

# Medienlinguistische Methodik

## Sprachmodelle

*basierend auf:*

Zweig, K. (2025). *Weiß die KI, dass sie nichts weiß?* München: Heyne.

Arne Rubehn

Lehrstuhl für Multilinguale Computerlinguistik

Universität Passau

13.01.2026



# Überblick

In dieser Sitzung beschäftigen wir uns mit der Funktionsweise von Sprachmodellen, ihren Fähigkeiten und ihren Einschränkungen.

**Wie erzeugen Sprachmodelle Texte?**

**Wie werden Sprachmodelle trainiert?**

**Was können Sprachmodelle, und was nicht?**



# Was ist KI?

Ist das KI?

- Chatbots (ChatGPT, Claude, Gemini, ...)



# Was ist KI?

Ist das KI?

- Chatbots (ChatGPT, Claude, Gemini, ...)
- **Taschenrechner**



# Was ist KI?

Ist das KI?

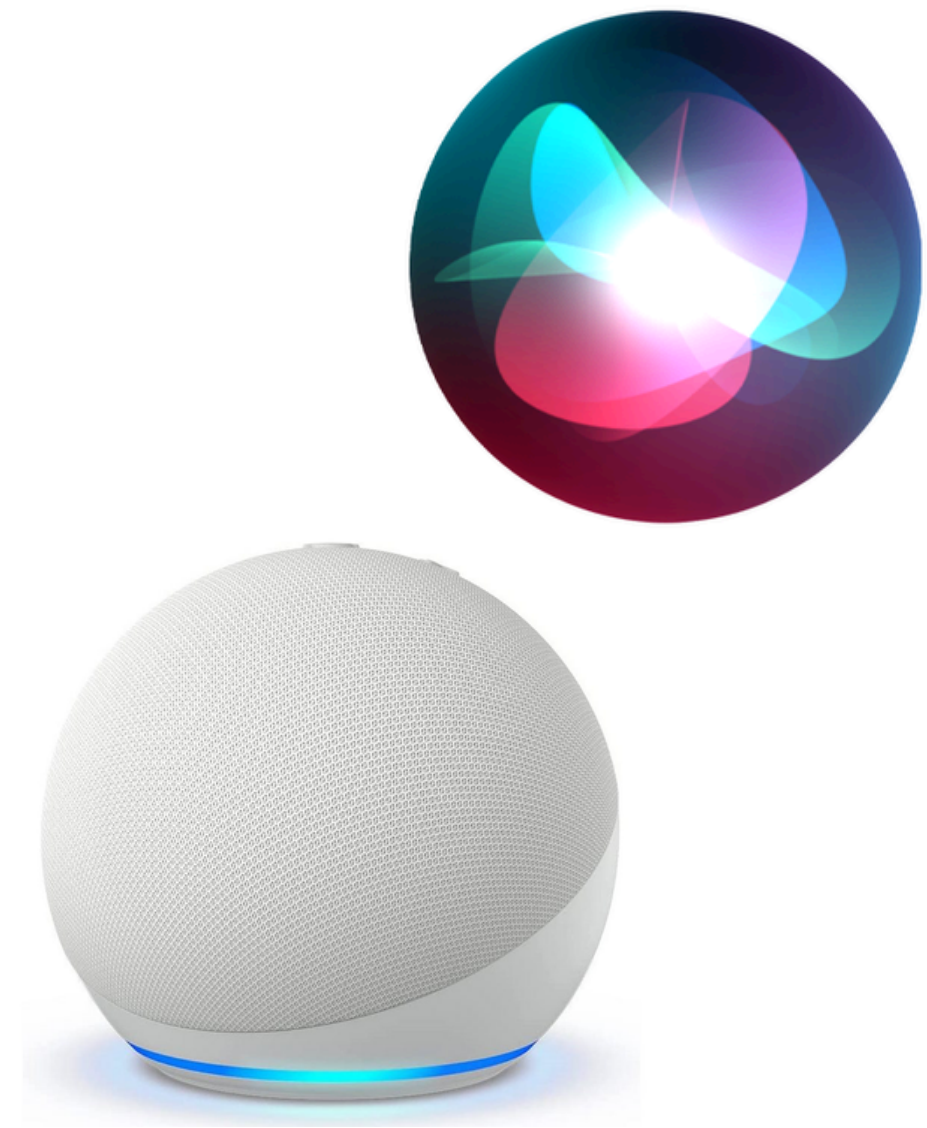
- Chatbots (ChatGPT, Claude, Gemini, ...)
- Taschenrechner
- **Computergesteuerte Gegner in Videospielen**



# Was ist KI?

Ist das KI?

- Chatbots (ChatGPT, Claude, Gemini, ...)
- Taschenrechner
- Computergesteuerte Gegner in Videospielen
- **Spracherkennung (Siri, Alexa, ...)**



# Was ist KI?

Ist das KI?

- Chatbots (ChatGPT, Claude, Gemini, ...)
- Taschenrechner
- Computergesteuerte Gegner in Videospielen
- Spracherkennung (Siri, Alexa, ...)
- **Schachroboter**



# Was ist KI?

Ist das KI?

- Chatbots (ChatGPT, Claude, Gemini, ...)
- Taschenrechner
- Computergesteuerte Gegner in Videospielen
- Spracherkennung (Siri, Alexa, ...)
- Schachroboter



**KI bezeichnet meistens neue Technologien, die zuvor nicht für möglich gehalten wurden. Der Begriff ist ein bewegliches Ziel.**



# Was ist KI?

Es gibt also **keine technisch saubere Definition** für den Begriff der Künstlichen Intelligenz – es ist eher ein subjektives, menschliches Urteil, welche Fähigkeiten wir bei einer Maschine als “intelligent” wahrnehmen.

Eine mögliche Umschreibung des Begriffes KI wäre:

(nach Marvin Minsky, einem der Urväter der KI-Forschung)

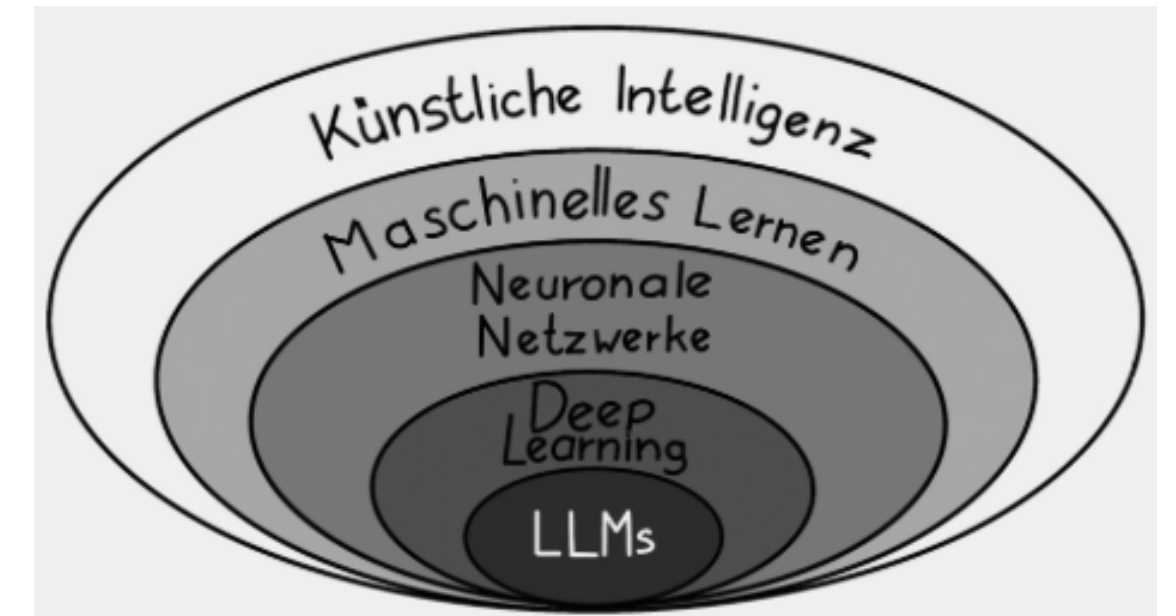


# Was ist KI?

Heutzutage bezeichnet KI meistens Modelle, die dem Paradigma des **Maschinellen Lernens** folgen – also statistische Muster aus Trainingsdaten lernen können.

Das war nicht immer der Fall: Lange waren **regelbasierte Systeme** der führende Ansatz.

Hierbei befolgte die Maschine streng ein Regelwerk, das von Menschen entworfen wurde. Diese Systeme sind daher **deterministisch** – derselbe Input führt immer zum selben Output.



# ELIZA

Auch die Sprachtechnologie folgte lange dem regelbasierten Paradigma. So auch Joseph Weizenbaums **ELIZA**, der wohl erste Chatbot.

Probiere ELIZA doch mal selbst aus: <https://jw.weizenbaum-institut.de/>

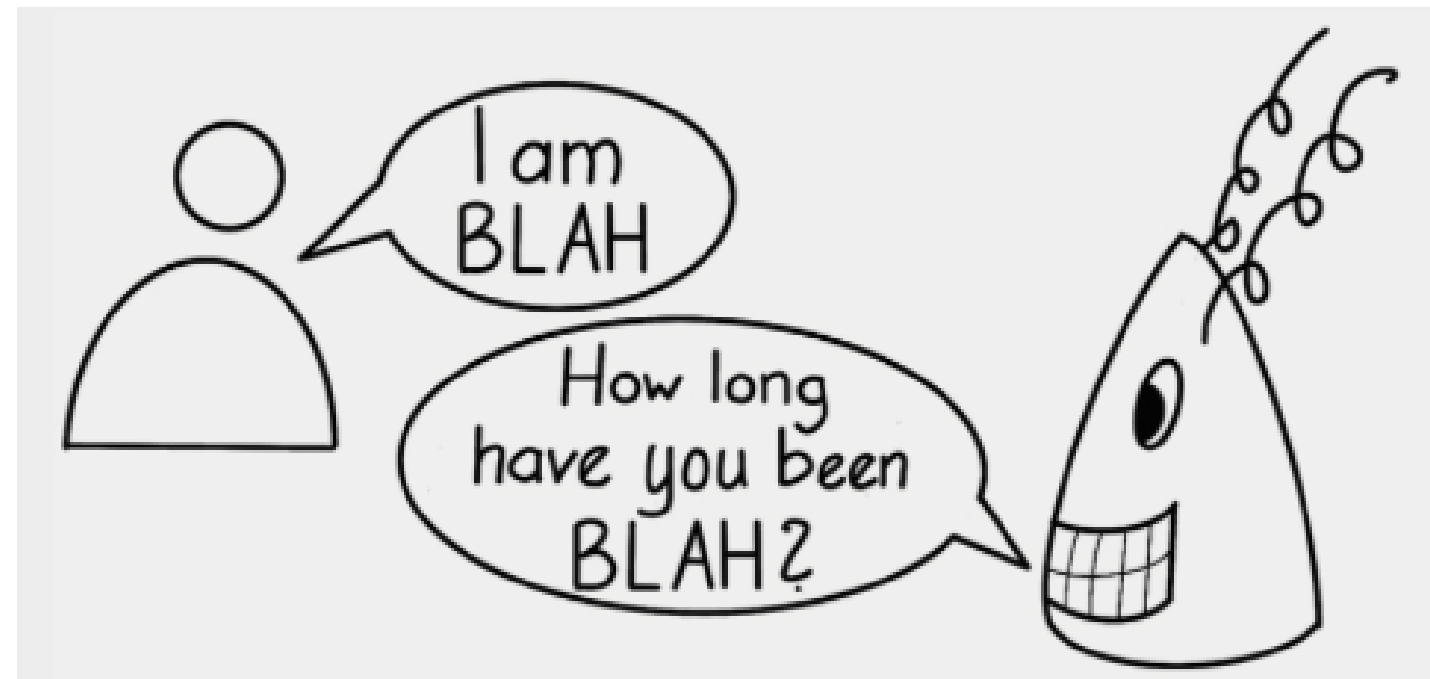
**Was fällt dir auf?**



# ELIZA

Auch die Sprachtechnologie folgte lange dem regelbasierten Paradigma. So auch Joseph Weizenbaums **ELIZA**, der wohl erste Chatbot.

Du wirst bemerkt haben, dass ELIZA ein ziemlich primitives System ist – im Grunde genommen werden nur Stichwörter erkannt und Aussagen zu Fragen umgestellt.



# ELIZA

Auch die Sprachtechnologie folgte lange dem regelbasierten Paradigma. So auch Joseph Weizenbaums **ELIZA**, der wohl erste Chatbot.

Du wirst bemerkt haben, dass ELIZA ein ziemlich primitives System ist – im Grunde genommen werden nur Stichwörter erkannt und Aussagen zu Fragen umgestellt.

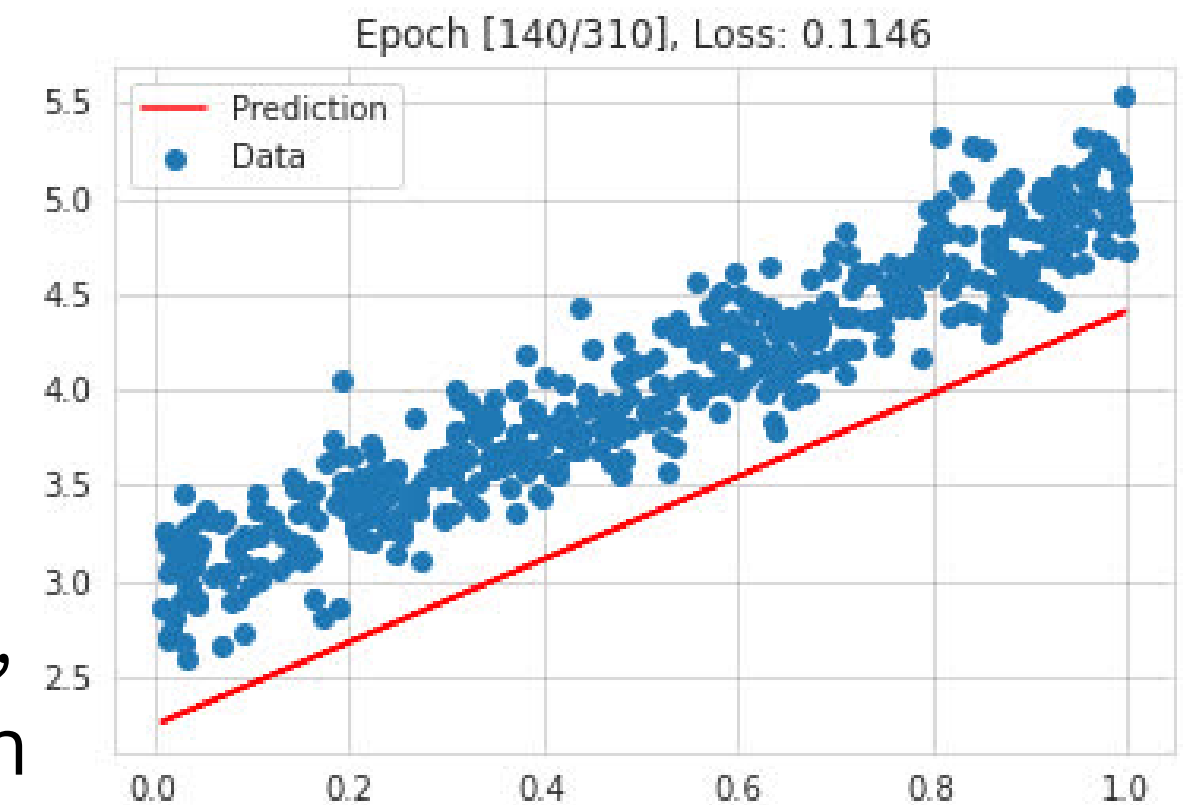
Viel interessanter als die Technologie selbst ist hier aber die **menschliche Reaktion** darauf: Nutzer:innen bauten extrem schnell eine Beziehung zu ELIZA auf, fühlten sich verstanden und teilten private Details!



# Maschinelles Lernen

Wie bereits erwähnt, basieren moderne Sprachmodelle nicht auf strikten Regeln, sondern auf gelernten **statistischen Mustern**. Menschliche Sprache ist scheinbar zu komplex, um sie anhand von klaren Regeln zu modellieren.

Solche Modelle lernen statistische Zusammenhänge in den Daten, indem sie eine **Kostenfunktion** minimieren, die angibt, wie stark die Vorhersagen des Modells von den beobachteten Daten abweichen.



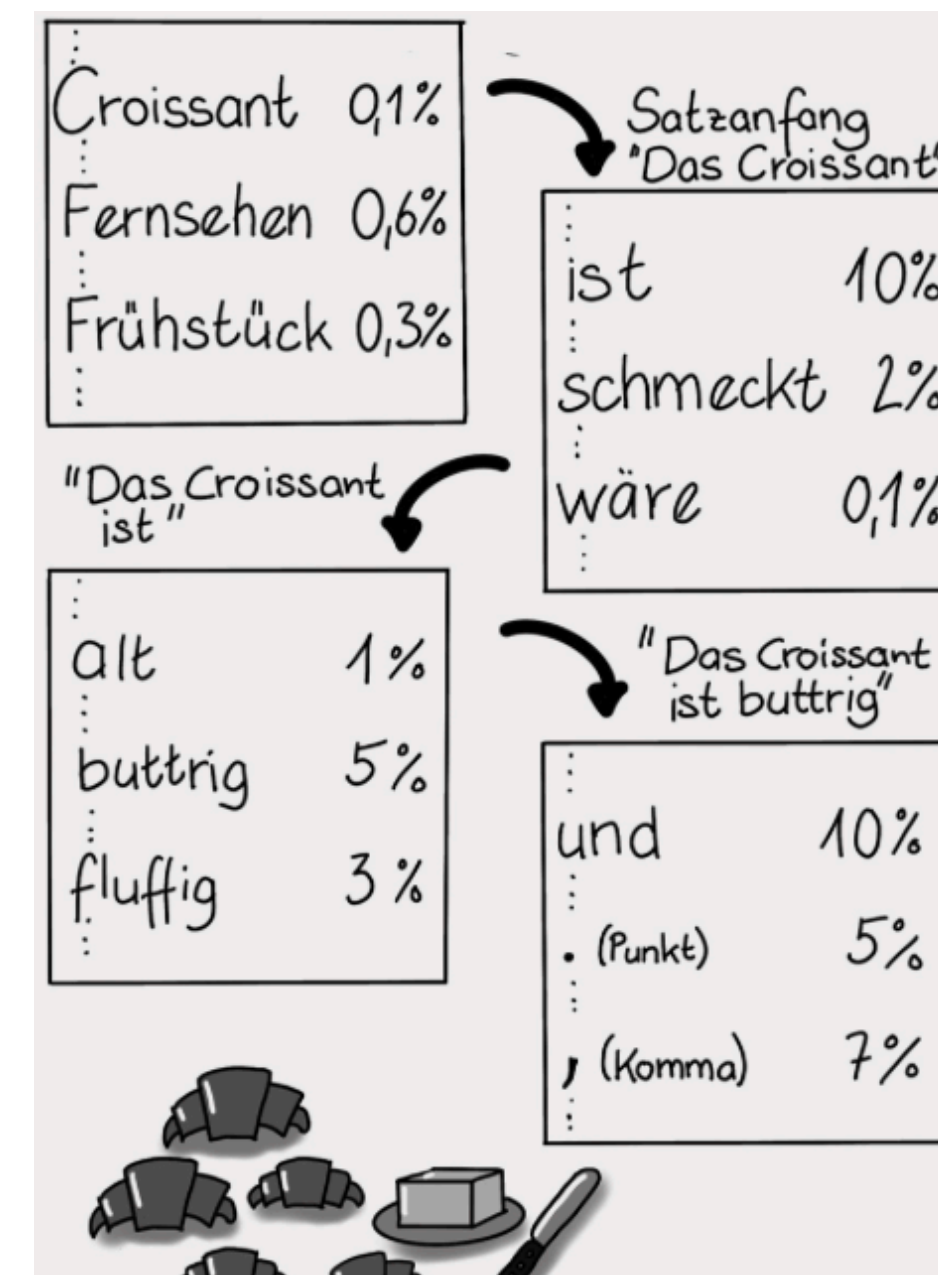
# Texterzeugung

Im Kern ist der Mechanismus, mit dem Sprachmodelle Texte erzeugen, denkbar einfach:

**Generiere das Wort, das basierend auf dem bisherigen Kontext das wahrscheinlichste ist.**

Das generierte Wort wird an den Kontext angehängt, der ganze Prozess startet von vorne.

**That's it. Das ist die Magie.**



# Klingt zu einfach?

Jein.

Natürlich gibt es einige Details, die beachtet werden müssen, damit das so funktioniert.

Dennoch lässt sich die Texterzeugung von Sprachmodellen auf eine (sehr komplexe) Wahrscheinlichkeitsrechnung herunterbrechen, welches Token als nächstes folgt.

**Aber wie berechnet das Modell, was wahrscheinlich ist?**



# Pretraining

Eine entscheidende Stärke der Sprachmodelle ist, dass die Trainingsprozedur in zwei Schritte unterteilt wird.

Im ersten Schritt werden generelle sprachliche Strukturen aus großen Textmengen gelernt.

Hierbei werden dem Modell “Satzanfänge” präsentiert; es soll dann entscheiden, mit welchem Token es mutmaßlich weitergeht.

***z.B.:** Ich gehe heute ins ??? [Kino / Theater / Schwimmbad / ...]*



# Word Embeddings

Es ist grundsätzlich möglich, beliebig viele neue Sätze zu bilden. Ein gutes Sprachmodell muss das auch abdecken können – wir kommen also mit dem bloßen Zählen von Satzanfängen nicht weit.

Stattdessen muss das Modell irgendwie lernen, welche Strukturen wie miteinander kombiniert werden können oder durcheinander ersetzt werden können.

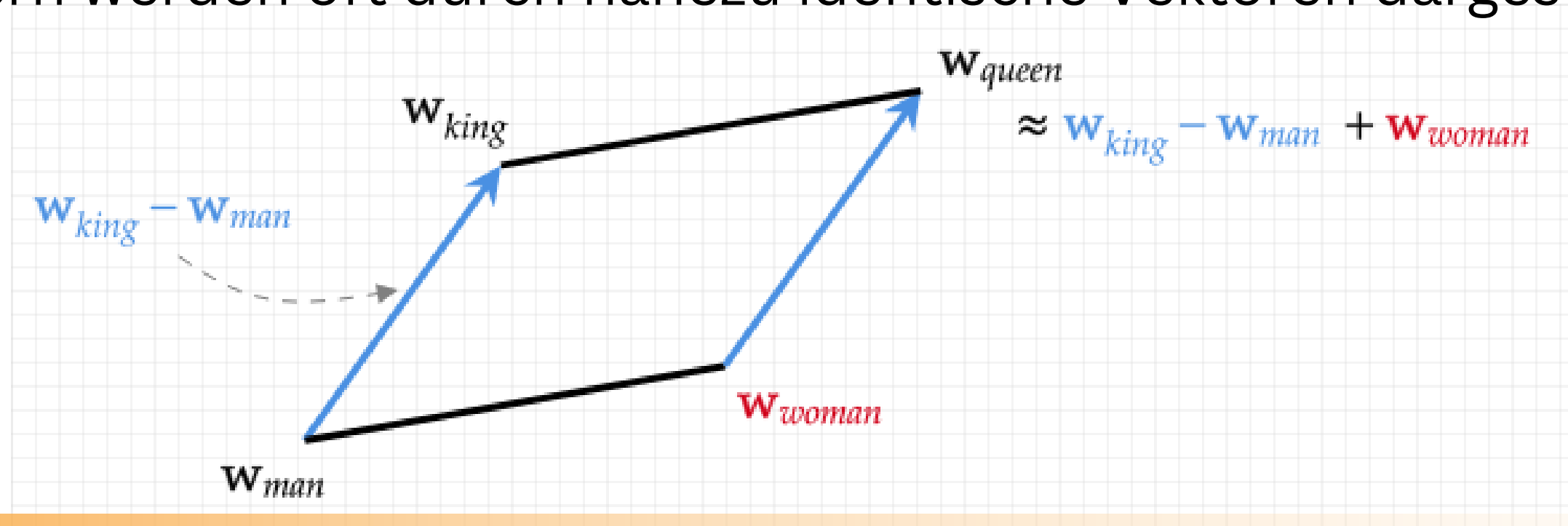
Das ist durch **Word Embeddings** (auch dt. *Worteinbettungen*) möglich: Tokens werden in einem hochdimensionalen Raum so angeordnet, dass wichtige Beziehungen zueinander mathematisch modelliert werden.



# Word Embeddings

Word Embeddings werden anhand des **Kontextes** gelernt, in denen ein jeweiliges Wort vorkommt. Wörter, die in ähnlichen Kontexten vorkommen, werden dadurch nah zueinander gruppiert.

Hierbei ergeben sich bemerkenswerte Regelmäßigkeiten: Semantische Beziehungen zwischen Wörtern werden oft durch nahezu identische Vektoren dargestellt!



# Attention is all you need

...ist der Titel des Artikels, das die Transformer-Architektur vorstellt, auf der alle modernen Sprachmodelle wie GPT aufbauen. Hier wird klar: Der **Attention Mechanism** (*Aufmerksamkeitsmechanismus*) spielt eine zentrale Rolle.

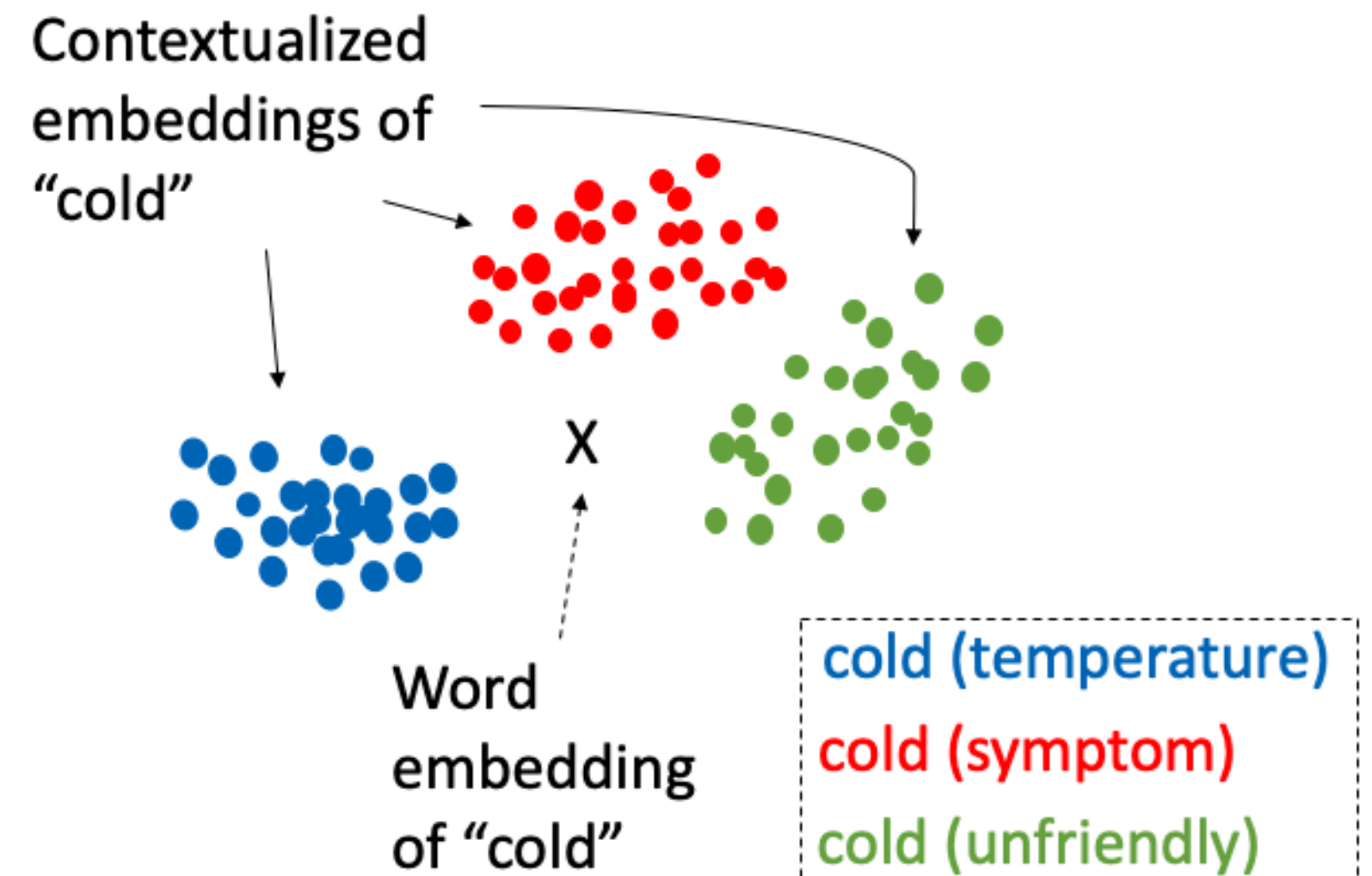
Die Idee ist ziemlich intuitiv: Wenn das Modell das nächste Wort voraussagen soll, sind nicht alle Wörter aus dem bisherigen Kontext gleich relevant. Durch **Attention-Blöcke** lernt das Modell, welche bisherigen Tokens für die Generierung des nächsten Tokens relevant sind.



# Attention is all you need

Die **Repräsentation** gesehener Tokens wird also dynamisch – je nach **Kontext** kann ein und dasselbe Token nun unterschiedlich dargestellt werden.

Das ist eine große Weiterentwicklung zu statischen Word Embeddings, in denen ja jedes Token nur genau eine Position im Raum zugewiesen wurde!



# Tokenisierung

Wir haben gelernt, dass sich Sprachmodelle im Grunde genommen eine kontextbasierte **Wahrscheinlichkeitsverteilung** über alle Wörter errechnet. Wie ist es dann also möglich, dass Sprachmodelle auch neue, ungesehene Wörter generieren?

Hier habe ich bislang eine Vereinfachung vorgenommen. Die Grundbausteine von Sprachmodellen sind nicht Wörter, sondern **Tokens**. Während häufige Wörter zwar typischerweise auch einzelnen Tokens entsprechen, können Tokens auch Bruchteile eines Wortes (*Subwords*) oder sogar einzelne Buchstaben sein.

Alle genannten Mechanismen assoziieren also **Tokens** miteinander, nicht Wörter.



# Visualisierungen

Es gibt interaktive Webseiten, die visualisieren, wie verschiedene Versionen von GPT Texte visualisieren, z.B.: <https://gpt-tokenizer.dev>

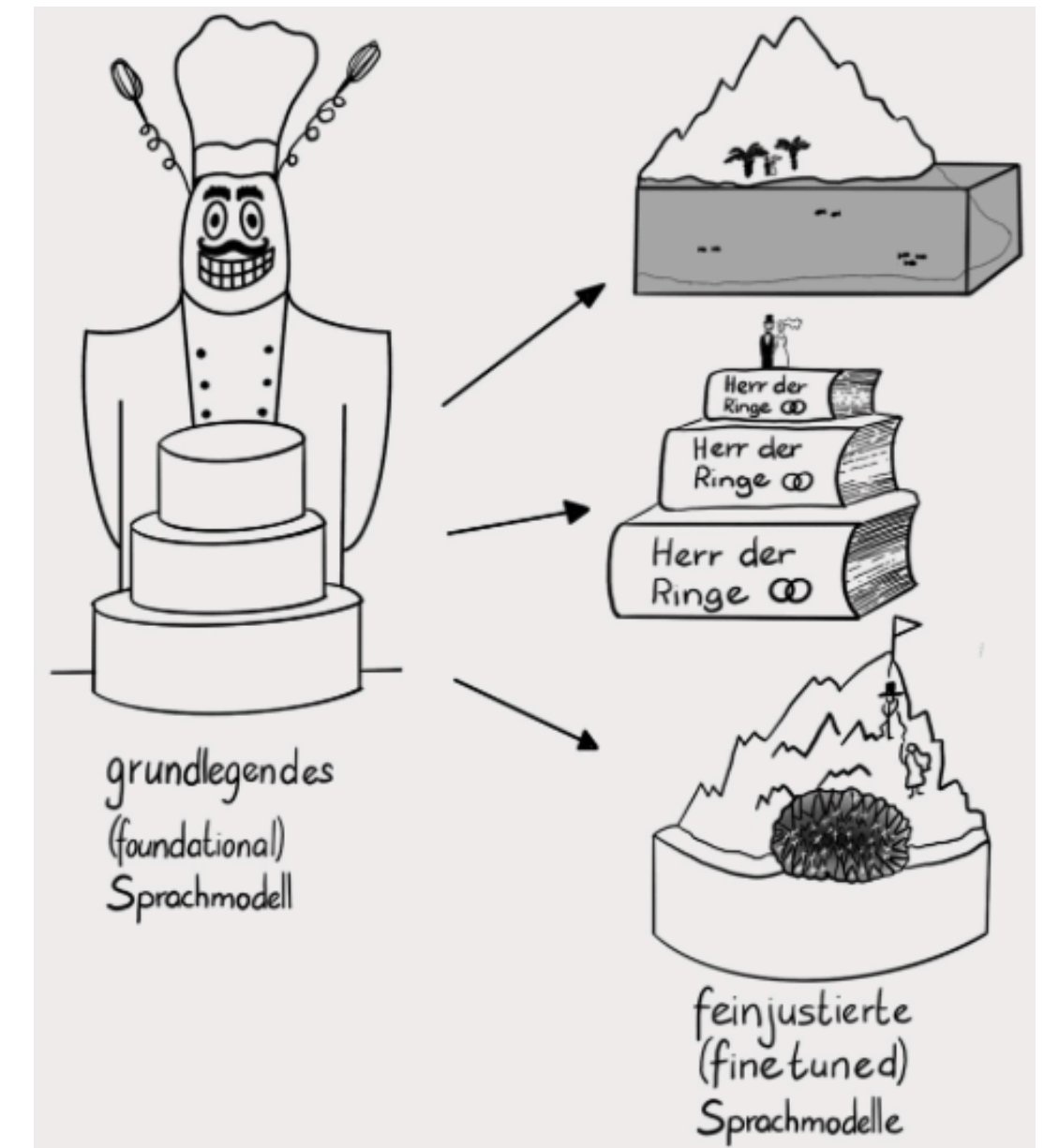
Das “technische Innenleben” von GPT-2 wird hier schön visualisiert: <https://poloclub.github.io/transformer-explainer/>



# Finetuning

Bislang war nur die Rede von generellen Sprachmodellen wie GPT (kurz für *Generative Pretrained Transformer*). Sie lernen taskunabhängig erstmal die Grundstrukturen der Sprache.

In einem darauffolgenden Schritt werden grundlegende Modelle dann **feinjustiert** (*finetuning*), um sie für bestimmte Aufgaben zu spezifizieren – etwa als Chatbots.



# Finetuning von Chatbots

Ein Sprachmodell alleine macht keinen Chatbot. Neben der grundlegenden Sprachfertigkeit muss ein Chatbot z.B. auch lernen, wie ein Dialog strukturiert ist und wie er sich zu “benehmen” hat. Das erfolgt grob gesagt in zwei Schritten:

**Überwachtes Finetuning** anhand spezifischer Trainingsdaten in Dialogform.

**Reinforcement Learning** mit menschlichem Feedback, bei dem das Modell sich zwischen verschiedene mögliche Antworten auf ein Prompt entscheiden muss. Ein Mensch bewertet, ob das eine gute Wahl war – entsprechend wird das Modell mathematisch “belohnt” oder “bestraft”.



# Was wissen Sprachmodelle?

Sprachmodelle lernen **Assoziationen** zwischen Wörtern (bzw. Tokens), die sie wiedergeben können. Durch die Größe der Modelle und die Menge an Trainingsdaten können so komplexe Muster gelernt werden, dass Chatbots auf die meisten Fragen tatsächlich passende Antworten finden.

**Aber Achtung:** Modelle haben keinen tatsächlichen Wissensstand (es gibt keine Wissensdatenbank o.Ä. im Hintergrund)! “Wissen” kann in Sprachmodellen nur indirekt über die Trainingsdaten aufgenommen werden. Modelle können also gar nicht wissen, was sie nicht wissen – es wird sehr gerne dann mit viel Selbstbewusstsein etwas erfunden (sog. **Halluzinationen**).



# Was wissen Sprachmodelle?

**Kontext:** Ein Freund konsultiert die Meta AI für die Zusammenstellung seines Kaders im Fantasy Manager “Kickbase”.

Nahezu alle genannten Spieler werden Vereinen zugeordnet, bei denen sie vor ca. 3 Jahren gespielt haben – ein offensichtliches Anzeichen für das Alter der Trainingsdaten.

Noch viel interessanter: Phillip Hofmann hat nie bei Bayer Leverkusen gespielt – und es **gab nie** einen Profifußballer namens Matthieu de Ronde!

 Meta AI

 Okay, lass uns schnell Ersatz finden!

Hier sind einige Optionen:

- **Stürmer:**
- Niclas Füllkrug spielt jetzt bei West Ham United, aber ein Ersatz könnte Marvin Ducksch (Werder Bremen) oder Phillip Hofmann (Bayer Leverkusen) sein.
- **Mittelfeld:**
- Leroy Sané spielt bei Galatasaray Istanbul, aber ein Ersatz könnte Xavi Simons (RB Leipzig) oder Florian Wirtz (Bayer Leverkusen) sein, wenn er fit ist.
- Kingsley Coman ist bei Bayern München, aber ein Ersatz könnte Julian Brandt (Borussia Dortmund) oder Jamal Musiala (Bayern München) sein, wenn er fit ist.
- **Abwehr:**
- Matthijs de Ligt spielt bei Bayern München, aber ein Ersatz könnte Dayot Upamecano (RB Leipzig) oder Benjamin Pavard (Bayern München) sein.
- Jeremie Frimpong ist bei Bayer Leverkusen, aber ein Ersatz könnte Nico Schlotterbeck (Borussia Dortmund) oder Matthieu de Ronde (keine aktuelle Info) sein.

Einige mögliche Ersatzspieler:

- **Mittelfeld:** Xavi Simons, Julian Brandt, Jamal Musiala
- **Angriff:** Marvin Ducksch, Phillip Hofmann
- **Abwehr:** Dayot Upamecano, Nico Schlotterbeck

Was meinst du? Sollen wir noch mehr Optionen suchen? 🤔 1 2 3

13:52



# Was wissen Sprachmodelle?

Dass Sprachmodelle faktisches Wissen nicht direkt abrufen können, ist ein Problem, das KI-Entwickler:innen schnell aufgefallen ist.

Moderne Systeme behelfen sich mit einem vorgelagerten Modell, das bei einer entsprechenden Anfrage eine **Internetsuche** durchführt und relevante Texte extrahiert. Diese werden dann – unsichtbar für User – dem **Kontext** des Sprachmodelles **angefügt**, sodass sich dieses auf die darin enthaltenen Informationen beziehen kann.

Dieser Mechanismus nennt sich **Retrieval-Augmented Generation (RAG)**.



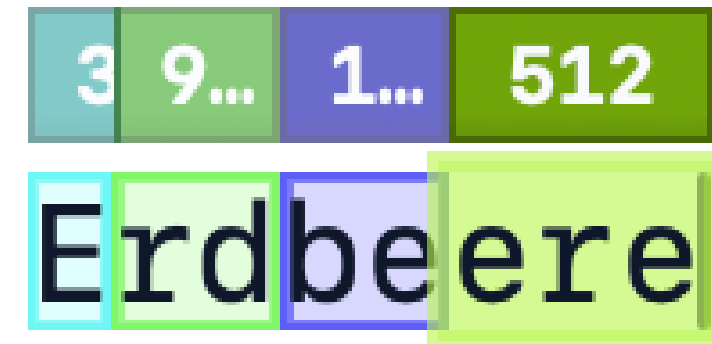
# Erdbeeren im Winter

Bis vor Kurzem konnte man Sprachmodelle auch noch mit einer sehr einfachen Anfrage hinters Licht führen: Indem man es einfach die **Buchstaben** eines Wortes zählen ließ.

Wie oft kommt der Buchstabe R in "Erdbeere" vor?

Der Buchstabe R kommt in „Erdbeere“ dreimal vor. ✓

(generiert am 09.10.2025  
mit ChatGPT-4o)



Für Sprachmodelle sind **Tokens** die kleinste Recheneinheit; diese werden intern als **ganze Zahlen** dargestellt – Sprachmodelle können gar nicht in Tokens “hineinsehen”!



# Stützräder für Sprachmodelle

Das “Erdbeerproblem” scheint in ChatGPT-5 gelöst zu sein – dieser Chatbot kann nämlich bei Bedarf kleine Python-Skripte ausführen, die solche Aufgaben einfach lösen können (und damit die richtige Antwort versteckt dem Prompt anfügen).

Das ist kein Anzeichen dafür, dass die KI “intelligenter” wird. Viel mehr sind es die Entwickler:innen, die die Schwachstellen ihrer Sprachmodelle erkennen und **Behilfskonstrukte** bauen, um der KI zu helfen, die richtige Antwort zu geben.

Dafür werden **relevante Informationen** anderswo erzeugt und unsichtbar an den **Prompt angefügt**.

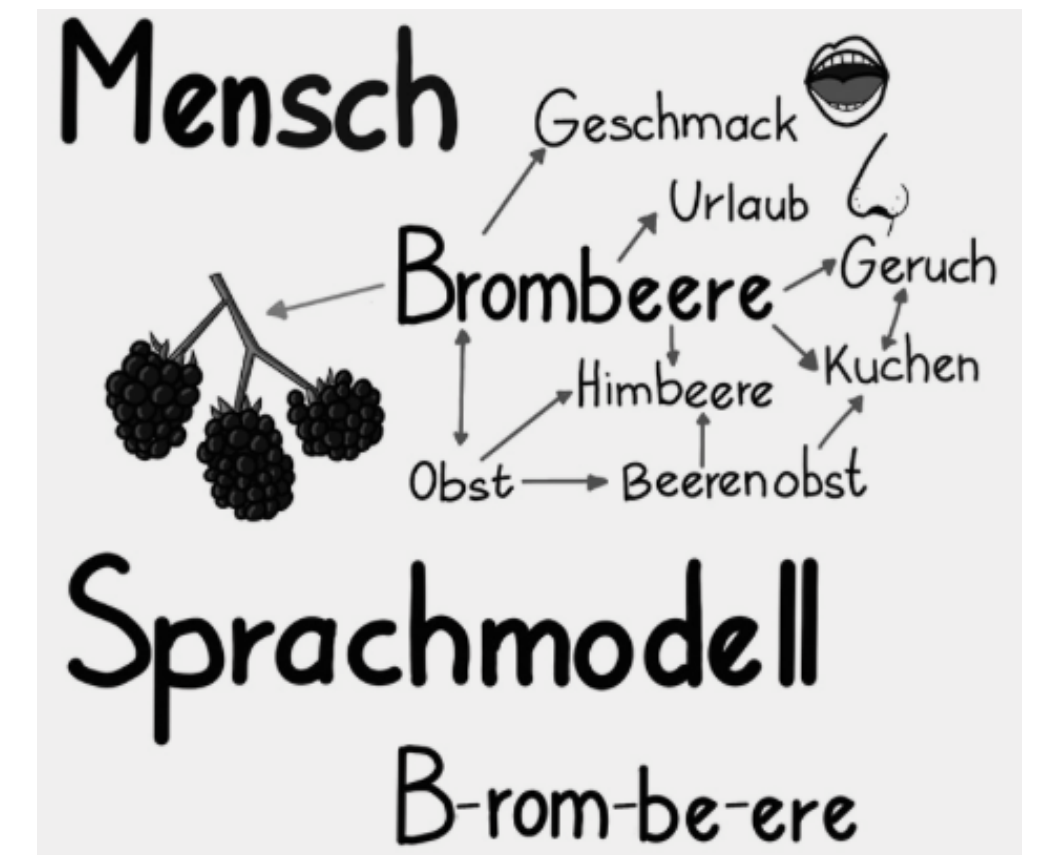


# Grounding Problem

Menschen handeln und sprechen in der echten Welt – ihre Aktionen und Worte haben reelle Auswirkungen auf diese Welt und auf ihr eigenes Leben.

Diese Komponente fehlt Sprachmodellen komplett, das bezeichnet man als **Grounding Problem**.

Emily Bender und Alexander Koller beschreiben das Problem anhand eines metaphorischen Oktopus, der menschliche Sprachstrukturen lernt: <https://aclanthology.org/2020.acl-main.463.pdf> (§ 4, S. 5188f.)

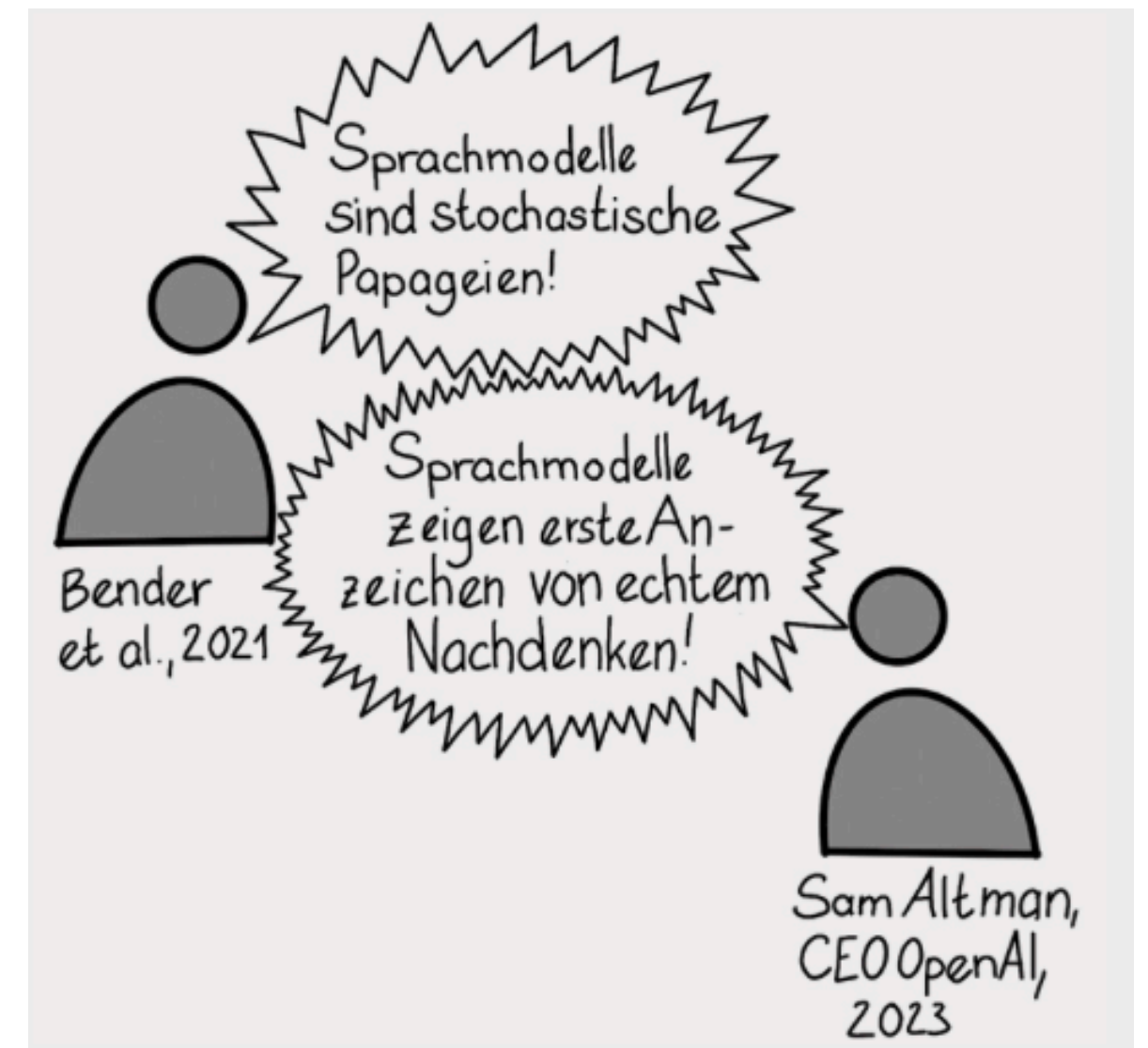


# Plappernde Papageien oder Superintelligenz?

Emily Bender gilt als eine der stärksten Kritikerinnen von Sprachmodellen und bezeichnet sie als **stochastic parrots** – Papageien, die lediglich das Gelernte nachplappern, ohne in irgendeiner Form zu verstehen, was sie da gerade sagen.

Andere Akteure, z.B. OpenAI-CEO Sam Altman, sehen Sprachmodelle teilweise sogar als Vorstufe zu **Artificial General Intelligence (AGI)**.

Yann LeCun (ehem. MetaAI) äußert sich auch eher skeptisch:  
[https://www.youtube.com/watch?v=4\\_gg83s\\_Do](https://www.youtube.com/watch?v=4_gg83s_Do)



# Ockhams Rasiermesser

In der Bewertung hilft uns ein alter Freund, Ockhams Rasiermesser – das Prinzip, dass zwischen zwei konkurrierenden Hypothesen diejenige zu bevorzugen sei, die mit **weniger** (insbes. nicht prüfbar) **Annahmen** auskommt.

Alle Fähigkeiten von aktuellen Sprachmodellen lassen sich durch die **technologischen Gegebenheiten erklären**. Diese bemerkenswerten Fähigkeiten sind zwar ein Beleg für die technische Meisterleistung vieler heller Köpfe, aber kein Hinweis auf tatsächliche Intelligenz in Sprachmodellen.



# Abschließende Gedanken

Moderne Sprachmodelle zeigen beeindruckende Fähigkeiten und entwickeln sich rasant weiter. Dennoch möchte ich davor warnen, ihnen blind zu vertrauen oder sie zu anthropomorphisieren.

Sprachmodelle haben keinen Bezug zur realen Welt, es kann keine **Referenzialität** im wahren Sinne stattfinden. Dementsprechend besteht auch keine kommunikative Intention – diese wird erst durch die menschliche Interpretation verliehen (vgl. ELIZA)!

Auch wenn Sprachmodelle immer besser werden, sollte man als Nutzer:in in der Lage sein, das Output **kritisch einzuordnen** und ggf. zu korrigieren.

