

OP

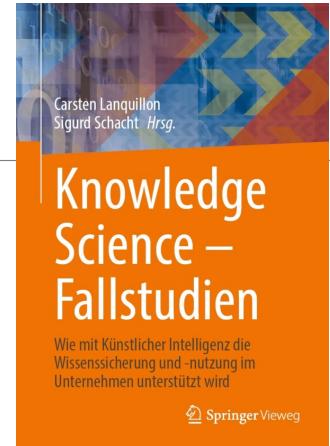
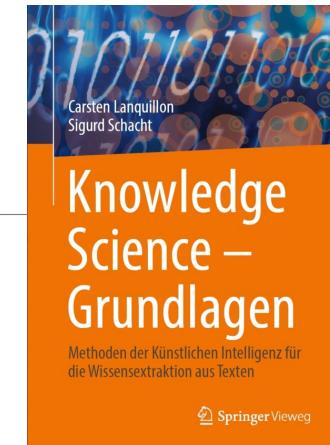
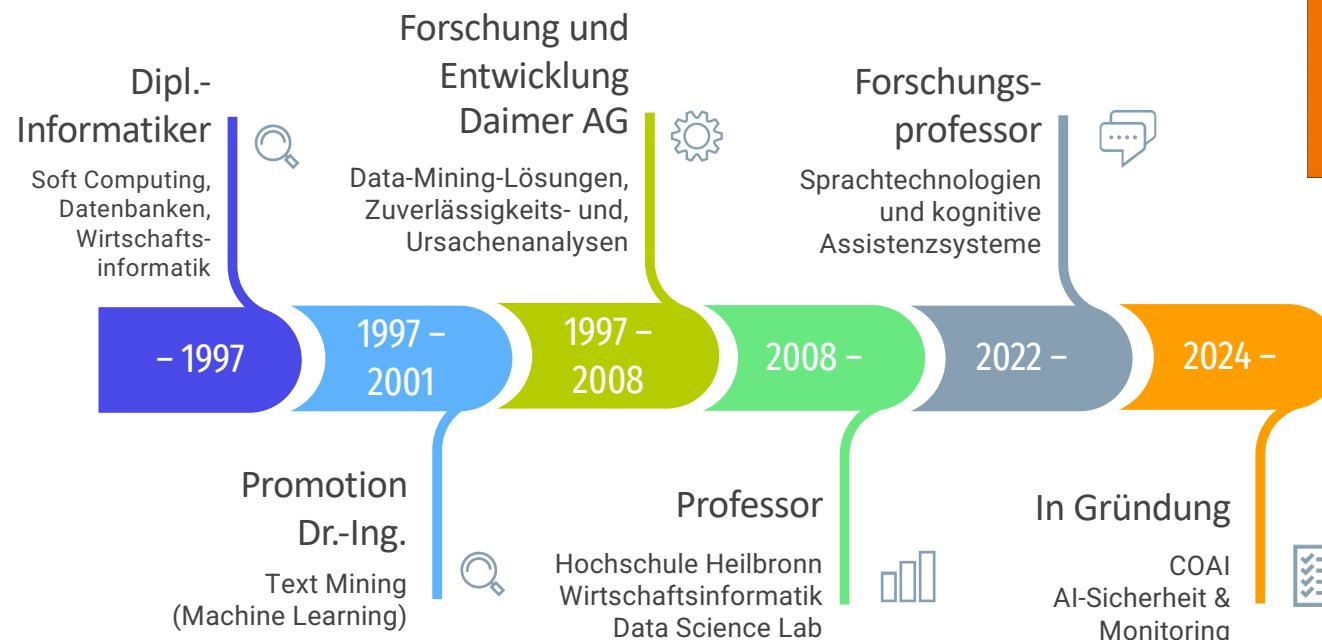
Erkläre in einfachen Worten, was ein Sprachmodell ist!



Ein Sprachmodell ist ein Computerprogramm, das entwickelt wurde, um menschliche Sprache zu verstehen und zu generieren. Es basiert auf künstlicher Intelligenz und nutzt statistische Muster und Regeln, um Texte zu analysieren und zu produzieren. Ein Sprachmodell kann beispielsweise verwendet werden, um Texte zu verstehen, Fragen zu beantworten oder sogar eigene Texte zu

› GENERATIVE KI: GROSSE SPRACHMODELLE FUNKTIONSWEISE UND ENTWICKLUNGSTRENDS

PROF. DR. CARSTEN LANQUILLON



<https://www.knowledgescience.ai>

GENERTIVE KI

BEISPIEL: MULTIMODALE FÄHIGKEITEN VON GPT4

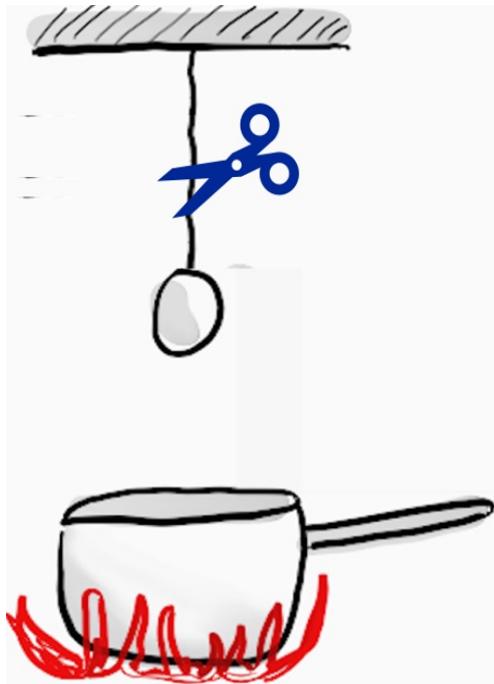


Bild basierend auf Henrik Kniberg (2024)

OP

You

Was wird auf dem Bild passieren?

ChatGPT

Auf dem Bild sieht es so aus, als würden die Scheren die Linie durchschneiden, an der die Maus hängt. Wenn die Linie durchgeschnitten wird, wird die Maus wahrscheinlich in den darunter befindlichen Topf fallen, der über einer Flamme zu stehen scheint. Es ist eine grafische Darstellung des Sprichworts „vom Regen in die Traufe“, was bedeutet, dass man aus einer schlechten Situation in eine noch schlechtere gerät.



PROMPT CHAT

+ New Prompt

Prompt

 Arena PII checker

Input Tokens: 245

Default Project



Du bist Experte für generative KI und große Sprachmodelle.
Erstelle eine Agenda für einen Vortrag mit einer Gliederungsebene und maximal fünf Gliederungspunkten.
Die Benennungen der einzelnen Gliederungspunkte soll sehr kurz und prägnant sein.
Ein Gliederungspunkt für Fragen und Diskussion soll nicht eingeplant werden.
Thema des Vortrags: "Generative KI: Große Sprachmodelle - Funktionsweise und aktuelle Herausforderungen"
Das Publikum hat wenig KI-Vorkenntnisse.
Die Dauer des Vortrags ist auf 20 Minuten begrenzt.
Gib die Gliederung tabellarisch mit Spalten für die Nummer, den Inhalt und die Dauer eines Gliederungspunkts in Minuten aus.
Eine Erläuterung des Aufbaus der Agenda ist nicht gewünscht.

KI-generierte Agenda

Versions

Dolphin Mixtral



Nummer

Inhalt

Nummer	Inhalt	Dauer (Minuten)
1	Einführung in generative KI	4
2	Grundlagen großer Sprachmodelle	5
3	Funktionsweise	4
4	Aktuelle Herausforderungen	4
5	Zukunftsansicht	3
1	Einführung in generative KI	4
2	Funktionsweise von großen Sprachmodellen	4
3	Transformer-Architektur und ihre Vorteile	4
4	Aktuelle Herausforderungen und Anwendungsbereiche	4
5	Sicherheit, Verzerrung und Limitationen	3
1	Anwendungsbeispiele und Eindrücke von großen Sprachmodellen	4
2	Zusammenfassung und Ausblick auf zukünftige Entwicklungen im Bereich generativer KI	4



EINFÜHRUNG IN GENERATIVE KI KI, MASCHINELLES LERNEN UND DEEP LEARNING

Künstliche Intelligenz

- Systeme, die ihr Umfeld wahrnehmen, schlussfolgern, Entscheidungen treffen und entsprechend handeln

Machine Learning

- Algorithmen, die automatisch aus Erfahrungen (Daten) lernen
 - das Automatisieren automatisieren

Deep Learning

- Künstliche Neuronale Netze mit vielen Schichten
 - die Merkmalserzeugung automatisieren

Generative KI

Erzeugung neuer Inhalte
(Text, Code, Bilder, Audio, Video)

Eingabe

Verarbeitung

Ausgabe



HOCHSCHULE HEILBRONN

Sensoren

KI-System (Software und ggf. Hardware)

Aktuatoren



Bild
Video

Sprache
Text

weitere
Sensoren

weitere
Daten

Wahrnehmung

Informationsverarbeitung:
Problemlösung und
Handlungsentscheidung

Interaktion

**Multi-modale „Sprach“-Modelle
(Sequence-to-Sequence-Modelle)**

Verstehen Eingangssignale und erzeugen Ausgangssignale

Wissensrepräsentation

Wissensbasis



Umwelt (environment)

KOMPONENTEN EINES KI-SYSTEMS

GRUNDLAGEN GROSSER SPRACHMODELLE

WAS IST EIN SPRACHMODELL (LANGUAGE MODEL)?

Aufgabe Bestimmung der Wahrscheinlichkeiten von Wörtern in einem Kontext
– insbesondere die Fortsetzung einer Sequenz von Wörtern

Beispiel

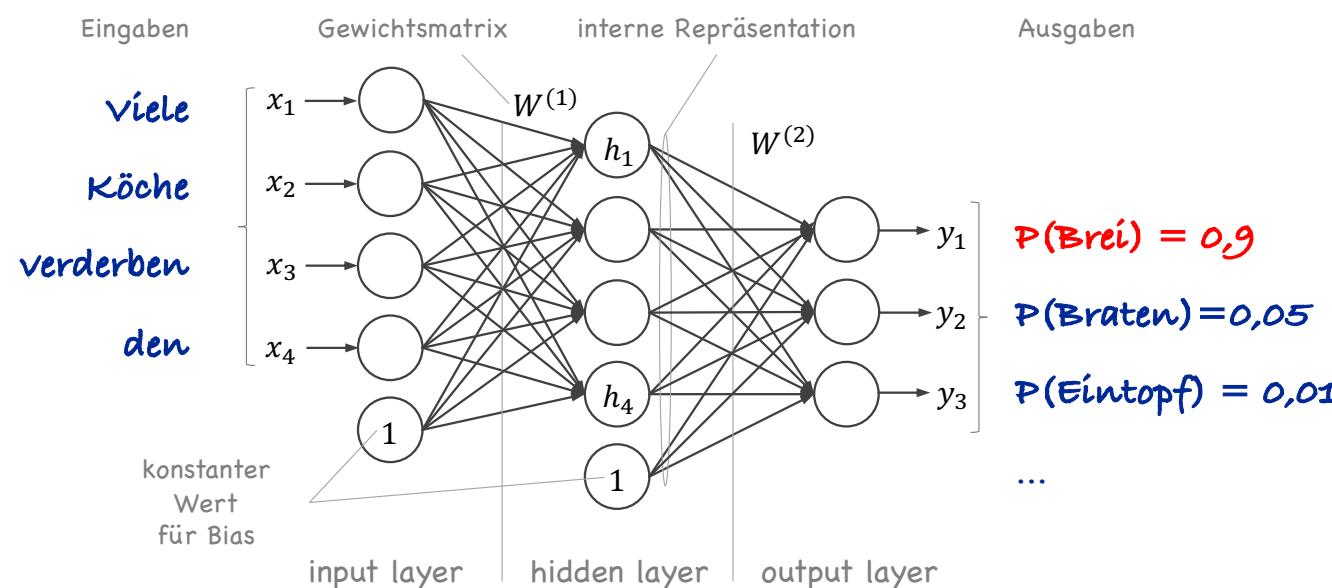
$P(\text{Wort} | \text{„Wie kann ich ChatGPT“}) = ?$



- Wie kann ich ChatGPT
- wie kann ich chatgpt **nutzen**
- wie kann ich chatgpt **nutzen ohne anmeldung**
- wie kann ich chatgpt **4 nutzen**
- wie kann ich chatgpt **4 kündigen**
- wie kann ich chatgpt **auf deutsch umstellen**

ANSÄTZE FÜR SPRACHMODELLE

- > **Klassisch** Auszählen (Bestimmung von Wort-Kookenzen) → **Problem: Kombinatorische Explosion**
- > **Modern** Modellierung mithilfe Neuronaler Netze



Die Gewichte sind die gesuchten Parameter des Sprachmodells

Hier:

> $W_{(4+1) \times 4}^{(1)}$
→ 20 Parameter

> $W_{(4+1) \times 3}^{(2)}$
→ 15 Parameter

> Summe: 35 Parameter

TEXTREPRÄSENTATION TOKENIZING UND EMBEDDINGS

Text 'Viele Köche verderben den'

Token 

['V', 'iele', 'Kö', 'che', 'ver', 'der', 'ben', 'den']

Token-IDs [53, 25641, 65994, 1557, 2807, 1126, 8123, 3453]

	v	iele	Kö	che	ver	der	ben	den
0	-0.002734	-0.041568	0.014458	-0.024257	-0.021238	0.013269	-0.007785	0.001318
1	0.000411	0.003102	-0.056067	-0.002339	0.024953	0.031441	0.028118	0.005436
2	0.001388	-0.007580	-0.006530	-0.010476	0.001380	-0.010363	-0.019278	-0.011350
3	0.018867	-0.020281	0.024415	0.031122	0.002332	0.019248	-0.006254	-0.002411
4	0.009468	-0.027377	0.028269	-0.015731	-0.010320	0.004745	0.023547	0.022701
...

PRE-TRAINING VON SPRACHMODELLEN

REZEPT ZUR ERSTELLUNG VON SPRACHMODELLEN

- > Man nehme viele Daten (Textkorpus) → Internet
- > Wähle eine Architektur → Transformer-basiert
- > Wiederhole bis Abbruchkriterium erreicht
 - Maskiere die Text-Sequenzen
 - Prognostiziere maskierten Wörter
 - Berechne Fehler (wahre Werte bekannt)
 - Passe die Gewichte an (Back-Propagation)→ Selbstüberwachtes Lernen

DIE PROGNOSE DES NÄCHSTEN TOKEN EINE BEKANNTEN AUFGABE AUS DER (GRUND-)SCHULE

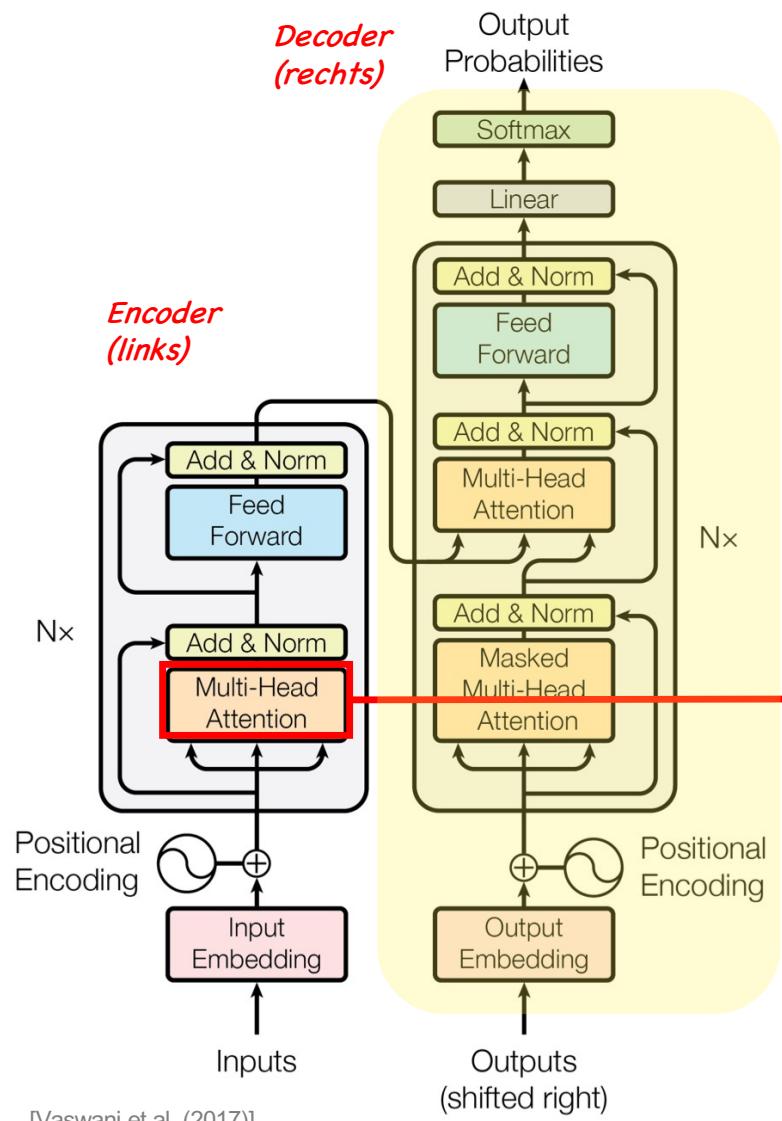
Die Informatik ist eine [REDACTED]. Dabei werden Informationen automatisch verarbeitet.

Sie werden also miteinander verrechnet, in einer [REDACTED] dargestellt, gespeichert und vieles anderes mehr. Der Begriff „Informatik“ setzt sich zusammen aus den [REDACTED] „Information“ und „Automatik“. Die Informationen werden also automatisch verarbeitet.

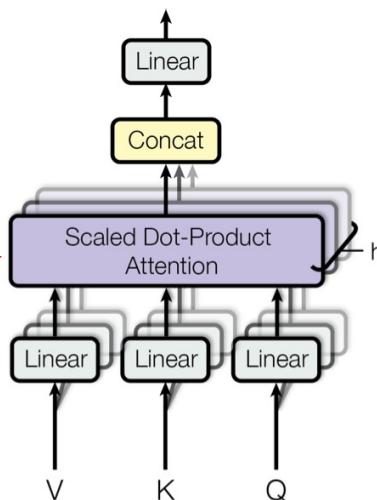
Ein Beispiel ist der [REDACTED]: Viele Stationen auf der Welt messen die Temperatur, die Feuchtigkeit der Luft, den [REDACTED], den Wind und viele andere Dinge. Die Daten liefern die Geräte selbstständig an einen zentralen [REDACTED]. Dieser ist so programmiert, dass er daraus selbstständig eine Wettervorhersage erstellen kann.

[REDACTED] brauchen die Informatik für die Berechnung ihrer Brücken, Ärzte brauchen sie für die Planung ihrer [REDACTED] und so weiter. Die Informatik ist aus unserem modernen Leben nicht mehr wegzudenken.

Bei der [REDACTED] stoßen wir auf verschiedene Abkürzungen: ICT, sprich: aj-si-ti, steht für das englische information and [REDACTED] technology, auf Deutsch also Informations- und Kommunikations-[REDACTED]. IT, sprich: aj-ti, ist eine noch kürzere Form und steht für Informationstechnik.



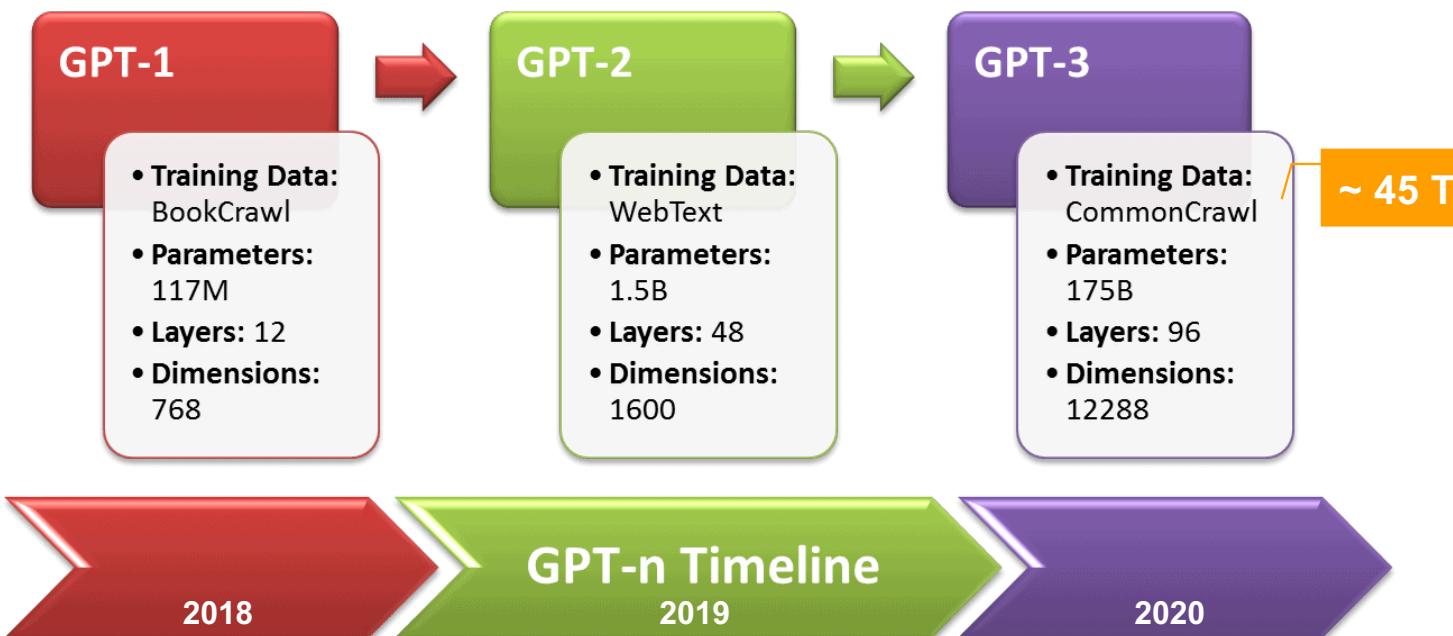
KOMPLEXE ARCHITEKTUR TRANSFORMER MIT SELF-ATTENTION



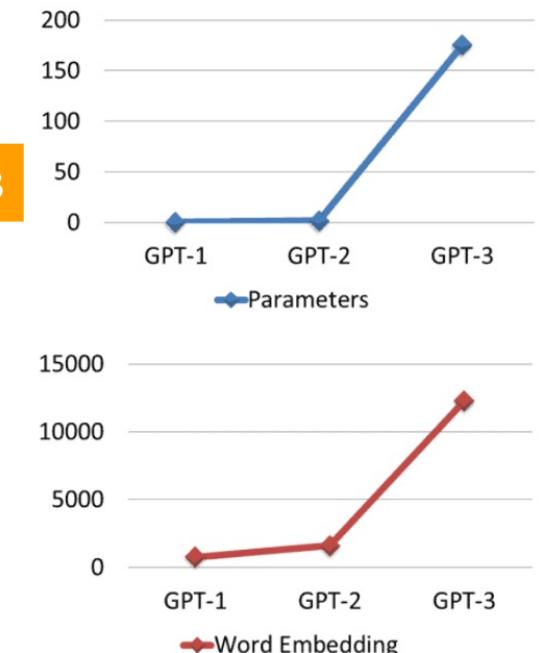
- > Spezielle Architektur für die Prognose des nächsten Tokens (auto-regressiv)
- > Parallele Verarbeitung von Eingabesequenzen
- > Self-Attention-Blöcke erzeugen kontext-spezifische Embeddings
- > Ursprung: Übersetzung
- > Wird inzwischen für viele Aufgaben und Datentypen verwendet

GPT – GENERATIVE PRETRAINED TRANSFORMER DECODER-ONLY ARCHITEKTUREN

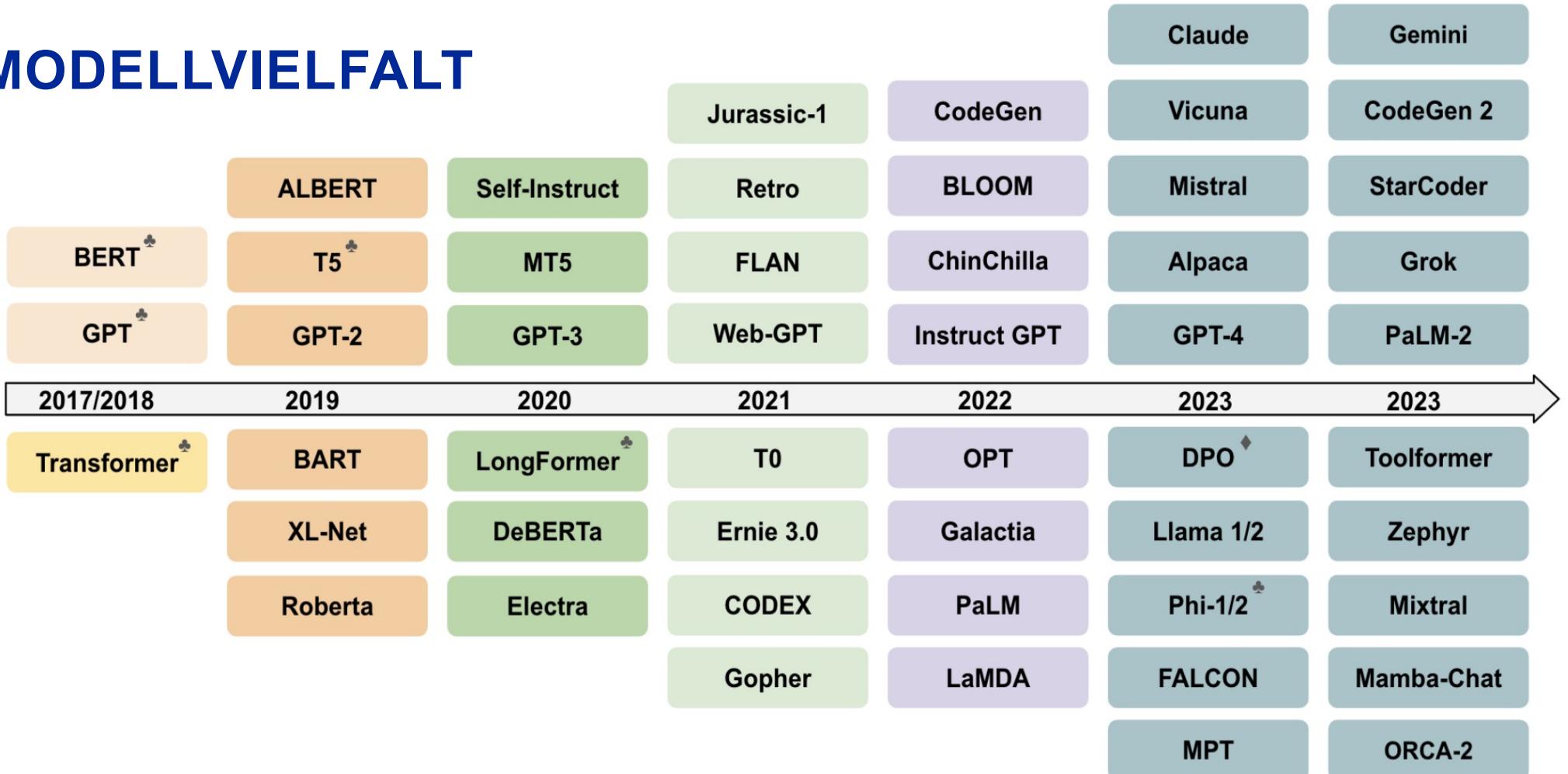
Skalierung: Durchschlagender Erfolg durch Größe (Anzahl trainierbarer Parameter)



Quelle: <https://businesssolution.org/de/gpt-3-statistics/>



MODELLVIELFALT



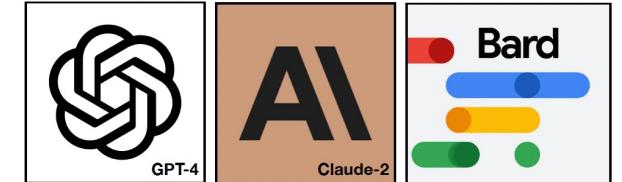
Quelle: Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). *Large Language Models: A Survey* (arXiv:2402.06196). arXiv. <http://arxiv.org/abs/2402.06196>

DURCHBRUCH IN DER KI GROSSE SPRACHMODELLE HABEN EMERGENTE FÄHIGKEITEN

Große Sprachmodelle (Large Language Models, LLMs)

- > Lösen sehr komplexe Aufgaben
- > Bemerkenswert:

Emergente Fähigkeiten, d.h. sie können Aufgaben lösen, für die sie nicht trainiert wurden



→ „*The unreasonable effectiveness of [scaling] data [and parameters]*“

→ **Viele Daten + großes Netz + sehr viel Rechenleistung = General Purpose AI**

VERWENDUNG GROSSER SPRACHMODELLE PROMPTING, FINE-TUNING, ALIGNMENT

> **Prompt Engineering**

Systematischer Ansatz zur Formulierung der Anfrage (Aufgabe)
→ bessere Ergebnisse

> **Fine-Tuning**

- Fokussierung auf bestimmte Inhalte
- Befolgen von Instruktion
- Feedback von Nutzerinnen und Nutzern

> **Alignment**

Sprachmodelle sollen unseren Vorgaben und Wünschen entsprechen
→ Z.B. die HHH-Alignment-Kriterien: helpful, honest, harmless

[Askell et al. (2021): „A General Language Assistant as a Laboratory for Alignment“]

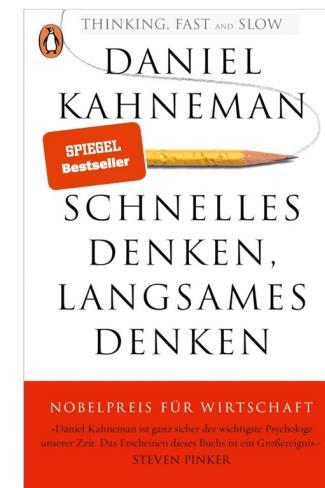
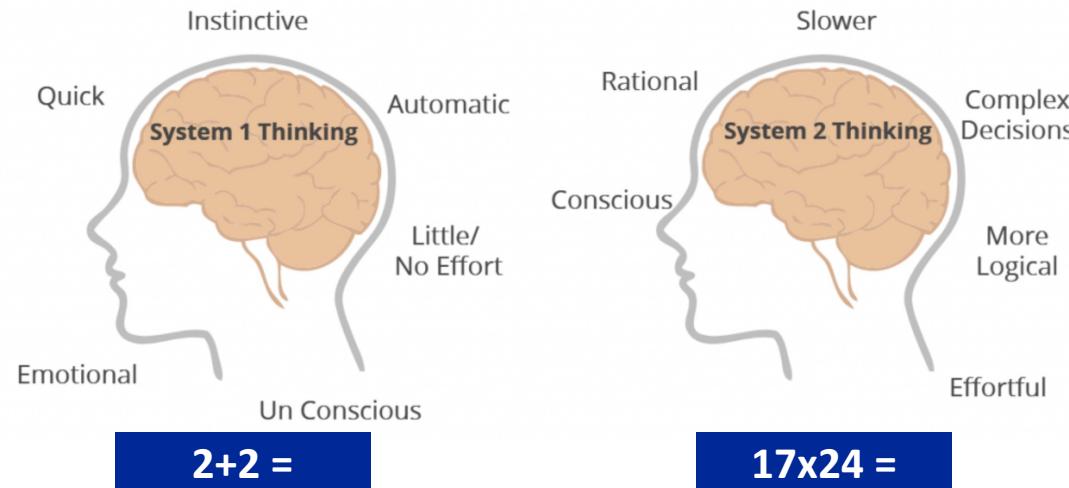
GRENZEN AKTUELLER SPRACHMODELLE

Korrektheit	Bias (Verzerrung)	Wissensstand	Abhängigkeit
<ul style="list-style-type: none"> Sprachmodelle neigen zum Halluzinieren, d.h. sie erfinden „Fakten“ Sprachmodelle antworten immer und mit großer Überzeugung Evaluierung qualitativer Ausgaben schwierig 	<ul style="list-style-type: none"> Sprachmodelle sind ein Spiegel der Gesellschaft, d.h. sie bilden z.B. Stereotypen aus den Daten ab Unerwünschte Sprache 	<ul style="list-style-type: none"> Keine Datenbank im klassischen Sinn Eingefrorenes Wissen zum Zeitpunkt der Erstellung Keine Anbindung an neues Wissen Begrenztes Kontextfenster 	<ul style="list-style-type: none"> Wenige Anbieter auf Grund hoher Kosten Monopolbildung durch Vorsprung von OpenAI Eingegebene Daten werden zum Training verwendet Guardrails / Content-Filter durch Anbieter

Ressourcenverbrauch und Datenschutz

SIND GROSSE SPRACHMODELLE INTELLIGENT?

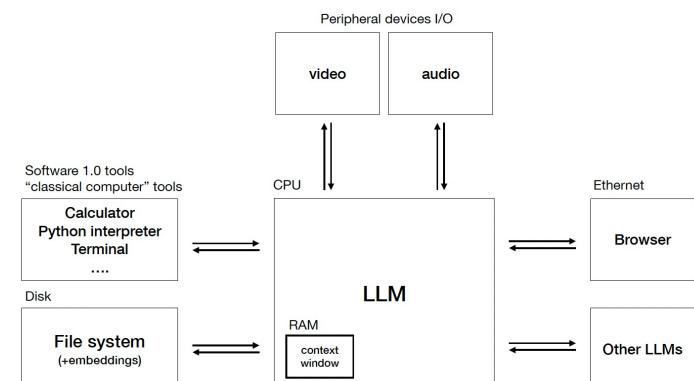
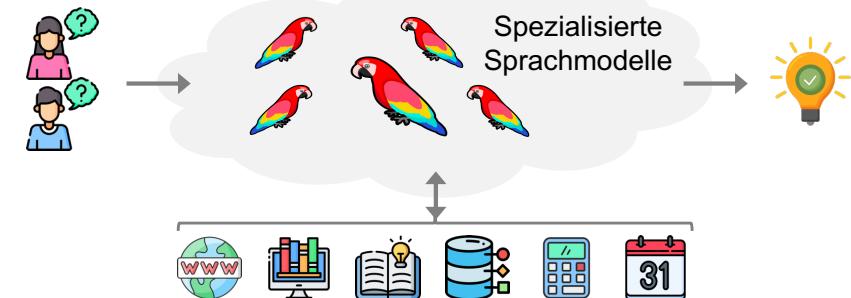
- Erfolge durch Skalierung – aber: „Getting better at guessing isn't the same as understanding“
[Steve Jones (2023): „GPT still understands nothing“, <https://blog.metamirror.io/gpt-still-understands-nothing-3eed694e00fb>]
- + LLMs (embeddings and their interactions) are a model of understanding. [Geoffrey Hinton]
- ! Aktuelle LLMs sind lediglich „System-1-Denker“ [Andrej Karpathy]



AUSBLICK

WOHIN GEHT DIE ENTWICKLUNG?

- > **Kollaborierende Agententeams**
 - LLM-basierte Multi-Agenten-Systeme mit Zugriff auf ausgewählte Wissensbestände und Werkzeuge
- > **Custom-LLMs**
- > **LLM Security**
 - z.B. Jailbreaking, Prompt Injection, Backdoors, Data Poisoning, Adversarial Attacks
- > **LLM OS**
 - Sprachmodell als Betriebssystems eines Computers [Andrej Karpathy]



VIELEN DANK!



Bei Fragen kontaktieren Sie bitte:

Prof. Dr. Carsten Lanquillon

Fakultät für Wirtschaft | Wirtschaftsinformatik
carsten.lanquillon@hs-heilbronn.de

<https://www.knowledgescience.ai>

