

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 = **N**atural **L**anguage **P**rocessing
 - Alles, was Computer mit Sprache machen können
- LLM = **L**arge **L**anguage **M**odel
 - KI-Systeme, die menschliche Sprache verstehen und generieren können
- GPT = **G**enerative **P**re-trained **T**ransformer
 - 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."

INPUT

"I"
"I went"
"I went to"
"I went to the"
"I went to the financial"
"I went to the financial bank"
"I went to the financial bank to"
"I went to the financial bank to deposit"
"I went to the financial bank to deposit money"



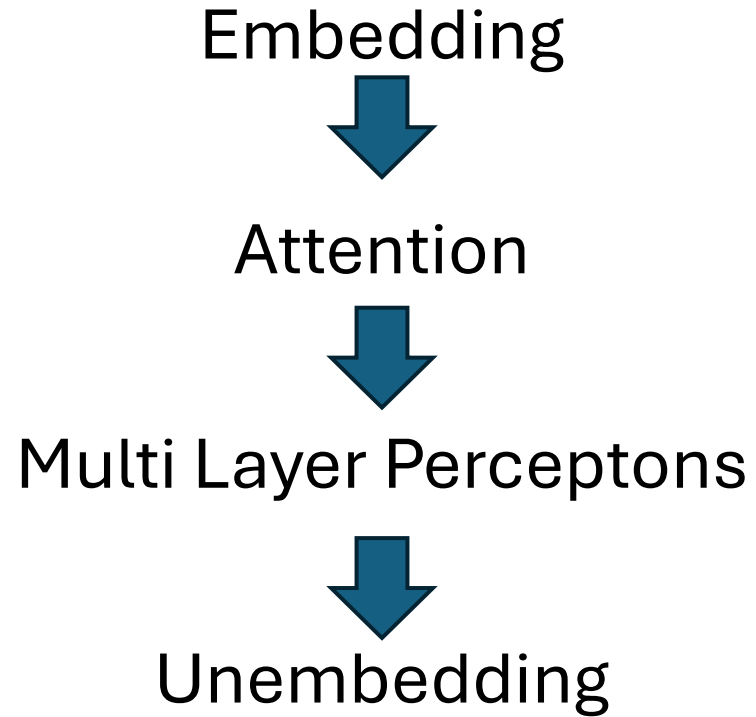
TARGET

"went"
"to"
"the"
"financial"
"bank"
"to"
"deposit"
"money"
."

Parameter Übersicht für Chat-GPT

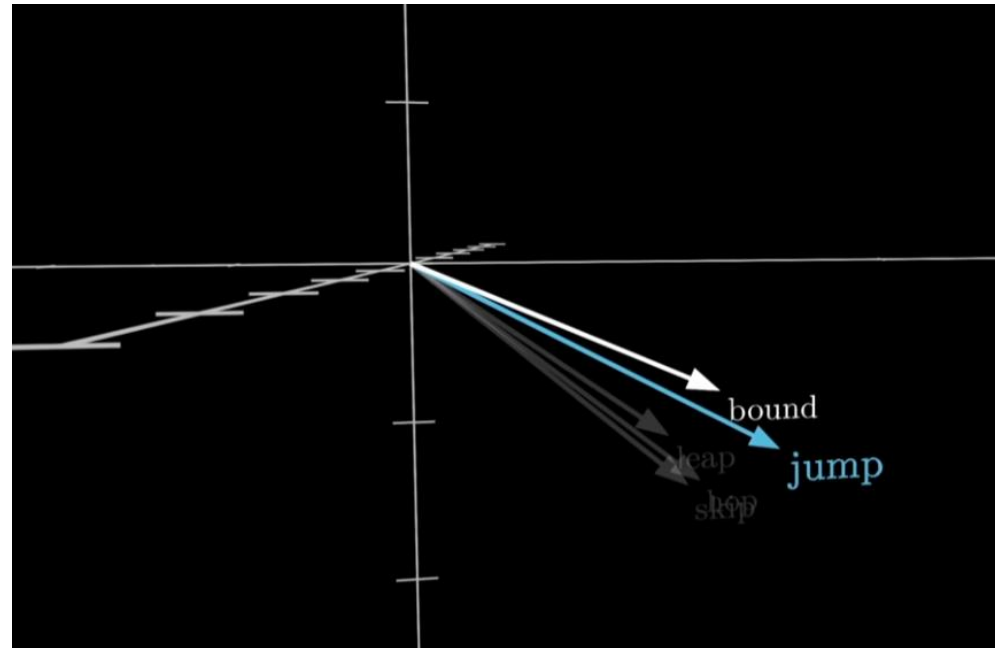
Modell	Architektur	Anzahl Parameter	Trainingsdaten	Veröffentlichungsdatum
GPT-1	12-stufiger 12-Kopf-Transformer-Decoder (kein Encoder), gefolgt von Linear-Softmax	117 Millionen	BookCorpus: 4,5 GB Text aus 7000 unveröffentlichten Büchern verschiedener Genres.	11. Juni 2018 ^[16]
GPT-2	GPT-1, jedoch mit 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 ^[18]
GPT-3	GPT-2, jedoch mit Modifikation, um eine größere Skalierung zu ermöglichen	175 Milliarden ^[20]	499 Milliarden Token, bestehend aus CommonCrawl (570 GB), WebText, englischer Wikipedia und zwei Buchkorpora (Books1 und Books2).	28. Mai 2020 ^[19]
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. ^[15]	nicht veröffentlicht	nicht veröffentlicht	14. März 2023
GPT-4o	unbekannt			13. Mai 2024 ^[21]
GPT-5				07. August 2025 ^[22]

Transformer



Transformer

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



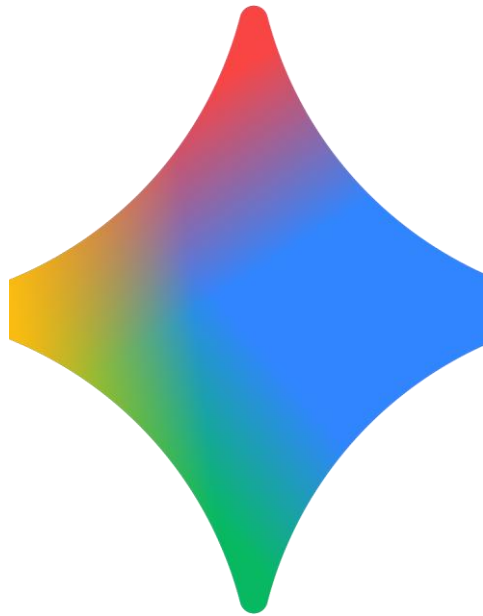
Transformer

- 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)

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

Kurzfassung: **extrem unsicher**, aber ein paar halbwegs solide Trends kann man schon skizzieren:

- **AI überall:** Nicht "Skynet-Drama", eher boring stuff wie automatisierte Workflows, Code-Assists, medizinische Diagnostik. Viel Regulierung, viel Streit darüber.
- **Energie-Shift:** Mehr Solar/Wind, Batteriespeicher, vermutlich erste kommerzielle Fusionsanlagen *vielleicht* im Pilotbetrieb — aber eher "Don't hold your breath".
- **Klimafolgen:** Häufigere Extremwetter, Klimaanpassung wird zum daily business: Kühlung in Städten, Wassermanagement, resilientere Infrastruktur.
- **Geo-Politics:** USA, China, EU — weiterhin Konkurrenz statt globaler Harmonie. Supply-Chains werden lokaler/robuster.
- **Digital Life:** AR/VR/Spatial Computing deutlich verbreiteter, aber nicht so allgegenwärtig wie Marketing-Videos versprechen.
- **Work Culture:** Remote/hybrid bleibt, aber Unternehmen versuchen wieder mehr Kontrolle reinzudrücken.
- **Meme-Forecast:** Wahrscheinlich weiterhin "AI did this" vs. "No dude that's Photoshop".

Mehr als "educated guesses" ist das nicht — alles andere wäre crystal-ball-BS. Dacă du willst, kann ich einzelne Bereiche deeper analysieren.

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

Kurz und realistisch, ohne Sci-Fi-Hype:

In 10 Jahren wirst du wahrscheinlich eine Welt sehen, in der **KI überall eingebaut** ist – nicht als magic brain, sondern als ziemlich nerviges, aber nützliches Toolset im Alltag. **Klimafolgen** werden sichtbarer und politisch noch mehr Stress erzeugen. **Energie** wird stärker Richtung erneuerbar kippen, aber nicht überall gleich schnell. **Geopolitik** bleibt messy: USA–China bleibt der zentrale Konflikt, Europa weiter eher reaktiv.

Arbeitswelt: Viel Automatisierung, aber nicht als "Robots replace everyone", sondern als "du brauchst mehr Tech-Literacy, sonst wirst du langsam irrelevant".

Digitales Leben: Mixed Reality und KI-Assistants deutlich präsenter, aber keine Matrix.

Gesellschaft: Mehr Polarisierung, aber auch mehr Druck auf Regulierung und digitale Ethik.

Und ja, alles mit ordentlicher uncertainty range – das sind Trends, keine Garantien.

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.

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 resilienter – 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=PLXV9Vh2jYcjbnv67sXNDJiO8MWLA3ZJKR&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://arxiv.org/pdf/2510.10991>
- <https://arxiv.org/pdf/2203.02155>