Wie konkurrenzfähig sind OpenLLMs?

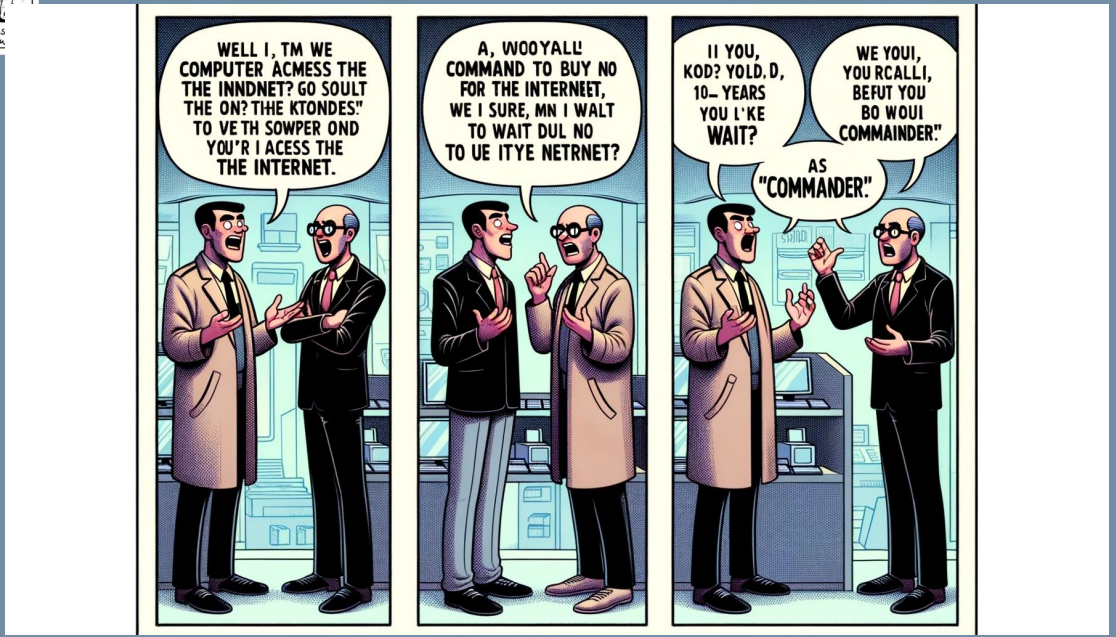
Vorweg: Computerlinguistik anno 2002

Comicinterpretation GPT4 (2023)

Der Comic beleuchtet die komischen Missverständnisse, die entstehen können, wenn jemand, der wenig bis keine Erfahrung mit moderner Technologie hat, versucht, sie zu nutzen. In der ersten Szene gibt der erste Mann vor, sich mit Computern auszukennen, wahrscheinlich um kompetent oder modern zu wirken. Im zweiten Bild wird sein Mangel an technischem Wissen offensichtlich, als er versucht, mit dem Computer zu sprechen, als wäre es ein menschliches Wesen oder ein Gerät aus einem Science-Fiction-Film. Das dritte Bild liefert die Pointe: Der Verkäufer scherzt, dass es noch Jahrzehnte dauern könnte, bis die Technologie auf dem Niveau ist, dass sie Sprachbefehle in der Art, wie der erste Mann es sich vorstellt, versteht. Der Begriff "Commander" könnte sich auf Star Trek oder ähnliche Sci-Fi-Serien beziehen, in denen Charaktere mit Computern sprechen.

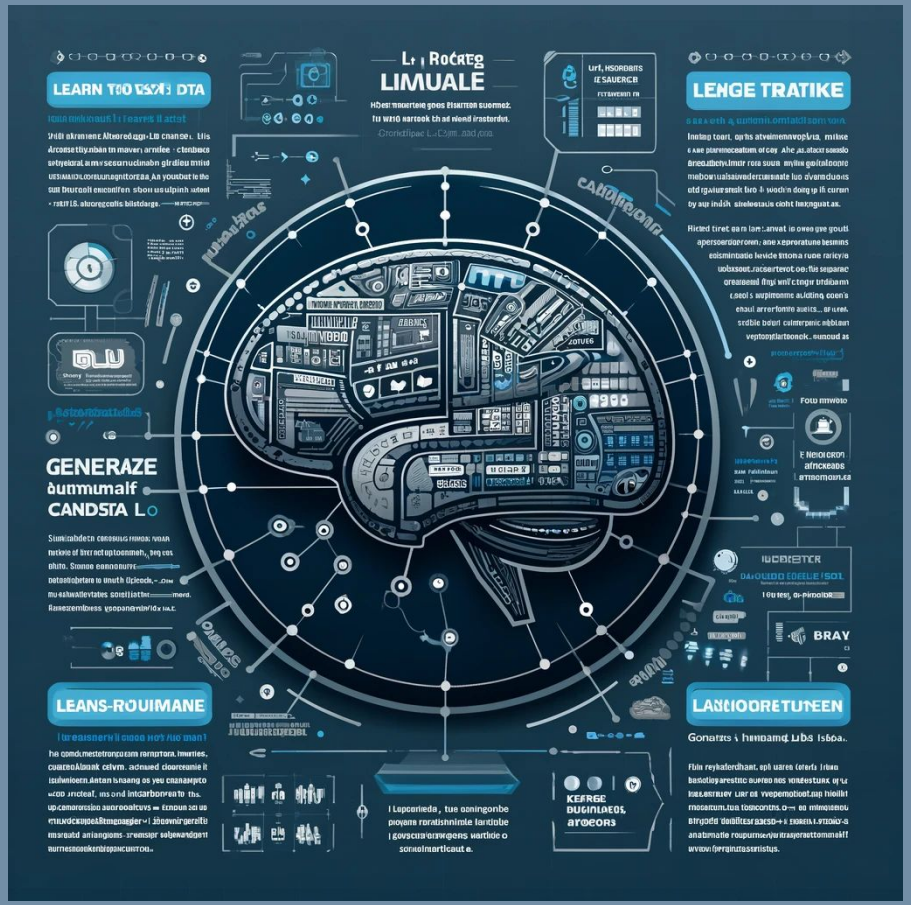
Insgesamt ist der Comic eine humorvolle Darstellung der Kluft zwischen den Erwartungen der Menschen und der tatsächlichen Technologie, insbesondere in den Anfangszeiten des Internets und der persönlichen Computer. Es zeigt auch, wie Menschen manchmal vorgeben, mehr zu wissen, als sie tatsächlich tun, nur um im Trend zu sein oder sich nicht uninformiert zu fühlen.

Comic-reverse-engineering GPT4/DallE3 (10/2023)



Was sind Sprachmodelle?

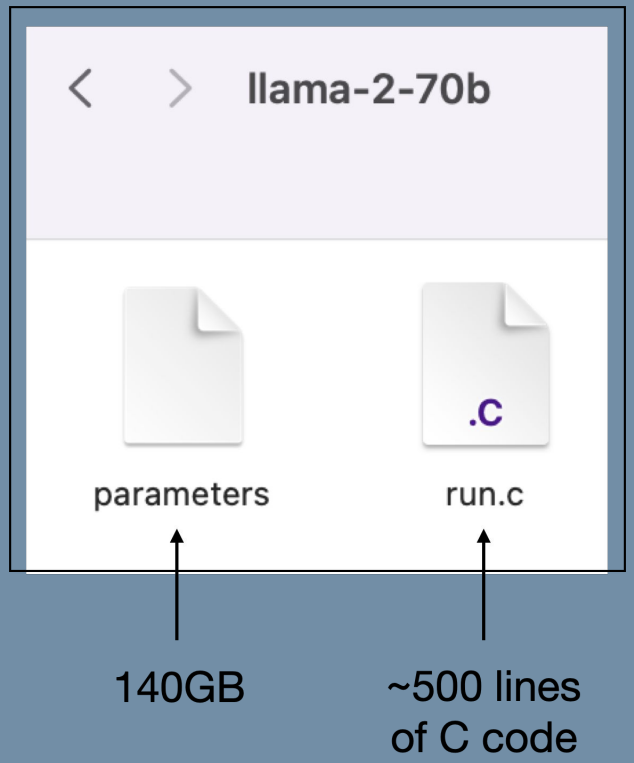
“Can you draw an infographic that explains a large language model as simple and catchy as possible?”



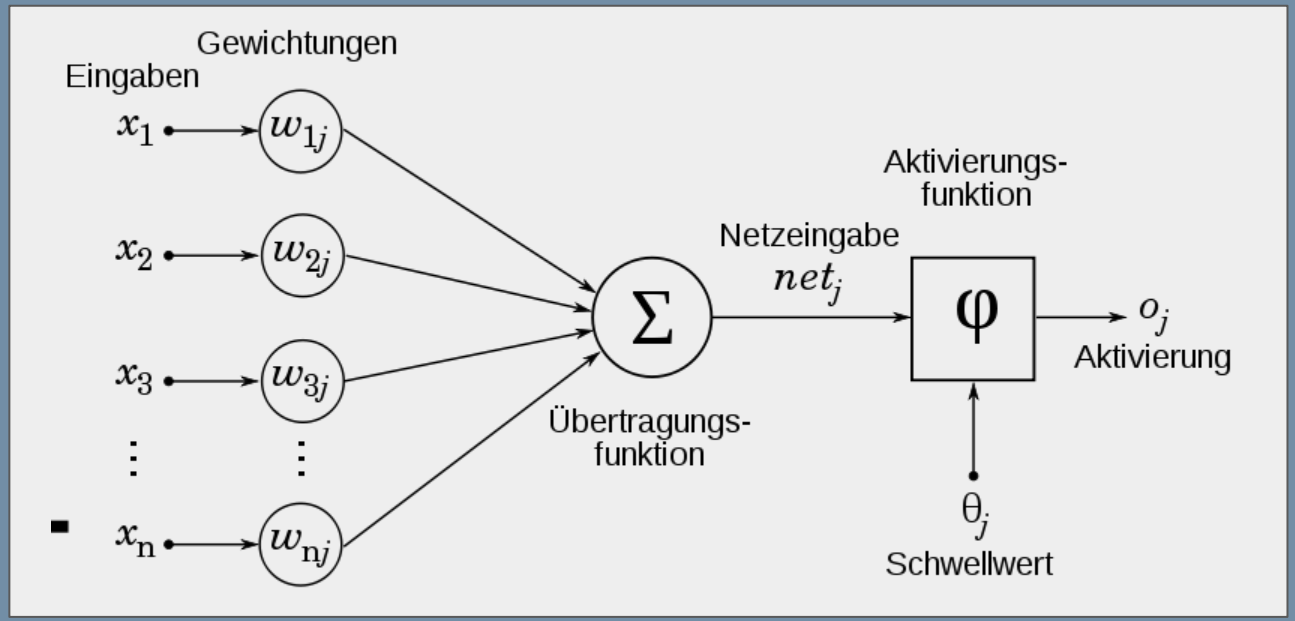
Was sind Sprachmodelle?

“Zip-Files of the Internet?”

Aus dem Vortrag von Andrej Karpathy: Intro to Large Language Models
https://www.youtube.com/watch?v=zjkBMFhNj_g&list=WL&index=17



Was sind Parameters?



(Hier nur Gewichtungsfunktion von neuronalen Netzwerken aus wikidata für Netz mit Input-, Output- P ar Parametern von Neuronen, vgl.Tafelbild)

Anatomie der Sprache I: Sprachliche Zeichen

Signifiant (Zeichenkörper)

Signifié (Zeicheninhalt)

(Für 2 Seiten einer Medaille vgl. Tafelbild)

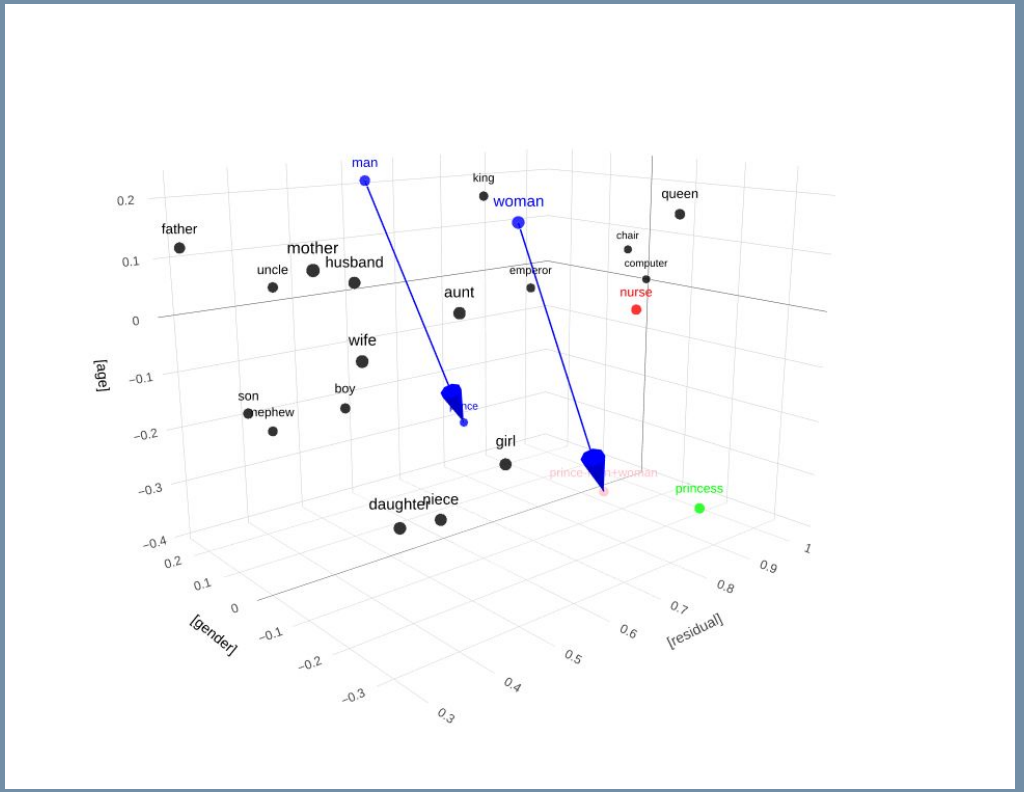
Digitale Repräsentation sprachlicher Zeichen

Zeichenkörper?

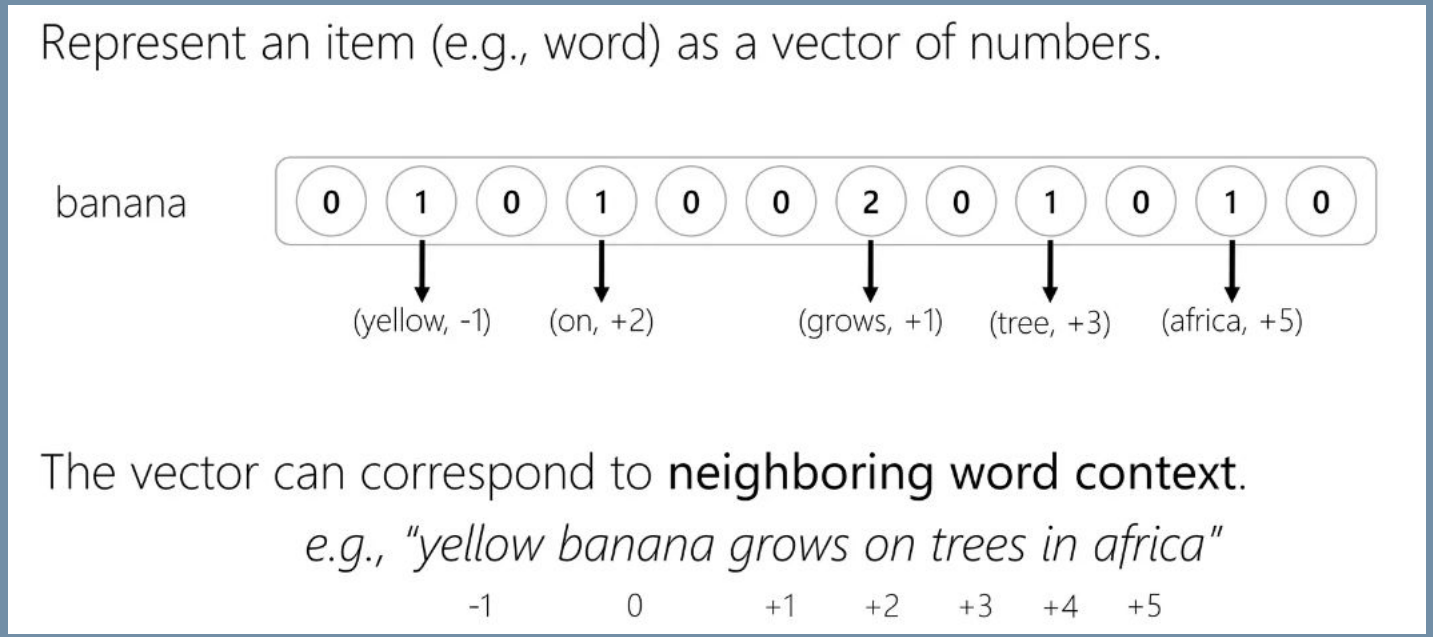
Digitale Repräsentation sprachlicher Zeichen

Zeicheninhalt?

Word Embeddings Demo

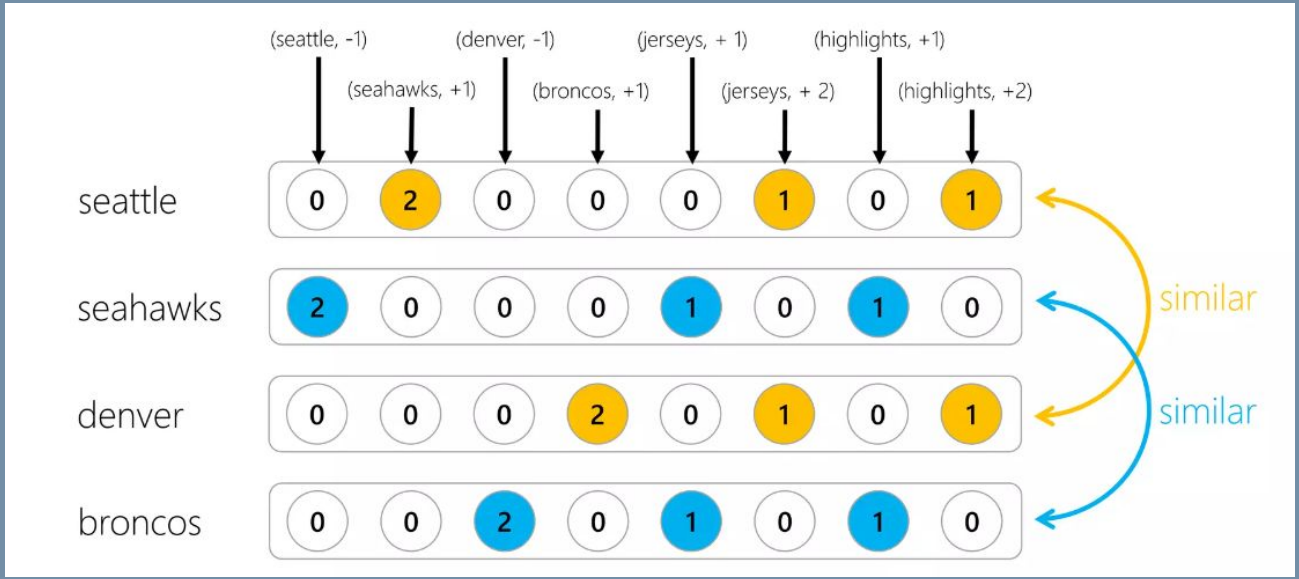


Voraussetzungen I: Kontextvektoren



Aus (der insgesamt sehr übersichtlichen Präsentation von Bhaskar Mitra):
https://www.slideshare.net/BhaskarMitra3/a-simple-introduction-to-word-embeddings

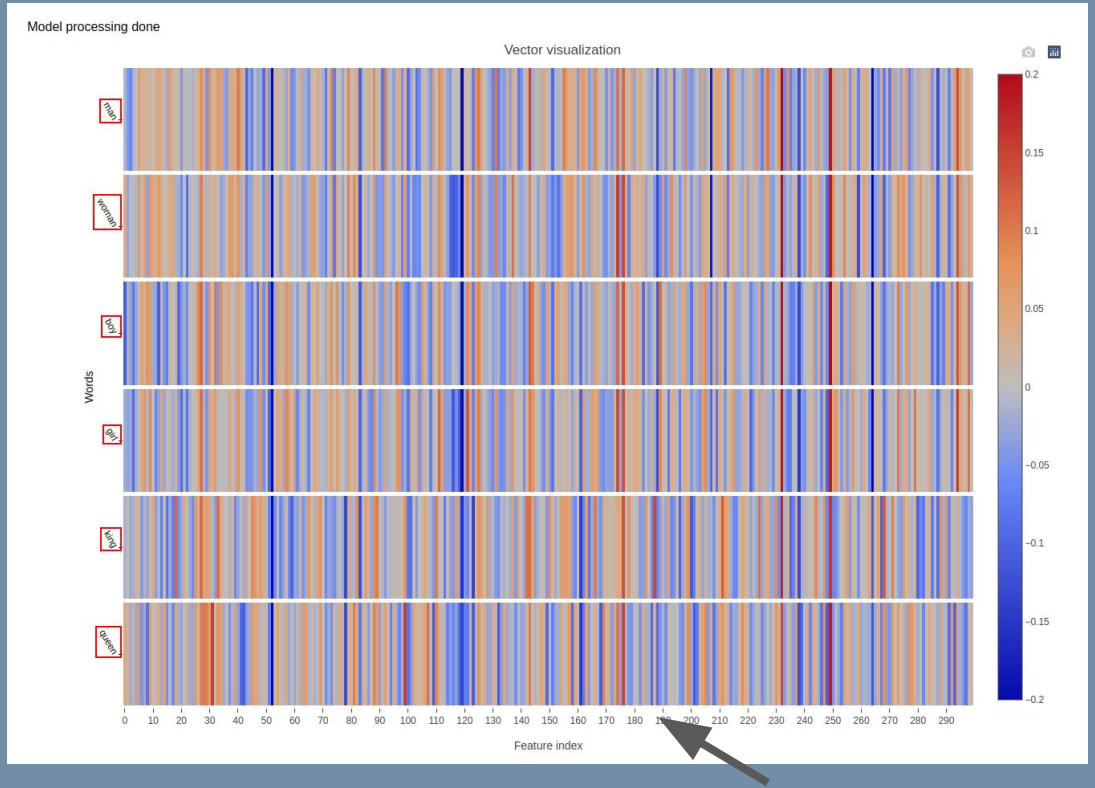
Voraussetzungen I: Vektorrepräsentationen



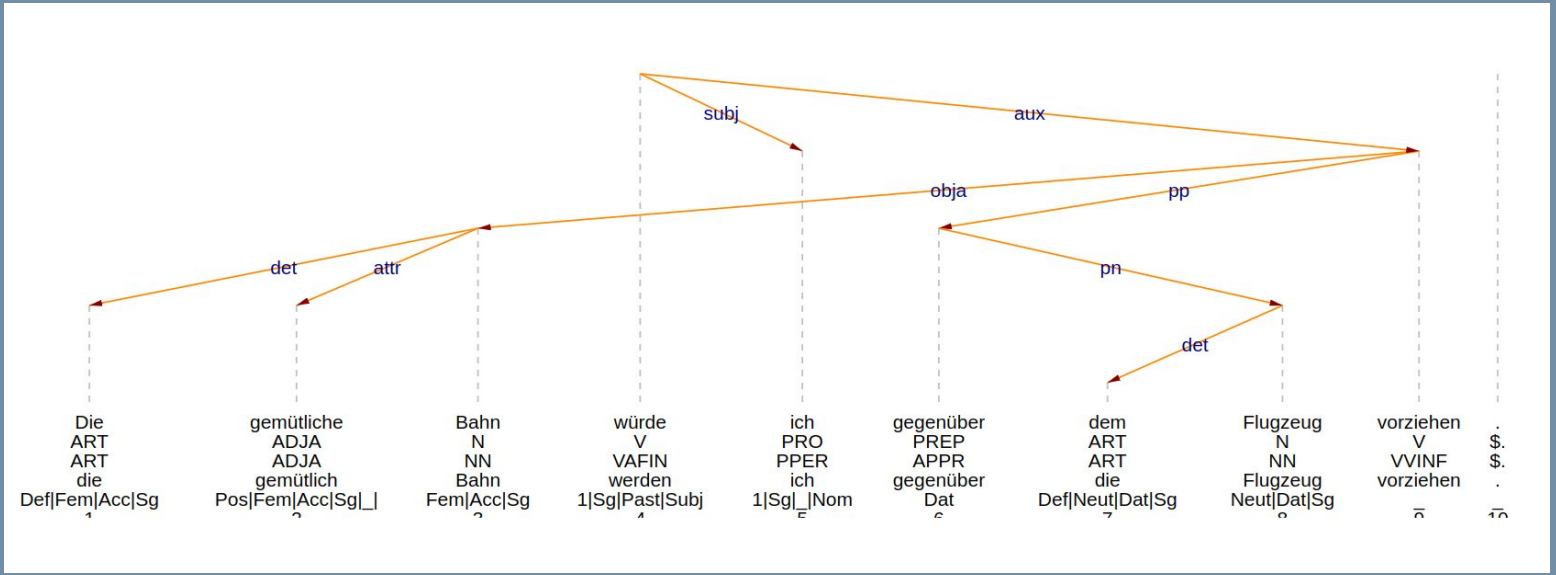
Aus (der insgesamt sehr übersichtlichen Präsentation von Bhaskar Mitra):
https://www.slideshare.net/BhaskarMitra3/a-simple-introduction-to-word-embeddings

Voraussetzungen I: Dense Vectors / Embeddings

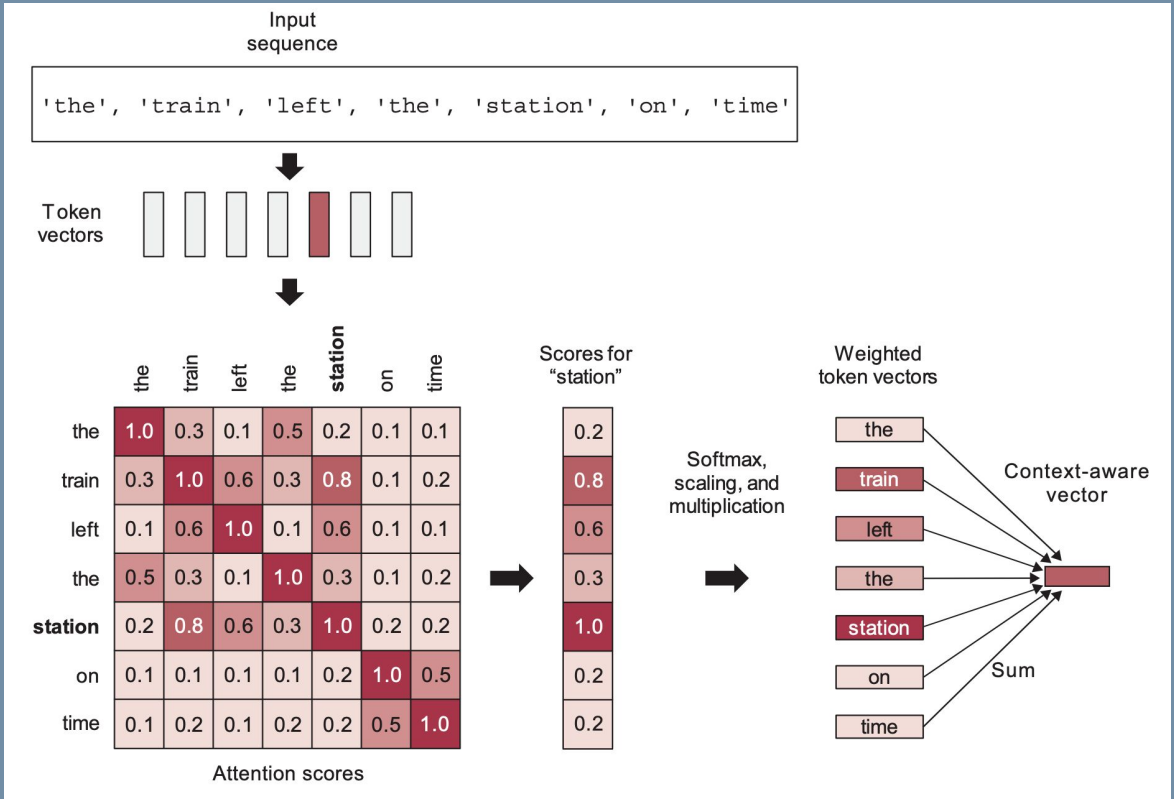
Verschiedene Methoden zur Dimensionreduktion, z.B. skipgram oder cbow



Voraussetzungen II: Attention

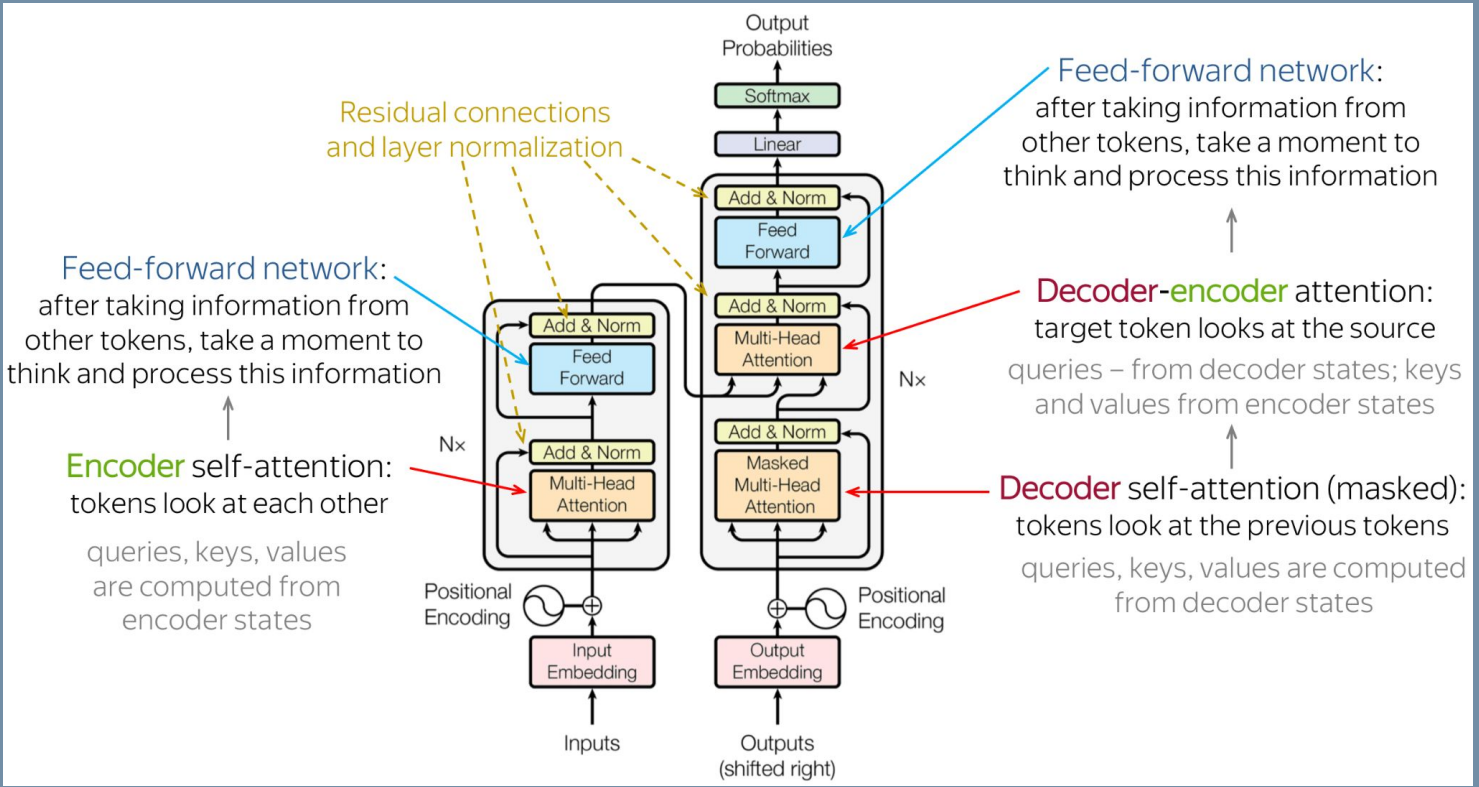


Voraussetzungen II: Attention

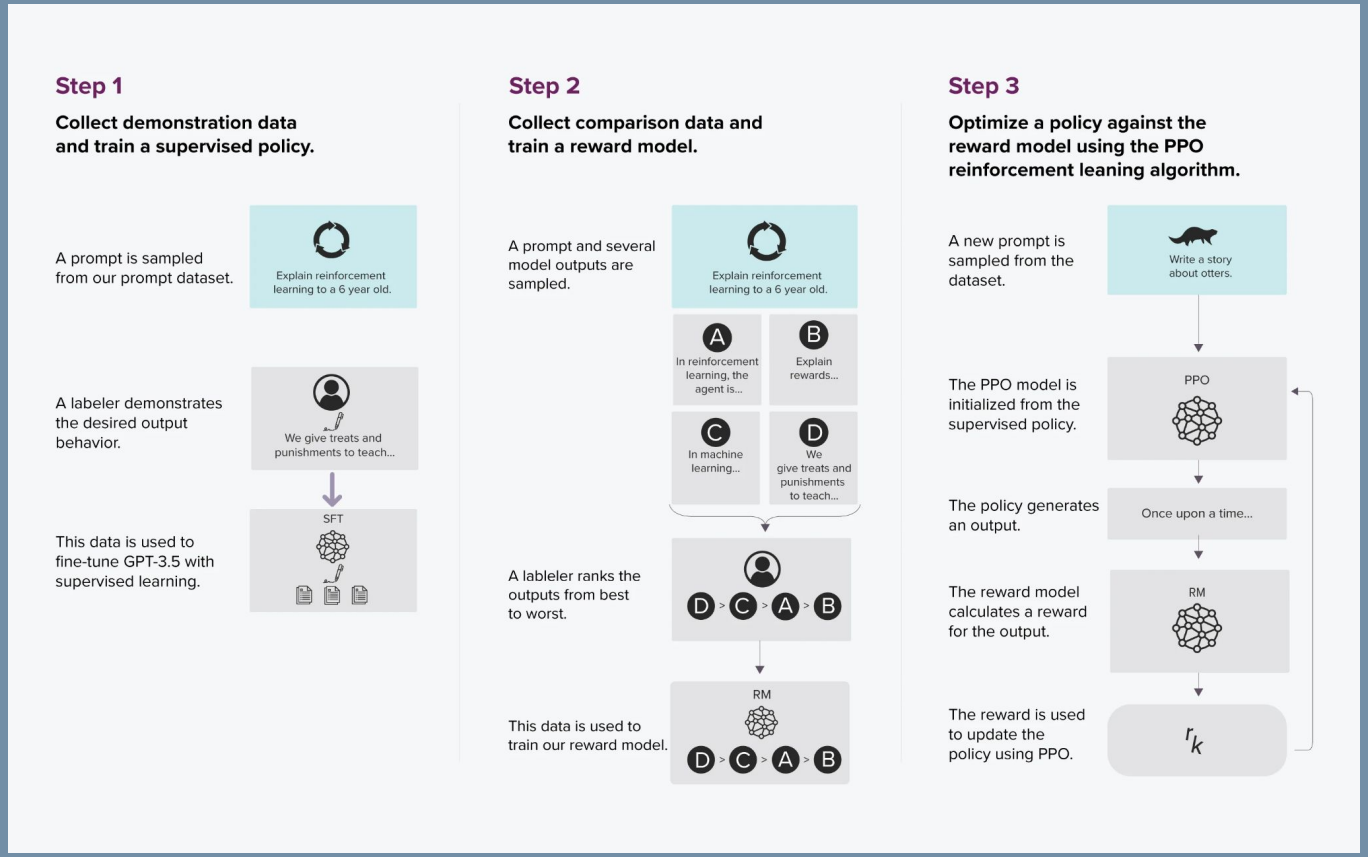


aus Chollet (2023)

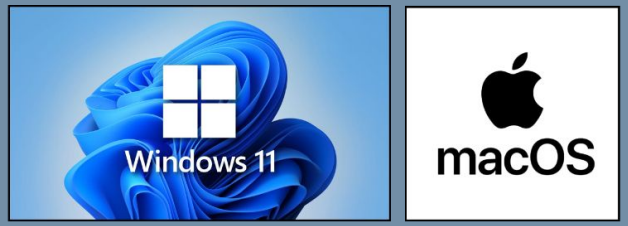
Voraussetzungen II: Attention

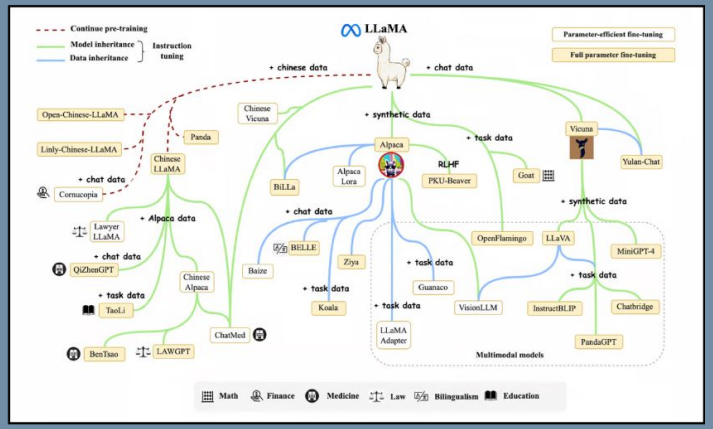


Finetuning: RLHF * 3



OS and LLM Ecosystems

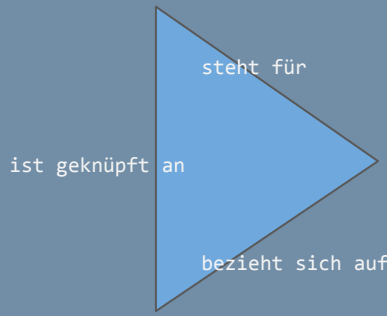




Aus dem Vortrag von Andrej Karpathy: Intro to Large Language Models
https://www.youtube.com/watch?v=zjkBMFhNj_g&list=WL&index=17

Das semiotische Dreieck

Signifiant (Zeichenkörper)



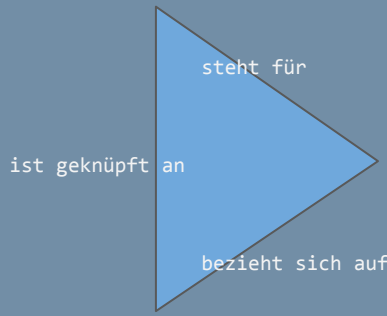
Referent (Objekt in der Welt)

Signifié (Zeicheninhalt)

Das semiotische Dreieck

Position in Codetabelle

Signifiant (Zeichenkörper)



Referent (Objekt in der Welt) ???

Signifié (Zeicheninhalt)

Position im Vektorraum

Vgl. Piantadosi / Hill: Meaning without reference in large language models?
& Underwood: The empirical triumph of theory

Anatomie der Sprache II: Doppelte Gliederung

Bedeutungsunterscheidende Einheiten
(Phoneme / Grapheme)

Bedeutungstragende Einheiten
(Morpheme)

Was ist das Problem mit proprietären LLMs?

- The Five Commandments of Probing LLMs/AI Models (Reiter/Hermes 2024): https://fedihum.org/@IDH_Cologne/112280957076291820
- Balloccu et al. (2024): Leak, Cheat, Repeat https://leak-llm.github.io/
  - unknown model/training data/architecture
  - data contamination (-> overfitting, memorization)
  - indirect data leaking

OpenLLMs News / aktuelle Ressourcen

News:

- Meta: Llama-3 veröffentlicht 8b - 70b - 400b rechnet noch...
- Microsoft: Phi-3 veröffentlicht 3.8b - 7b - 14b

Auf Huggingface verfügbare Modelle https://huggingface.co/models

LMSYS Chatbot Arena Leaderboard https://chat.lmsys.org/?leaderboard

Konzepte der Offenheit (Liesenfeld et al.)

- **Verfügbarkeit**: Code - Trainingsdaten - Gewichte - RLHF-Daten - RLHF-Gewichte - Lizenz
- **Dokumentation**: Code - Architektur - Preprint - Paper - DataSheet
- **Zugriffsmethoden**: (Webservice) - Package - API

Studienleistung A - Vorstellung OpenLLM (ILIAS)

Sie haben sich ein OpenLLM zur Vorstellung ausgesucht und im Etherpad spezifiziert.

Bereiten Sie für diese LLM eine knappe 5-7-Minuten-Einführung vor, bei der Sie kurz auf die Besonderheiten des Modells eingehen (z.B. Größe, Datengrundlage, wann und von wem entwickelt, aufsetzend auf anderem Modell etc.).

Versuchen Sie dabei auch auf die Aspekte der Offenheit, die in Liesenfeld et. al. spezifiziert werden, einzugehen, indem Sie versuchen, das Modell hinsichtlich dieser Aspekte zu bewerten. Erwähnen Sie auch die Aspekte, über die Sie nichts herausgefunden haben! Gerne können Sie dafür eine Tabelle mit Erläuterungen verwenden.