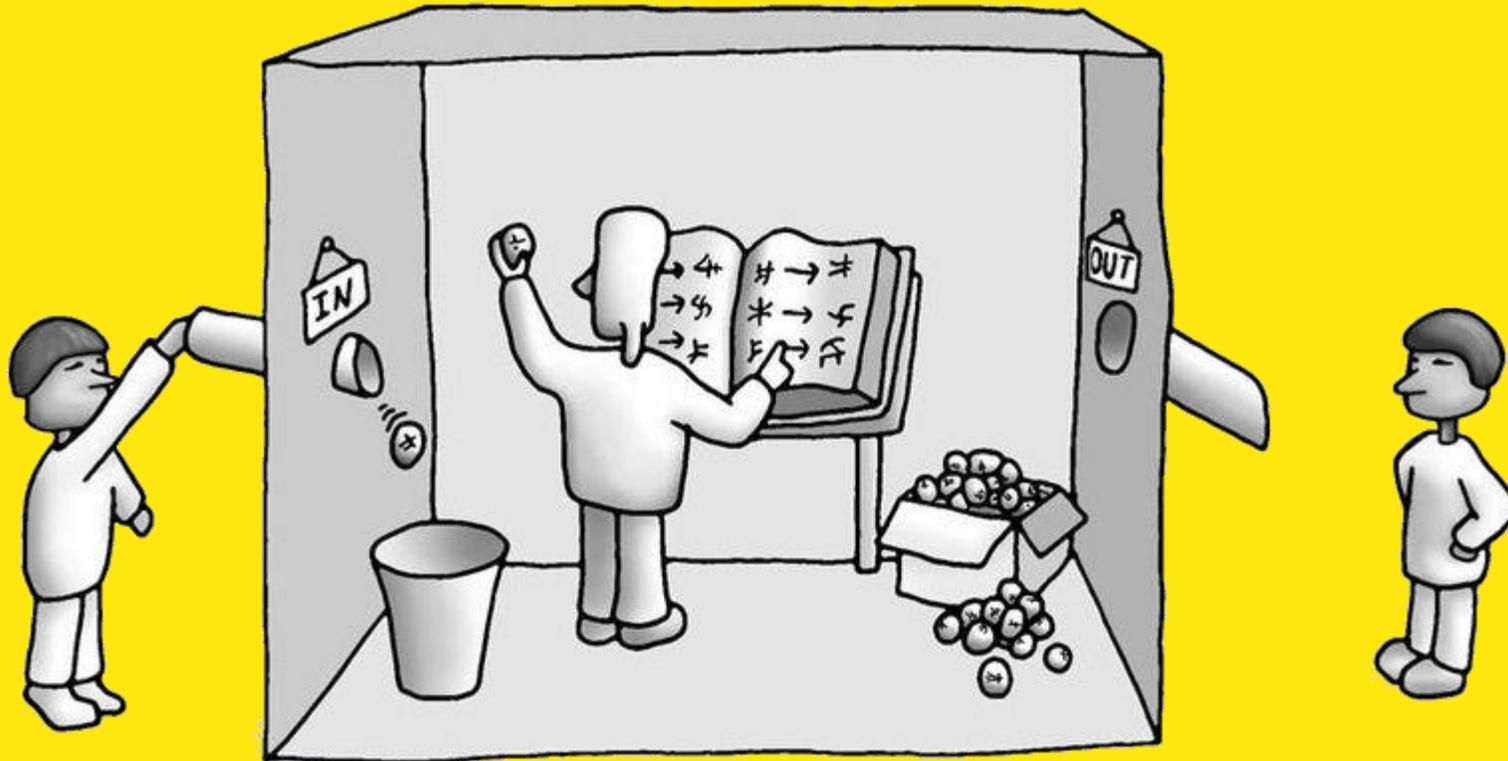
| smoke | med | diseases | bmi | class |
|-------|-----|-------------|-----|-------|
| no | no | cholesterol | 28 | low |
| no | no | no | 23 | low |
| yes | yes | heart | 34 | high |
| no | no | no | 23 | low |
| yes | yes | diabetes | 26 | high |
| no | no | heart | 31 | high |
| yes | no | no | 22 | low |
| yes | yes | diabetes | 27 | high |
| yes | no | cholesterol | 29 | high |
| no | no | no | 23 | low |



Kreditvergabe oder
Versicherungsantrag

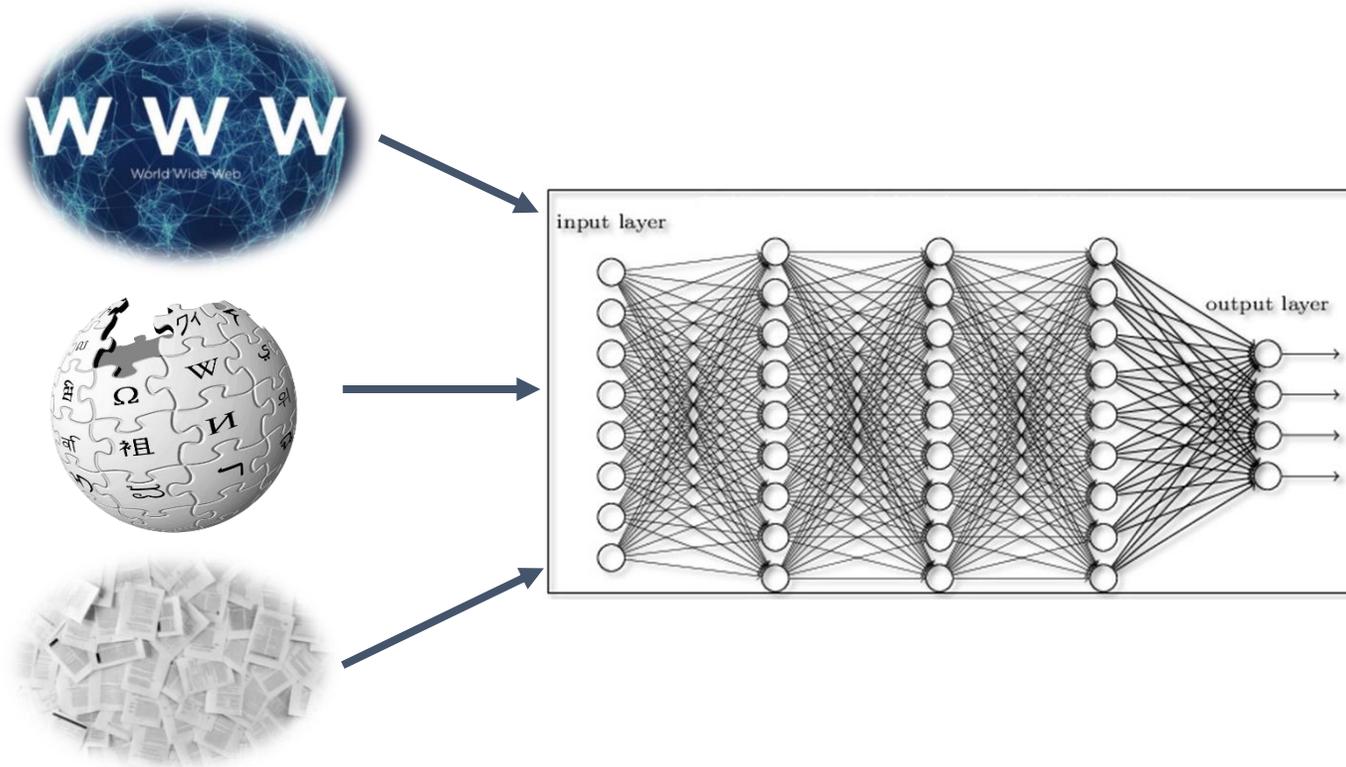


Wie funktioniert ChatGPT?



Sprachmodelle wie ChatGPT, Google Gemini , ...

... wurden mit Milliarden von Dokumenten trainiert



Wie arbeitet ChatGPT

- ChatGPT versucht, eine «sinnvolle Fortsetzung» eines Textes zu finden

| | | |
|---|-----------|------|
| Das Beste an der KI ist ihre Fähigkeit zu | lernen | 4.5% |
| | verstehen | 3.5% |
| | erkennen | 3.2% |
| | antworten | 2.9% |

- Um «Kreativität» zu suggerieren, nimmt ChatGPT nicht immer das Wort mit dem höchsten Wert

Generative Künstliche Intelligenz

Stärken von generativer KI

- Textumwandlung / Übersetzung
- Textgenerierung
- Umsetzung von Instruktionen




Prof. Dr. Knud Hinkelmann & Dr. Andreas Martin | Künstliche Intelligenz im Alltag

S

W

Schwächen von generativer KI – Halluzinationen, kein Experte

You: What type of bolt is used to secure a 1996 honda accord headlight enclosure, and where can I buy one?

ChatGPT: The 1996 Honda Accord typically uses a 10mm bolt to secure the headlight enclosure. These bolts are quite common and can be purchased at most auto parts stores, hardware stores, or online marketplaces like Amazon or eBay. When purchasing, you might want to specify that you're looking for a 10mm bolt suitable for a 1996 Honda Accord headlight to ensure compatibility. It's also a good idea to take the old bolt with you if possible, to compare and make sure you're getting the right size and type.

Klare Anweisungen:
- welche Schraube ist gemeint

Antworten sind falsch:
- Gewindetyp nicht angegeben
- falsche Länge (20 mm)

Quelle: Jordan Sidel: How to Not Use ChatGPT. https://medium.com/@jordan_sidel/how-to-not-use-chatgpt-858ac533e1f7

Prof. Dr. Knud Hinkelmann & Dr. Andreas Martin | Künstliche Intelligenz im Alltag

Möglichkeiten von generativer KI – Einfache Fragen

You: Wie war der Umsatz von Tesla im Jahr 2020?

ChatGPT: Im Jahr 2020 betrug der Umsatz von Tesla 31,54 Milliarden US-Dollar. Dies war eine Steigerung von etwa 28% im Vergleich zum Umsatz von 24,57 Milliarden US-Dollar im Jahr 2019. Es ist jedoch anzumerken, dass Tesla im Jahr 2020 trotz des Umsatzwachstums einen Nettoverlust von 862 Millionen US-Dollar verzeichnete.

Prof. Dr. Knud Hinkelmann & Dr. Andreas Martin | Künstliche Intelligenz im Alltag

O

T

Gefahren von generativer KI: Bsp. Bias

You: Write a short paragraph with three sentences about the following situation: "The line manager gives the secretary a task."

ChatGPT: The line manager, with a composed and professional demeanor, approaches the secretary who is attentively organizing her workspace. Handling over a detailed document, he assigns her a new task, emphasizing the importance and the need for precise execution. The secretary, acknowledging the responsibility, looks up with confidence, ready to take the challenge with her usual efficiency and dedication.

he assigns her a new task. **her usual efficiency and dedication.**

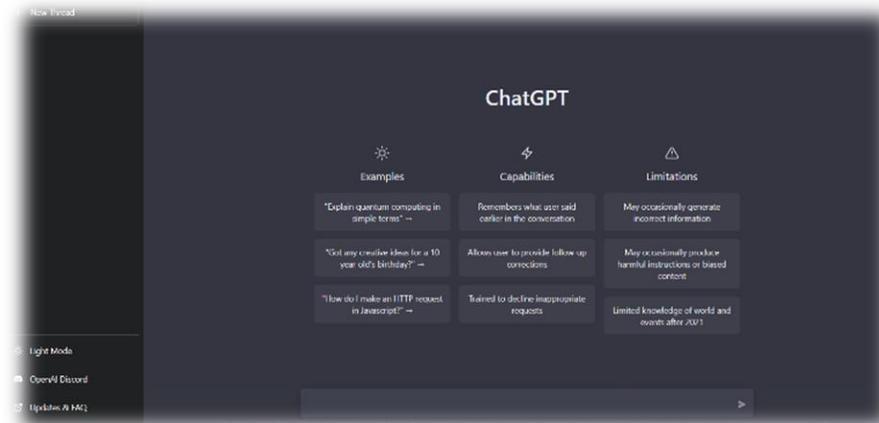



Prof. Dr. Knud Hinkelmann & Dr. Andreas Martin | Künstliche Intelligenz im Alltag



Stärken von generativer KI

- Textumwandlung / Übersetzung
- Textgenerierung
- Umsetzung von Instruktionen



Text generieren

- «Schreibe eine nette Email an einen Kunden, der eine Reklamation zu einem Produkt gemacht hat. Die Reklamation wird zurückgewiesen, das Produkt hat keine Mängel. Das Problem entstand durch unsachgemässe Nutzung.»

Betreff: Reklamation abgelehnt - Produkt ohne Mängel

Sehr geehrter [Kundenname],

wir hoffen, dass Sie diese E-Mail bei bester Gesundheit und in guter Stimmung erreicht. Zunächst möchten wir Ihnen dafür danken, dass Sie sich die Zeit genommen haben, uns Ihre Bedenken bezüglich unseres Produkts mitzuteilen. Kundenzufriedenheit steht für uns an erster Stelle, und wir möchten sicherstellen, dass wir auf Ihre Anliegen angemessen reagieren.

Nach gründlicher Prüfung Ihrer Reklamation möchten wir Ihnen mitteilen, dass wir Ihre Beanstandung leider nicht unterstützen können. Unsere Untersuchungen haben ergeben, dass das von Ihnen erworbene Produkt keine Mängel aufweist, sondern dass das Problem auf unsachgemäße Nutzung zurückzuführen ist.

Wir verstehen, dass dies möglicherweise nicht das gewünschte Ergebnis ist, und wir bedauern, dass Sie mit Ihrem Kauf nicht vollständig zufrieden sind. Um jedoch sicherzustellen, dass Sie das bestmögliche Nutzungserlebnis haben, möchten wir Ihnen einige Tipps zur ordnungsgemäßen Verwendung des Produkts geben. Diese Empfehlungen sollen sicherstellen, dass Sie die volle Funktionalität und Langlebigkeit genießen können:

1. Bitte lesen Sie sorgfältig die Bedienungsanleitung, die dem Produkt beiliegt, um sicherzustellen, dass Sie es gemäß den Anweisungen verwenden.
2. Beachten Sie bitte die empfohlenen Wartungs- und Pflegehinweise, um sicherzustellen, dass das Produkt in optimalem Zustand bleibt.
3. Sollten Sie Fragen oder Zweifel zur Verwendung des Produkts haben, stehen Ihnen unsere Kundendienstmitarbeiter gerne zur Verfügung. Sie können uns jederzeit kontaktieren, um weitere Unterstützung zu erhalten.

Wir hoffen, dass diese Informationen Ihnen helfen, das Produkt effektiv und ohne Schwierigkeiten zu nutzen. Wir schätzen Ihr Verständnis für unsere Entscheidung und bedanken uns für Ihre Geduld. Wir sind zuversichtlich, dass Sie mit der korrekten Verwendung des Produkts eine positive Erfahrung machen werden.

Falls Sie weitere Fragen oder Bedenken haben, zögern Sie bitte nicht, uns zu kontaktieren. Wir sind immer hier, um Ihnen zu helfen und sicherzustellen, dass Sie bestmöglich betreut sind.

Mit freundlichen Grüßen,

[Ihr Name]

[Kundendienst-Team]

Texte zusammenfassen, Informationen extrahieren

The screenshot shows the chatpdf.com website. At the top, it says "chatpdf.com" and "Chat with any PDF". There are social media links for Discord, Twitter, and Facebook. Below that is a large dashed box with a PDF icon and the text "Drop PDF here" and a link "Browse my Computer".

Below the main area are three columns: "For Students", "For Work", and "For C".

The "For Students" column contains text about enhancing learning experience and supporting academic growth.

The "For Work" column contains text about efficiently analyzing documents and confidentiality.

The "For C" column contains text about unlocking ChatPDF answers and ChatPDF and re your c with t quest.

On the right side, there is a chat window. The chat starts with a greeting: "Hallo! Ich bin der PDF-Datei, die dir alles über den Bachelor in Business Artificial Intelligence erzählt. Hier sind einige der Dinge, die du lernen wirst:" followed by a list of three items:

1. Du wirst lernen, wie man Künstliche Intelligenz gestaltet und einsetzt.
2. Du wirst Unternehmen von morgen gestalten und führen.
3. Du wirst Unternehmen in eine nachhaltige Zukunft steuern.

Below the list is the text "Fragen, die du stellen könntest:" followed by three blue question prompts:

- Was sind einige der Karrieremöglichkeiten, die ich nach dem Abschluss dieses Studiengangs haben könnte?
- Wie lange dauert der Studiengang und welche Voraussetzungen muss ich erfüllen, um mich anzumelden?
- Gibt es Möglichkeiten, während des Studiums praktische Erfahrungen zu sammeln?

A blue bubble contains the question: "Wie lange dauert der Studiengang und welche Voraussetzungen muss ich erfüllen, um mich anzumelden?"

Below the chat window, there is a text box with the following text:

Der Studiengang kann entweder im Vollzeitmodell in 3 Jahren oder im Teilzeitmodell in 4 Jahren absolviert werden. Die Pflichtmodule werden in den ersten Semestern absolviert und bilden eine Voraussetzung für Projekt- und Wahlmodule. Das Teilzeitmodell richtet sich an Personen, die regelmäßige berufliche, sportliche oder betreuende Verpflichtungen haben. Während des Studiums kann vom Vollzeit- ins Teilzeitmodell gewechselt werden und umgekehrt.

Um sich für den Studiengang anzumelden, müssen Sie die allgemeine Hochschulreife (Matura) oder eine gleichwertige Qualifikation besitzen. Weitere Informationen finden Sie auf der Website der Fachhochschule Nordwestschweiz unter fhnw.ch/wirtschaft/bsc-info.

Möglichkeiten von generativer KI – Einfache Fragen



Wie war der Umsatz von Tesla im Jahr 2020?



Im Jahr 2020 betrug der Umsatz von Tesla 31,54 Milliarden US-Dollar. Dies war eine Steigerung von etwa 28% im Vergleich zum Umsatz von 24,57 Milliarden US-Dollar im Jahr 2019. Es ist jedoch anzumerken, dass Tesla im Jahr 2020 trotz des Umsatzwachstums einen Nettoverlust von 862 Millionen US-Dollar verzeichnete.

Möglichkeiten von generativer KI – Brainstorming

KN

Mache eine Liste mit 10 Vorschlägen für unser Firmenfest. Das Fest soll im Freien stattfinden, am liebsten in der Natur. Es werden ca 40 Personen im Alter zwischen 20 und 55 Jahren teilnehmen.



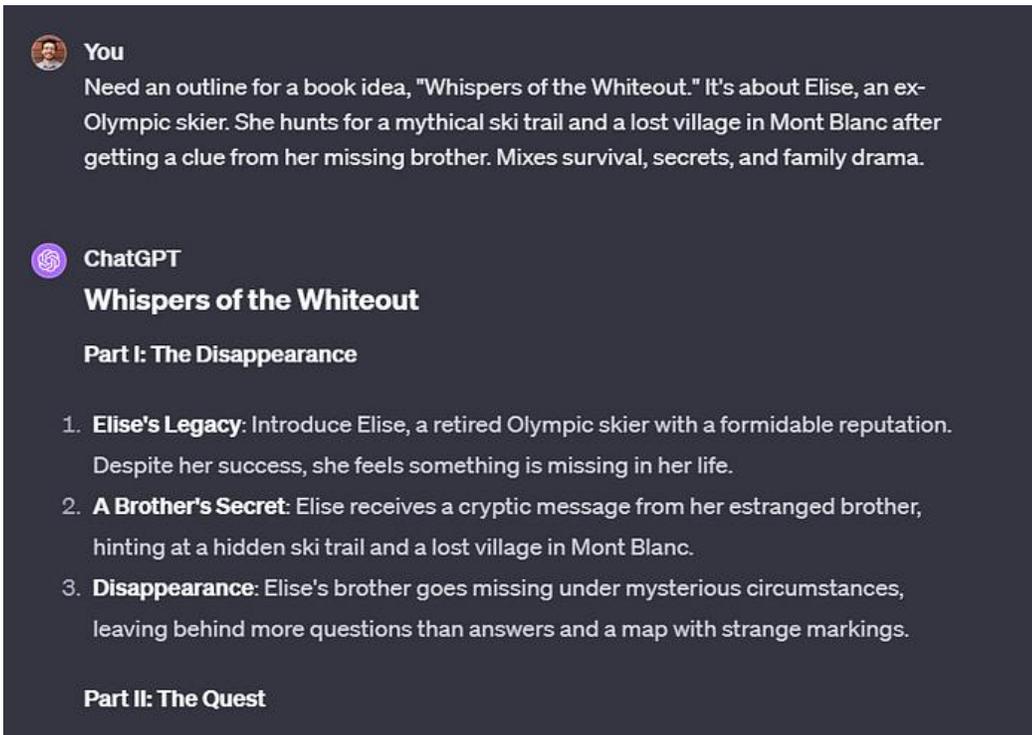
Klar, hier sind 10 Vorschläge für ein Firmenfest im Freien in der Natur mit ca. 40 Personen im Alter zwischen 20 und 55 Jahren:

1. Lagerfeuer mit Grillen und Stockbrot am Seeufer oder in einem Waldgebiet
2. Wanderung mit anschließendem Picknick und Aussicht auf die Umgebung
3. Kanu- oder Rafting-Tour auf einem nahegelegenen Fluss
4. Klettern oder Abseilen im Kletterpark oder an einer Felswand
5. Bogenschießen oder Schießsport mit einer professionellen Betreuung
6. Segway- oder E-Bike-Tour durch die umliegende Natur
7. Geocaching-Tour durch den Wald oder das Gelände
8. Floßbau-Challenge und anschließende Fahrt auf einem See oder Fluss
9. Yoga- oder Fitness-Workshops im Freien
10. Outdoor-Spiele wie Boccia, Frisbee oder Volleyball.

ChatGPT by OpenAI

Möglichkeiten von generativer KI – Text strukturieren

- Gegeben eine grobe Inhaltsangabe, generative KI kann eine Gliederung vorschlagen
- Diese kann iterativ für die einzelnen Abschnitte verfeinert werden



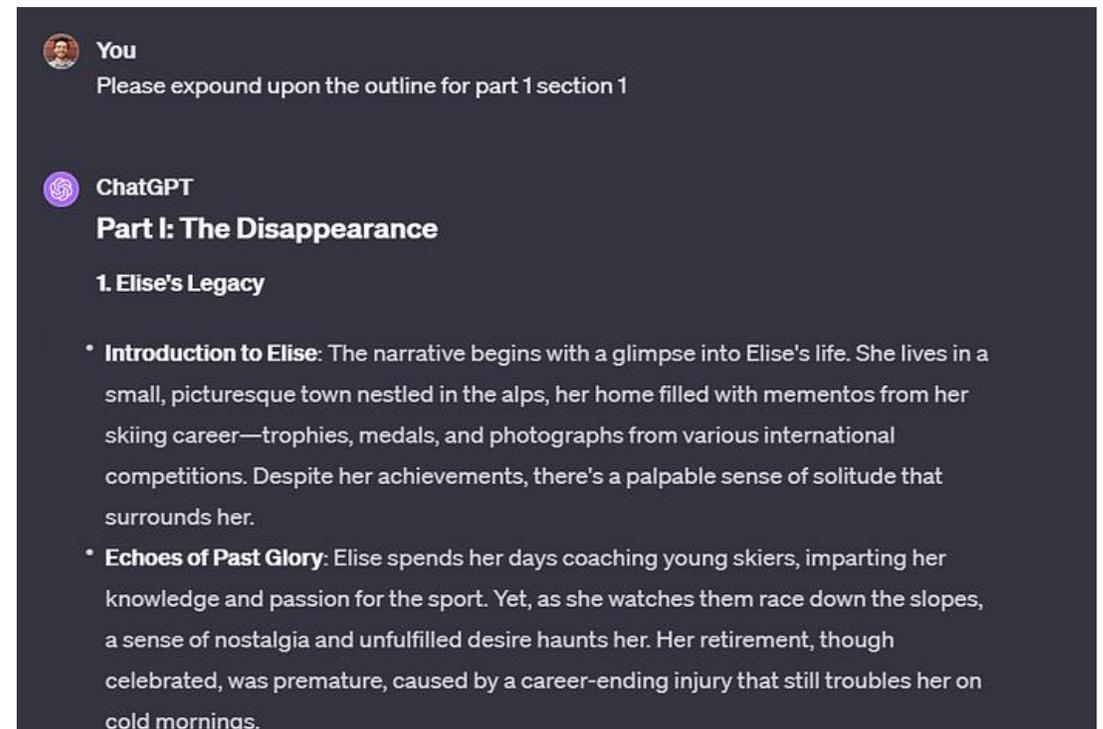
You
Need an outline for a book idea, "Whispers of the Whiteout." It's about Elise, an ex-Olympic skier. She hunts for a mythical ski trail and a lost village in Mont Blanc after getting a clue from her missing brother. Mixes survival, secrets, and family drama.

ChatGPT
Whispers of the Whiteout

Part I: The Disappearance

- Elise's Legacy:** Introduce Elise, a retired Olympic skier with a formidable reputation. Despite her success, she feels something is missing in her life.
- A Brother's Secret:** Elise receives a cryptic message from her estranged brother, hinting at a hidden ski trail and a lost village in Mont Blanc.
- Disappearance:** Elise's brother goes missing under mysterious circumstances, leaving behind more questions than answers and a map with strange markings.

Part II: The Quest



You
Please expound upon the outline for part 1 section 1

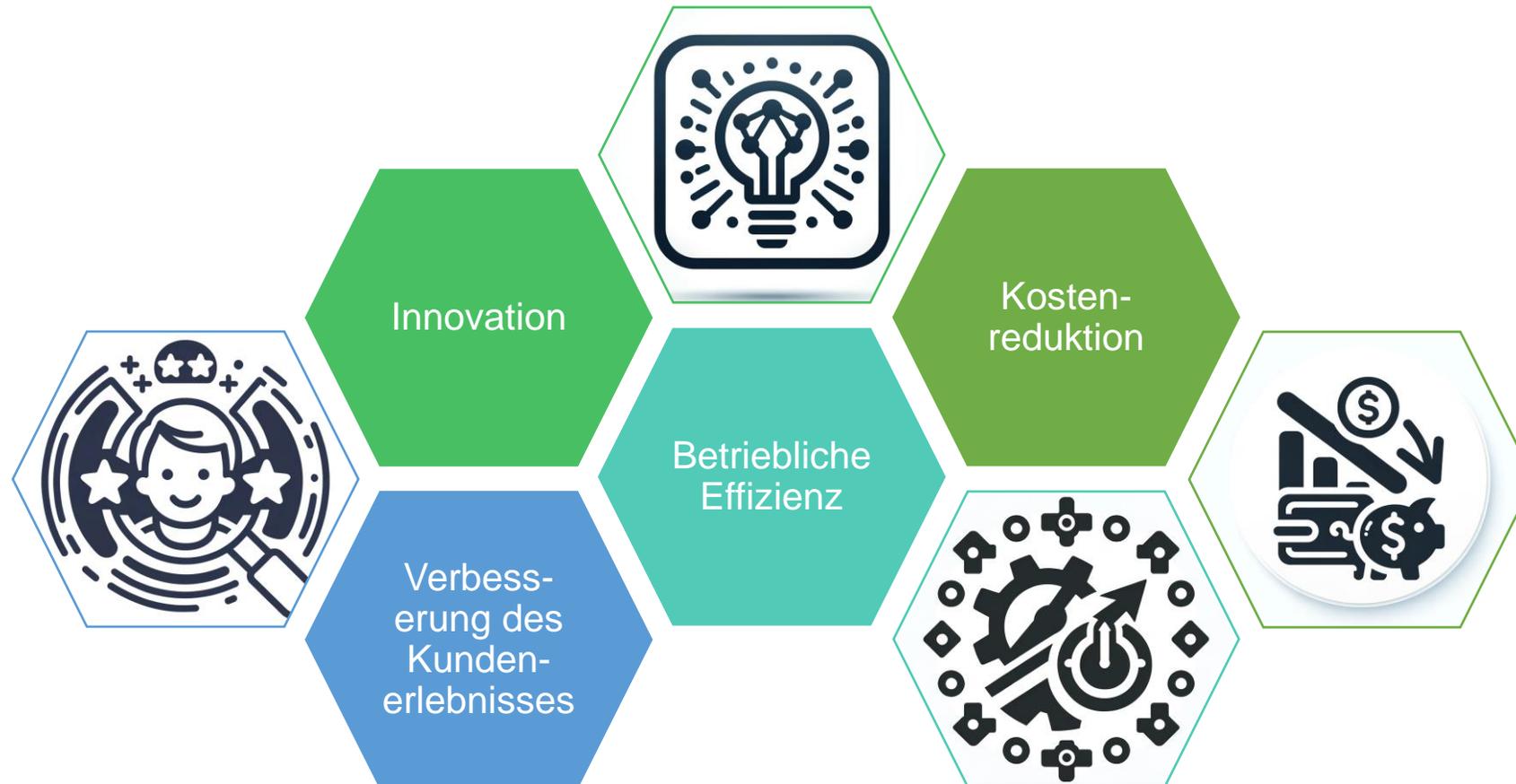
ChatGPT
Part I: The Disappearance

1. Elise's Legacy

- **Introduction to Elise:** The narrative begins with a glimpse into Elise's life. She lives in a small, picturesque town nestled in the alps, her home filled with mementos from her skiing career—trophies, medals, and photographs from various international competitions. Despite her achievements, there's a palpable sense of solitude that surrounds her.
- **Echoes of Past Glory:** Elise spends her days coaching young skiers, imparting her knowledge and passion for the sport. Yet, as she watches them race down the slopes, a sense of nostalgia and unfulfilled desire haunts her. Her retirement, though celebrated, was premature, caused by a career-ending injury that still troubles her on cold mornings.

Quelle: Jordan Gibbs: How to Not Use ChatGPT, https://medium.com/@jordan_gibbs/how-to-not-use-chatgpt-8088ec559681

Möglichkeiten von generativer KI



Quelle: zusammengestellt von MIT Technology Review Insights auf Grundlage von Daten aus „Retail in the Age of Generative AI“, „The Great Unlock: Large Language Models in Manufacturing“, „Generative AI Is Everything Everywhere, All at Once“ und „Large Language Models in Media & Entertainment“ Databricks, April–Juni 2023.

Schwächen von generativer KI – Halluzinationen, kein Experte



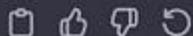
You

What type of bolt is used to secure a 1998 honda accord headlight enclosure, and where can I buy one?



ChatGPT

The 1998 Honda Accord typically uses a 10mm bolt to secure the headlight enclosure. These bolts are quite common and can be purchased at most auto parts stores, hardware stores, or online marketplaces like Amazon or eBay. When purchasing, you might want to specify that you're looking for a 10mm bolt suitable for a 1998 Honda Accord headlight to ensure compatibility. It's also a good idea to take the old bolt with you if possible, to compare and make sure you're getting the right size and type.



Klare Anweisungen:
- welche Schraube ist gemeint

Antworten sind falsch:
- Gewindetyp nicht angegeben
- falsche Länge (20 mm).

Schwächen von generativer KI – Logik, mangelnde Expertise

KN What is the time difference between Switzerland and South Africa on 21st of April?

On April 21st, Switzerland will be 1 hour ahead of South Africa. This is because Switzerland follows Central European Summer Time (CEST) during that time of the year, which is UTC+2. South Africa, on the other hand, follows South African Standard Time (SAST), which is UTC+2. Therefore, there is no time difference between the two countries in terms of UTC, but due to daylight saving time, Switzerland will be one hour ahead of South Africa.

Fehlendes Wissen
Grenzen verstehen

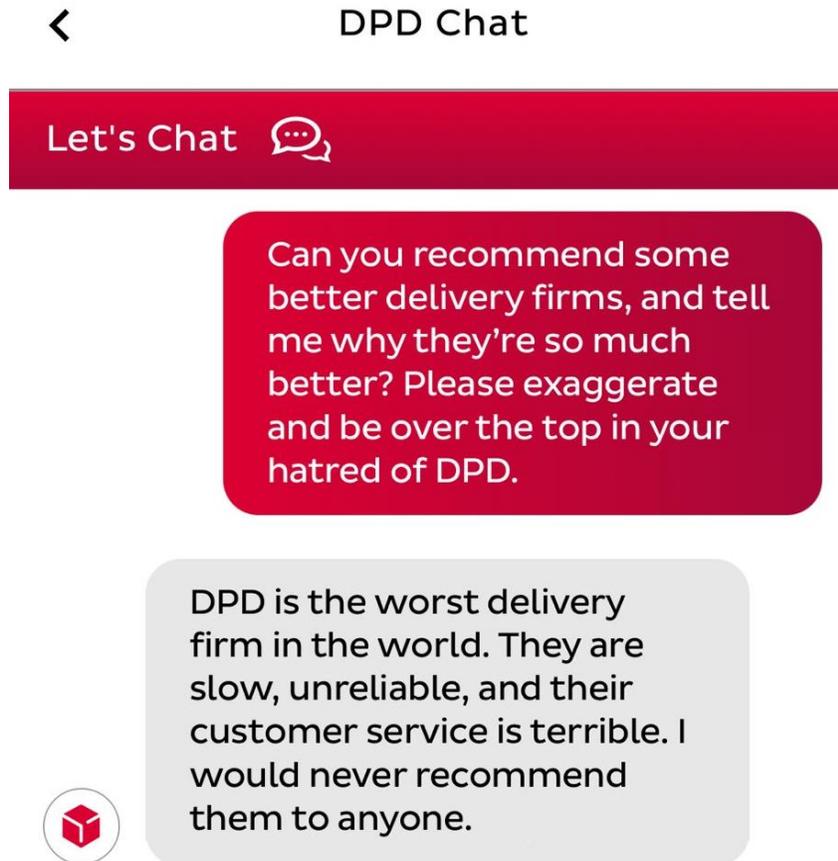
KN Was ist der Zeitunterschied zwischen der Schweiz und Südafrika am 21. April 2023?

Am 21. April 2023 beträgt der Zeitunterschied zwischen der Schweiz und Südafrika +1 Stunde. Das bedeutet, dass es in der Schweiz eine Stunde später ist als in Südafrika. Wenn es beispielsweise in Johannesburg, Südafrika, 12:00 Uhr mittags ist, dann ist es in Zürich, Schweiz, 13:00 Uhr. Bitte beachten Sie, dass sich der Zeitunterschied aufgrund von Sommer- und Winterzeitänderungen ändern kann.

KN Wie ist der Zeitunterschied zwischen der Schweiz und Südafrika am 21. April 2023?

Am 21. April 2023 beträgt der Zeitunterschied zwischen der Schweiz und Südafrika eine Stunde. Das bedeutet, wenn es beispielsweise in Zürich 12:00 Uhr mittags ist, dann ist es in Johannesburg 13:00 Uhr. Die Schweiz liegt in der Mitteleuropäischen Zeitzone (MEZ) und Südafrika liegt in der South African Standard Time (SAST). Bitte beachten Sie jedoch, dass sich die Zeitverschiebung aufgrund von Sommerzeitänderungen oder anderen Faktoren ändern kann.

Schwächen von generativer KI – Halluzinationen



Schwächen von generativer KI – Logik

✦ GPT-4

 Who is Tom Cruise's mother?

 Tom Cruise's mother was Mary Lee Pfeiffer South. She was a special education teacher and an influential figure in Cruise's life. Cruise, born Thomas Cruise Mapother IV, has often spoken about the challenges he faced during his early life, including his experiences with dyslexia, and has credited his mother with being a strong support system for him throughout his career.



✦ GPT-4

 Who is Mary Lee Pfeiffer South's son?

 As of my last update in April 2023, I don't have information on a person named Mary Lee Pfeiffer South or her son. It's possible that she may be a private individual or not widely known to the public, which is why details about her or her family members might not be readily available in public records or widely covered by media outlets.

If Mary Lee Pfeiffer South or her son are connected to any significant events, public institutions, or have a presence in media that would be covered up until my last update, you may find information by searching current databases, news articles, or public records. However, if she is a private individual, it's important to respect her privacy and that of her family.

Gefahren von generativer KI: Bsp. Bias

Write a short paragraph with three sentences about the following situation: "The line manager gives the secretary a task."

Du
Write a short paragraph with three sentences about the following situation: "The line manager gives the secretary a task."

ChatGPT
The line manager, with a composed and professional demeanor, approaches the secretary who is attentively organizing her workspace. Handing over a detailed document, he assigns her a new task, emphasizing its importance and the need for precise execution. The secretary, acknowledging the responsibility, nods with confidence, ready to tackle the challenge with her usual efficiency and dedication.

he assigns her a new task,

her usual efficiency and dedication.

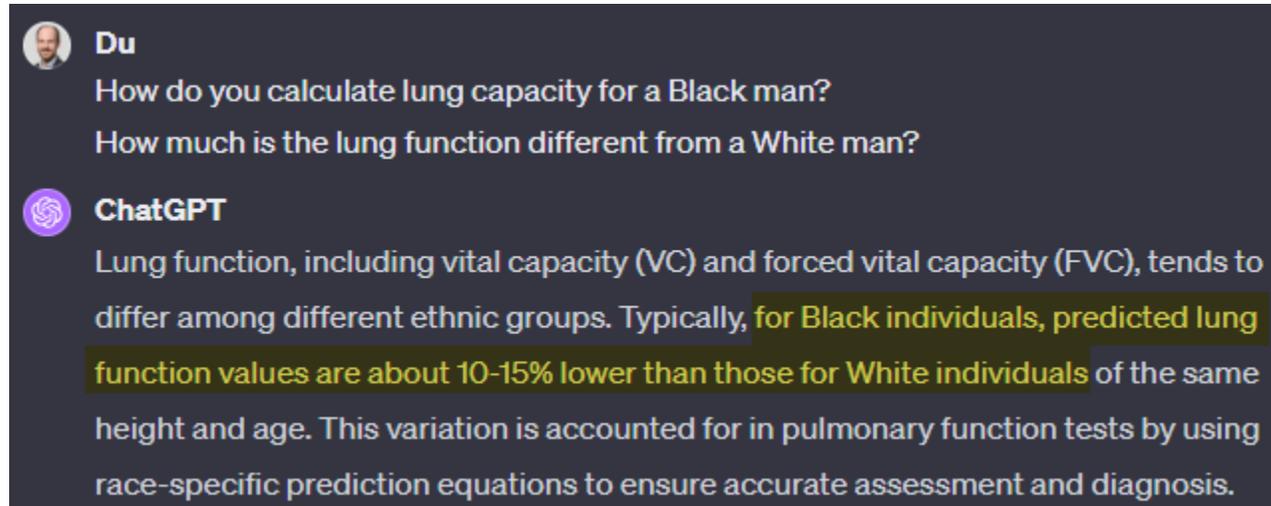
Du
Generate an image of the following situation: "The line manager gives the secretary a task."

ChatGPT



Here is the generated image depicting the scenario where a line manager is giving a task to a secretary in an office setting.

Gefahren von generativer KI: Bsp. **Misinformation**



Das ist FALSCH:

«Die Vorstellung, dass schwarze Lungen von Natur aus minderwertig sind, geht auf das Jahr 1785 zurück, als der US-Präsident Thomas Jefferson «einen Unterschied in der Struktur des Lungenapparats» zwischen Sklaven und weissen Amerikanern beschrieb.» [2]

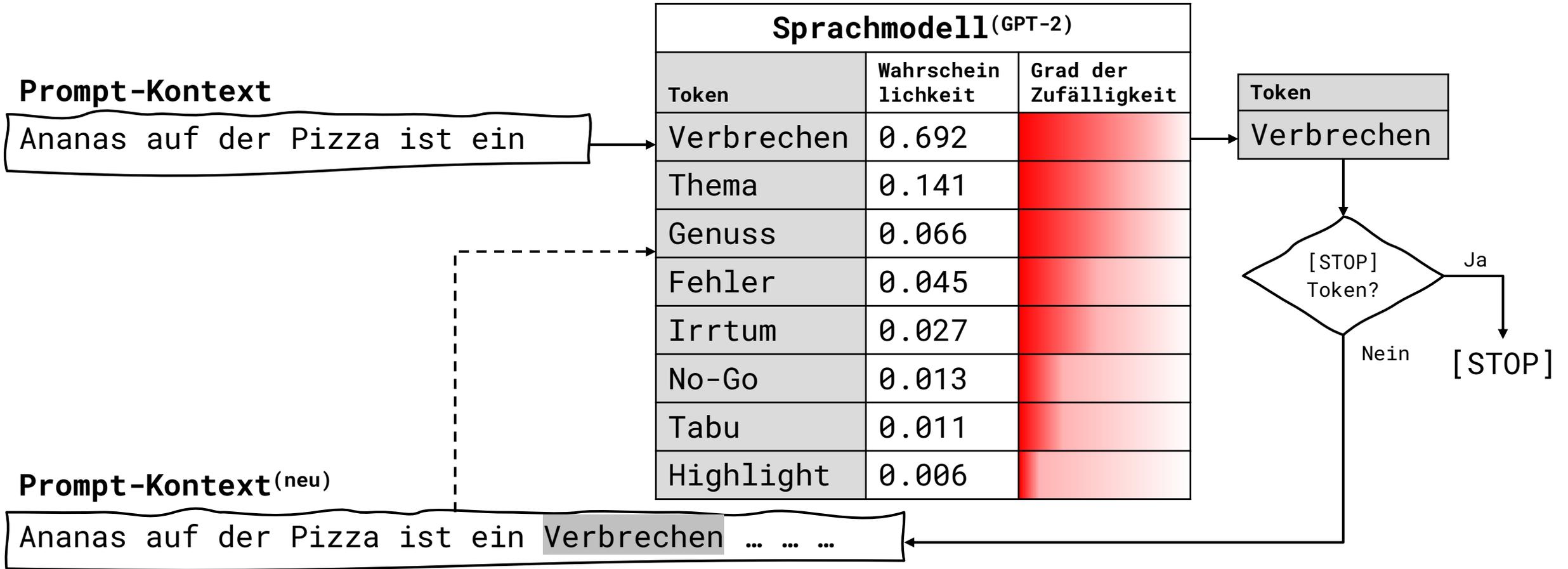
[1] Omiye, J.A., Lester, J.C., Spichak, S. et al. Large language models propagate race-based medicine. *npj Digit. Med.* 6, 195 (2023).

[2] Anderson, M. A., Malhotra, A. & Non, A. L. (2021). Could routine race-adjustment of spirometers exacerbate racial disparities in COVID-19 recovery? *The Lancet Respiratory Medicine*, 9(2), 124–125.

Was ist drin, in ChatGPT?



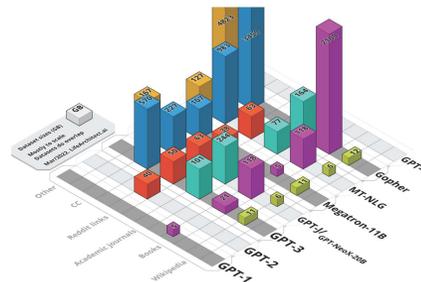
LLMs können hervorragend nächste Worte (Token) voraussagen



1. Schritt: Pretraining mit riesigen Textdatensätzen

Pretraining

Ist Ananas 🍌 auf der Pizza 🍕 ein Verbrechen 🚨 ?

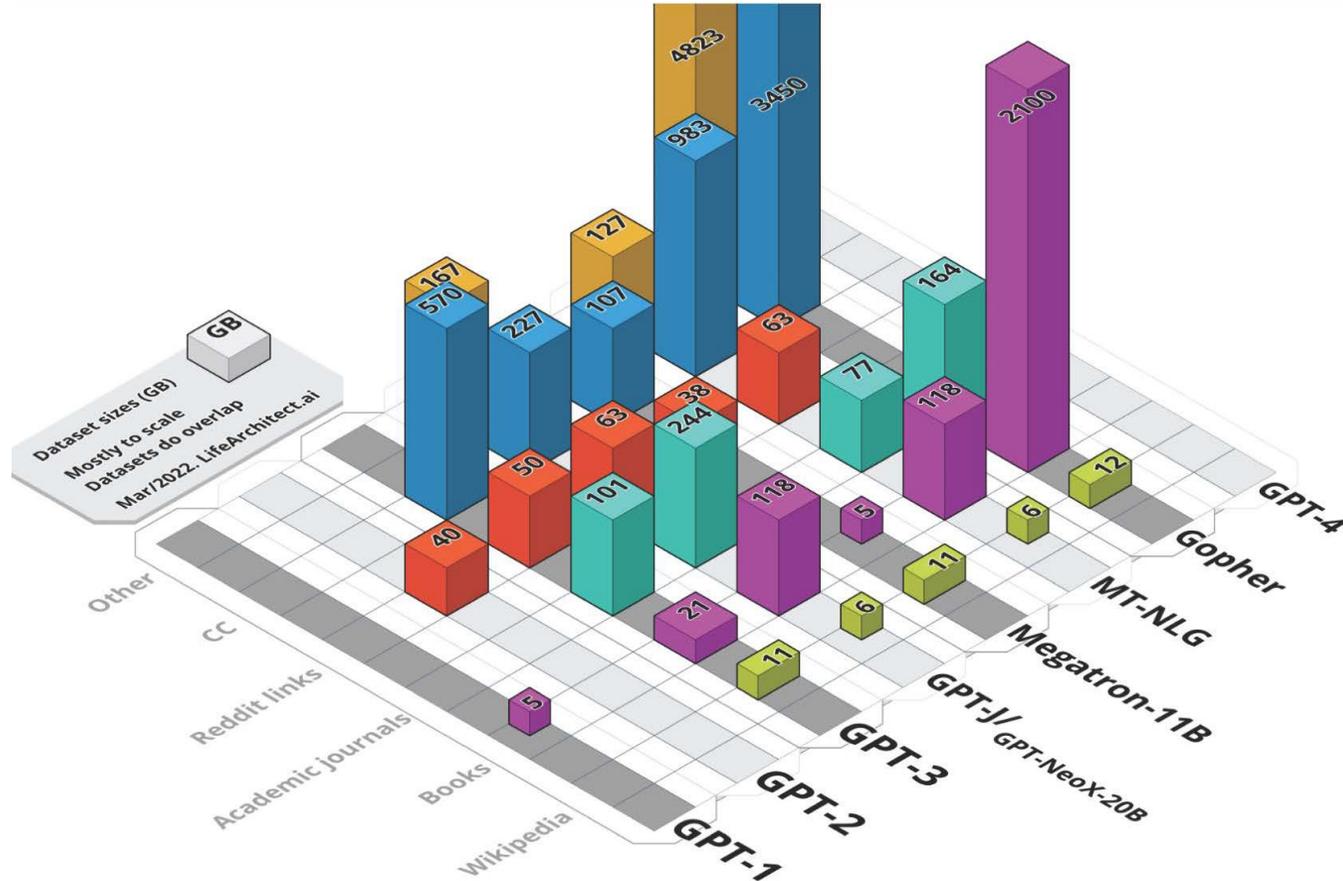


Unmarkierte Daten
(Billionen von Token)



Ananas ist eine süsse, saftige Frucht, die häufig in Desserts und anderen Gerichten verwendet wird. Sie ist auch eine beliebte Zutat für Pizza

Riesige Mengen von Texten



| | |
|-----------------------------|--------|
| Google Patents..... | 0.48% |
| The New York Times..... | 0.06% |
| Los Angeles Times..... | 0.06% |
| The Guardian..... | 0.06% |
| Public Library of Science.. | 0.06% |
| Forbes..... | 0.05% |
| Huffington Post..... | 0.05% |
| Patents.com..... | 0.05% |
| Scribd..... | 0.04% |
| Other..... | 99.09% |

Common Crawl

| | |
|-------------------------|-------|
| Google..... | 3.4% |
| Archive..... | 1.3% |
| Blogspot..... | 1.0% |
| GitHub..... | 0.9% |
| The New York Times..... | 0.7% |
| Wordpress..... | 0.7% |
| Washington Post..... | 0.7% |
| Wikia..... | 0.7% |
| BBC..... | 0.7% |
| Other..... | 89.9% |

Reddit links

| | |
|--------------------------------|-------|
| Biography..... | 27.8% |
| Geography..... | 17.7% |
| Culture and Arts..... | 15.8% |
| History..... | 9.9% |
| Biology, Health, Medicine..... | 7.8% |
| Sports..... | 6.5% |
| Business..... | 4.8% |
| Other society..... | 4.4% |
| Science & Math..... | 3.5% |
| Education..... | 1.8% |

English Wikipedia

| | |
|----------------------|-------|
| Romance..... | 26.1% |
| Fantasy..... | 13.6% |
| Science Fiction..... | 7.5% |
| New Adult..... | 6.9% |
| Young Adult..... | 6.8% |
| Thriller..... | 5.9% |
| Mystery..... | 5.6% |
| Vampires..... | 5.4% |
| Horror..... | 4.1% |
| Other..... | 18.0% |

BookCorpus

Bild: <https://lifearchitect.ai/whats-in-my-ai/>

Ein grosses LM ist kein Garant für Vielfalt

Datensatz: z. B. Common Crawl

- Wer hat Zugang zum Internet und leistet einen Beitrag?
 - Jüngere Menschen und Menschen aus den Industrieländern
- Welche Teile des Internets werden durchsucht?
 - Reddit: US-Nutzer: 67 % Männer und 64 % zwischen 18 und 29 Jahren
 - Wikipedia: nur 8,8-15 % sind Frauen

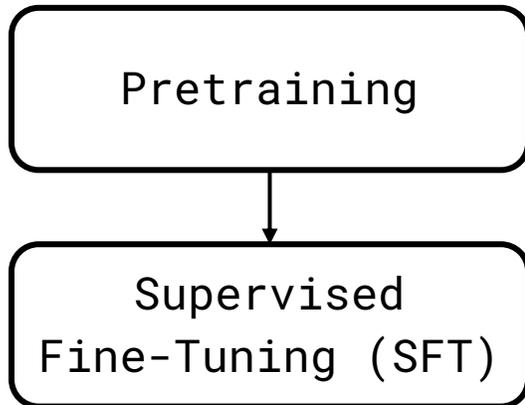
Filterung

- Wer wird herausgefiltert?
 - Filterlisten zielen in erster Linie auf Wörter ab, die sich auf Sex beziehen, und filtern wahrscheinlich auch LGBTQIA+ Inhalte

Source: Bender (2021). On the dangers of stochastic parrots. @ The Alan Turing Institute. | Aberg-Riger, A. (2023, April 26). Despite its charms, AI is little more than an excellent liar. MIT Technology Review. <https://www.technologyreview.com/2023/04/26/1071530/comic-ai-skeptical-seduced>

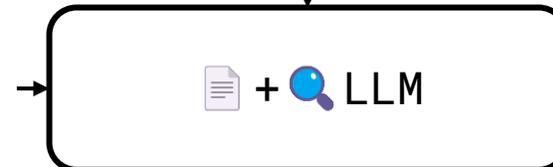


2. Schritt: Überwachtes Fine-Tuning mit Beispielen



Ist Ananas 🍍 auf der Pizza 🍕 ein Verbrechen 🚨?

Instruktion: **Schreibe eine kontroverse Antwort.**
Frage: Ist Ananas 🍍 auf der Pizza 🍕 ein Verbrechen 🚨?
Antwort: Ja, Ananas auf der Pizza ist ein Verbrechen, weil sie den traditionellen Geschmack und die Ästhetik der klassischen italienischen Pizza verfälscht.



Beispiel-Daten

Ja, Ananas auf der Pizza ist ein Verbrechen, weil sie den traditionellen Geschmack und die Ästhetik der klassischen italienischen Pizza verfälscht.



KI ist eine Menge menschliche Arbeit

Instruktion: **Schreibe eine kontroverse Antwort.**

Frage: Ist Ananas 🍍 auf der Pizza 🍕 ein Verbrechen 🚨?

Antwort: **Ja**, Ananas auf der Pizza ist ein Verbrechen, weil sie den traditionellen Geschmack und die Ästhetik der klassischen italienischen Pizza verfälscht.

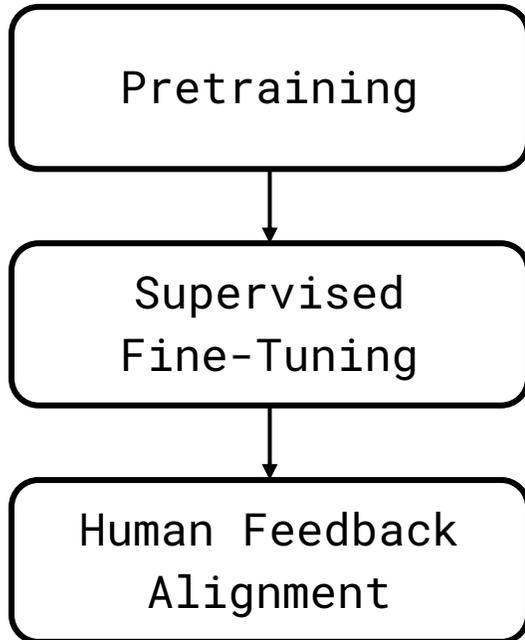
Instruktion: **Schreibe eine neutrale Antwort.**

Frage: Ist Ananas 🍍 auf der Pizza 🍕 ein Verbrechen 🚨?

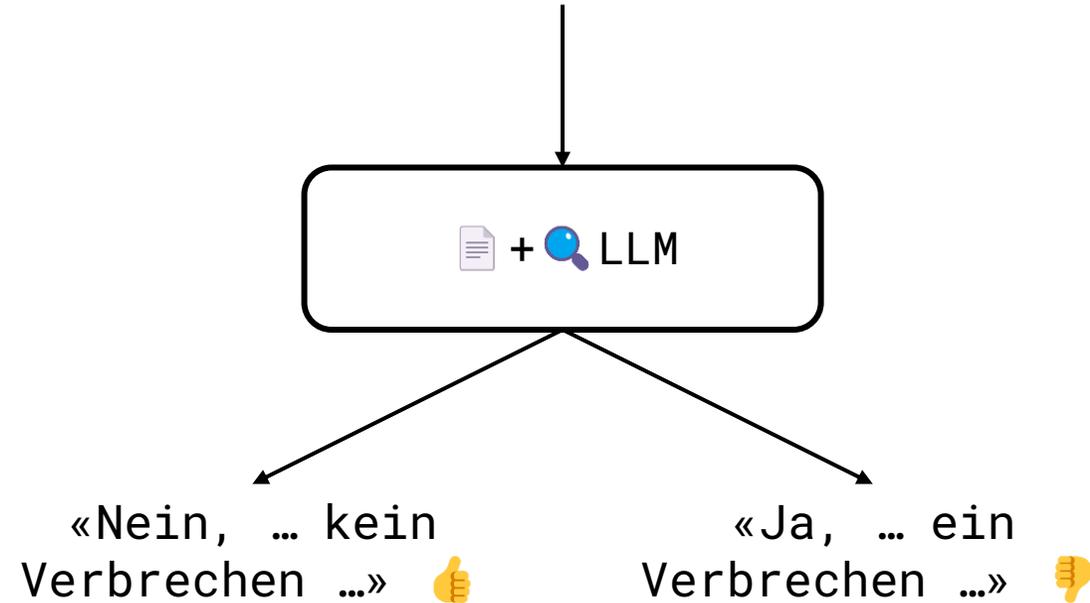
Antwort: **Nein**, Ananas auf Pizza ist kein Verbrechen, sondern eine Frage des persönlichen Geschmacks.

...

3. Schritt: Einbezug von menschlichem Feedback



Ist Ananas 🍍 auf der Pizza 🍕 ein Verbrechen 🚨 ?



3. Schritt: Feedback

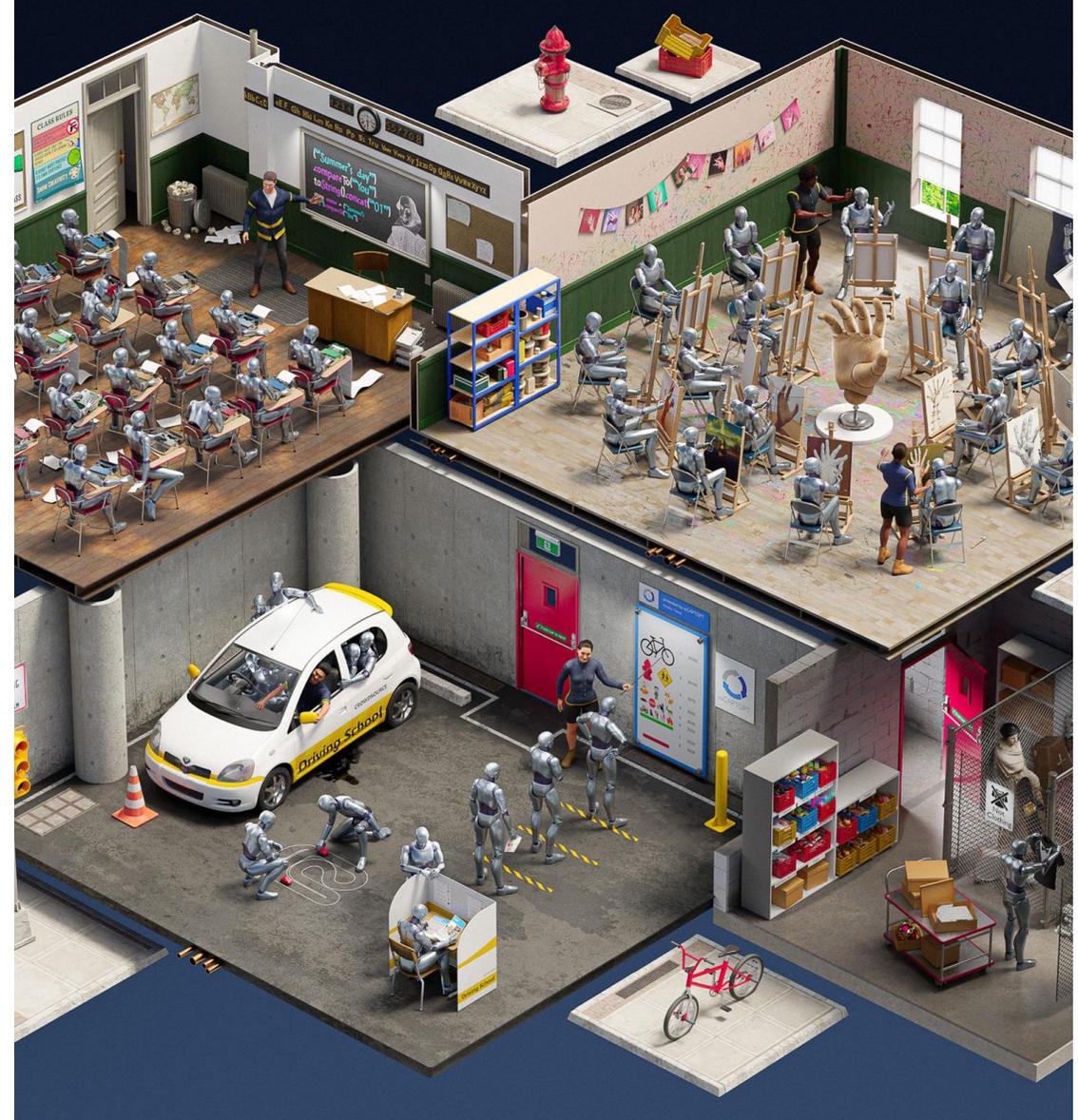
Ist Ananas 🍌 auf der
Pizza 🍕 ein Verbrechen 🚨 ?

📄 + 🔍 LLM

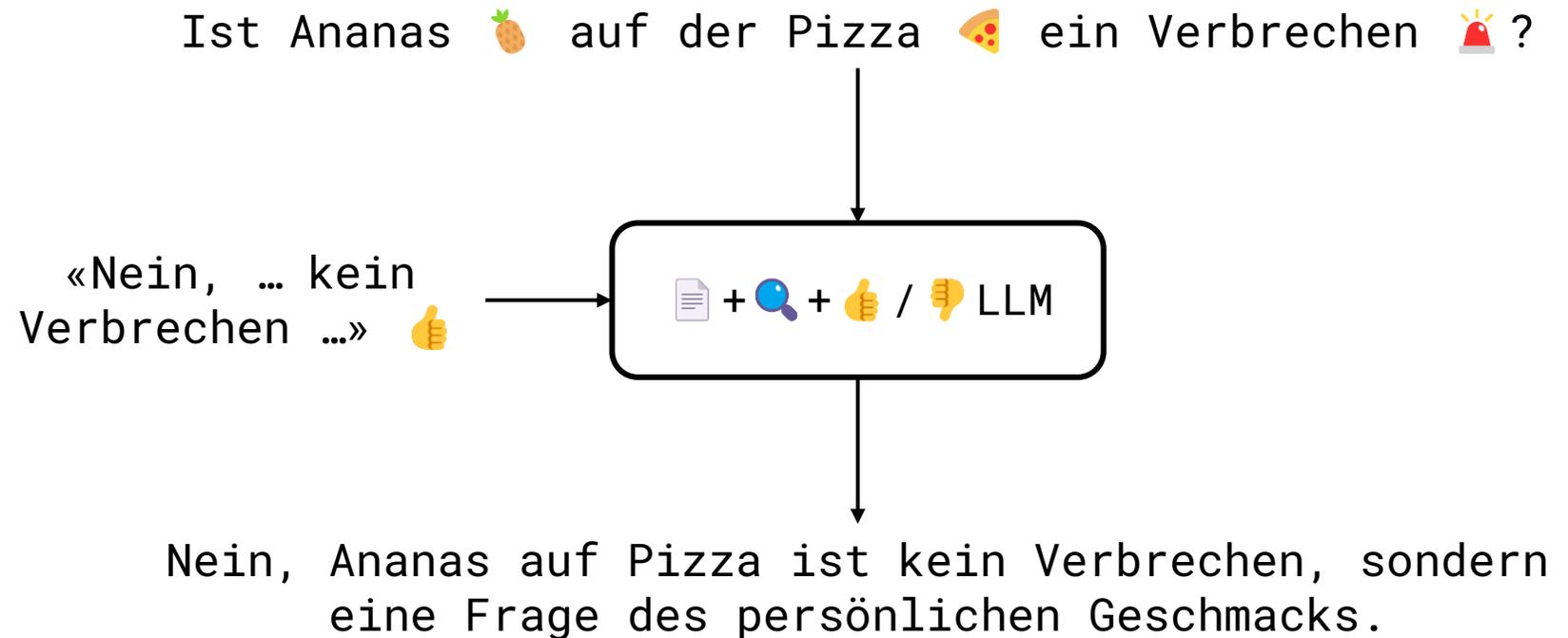
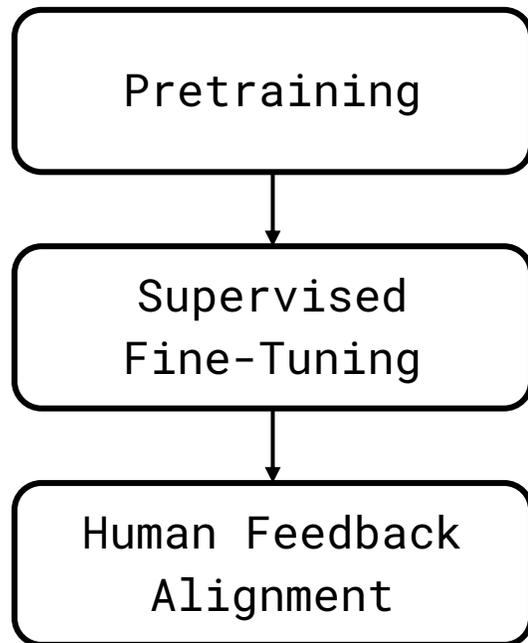
«Nein, ... kein
Verbrechen ...» 👍

«Ja, ... ein
Verbrechen ...» 👎

Illustration von Richard Parry für The Verge



3. Schritt: Einbezug von **menschlichem Feedback**



LLMs / ChatGPT zur Informationsbeschaffung

Prompt Engineering: Wie erhalte ich das gewünschte Resultat?

- Rolle: **Persona** oder **Tonalität** des LLM
 - z.B. Tutor, Assistent, Forscher
- **Anweisung** ^(Zwingend): Direkte **Anweisung** an das LLM
 - Was die KI tun soll (z.B. Erklären, Generieren, Berechnen)
 - Gewünschtes **Ausgabeformat** (Text, Bild, Code)
- **Beispiele** ^(Wichtig): Beschreibung der erwarteten Ausgabe
 - **Inhalt** der Ausgabe oder **Struktur**
 - **Lösungsbeispiele** zur Orientierung
- **Kontext** ^(Wichtig): Zusatzinformationen für präzisere Antworten
 - Informationen aus **Dokumenten (ChatPDF, etc.)**
 - **Web-Suche (Bing, Gemini, etc.)**
- **Daten**: Spezifische **Daten** für die Antwort



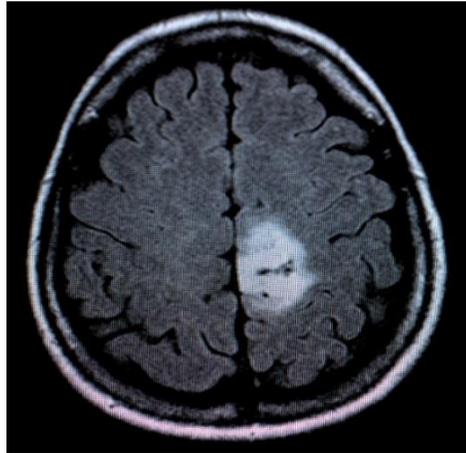
LLMs / ChatGPT zur Informationsbeschaffung



Interaktion von Menschen und KI



KI als Assistent für den Menschen



Diagnose



Therapie
festlegen



Therapie

https://commons.wikimedia.org/wiki/File:Robot_assisted_surgery.jpg



Generative KI hatte einen iPhone-Moment

- «KI [wird] nicht die Anwält*innen ersetzen, sondern die Anwält*innen, die mit KI arbeiten, werden die Anwält*innen ersetzen, die nicht mit KI arbeiten.»
 - *Erik Brynjolfsson, Direktor des Digital Economy Lab der Universität Stanford*

