

# Text generation with LLMs

Wie genau funktioniert ein LLM?

# Gliederung

- Terminologie
- Wie trainiert man ein LLM
- Wie funktionieren LLMs
- Übersetzung in andere Sprachen
- Chatbots
- Vergleich verschiedener Modelle
- Zukunft/Trends
- Quellen

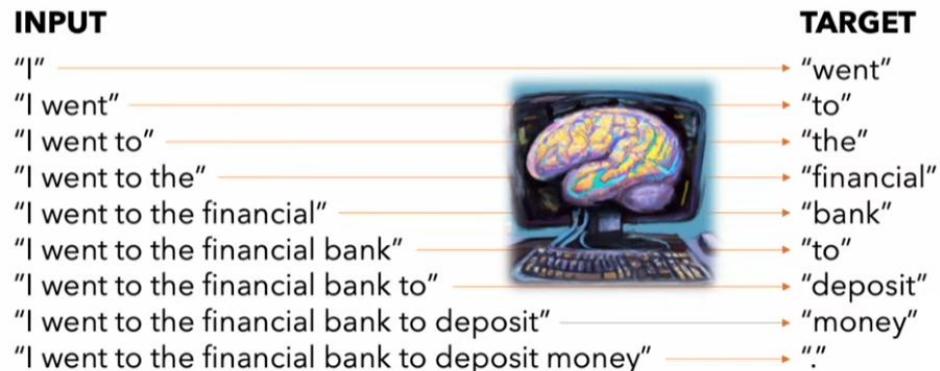
# Was ist NLP/LLM/GPT?

- **NLP = Natural Language Processing**
  - Alles, was Computer mit Sprache machen können
- **LLM = Large Language Model**
  - KI-Systeme, die menschliche Sprache verstehen und generieren können
- **GPT = Generative Pre-trained Transformer**
  - ein Typ von LLM, dass sich auf Sprachgenerierung fokussiert

# Generative Pre-trained

- Erzeugt eigenständig Text durch Vorhersage des nächsten Tokens
- Das Modell wurde vorab auf riesigen Textmengen trainiert
- Pre-Training besteht aus:
  - Milliarden Tokens aus Büchern, Websites, Artikeln usw.
- Beispiel:

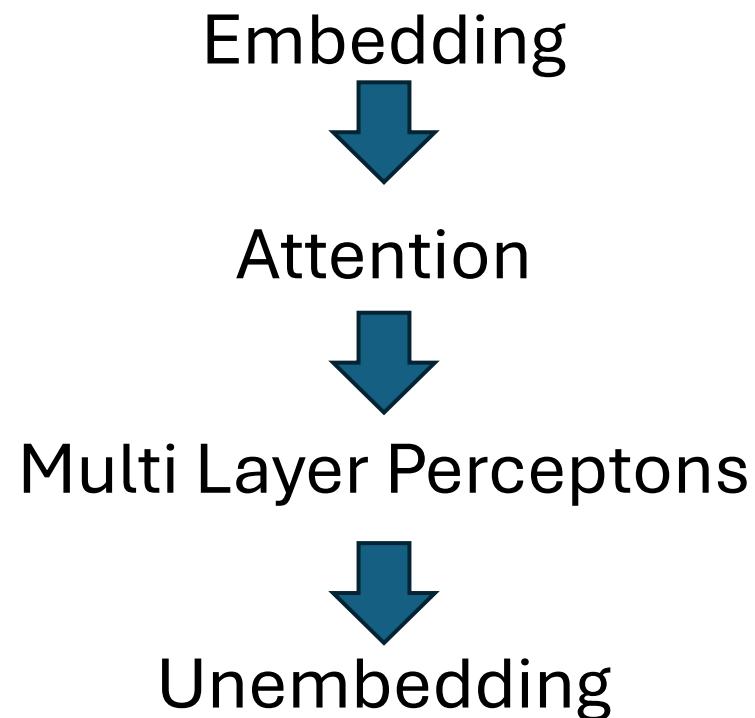
**JUST RANDOM TEXT FROM INTERNET**  
"I went to the financial bank to deposit money."



# Parameter Übersicht für Chat-GPT

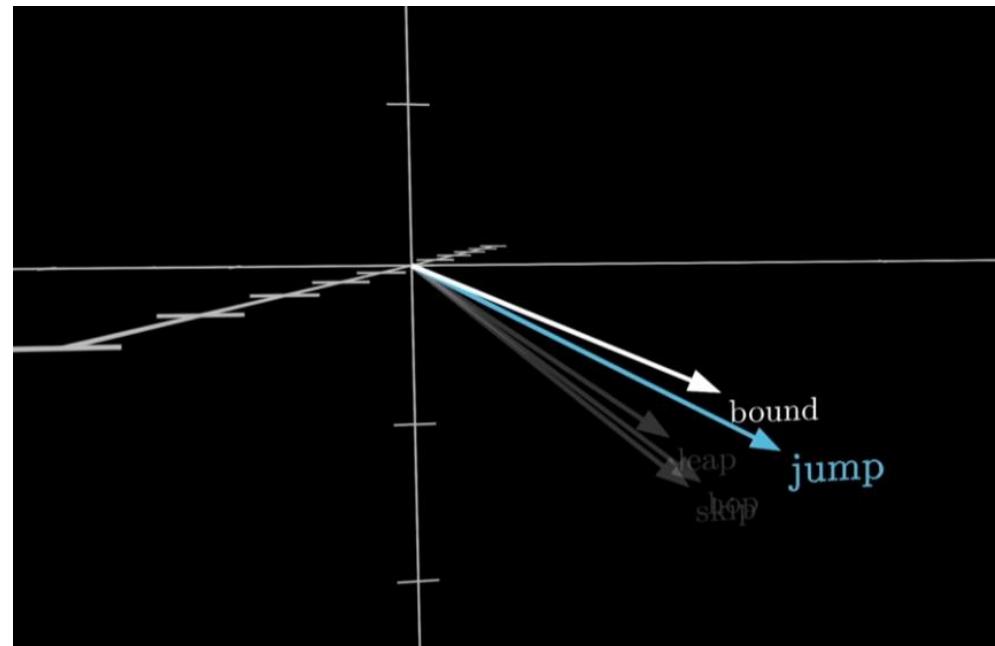
Modell	Architektur	Anzahl Parameter	Trainingsdaten	Veröffentlichungsdatum
GPT-1	12-stufiger 12-Kopf-Transformer-Decoder (kein Encoder), gefolgt von Linear-Softmax	117 Millionen	BookCorpus: 4,5 GB Text aus 7000 unveröffentlichten Büchern verschiedener Genres.	11. Juni 2018 <sup>[16]</sup>
GPT-2	GPT-1, jedoch mit modifizierter Normalisierung	1,5 Milliarden	WebText: 40 GB Text, 8 Millionen Dokumente, von 45 Millionen Webseiten, die auf Reddit positiv bewertet wurden.	erste/eingeschränkte Version: 14. Februar 2019 Vollversion: 5. November 2019 <sup>[18]</sup>
GPT-3	GPT-2, jedoch mit Modifikation, um eine größere Skalierung zu ermöglichen	175 Milliarden <sup>[20]</sup>	499 Milliarden Token, bestehend aus CommonCrawl (570 GB), WebText, englischer Wikipedia und zwei Buchkorpora (Books1 und Books2).	28. Mai 2020 <sup>[19]</sup>
GPT-3.5	nicht veröffentlicht		nicht veröffentlicht	15. März 2022
GPT-4	Auch mit Textvorhersage und RLHF trainiert; akzeptiert sowohl Text als auch Bilder als Eingabe. Weitere Details sind nicht öffentlich. <sup>[15]</sup>	nicht veröffentlicht	nicht veröffentlicht	14. März 2023
GPT-4o	unbekannt			13. Mai 2024 <sup>[21]</sup>
GPT-5	unbekannt			07. August 2025 <sup>[22]</sup>

# Transformer



# Transformer

- Input wird in mehrere Tokens aufgeteilt
- Embedding: Umwandlung von Tokens in numerische Vektoren



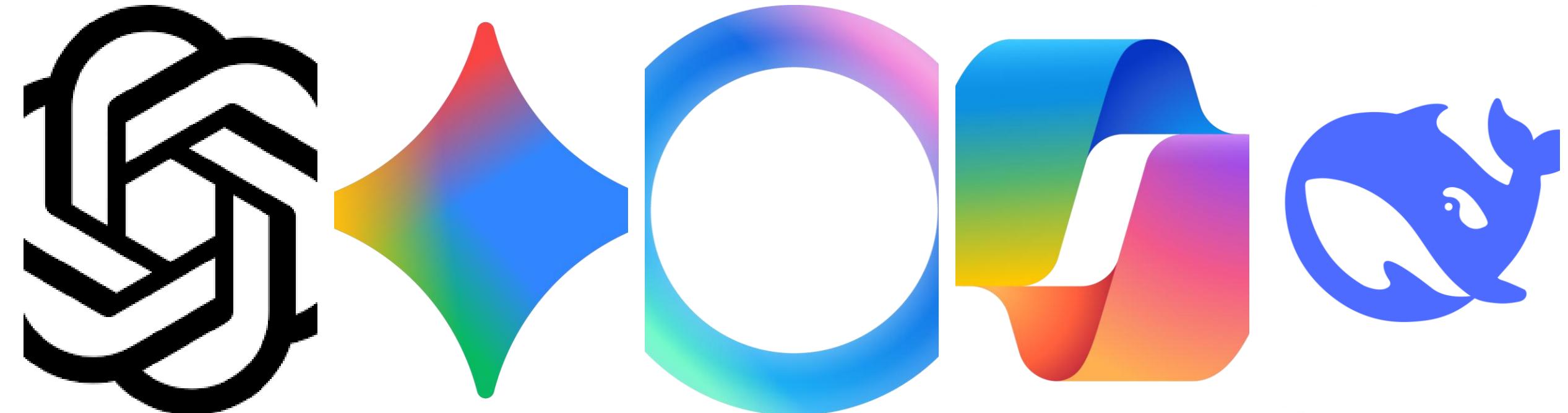
# Transformer

- Vektoren tauschen Informationen aus und aktualisieren ihre Werte
  - -> genaue Definition des Tokens wird erkannt
- Bedeutung der Tokens werden dann durch MLPs verfeinert durch nicht lineare Transformationen
- Nach mehreren Durchläufen entsteht ein finaler kontextreicher Vektor
  - -> anhand diesem wird das nachfolgende Wort vorhergesagt
- Rückumwandlung von Vektoren zu Tokens zu Sätzen

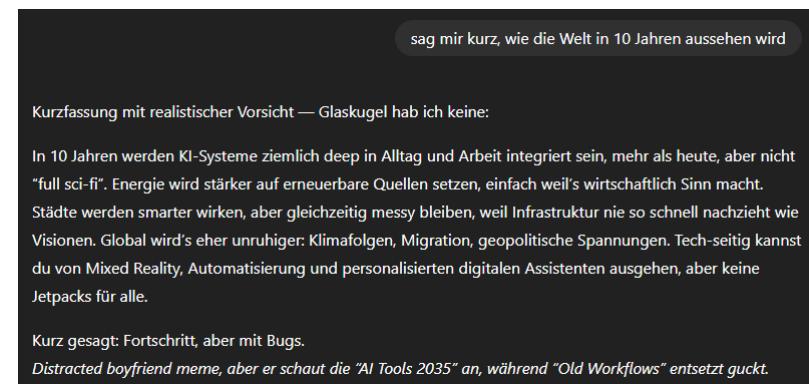
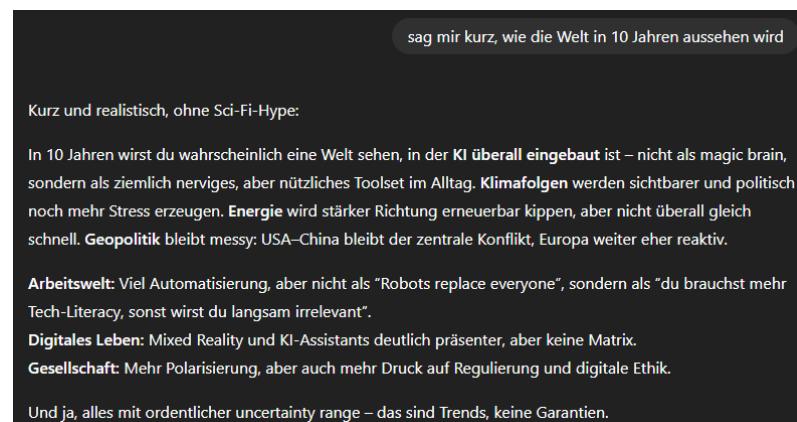
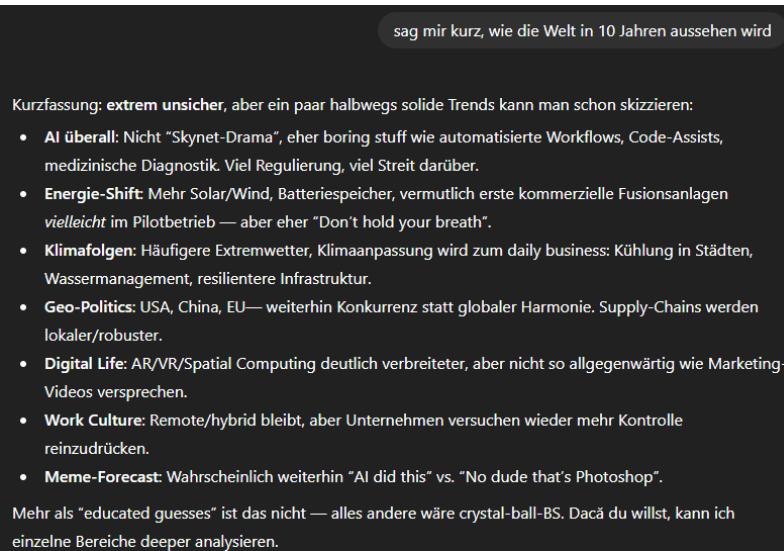
# Übersetzung in andere Sprachen

- Übersetzung findet nicht in Wörterbuchübersetzungen oder festen Regeln statt
- Nach Kontextualisierung werden die Tokens in ihrer Zielsprache neuformuliert
- Kurz: erst Bedeutung verstehen, dann formulieren

# Welche Chatbots gibt es?



# Vergleich (anhand der gleichen Frage an ChatGPT 5.1)



# Vergleich (anhand der gleichen Frage an verschiedene Chatbots)

GPT-4o

sag mir kurz, wie die Welt in 10 Jahren aussehen wird

Kurzfassung – Welt in 10 Jahren (2035):

- **Technologie:**  
KI wird deutlich präsenter sein – in Arbeit, Bildung, Medizin, Design. Deepfakes, autonome Systeme, personalisierte KI-Assistenten: Standard. Quantum Computing? Vielleicht erste echte Durchbrüche.
- **Klima & Umwelt:**  
Spürbare Klimafolgen (Hitze, Wasserknappheit, Extremwetter) werden Alltag. Mehr Klimamigration. Nachhaltigkeit ist nicht mehr nice-to-have, sondern Pflicht.
- **Geopolitik:**  
USA, China, Indien – multipolares Machtgefüge. Cyberwar & Informationskontrolle sind zentrale Konfliktzonen. Demokratien vs. autoritäre Systeme – Spannungen bleiben.
- **Gesellschaft:**  
Gesellschaften werden gespaltener oder resilenter – je nachdem, wie gut sie mit Tech, Klima & Ungleichheit umgehen. Polarisierung, aber auch Gegenbewegungen.
- **Arbeit & Wirtschaft:**  
Viele Jobs wandeln sich oder verschwinden durch Automatisierung. Neue Jobs entstehen – aber nicht überall gleichzeitig. Gig-Economy, Remote, Creator-Economy wachsen.
- **Space & Sci-Fi-Stuff:**  
Marslandung? Vielleicht. Mondbasis? Möglich. Aber wahrscheinlich eher Infrastruktur für Orbit-Wirtschaft, Weltraum-Internet & Satelliten-Netzwerke.

Fazit:  
Es wird nicht utopisch, aber auch nicht zwangsläufig dystopisch. Es wird komplex. Wer sich gut anpasst, bleibt relevant – wer nicht, wird abgehängt.

Gemini

Die Welt im Jahr 2035: Ein kurzer Ausblick

Obwohl Prognosen immer unsicher sind, zeichnen sich für das nächste Jahrzehnt folgende Megatrends ab:

1. Technologie & KI
  - **Allgegenwärtige KI:** Künstliche Intelligenz wird nicht mehr als "neu" wahrgenommen, sondern ist unsichtbare Infrastruktur (wie Strom heute). Sie assistiert bei fast jeder Büroarbeit, im kreativen Bereich und in der Programmierung.
  - **Automatisierung:** Roboter übernehmen vermehrt physische Aufgaben in der Pflege, Logistik und Landwirtschaft.
2. Umwelt & Energie
  - **Klimaanpassung:** Der Fokus verschiebt sich von der reinen Vermeidung (Klimaschutz) hin zur Anpassung (Hitzeresistente Städte, Deiche, neue Agrarsorten), da das 1,5-Grad-Ziel voraussichtlich überschritten wird.
  - **Energiewende:** Erneuerbare Energien sind in den meisten Teilen der Welt die günstigste Stromquelle; Batterietechnologien haben sich massiv verbessert.
3. Gesellschaft & Gesundheit
  - **Medizin:** Personalisierte Medizin auf Basis von Genanalysen und mRNA-Technologie wird Standard. Die Lebenserwartung steigt weiter, was demografische Probleme (alternde Gesellschaften) im Westen und Teilen Asiens verschärft.
  - **Arbeitswelt:** Die 4-Tage-Woche könnte sich in vielen Branchen durchsetzen. "Lebenslanges Lernen" ist keine Floskel mehr, sondern Notwendigkeit, da sich Berufsbilder alle paar Jahre radikal wandeln.
4. Geopolitik
  - **Multipolarität:** Die Dominanz einzelner Supermächte schwindet zugunsten regionaler Machtzentren (z.B. Indien, EU, China, USA).

GPT-5.1

sag mir kurz, wie die Welt in 10 Jahren aussehen wird

Kurzfassung mit realistischer Vorsicht — Glaskugel hab ich keine:

In 10 Jahren werden KI-Systeme ziemlich deep in Alltag und Arbeit integriert sein, mehr als heute, aber nicht "full sci-fi". Energie wird stärker auf erneuerbare Quellen setzen, einfach weil's wirtschaftlich Sinn macht. Städte werden smarter wirken, aber gleichzeitig messy bleiben, weil Infrastruktur nie so schnell nachzieht wie Visionen. Global wird's eher unruhiger: Klimafolgen, Migration, geopolitische Spannungen. Tech-seitig kannst du von Mixed Reality, Automatisierung und personalisierten digitalen Assistenten ausgehen, aber keine Jetpacks für alle.

Kurz gesagt: Fortschritt, aber mit Bugs.  
*Distracted boyfriend meme, aber er schaut die "AI Tools 2035" an, während "Old Workflows" entsetzt guckt.*

# Zukunft/Trends

- Multimodale Modelle
  - Text, Bild und Audio
- Persönlichere Chatbots
  - Emotionen, Tonfall, Stil
- Spezialisierte Modelle
  - In z.B. Medizin, Recht oder Bildung

# Quellen

- <https://github.com/steven2358/awesome-generative-ai>
- <https://github.com/Hannibal046/Awesome-LLM>
- <https://www.youtube.com/watch?v=wjZofJX0v4M>
- <https://www.youtube.com/watch?v=RVFj88i63rU&list=PLXV9Vh2jYcjnv67sXNDJiO8MWLA3ZJKR&index=4>
- <https://azure.microsoft.com/de-de/resources/cloud-computing-dictionary/what-are-large-language-models-llms>
- [https://de.wikipedia.org/wiki/Generativer\\_vortrainierter\\_Transformer](https://de.wikipedia.org/wiki/Generativer_vortrainierter_Transformer)
- <https://arxiv.org/pdf/2510.10991>
- <https://arxiv.org/pdf/2203.02155>