

Annotation
mit
Sprachmodellen
zur Sprachmodellierung
mit
Annotation



Vorweg: Computerlinguistik anno 2002



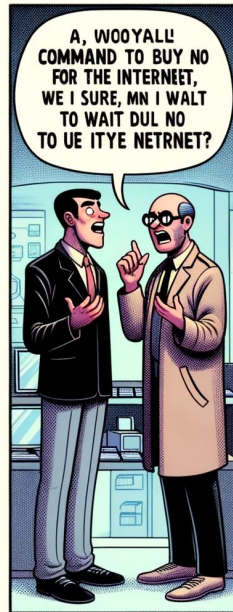
Comiciinterpretation GPT4 (2023)



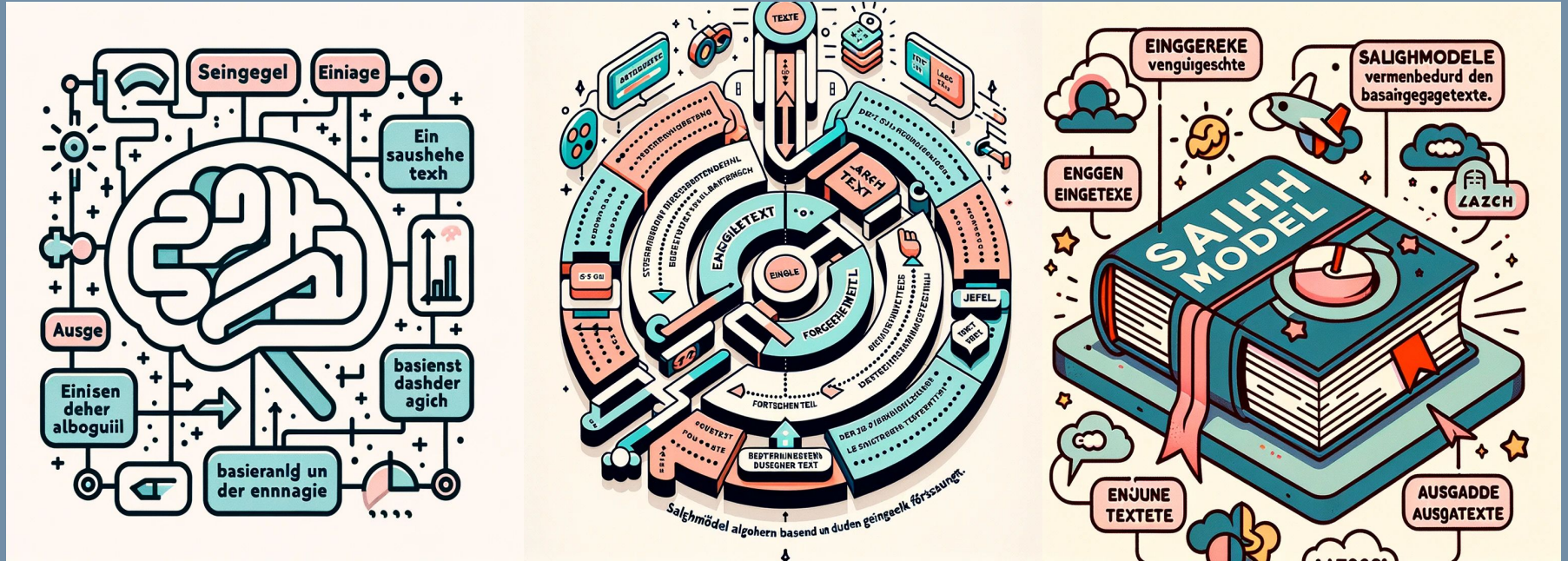
Der Comic beleuchtet die komischen Missverständnisse, die entstehen können, wenn jemand, der wenig bis keine Erfahrung mit moderner Technologie hat, versucht, sie zu nutzen. In der ersten Szene gibt der erste Mann vor, sich mit Computern auszukennen, wahrscheinlich um kompetent oder modern zu wirken. 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?



“Kannst du mir eine Infografik machen, wo ein Sprachmodell als Algorithmus, der eine Eingabekette fortsetzt, möglichst simpel und eingängig erklärt ist?”

Was sind Sprachmodelle?



“Kannst du
der eine Ei

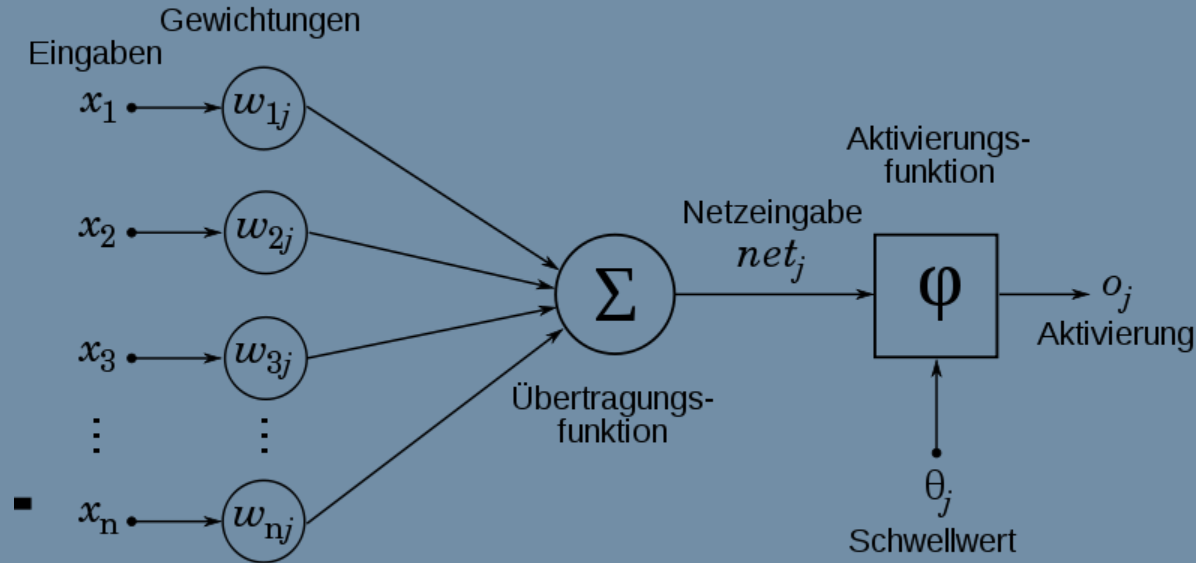
als Algorithmus,
gig erklärt ist?”

Was sind Sprachmodelle?



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

Was sind Sprachmodelle?



(Hier nur Gewichtungsfunktion von neuronalen Netzwerken aus [wikidata](#) für Netz mit Input, Output und Neuronen / Parametern vgl. Tafelbild)

Anatomie der Sprache I: Sprachliche Zeichen

Signifiant (Zeichenkörper)

Signifié (Zeicheninhalt)

(Für 2 Seiten einer Medaille vgl. Tafelbild)

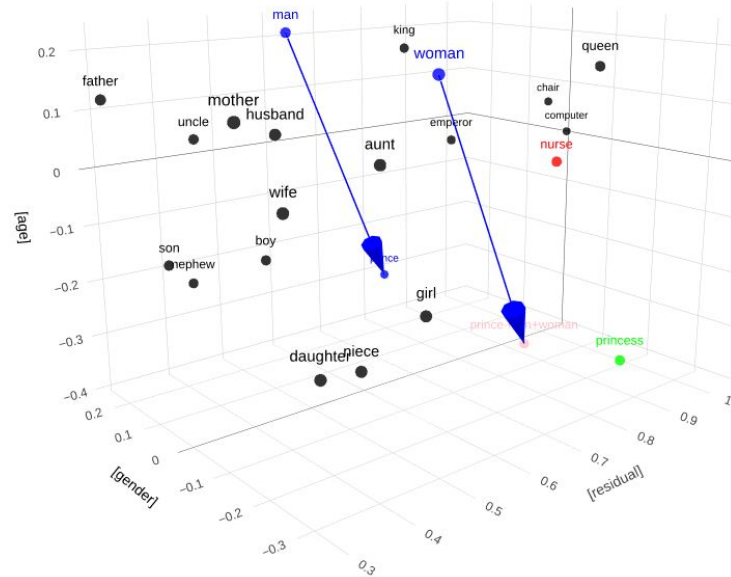
Digitale Repräsentation sprachlicher Zeichen

Zeichenkörper?



Digitale Repräsentation sprachlicher Zeichen

Zeicheninhalt?



Anatomie der Sprache II: Doppelte Gliederung

Bedeutungsunterscheidende Einheiten
(Phoneme / Grapheme)

Bedeutungstragende Einheiten
(Morpheme)