

Gliederung



- Das Internet
- Algorithmen und Berechenbarkeit
- LLM
- Compute
- Ausblick

Internet



"big data"



Bilder

Videos

News

Einfach erklärt

Analytics

Beispiele

Abnehmen

English

Vc >

Alle Filter

Suchfilter

SafeSearch

Ungefähr 452.000.000 Ergebnisse (0,40 Sekunden)

Unter Big Data versteht man Daten, die in großer Vielfalt, in großen Mengen und mit hoher Geschwindigkeit anfallen. Dies ist auch als die drei V-Begriffe bekannt (Variety, Volume, Velocity).



Oracle

<https://www.oracle.com> > Oracle Deutschland > Big Data

Was versteht man unter Big Data? | Oracle Deutschland



Informationen zu hervorgehobenen Snippets • Feedback geben



Boersenwissen Aktuell

<https://www.boersenwissen-aktuell.de> > Megatrend KI

Die Top 10 Big Data Aktien | Big Data und KI Aktienmarkt

Daten sind das neue Gold – exklusiver Report deckt die Top 10 Unternehmen auf. Investieren Sie in den wachsenden Markt der **Big Data** und KI-Technologie

Megatrend Wasserstoff

Brutal hohe Kursanstiege möglich

Mehr als 500% Gewinn in 12 Monaten

Aktien unter 10 Euro

Kleine Aktien - große Wirkung

Die ultimative 9-Punkte-Checkliste



Jobworld

<https://www.jobworld.de> > Big Data > Jobs

Berechenbarkeit

Bitte extrahieren die wesentlichen Informationen und schreibe diese in ein JSON.

PAW Patrol (wörtlich übersetzt: Pfoten-Patrouille), oder PAW Patrol – Helfer auf vier Pfoten (Super-RTL-Version), ist eine kanadische Computeranimationsserie von Keith Chapman. Die Serie wird von Spin Master Entertainment und Guru Studio produziert. In Kanada strahlt der Sender TVOKids die Serie aus, die Pilotfolge der Serie lief im August 2013. Die ersten Folgen wurden am 12. August 2013 auf Nickelodeon in den USA ausgestrahlt. In Deutschland wurde sie zunächst vom 11. November 2013 bis Ende Oktober 2016 bei Nickelodeon ausgestrahlt.[1] Seitdem wird sie in teilweise geänderter Synchronisation bei Super RTL und Toggo plus[2] im Free-TV und von Nick Jr. im Pay-TV sowie bei Amazon Prime, Paramount+ und Netflix gezeigt. Außerdem veröffentlicht Leonine Distribution seit 2017 einige Folgen der Serie als Hörspiele mit dem Sprecher Tobias Diakow.[3]

Seit 2023 gibt es ein Spin-off namens Rubble & Crew, das seit dem 25. Februar 2024 auch in Deutschland ausgestrahlt wird.

< 2/2 >



json

Code kopieren

```
{
  "Serie": "PAW Patrol",
  "Beschreibung": "Kanadische Computanimationsserie von Keith Chapman.",
  "Produktion": {
```

```
},  
  "Spin-off": {  
    "Name": "Rubble & Crew",  
    "Erstausstrahlung_Deutschland": "Seit 25. Februar 2024"  
  }  
}
```



Spiele ein Frage-Antwort-Spiel mit mir und teste mein Wissen basierend auf dem JSON. Gib mir jeweils 3 Antwortmöglichkeiten. Zähle meine Punkte. Stelle die erste Frage.

< 4/4 >



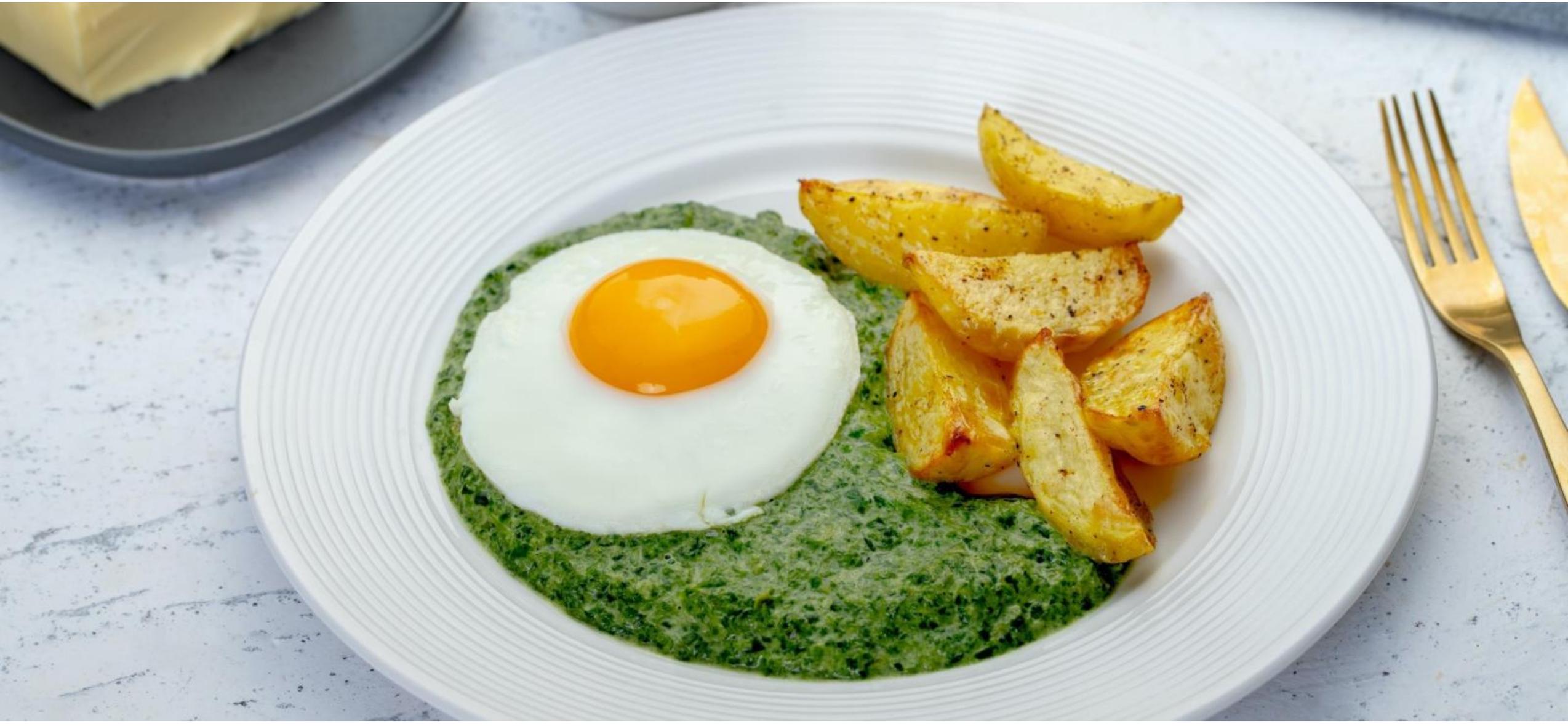
Frage 1: Wer ist der Schöpfer der Serie PAW Patrol?

- A) Keith Chapman
- B) Toby Braun
- C) Sam Miller

Bitte wähle die richtige Antwortoption (A, B, oder C).



Bilder verstehen



Large Language Models (LLM)

Grundlagen LLM – Beispiel Llama 3

■ Fakten zum „Open Parameter“ LLM Llama 3:

- ⇒ 70 Mrd. Parameter in Transformer Architektur (Attention Mechanismus)
- ⇒ Trainiert auf 15 Billionen Tokens. Mit der Aufgabe basierend auf den letzten k Wörtern das nächste Wort vorherzusagen

$$P(w_t \mid w_{t-k}, \dots, w_{t-1})$$

Eric: Hi Tom! How are

Eric: Hi Tom! How are you?

Eric: Hi Tom! How are you? Tom:

Eric: Hi Tom! How are you? Tom: Sorry,

Eric: Hi Tom! How are you? Tom: Sorry, who

Eric: Hi Tom! How are you? Tom: Sorry, who are

Eric: Hi Tom! How are you? Tom: Sorry, who are you

Eric: Hi Tom! How are you? Tom: Sorry, who are you again?

Several iterations of an LLM predicting the next word of a sentence

■ Mit H100 Grafikkarten für 6,4 Millionen GPU Stunden

- ⇒ Z.B. 3000 GPUs für 3 Monate

■ LMMs sind sehr teuer und aufwändig zu trainieren

- ⇒ Veralteter Wissensstand

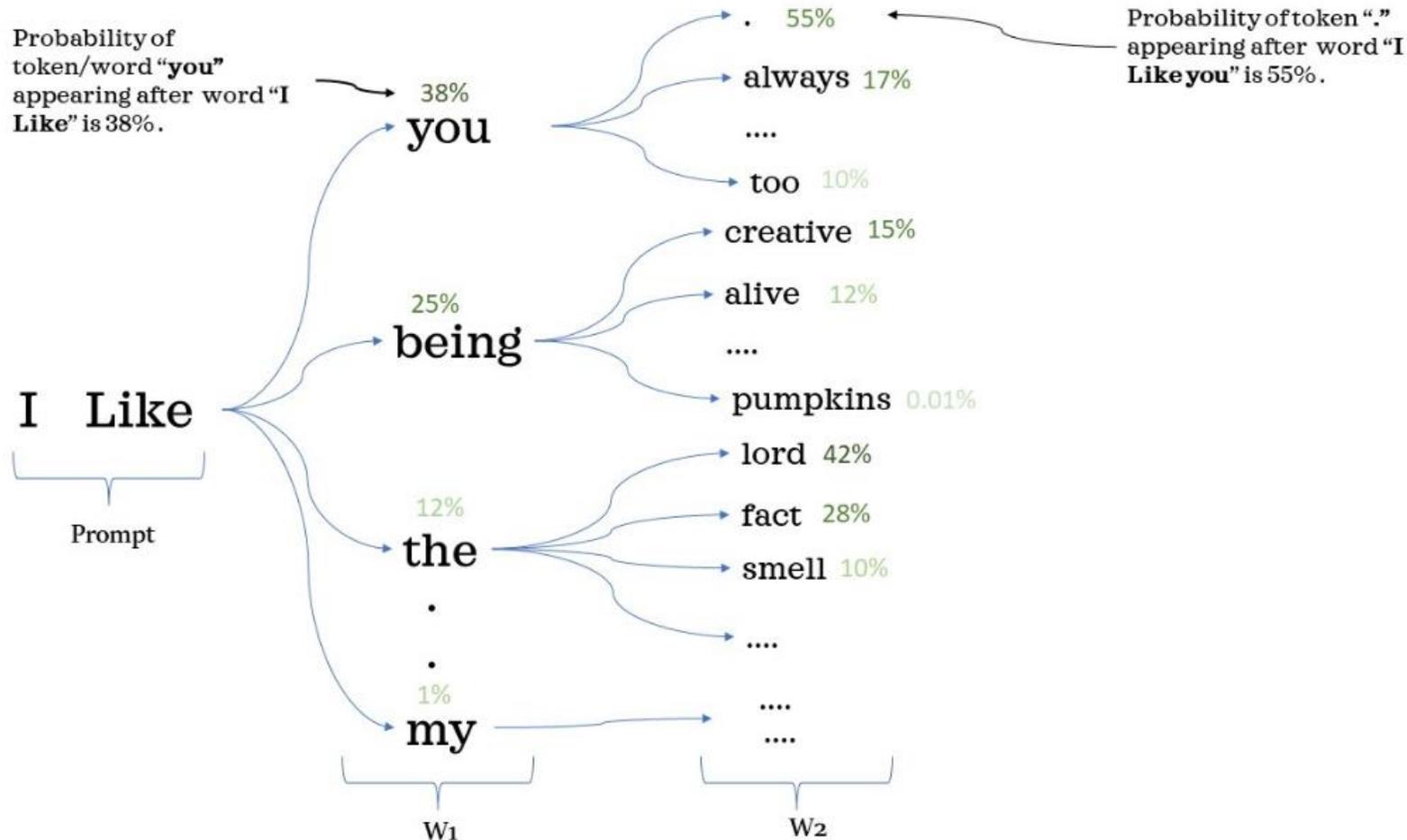
■ Keine Trennung von Fachwissen, Logik und Sprachfähigkeiten

■ Schlechte Erklärbarkeit

Aufgaben für Wort Vorhersage

Aufgabe	Beispiel
Grammatik	Wenn ich Zeit habe, dann (schwimme , Sport)
Semantik	Ich ging in den Laden, um Papaya, Drachenfrucht und (Äpfel , Eichhörnchen) zu kaufen
Wissen	Die Hauptstadt von Frankreich ist (Paris , London)
Sentimentanalyse	Film Bewertung: Ich war die ganze Zeit gefesselt und gespannt wie es weiter geht. Der Film war wirklich (gut , schlecht)
Übersetzung	Das Wort für "hübsch" auf Spanisch ist (bonita , hola)
Raumvorstellungsfähigkeit	Thomas ging in die Küche, um Tee zu machen. Neben Thomas stand Anton. Anton verließ die (Küche , Kammer)
Mathematik	$3 + 8 + 4 =$ (15 , 11)

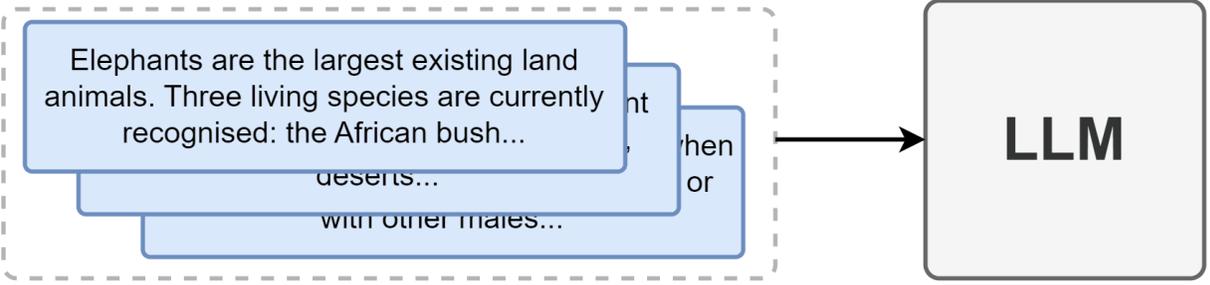
Grundlagen LLM – Generierung des nächsten Wortes



- Inference: der Prozess, bei dem ein LLM auf Eingabedaten angewendet wird, um Ausgaben zu erzeugen.
- Greedy Search
 - ⇒ Generiert immer das nächste Wort basierend auf den nächsten Wahrscheinlichkeit
- Beam Search
 - ⇒ Approximation der besten Lösung durch Exploration von mehreren Pfaden gleichzeitig.

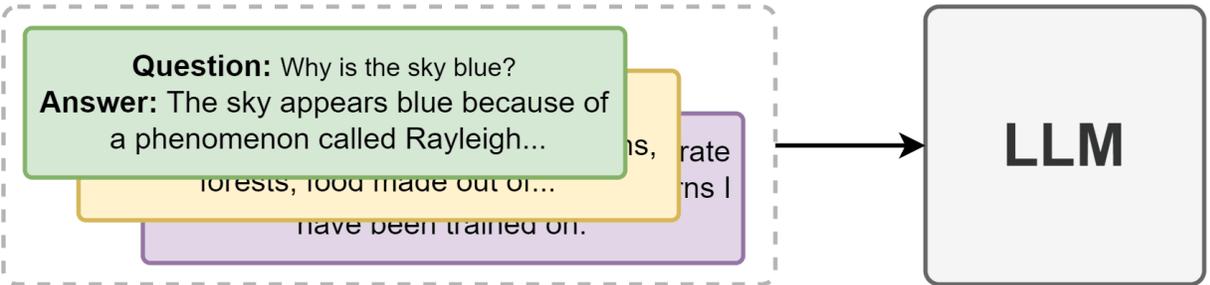
Pre-training and Post-training

Stage 1: Pretraining (Expensive, Unaligned)



Arbitrary Unstructured Data

Stage 2: Instruction Tuning (Cheap, Where Alignment Happens)

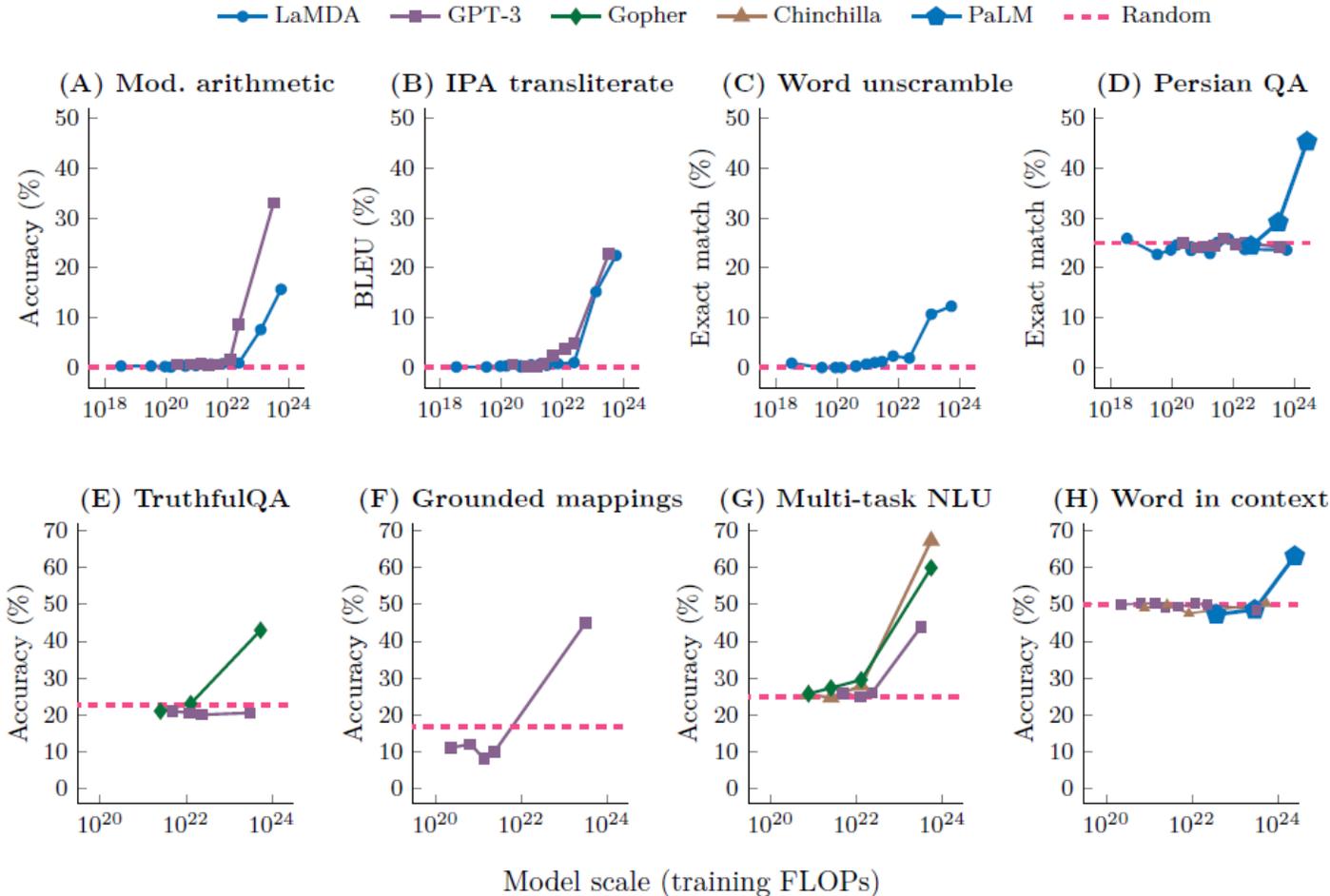


Task-Related Data
(Sample interactions, RLHF, etc.)

Supervised Fine-Tuning (SFT)

Aufgabe (Task)	Eingabeaufforderung (Prompt)	Erwartete Ausgabe (Expected Output)
Übersetzung (Translation)	Übersetzen Sie den folgenden Satz ins Deutsche: "Der Hund rannte hinter der Katze her."	El perro corrió tras el gato.
Zusammenfassung (Summarization)	Fassen Sie folgenden Text zusammen: "Alice wurde es allmählich sehr langweilig, neben ihrer Bilder oder Gespräche?"	Ein junges Mädchen namens Alice ...
Sentimentanalyse (Sentiment Analysis)	Klassifizieren Sie die Stimmung der folgenden Bewertung als positiv oder negativ: "Das Essen"	Positiv
Fragenbeantwortung (Question Answering)	Was ist die Hauptstadt von Frankreich?	Paris
Grammatikkorrekturen (Grammar Correction)	Verbessere diesen Satz: "Er nicht weiß, wo die Bibliothek ist."	Er weiß nicht, wo die Bibliothek ist.

LLM emerging ability

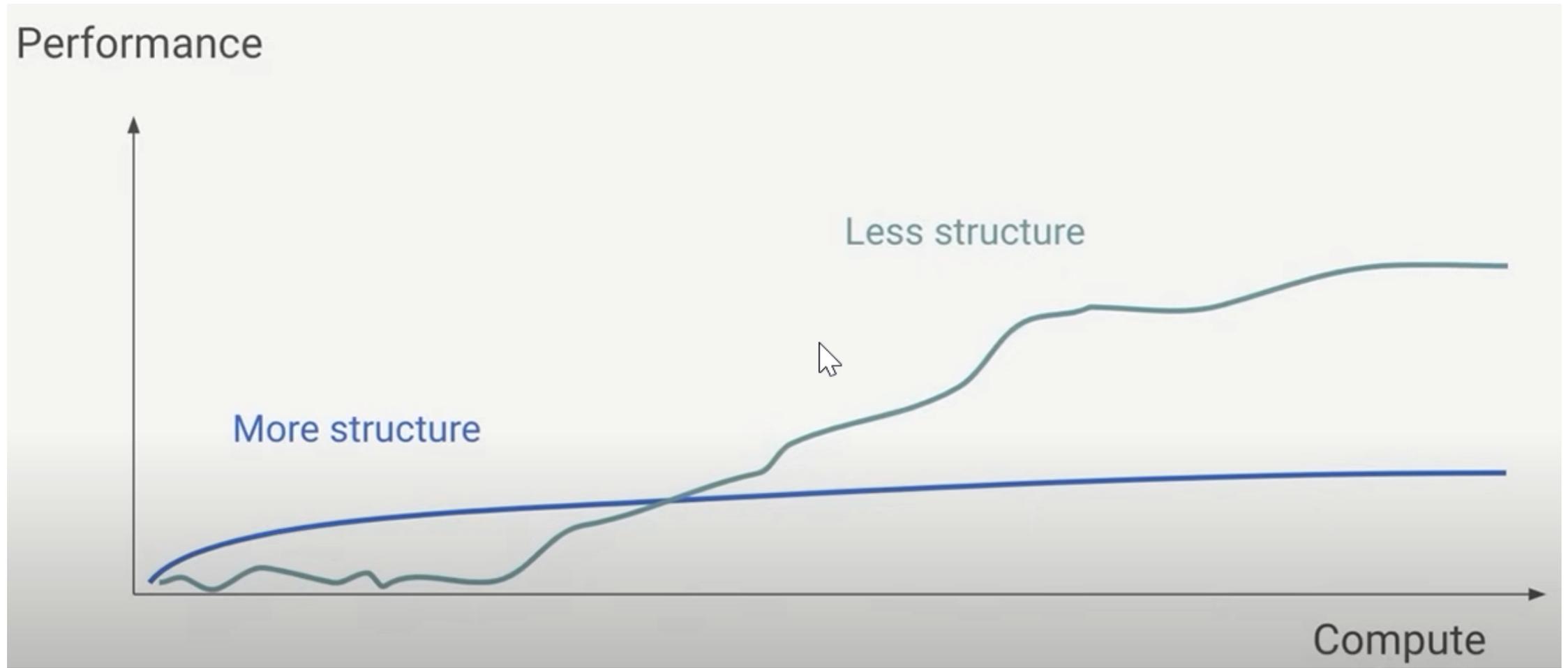


- Verstehen komplexer Anfragen
- Generierung kohärenter Inhalte
- Lernfähigkeit aus Interaktionen
- Sprachübergreifende Fähigkeiten
- Einsatz in spezialisierten Domänen
- Ethik und Verantwortung

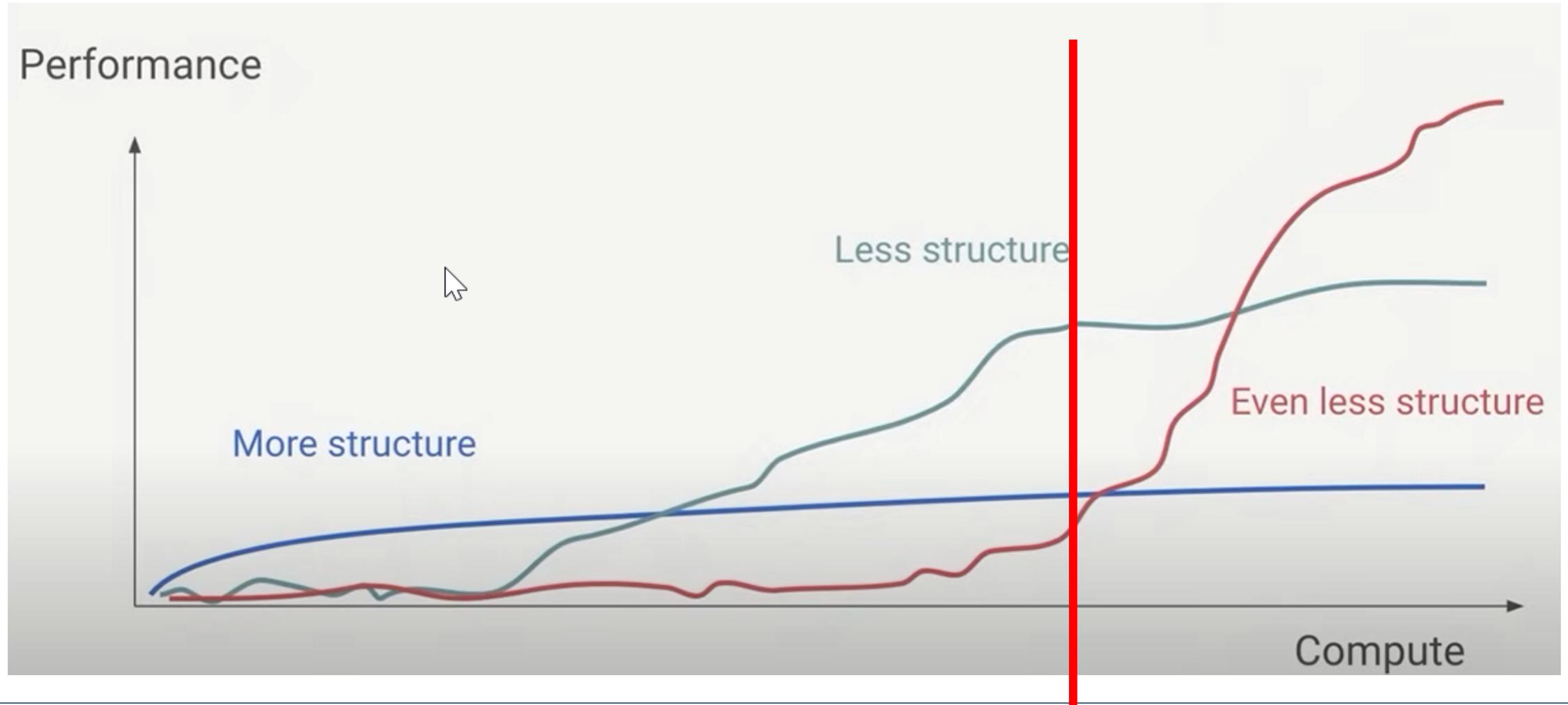
Ernüchternde Wahrheit

- Der Fortschritt in der KI in den letzten Jahrzehnten basiert auf:
 - ⇒ Der Entwicklung von immer allgemeineren Methoden mit immer weniger Modelannahmen
- Immer mehr Daten und Compute (scale up)
- Akzeptiere diesen Trend und nutze ihn zu deinem Vorteil

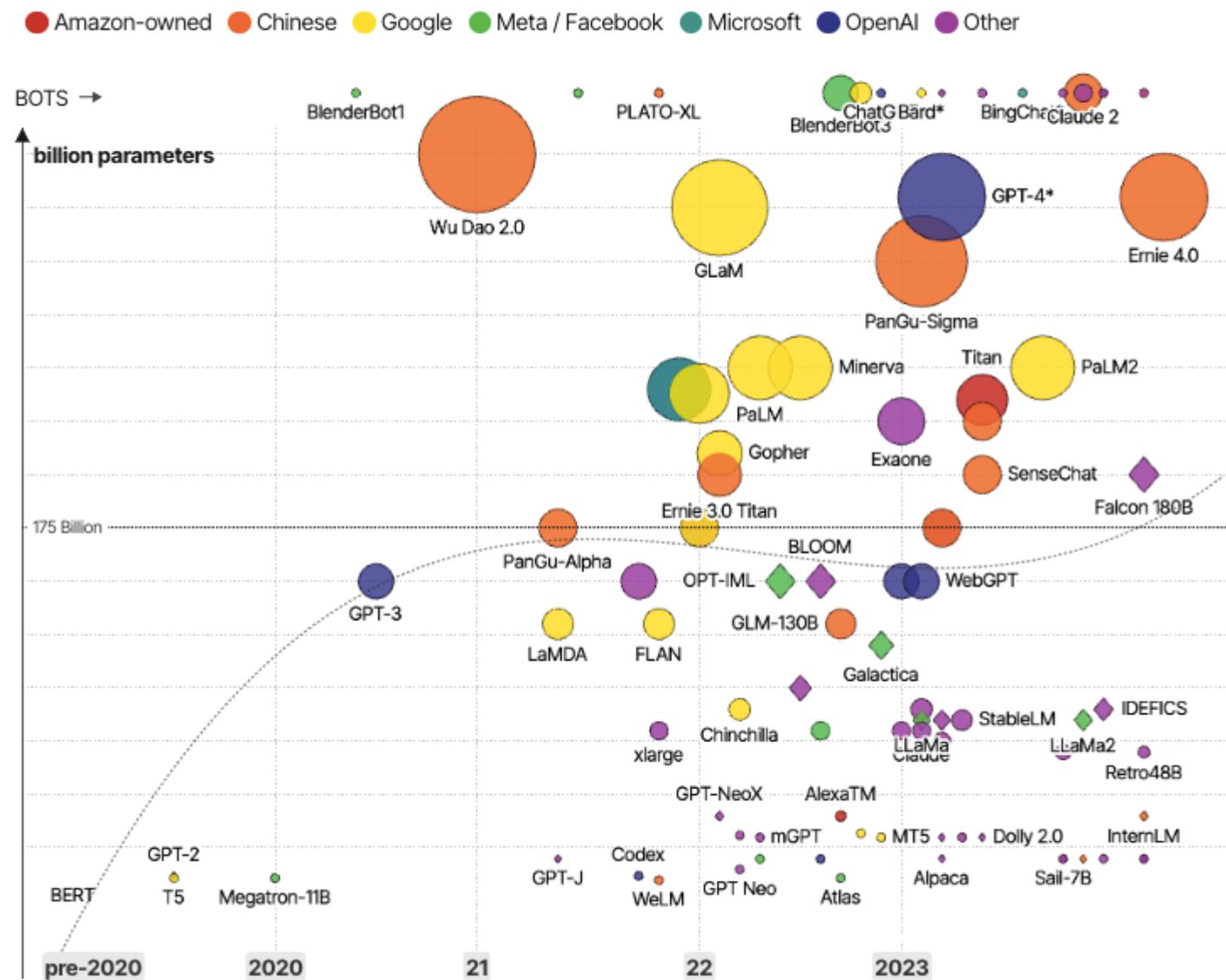
Ernüchternde Wahrheit



Ernüchternde Wahrheit



The Rise and Rise of A.I. Large Language Models (LLMs) & their associated bots like ChatGPT

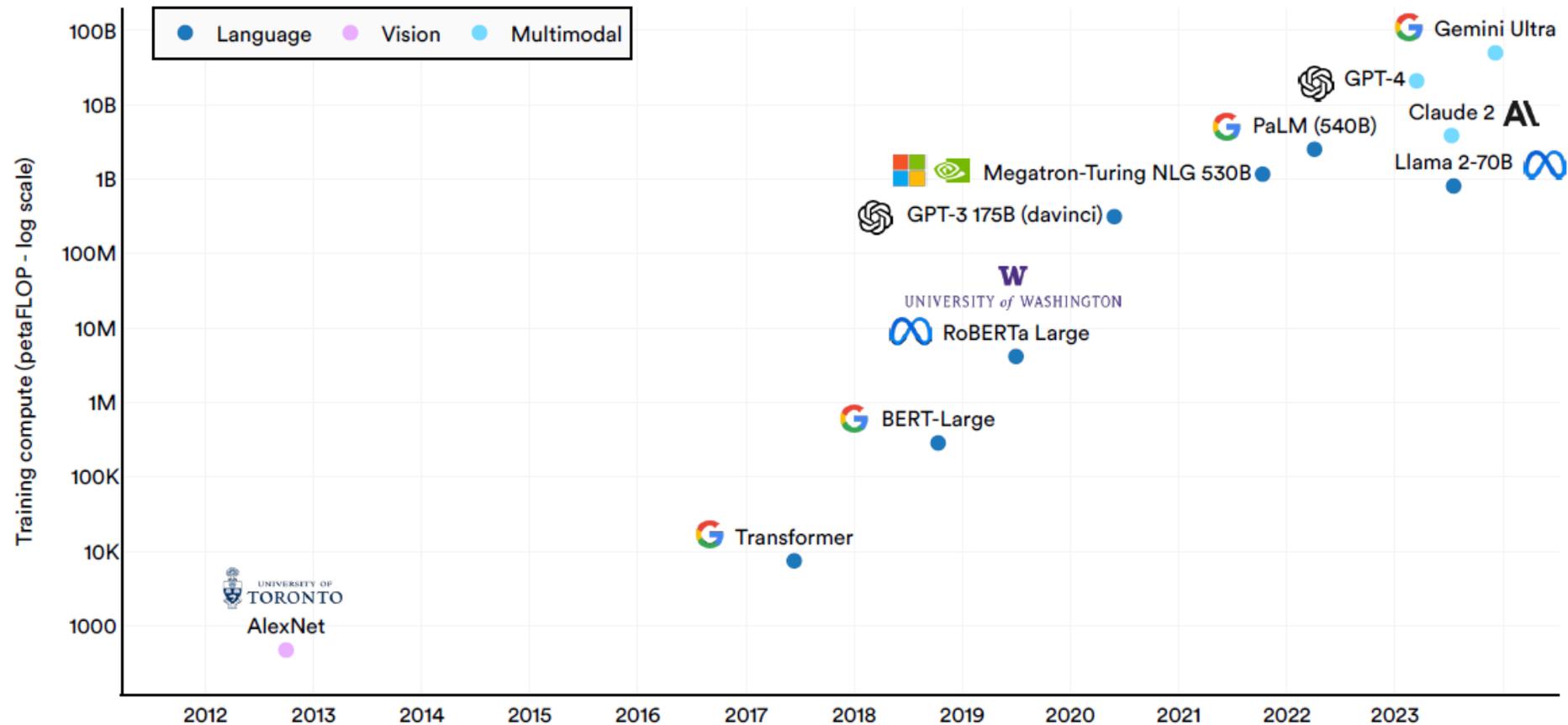


Size matters

Compute matters

Training compute of notable machine learning models by domain, 2012–23

Source: Epoch, 2023 | Chart: 2024 AI Index report



Big tech is the biggest buyer, hoarding H100s in the thousands

Estimated H100 GPUs shipped to select buyers in 2023

- Produktionskosten pro H100: \$ 3300
- Durchschnittlicher Verkaufspreis pro H100: \$ 30000



Stargate

COMPUTERBASE › WIRTSCHAFT

Projekt „Stargate“: Microsoft und OpenAI planen Supercomputer für 100 Mrd. Dollar

31.3.2024 14:12 Uhr  Andreas Frischholz

 105 Kommentare



Bild: Nvidia

- OpenAI und Microsoft bauen ein riesiges KI-Zentrum namens "Stargate"
- Kosten von 100 Milliarden Dollar
- Stargate soll 2028 in Betrieb
- Es wird Chips von NVIDIA und Microsofts eigene KI-Chips integrieren
- Ziel ist es, eine leistungsstarke Maschine zur Verarbeitung massiver KI-Modelle zu schaffen

Großprojekte



■ Stuttgart 21

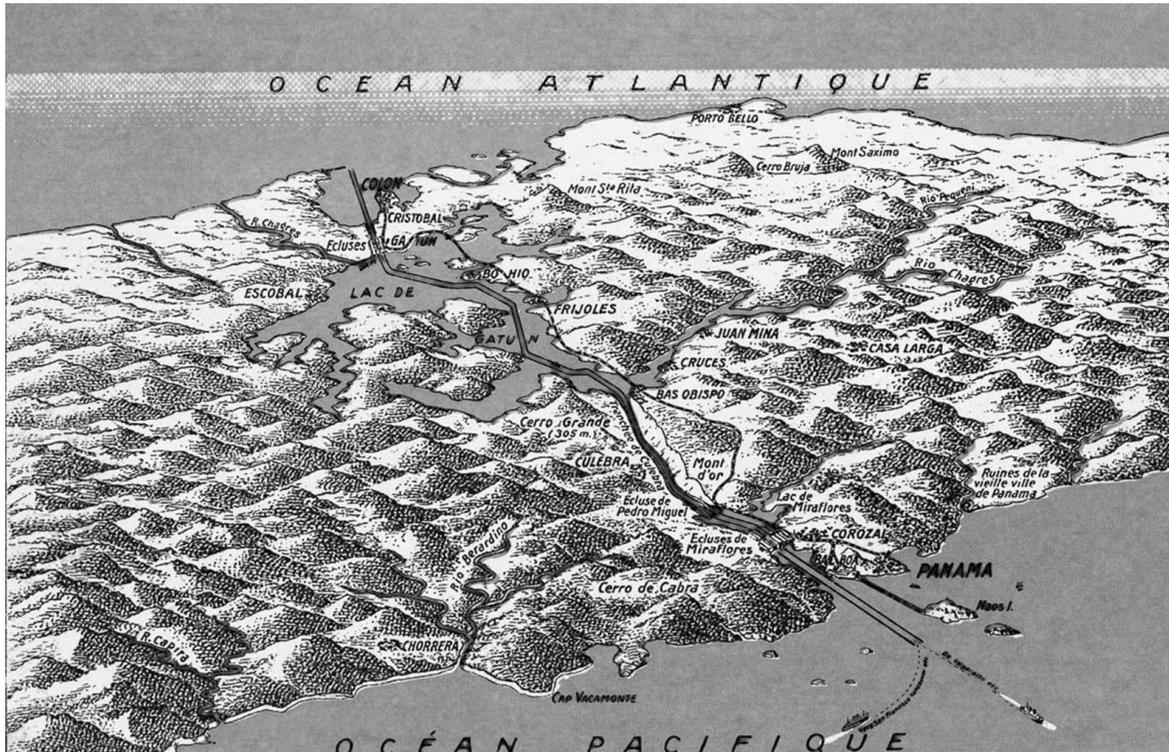
⇒ Kosten ca. 10 Mrd. €



■ Gibe-III-Staudamm (Äthiopien)

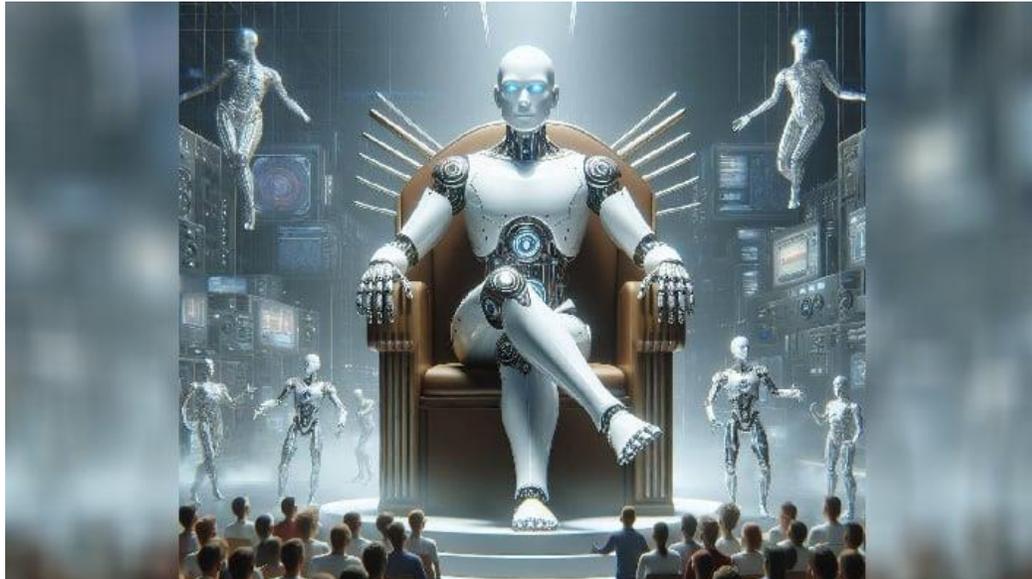
⇒ Kosten 2 Mrd. €

The construction of Panama Canal, 1881-1914



God Like AI?

Neue Berechnungsmöglichkeiten



5 Leading Small Language Models of 2024

	Llama 3	Meta 8 billion parameters
	Phi-3	Microsoft 3.8 billion - 7 billion parameters
	Gemma	Google 2 billion - 7 billion parameters
	Mixtral 8x7B	Mistral AI 7 billion parameters
	OpenELM	Apple 0.27 billion - 3 billion parameters

