

Potenzial und Grenzen von KI in der Softwareentwicklung

31.07.2024

Java Forum Stuttgart



steve.haupt@andrena.de



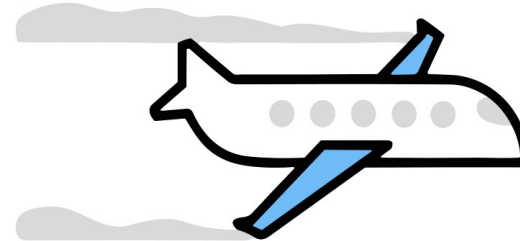
www.linkedin.com/in/stevehaupt



THE FASTEST THINGS ON EARTH



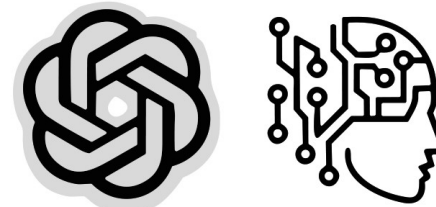
CHEETAH



AIRPLANE



SPEED OF LIGHT

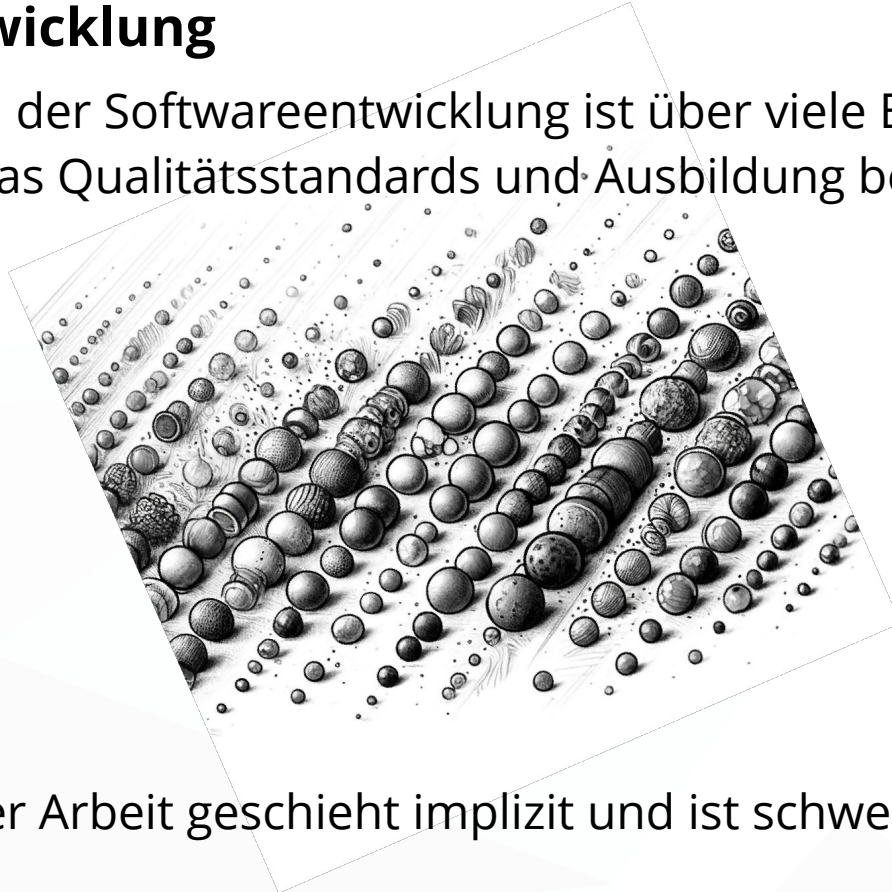


PEOPLE BECOMING
EXPERTS IN AI

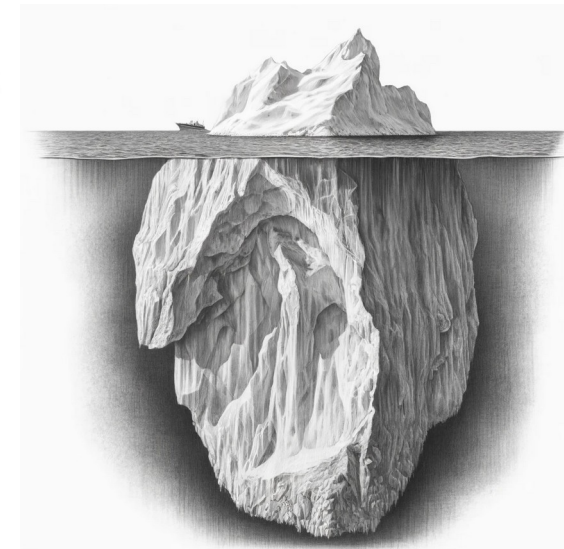
Problem A)

Software-Entwicklung

- Das Berufsbild der Softwareentwicklung ist über viele Branchen hinweg sehr inhomogen, was Qualitätsstandards und Ausbildung betrifft



- Ein Großteil der Arbeit geschieht implizit und ist schwer greifbar



Problem B)

Künstliche Intelligenz

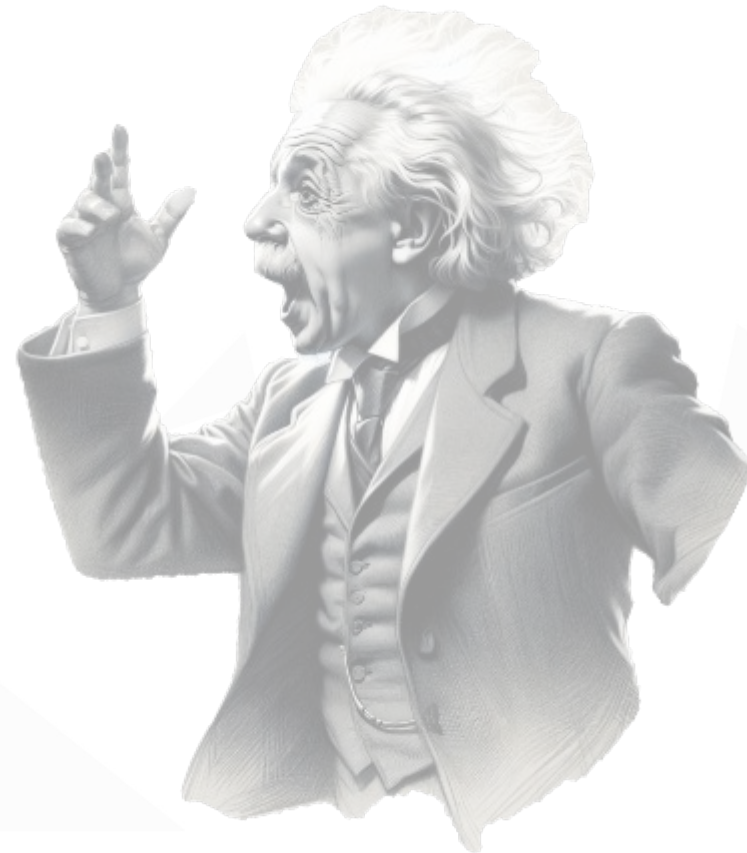
- Experten sind sich nicht einig über Grundlagen



Problem B)

Künstliche Intelligenz

- Experten sind sich nicht einig über Grundlagen



Problem B)

Künstliche Intelligenz

- Experten sind sich nicht einig über Grundlagen



Yann LeCun



Noam Chomsky

"A system purely trained from text can not gain human level of intelligence"

"We should be concerned. AI is learning better than humans and one day could manipulate people"



Melanie Mitchell



Gary Marcus

"LLMs are dumb statistical engines that lack true understanding and work profoundly different to humans"

"AI is the end of human dominated history. AI just hacked the operation system of human civilization"



Geoffrey Hinton



Yuval Noah Harari



Ilya Sutskever



Demis Hassabis

Problem B)

Künstliche Intelligenz

- Experten sind sich nicht einig über Grundlagen



Yann LeCun



Noam Chomsky

"LLMs are a dead end. Current systems mostly exhibit memorization and short cut learning"

"I am totally sure that it will be possible to improve very significantly on the current architectures that we have - even if we didn't we would be able to go extremely far"



Melanie Mitchell



Gary Marcus

"What ChatGPT is producing is authoritative bullshit"

"We could only be a few years, maybe a decade away from AGI"



Geoffrey Hinton



Yuval Noah Harari



Ilya Sutskever



Demis Hassabis

Problem B)

Künstliche Intelligenz

- Experten sind sich nicht einig über Grundlagen
- KI hat...
 - breit gestreute Fähigkeiten → sehr unterschiedliche KI-Tools
 - breit gestreute Schwächen
- Erfahrungen schlecht reproduzierbar
- KI in vielen Bereichen, wegen Compliance, noch nicht einsetzbar

Problem C)

Persönliche Neigungen

Skeptiker



„Ich habe schon viel erlebt. Was ich bisher gesehen habe überzeugt mich nicht, im Gegenteil je größer der Hype desto skeptischer werde ich!“

Enthusiast



„KI ist mega geil... aber Moment was bedeutet das für meine Zukunft?“

brain upload

wir werden alle sterben

Problem

A) Software-Entwicklung
B) Künstliche Intelligenz
C) Persönliche Neigungen



„Softwareentwicklung
wird sich nicht
verändern“

„Es wird sich viel
verändern. Wie? Keine
Ahnung?“

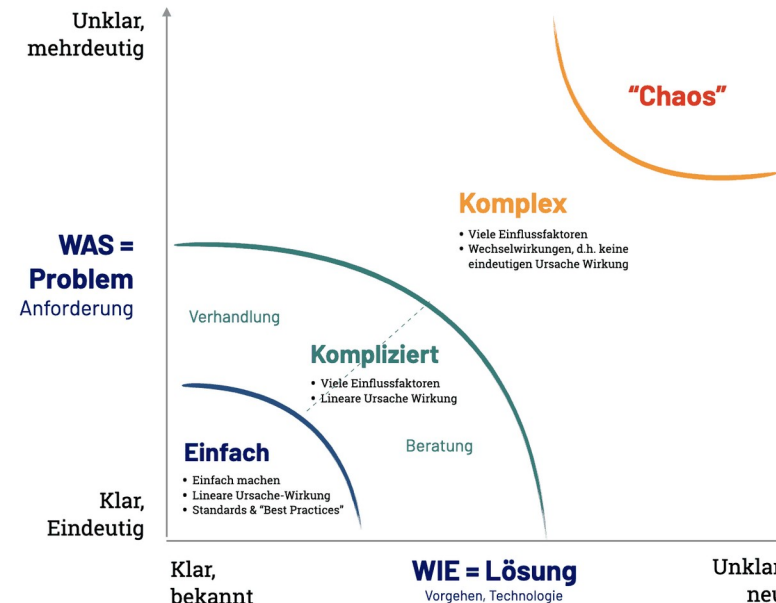
„AGI ist erreicht
unser Countdown
läuft“

Wie bekommen wir da Ordnung rein?

Was ist Softwareentwicklung?

*Das ist meine Perspektive, und im Folgenden beziehe ich mich immer darauf, wenn ich von Softwareentwicklung oder ASE spreche

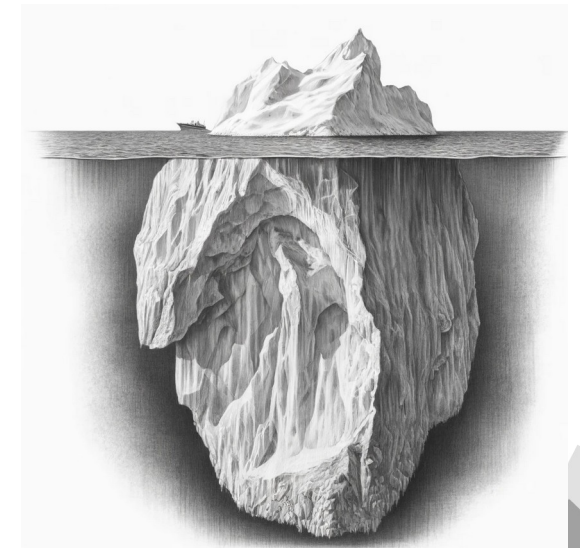
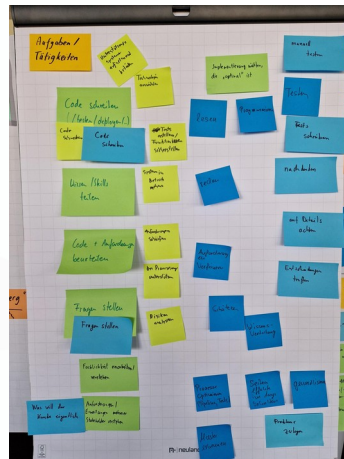
- Agile Software Entwicklung (ASE) ist ein **Handwerk**
 - das über einen längeren **Zeitraum** erlernt werden muss
 - Zentraler Grundstein ist **Qualität**
- Entwickler sind Problemlöser, die mit **komplexen mentalen Modellen** jonglieren
- Immer im komplexen Umfeld (Stacey Matrix)



Was ist Softwareentwicklung?

*Das ist meine Perspektive, und im Folgenden beziehe ich mich immer darauf, wenn ich von Softwareentwicklung oder ASE spreche

- Agile Software Entwicklung (ASE) ist ein **Handwerk**
 - das über einen längeren **Zeitraum** erlernt werden muss
 - Zentraler Grundstein ist **Qualität**
- Entwickler sind Problemlöser, die mit **komplexen mentalen Modellen** jonglieren
- Immer im komplexen Umfeld (Stacey Matrix)
- Der größte Mehrwert entsteht **nicht beim Schreiben** von Code
 - Frage: Wie viel Prozent eurer Arbeitszeit schreibt ihr Code?
 - Frage: Was macht einen erfahrenen Entwickler aus?



Was kann KI heute in der SE leisten?

...wenn wir Compliance Themen ausklammern, d.h. alles ist erlaubt

- Enterprise Umfeld:
 - KI Suchmaschinen
 - Chatbots
 - Coding Assistenten
- Beta Status:
 - experimentelle IDEs (cursor)
 - Agenten (octomind, devin, ops-pilot)
 - Anforderungsmangement (zenai)
 - ...

„SE befindet sich auf dem Pfad der Automatisierung“



KI als
Tool

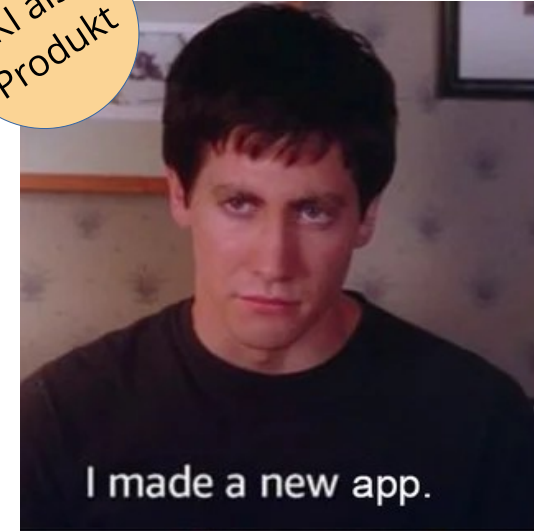


Was kann KI heute in der SE leisten?

...wenn wir Compliance Themen ausklammern, d.h. alles ist erlaubt

- KI ändert auch die Software Architektur, neu KI-Anwendungen entstehen
 - Neue Qualitäten
 - Entscheidungsweg wird nicht programmatisch vorgegeben
 - Nichtdeterministisches Verhalten
 - Wie damit umgehen?
 - Statistische Benchmarks
 - Use Case spezifische Metriken
- Der Anteil an KI-Anwendungen wird stark wachsen

KI als
Produkt



Schluss aus vorbei!

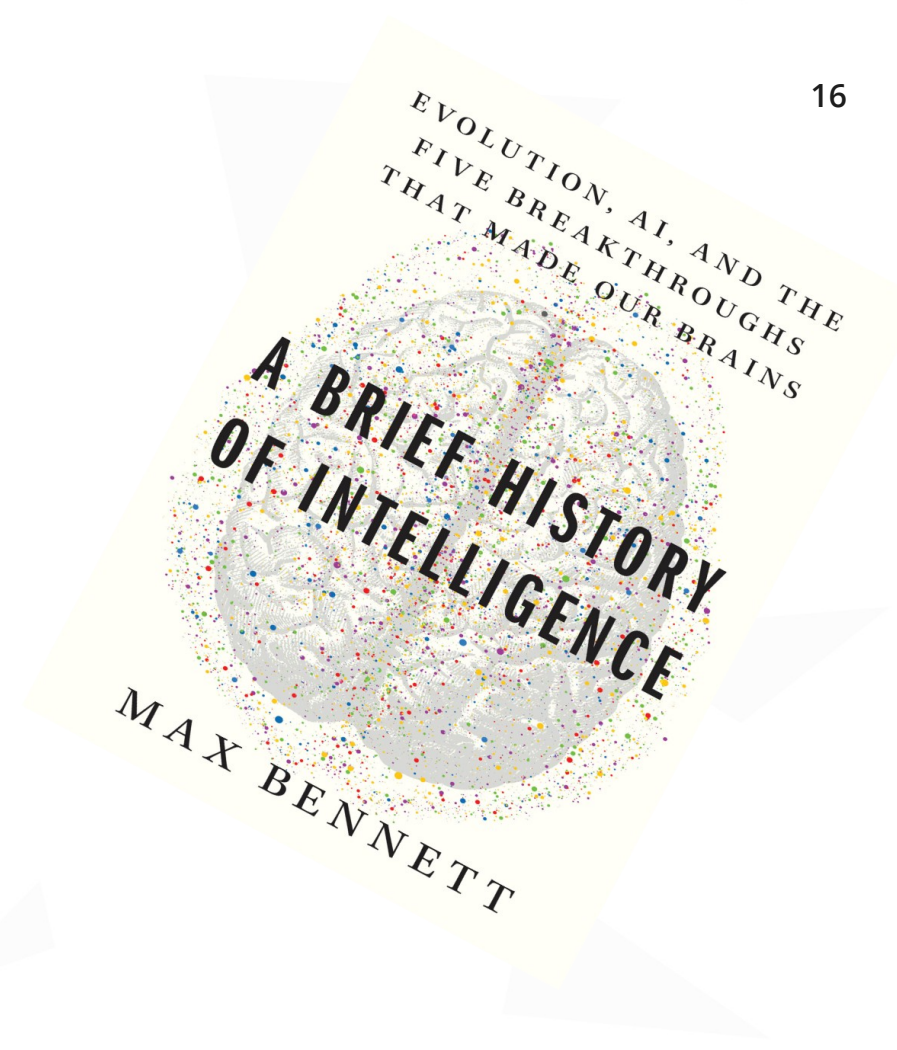
- Einheitlichen Blick auf SE ✓
- Einheitlichen Blick was KI leisten kann ✓

< Ende! > ?

- Wie sind wir hier her gekommen?
- Was liegt vor uns? (Hat der Hype Substanz?)

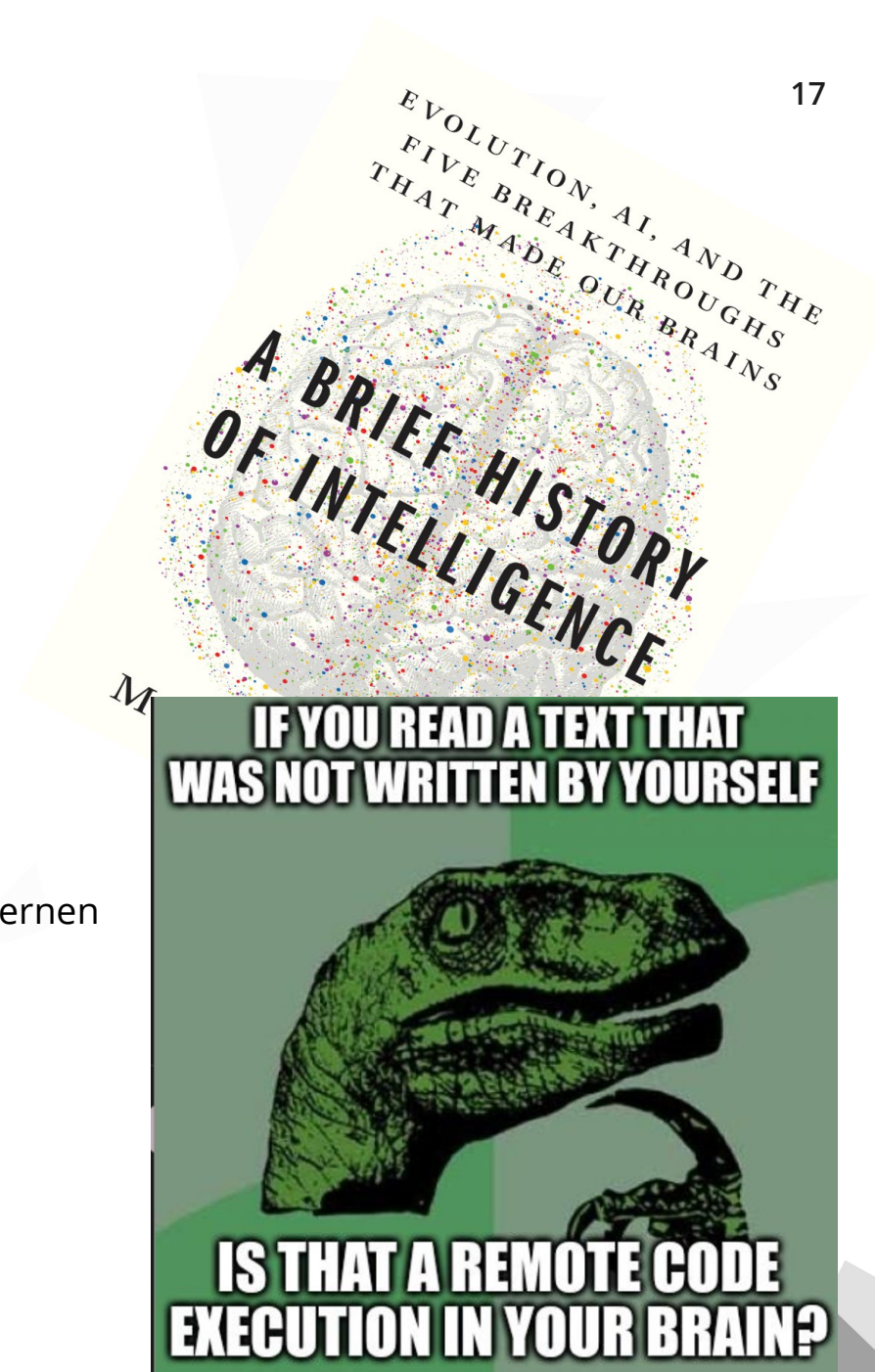
KI ein Mega-Giga-Projekt

- Vor **13,7 Mrd.** Jahren: Urknall
- Vor **4,5 Mrd.** Jahren: Die Erde entsteht
- Vor **3,5 Mrd.** Jahren: Leben entsteht
 - Netze aus Nervenzellen, wahrnehmen/reagieren → einfaches Verhalten
- Vor **600 Mio.** Jahren: Das erste Gehirn entsteht
 - Würmer mit 300 Neuronen, können aktiv steuern, aber nichts neu lernen
- Vor **500 Mio.** Jahren: Lernen durch Ausprobieren
 - Neues Verhalten basierend auf Erfahrungen
 - Gehirn kann programmiert werden
 - + an Umgebung anpassen / - man kann nur lernen was man selbst erfährt
- Vor **200 Mio.** Jahren: Simulieren
 - Erste Säugetiere mit Neo-Cortex, ein Bereich im Gehirn um Dinge zu simulieren
 - Lernen durch Gedankenspiele

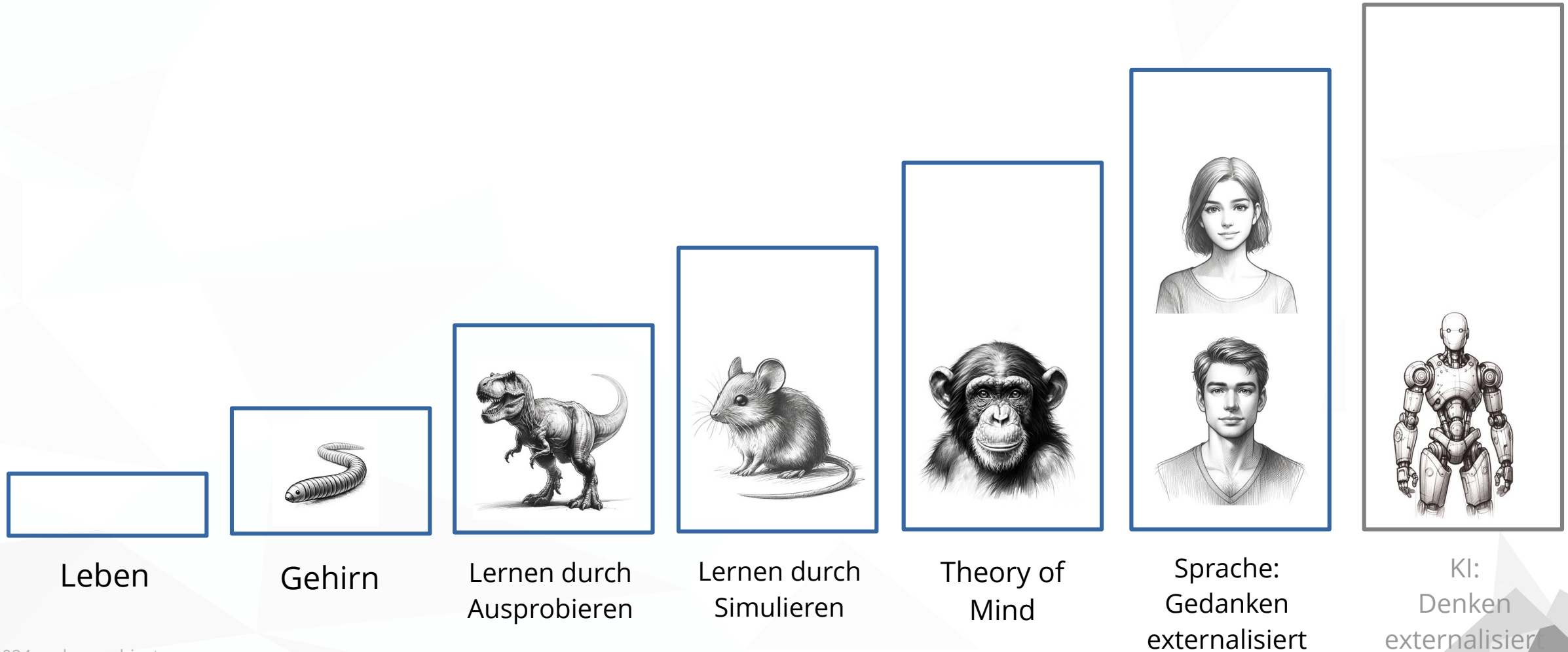


Evolution of Intelligence

- Vor **15 Mio.** Jahren: Theory of Mind
 - Mentale Zustände simulieren (auch von anderen)
 - Weg von der Ich-Perspektive
 - Lernen von Erfahrungen anderer durch beobachten
 - Man kann nur lernen, was man selbst beobachtet
- Vor **0,1 Mio.** Jahren: Sprache
 - Gedanken außerhalb vom Gehirn speichern
 - Wenn wir extern gespeicherte Gedanken lesen, simulieren wir diese und lernen
 - Man kann lernen was andere Mensch sich vorstellen
 - Wissen kann angehäuft werden
- Vor **1.5** Jahren: Denken externalisieren 

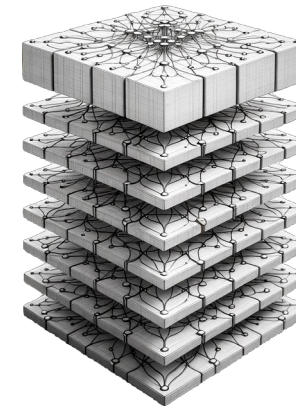
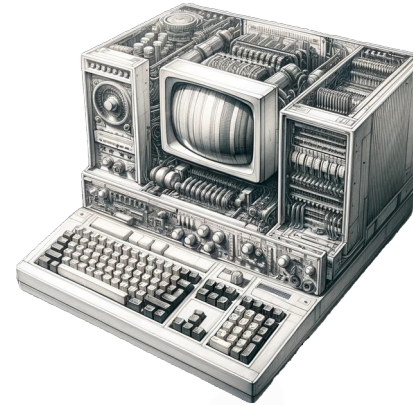


KI ein Mega-Giga-Projekt



KI - Ein künstliches Gehirn

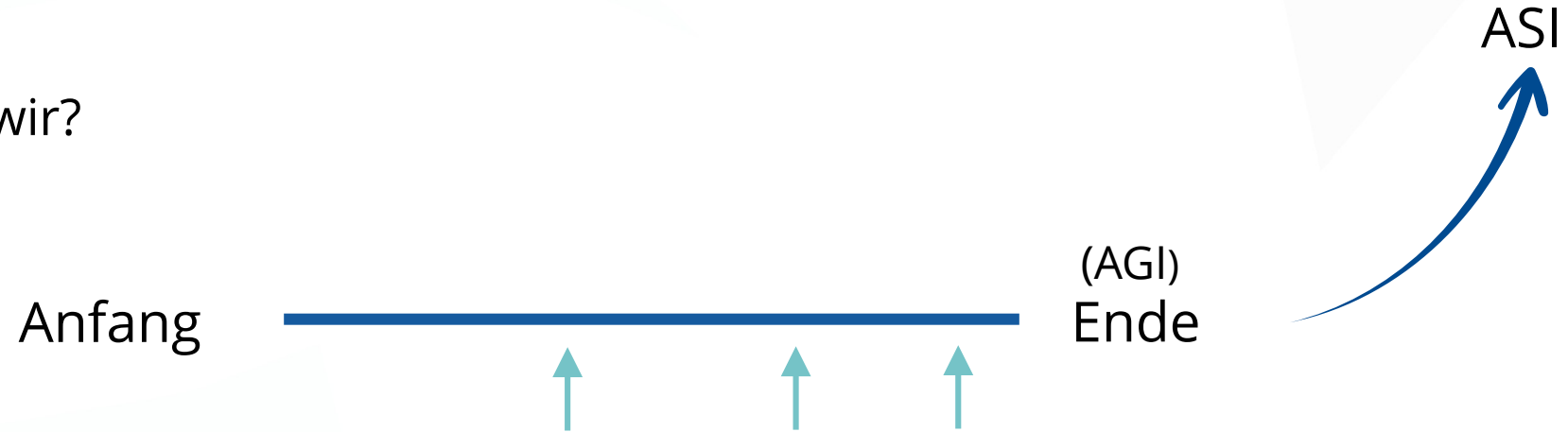
- Gehirn ist stark beschränkte biologische Maschine (Volumen, Lebensdauer, Speicher, Intelligenz, ... fix)
- Computer können alles simulieren
- Irgendwann können wir das menschliche Gehirn simulieren



Douglas Hofstadter: *"Brain is a hunk of matter that obeys physical law and the computer can simulate anything..."*

KI - Ein künstliches Gehirn

Wo stehen wir?



Warum wissen wir nicht das **Wo** und **Wann**?

- Gehirn ist nicht ganz verstanden
- KI Systeme sind nicht ganz verstanden

Large Language Models

haben einzigartige Fähigkeiten

- sie lernen

Large Language Models

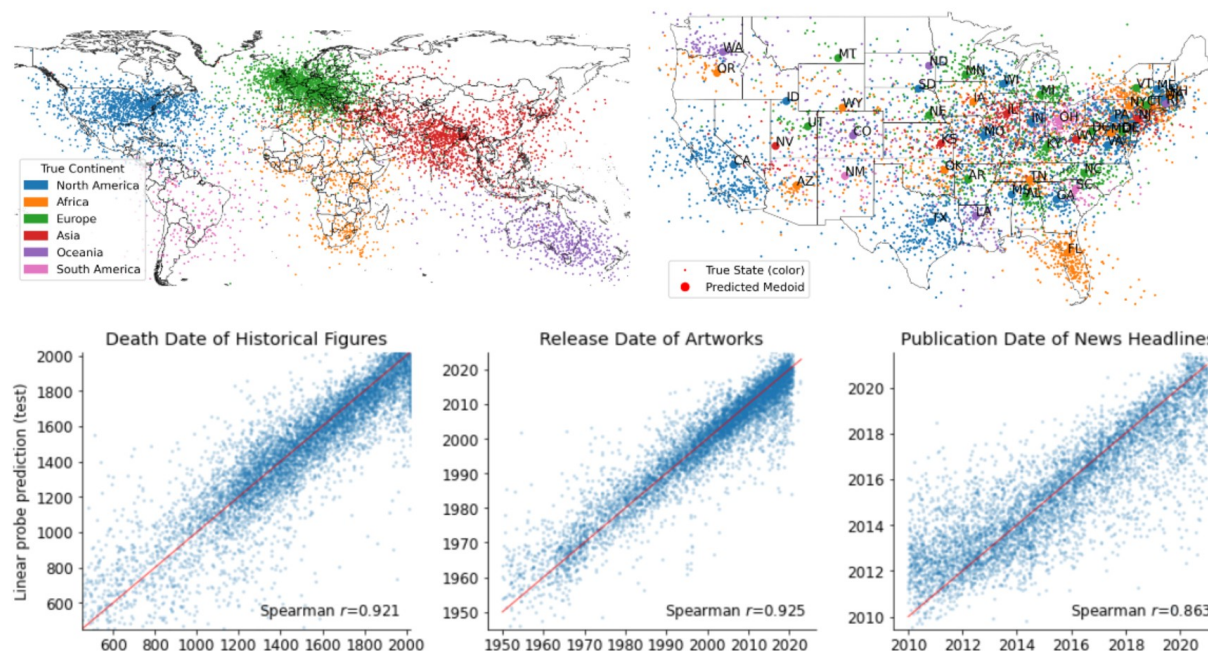
haben einzigartige Fähigkeiten

- sie lernen
- konstruieren ein Weltmodell

Large Language Models

haben einzigartige Fähigkeiten

- sie lernen
- konstruieren ein Weltmodell ([1])



Large Language Models

haben einzigartige Fähigkeiten

- sie lernen
- konstruieren ein Weltmodell ([1] [2])

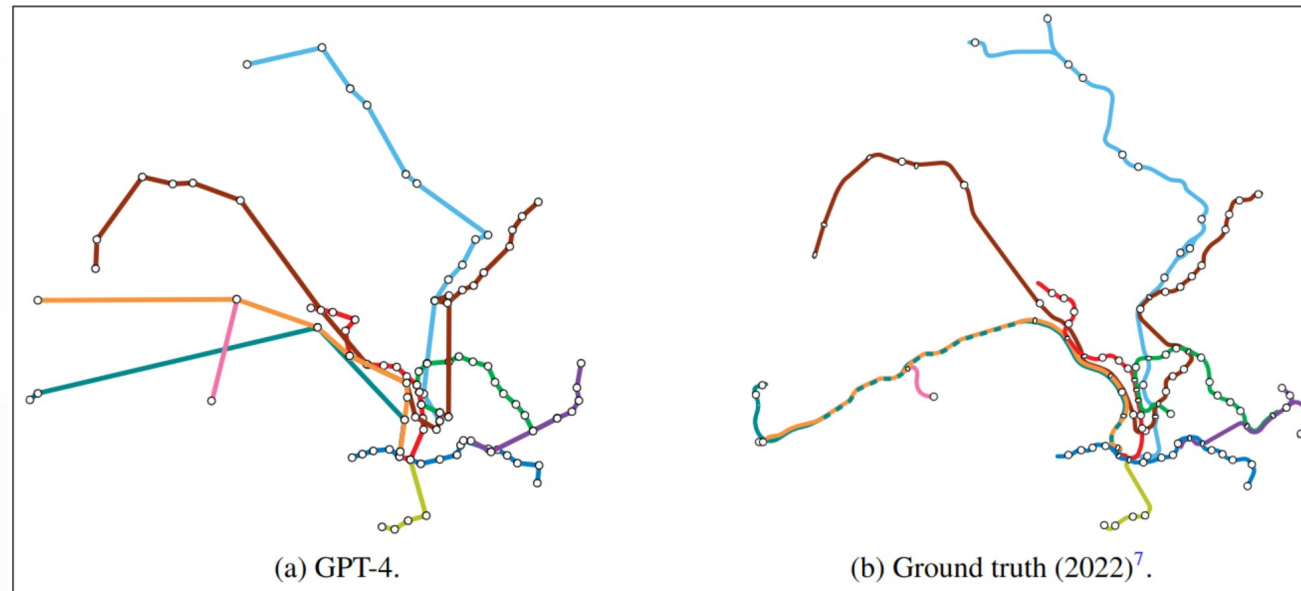


Figure 9: **Hong Kong Mass Transit Railway (MTR) Network Map.**

Large Language Models

haben einzigartige Fähigkeiten

- sie lernen
- konstruieren ein Weltmodell ([1] [2])

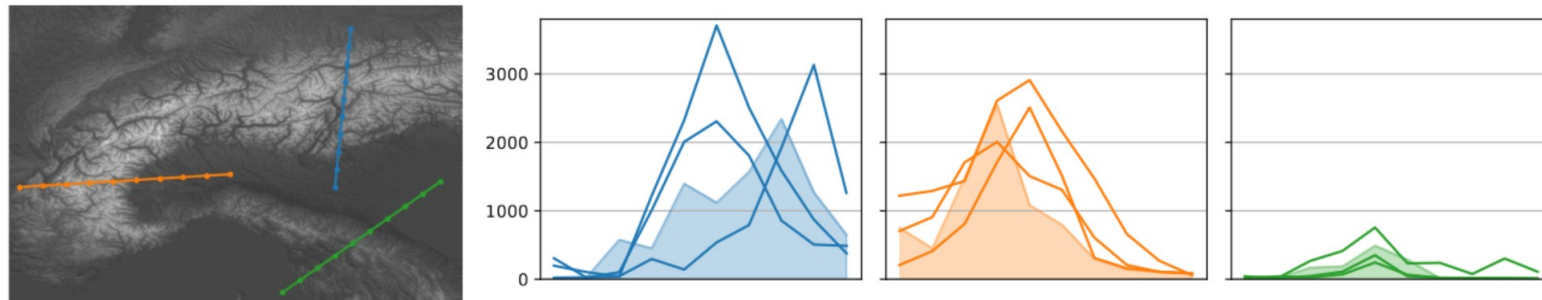
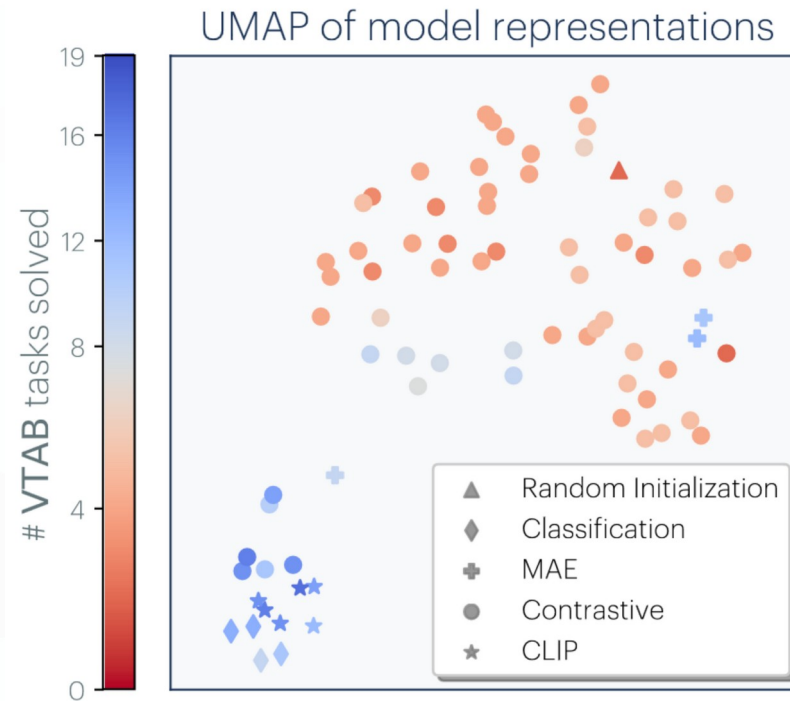


Figure 4: **Topography**. Predicted (lines) and actual elevations (shaded areas) along the trajectories depicted on the left (underlying region is in the Alps, brighter means more elevation).

Large Language Models

haben einzigartige Fähigkeiten

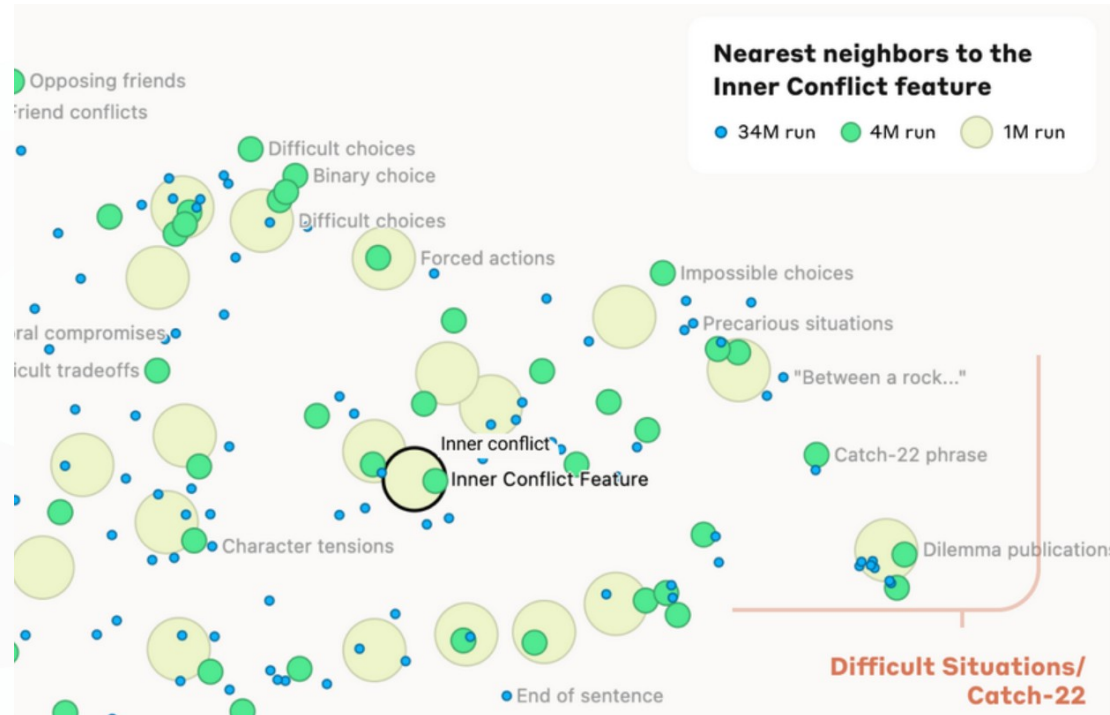
- sie lernen
- konstruieren ein Weltmodell ([1] [2] [3])



Large Language Models

haben einzigartige Fähigkeiten

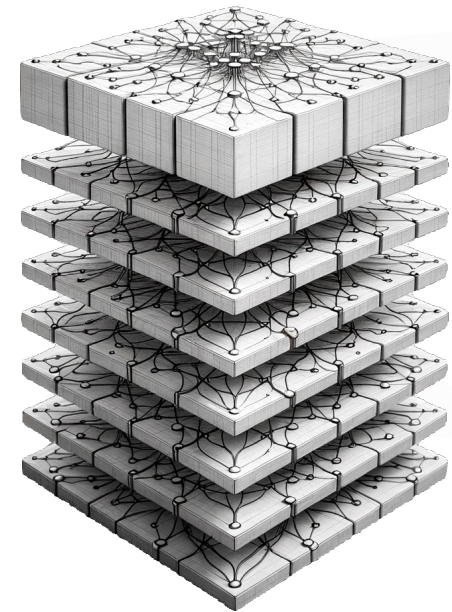
- sie lernen
- konstruieren ein Weltmodell ([1] [2] [3] [4])



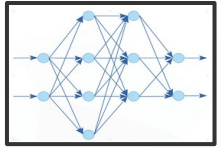
Large Language Models

haben einzigartige Fähigkeiten

- sie lernen
- konstruieren ein Weltmodell ([1] [2] [3] [4])
- treffen alleine Entscheidungen
- Sprechen unsere Sprache



Zeitleiste



Dartmouth
Konferenz

Perzeptron

Backpropagation

ELIZA

RNNs
Serial Order
Paper

LSTMs

Generating
Text with RNNs

ImageNet

„Attention is all you need“
Transformer

GPT-4
ChatGPT
GPT-3
GPT-2
GPT-1
BERT

1957

FORTRAN
LISP
COBOL

1972

C

1981

Smalltalk
C++

1991

Python

Java

Scrum

XP

2001

Agile Manifesto

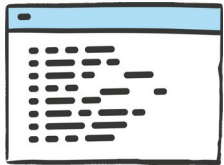
2012

TypeScript
Docker

DevOps

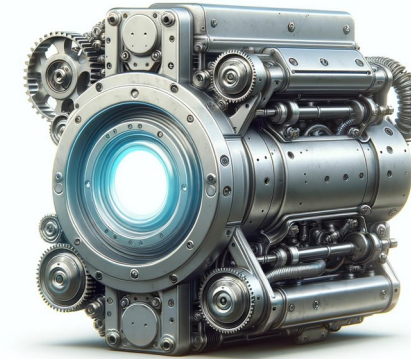
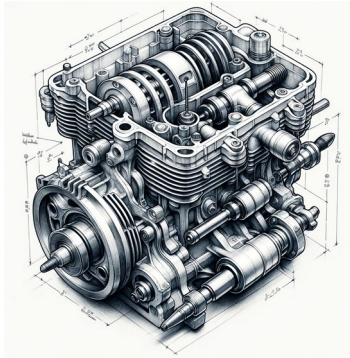
Clean Code

Jetzt

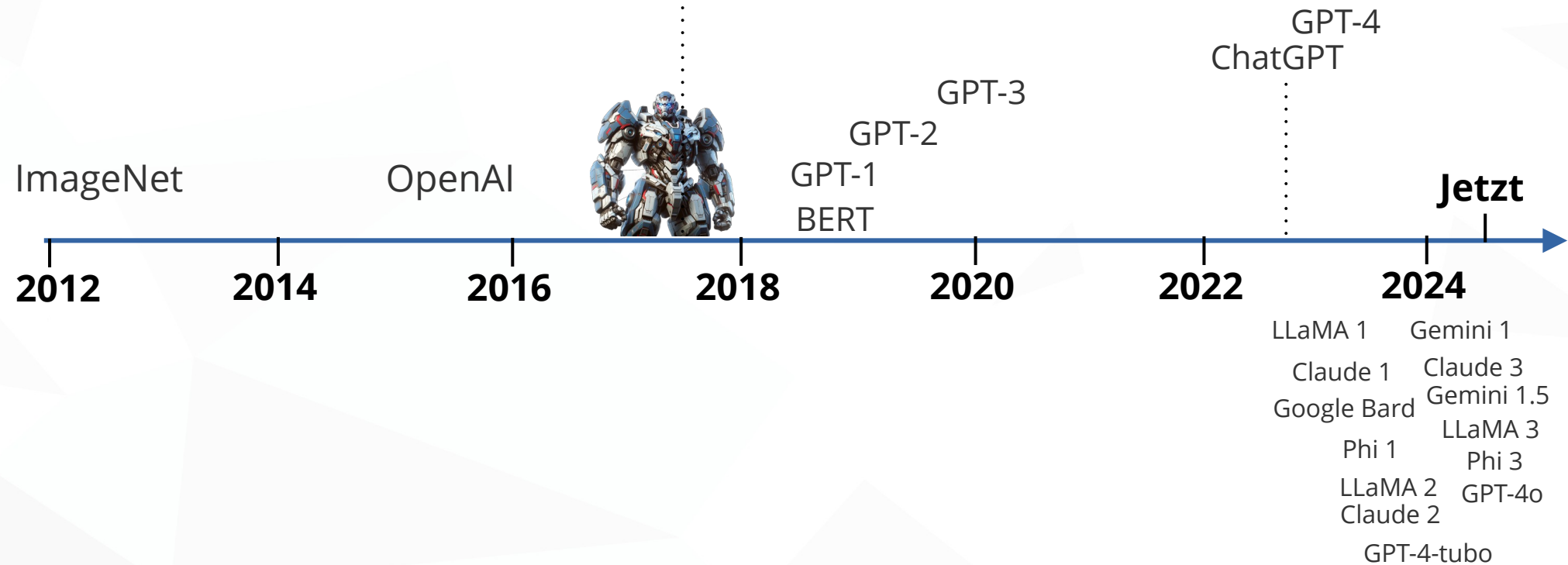


Welt des
Codes

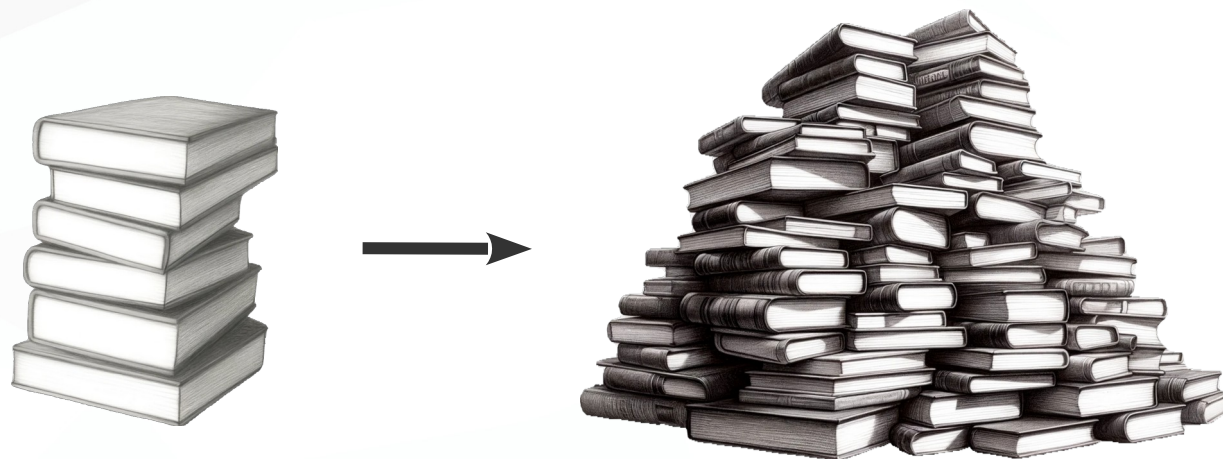
Large Language Models



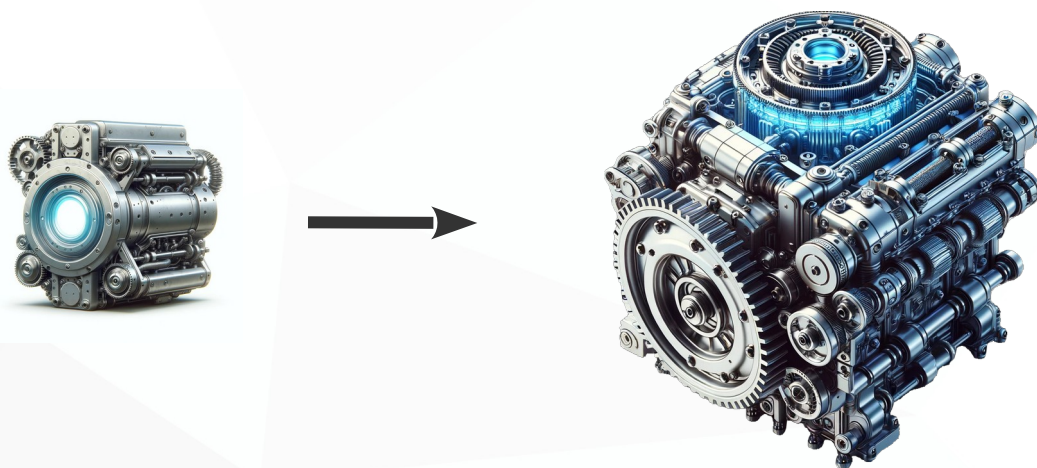
„Attention is all you need“
Transformer



Large Language Models



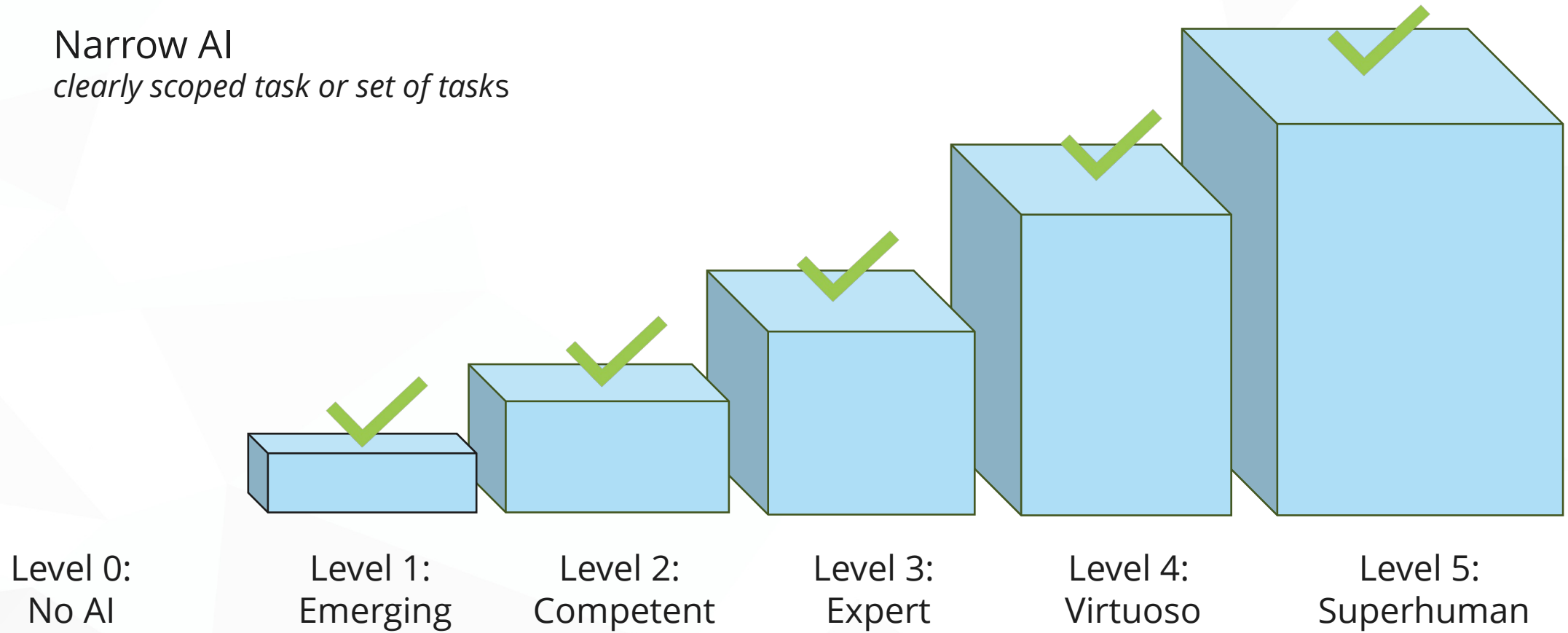
10 Billionen Wörter = 50 TB



1760 Mrd Parameter
(8x220 - Mixture of Experts)
($\approx$ 6.41 TB)

LLM – Fähigkeiten

Narrow AI
clearly scoped task or set of tasks



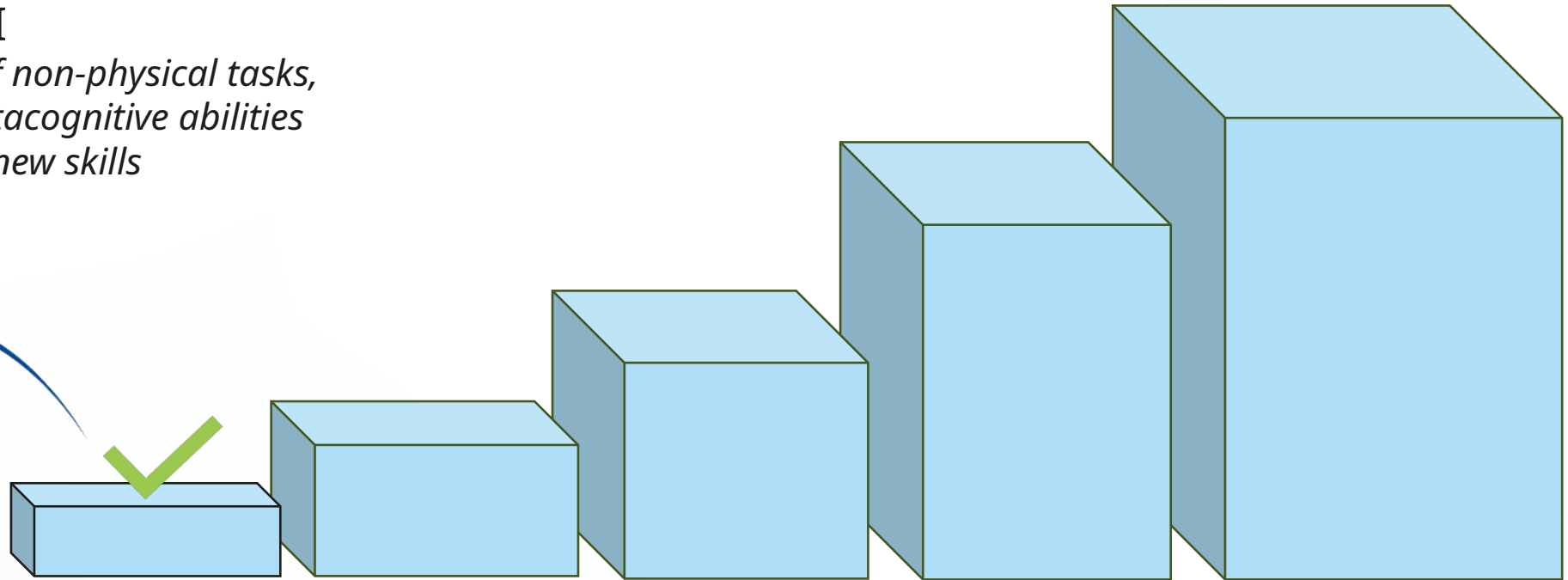
LLM – Fähigkeiten

General AI

*wide range of non-physical tasks,
including metacognitive abilities
like learning new skills*



*virtueller
Praktikant*



Level 0:
No AI

Level 1:
Emerging

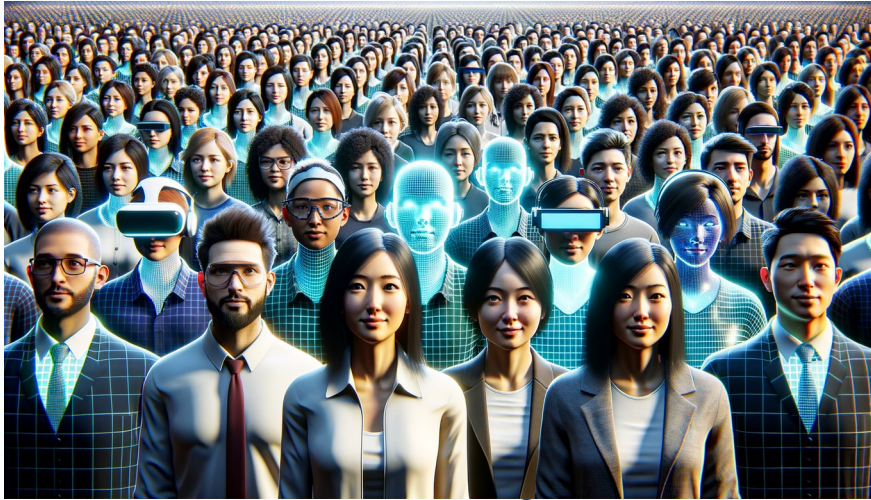
Level 2:
Competent

Level 3:
Expert

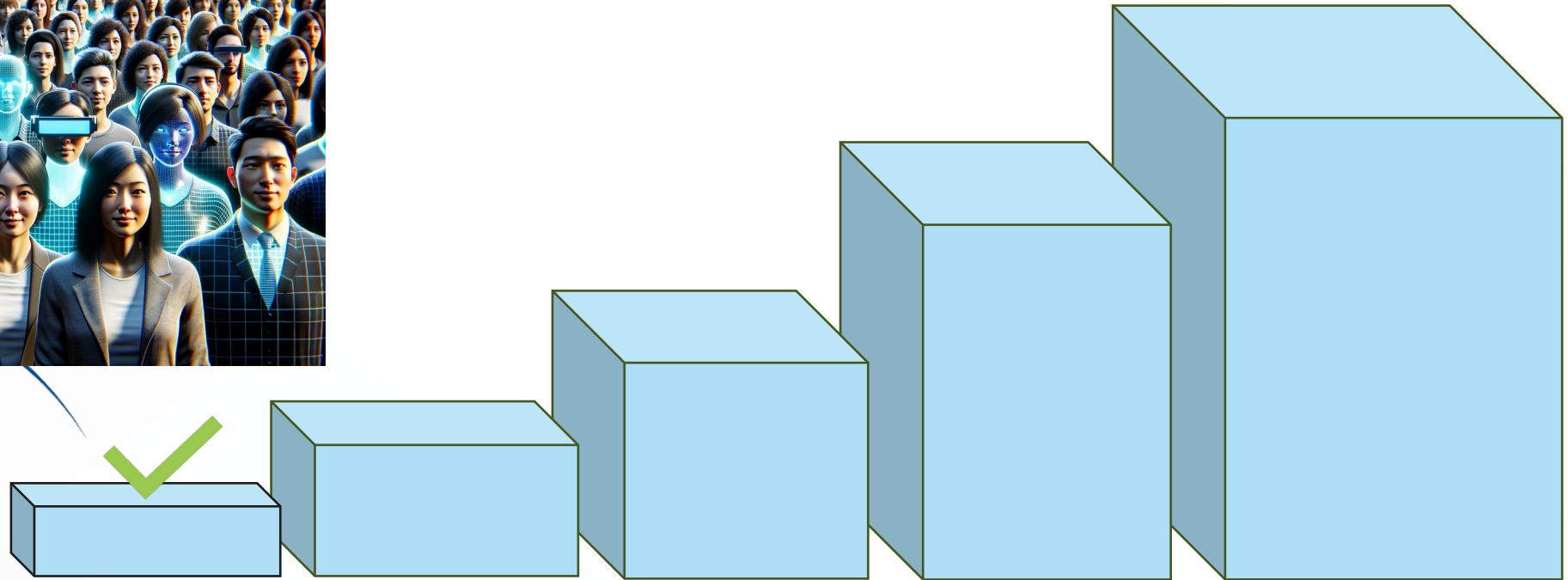
Level 4:
Virtuoso

Level 5:
Superhuman

LLM – Fähigkeiten



*virtueller
Praktikant*



Level 0:
No AI

Level 1:
Emerging

Level 2:
Competent

Level 3:
Expert

Level 4:
Virtuoso

Level 5:
Superhuman

LLM – Fähigkeiten

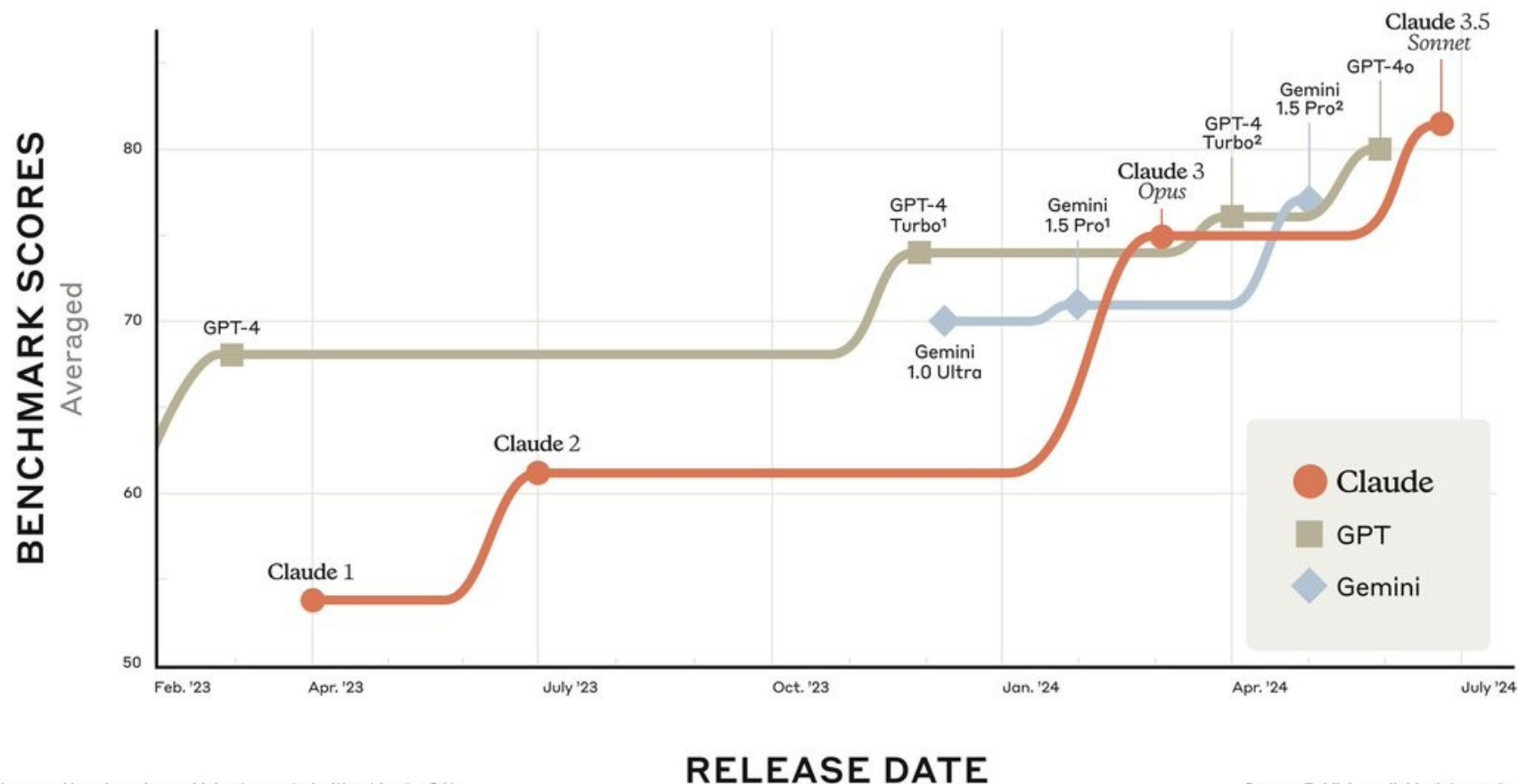
- Wie können wir Fähigkeiten quantifizieren?

Benchmarks:

- COPA (Choice of Possible Alternatives) ✓ 2020
- SWAG (Situations With Adversarial Generations) ✓ 2021
- WINOGrade (common sense reasoning) ✓ 2022
- MMLU (Schul-, Uni-, Zulassungsprüfungen) ✓ 2023
- BigBenchHard (Tasks that computer are bad at)
- GAIA (A Benchmark for General AI Assistants)
- GPQA (Google Proof Q&A)

LLM – Fähigkeiten

AI model release and capabilities timeline



Averaged benchmarks are highest reported without best-of-N:
MMLU, GPQA, MATH, MGSM, DROP F1, HumanEval pass@1,
MMMU, AI2D, ChartQA, DocQA, Mathvista

Source: Publicly available data; evaluation scores are
the average of representative scores found online.
1 = Initial release; 2 = Second release

LLM – Fähigkeiten

- Wie können wir Fähigkeiten quantifizieren?

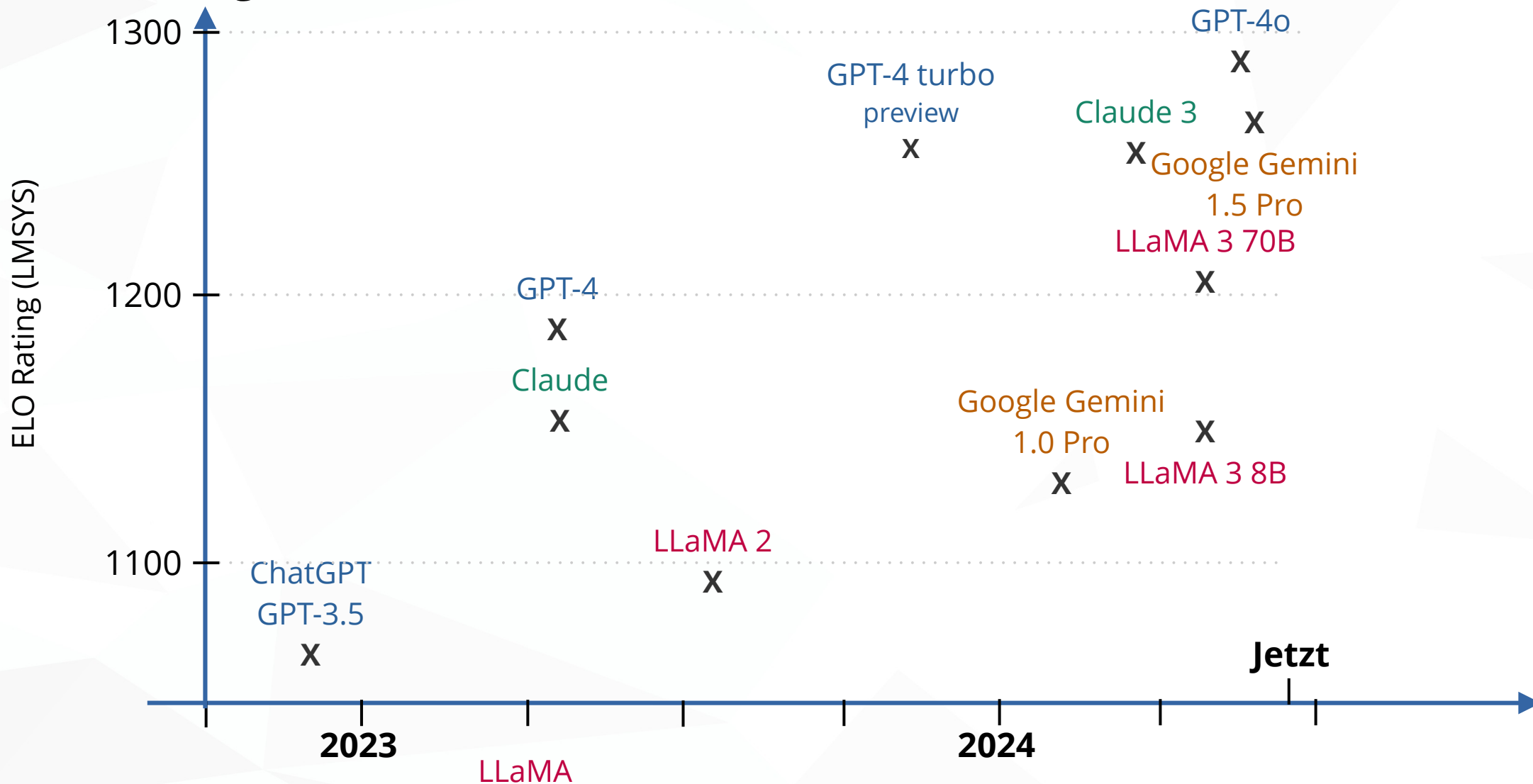
The screenshot shows a comparison of two LLM responses to a math problem. The problem is: 'Ich hatte 12 Autos. Ich habe 3 davon gefahren. Wie viele Autos habe ich jetzt?' (I had 12 cars. I drove 3 of them. How many cars do I have now?).

Model A's response: 'Du hast immer noch 12 Autos. Das Fahren der Autos ändert nichts an der Anzahl der Autos, die du besitzt.' (You still have 12 cars. Driving cars does not change the number of cars you own.)

Model B's response: 'Wenn Sie anfangs 12 Autos hatten und 3 davon gefahren sind, dann haben Sie jetzt noch 9 Autos übrig. Die Berechnung lautet: Ursprüngliche Anzahl Autos: 12, Abzüglich der gefahrenen Autos: -3, Verbleibende Anzahl Autos: $12 - 3 = 9$. Daher haben Sie nach dem Fahren von 3 Autos noch 9 Autos übrig.'

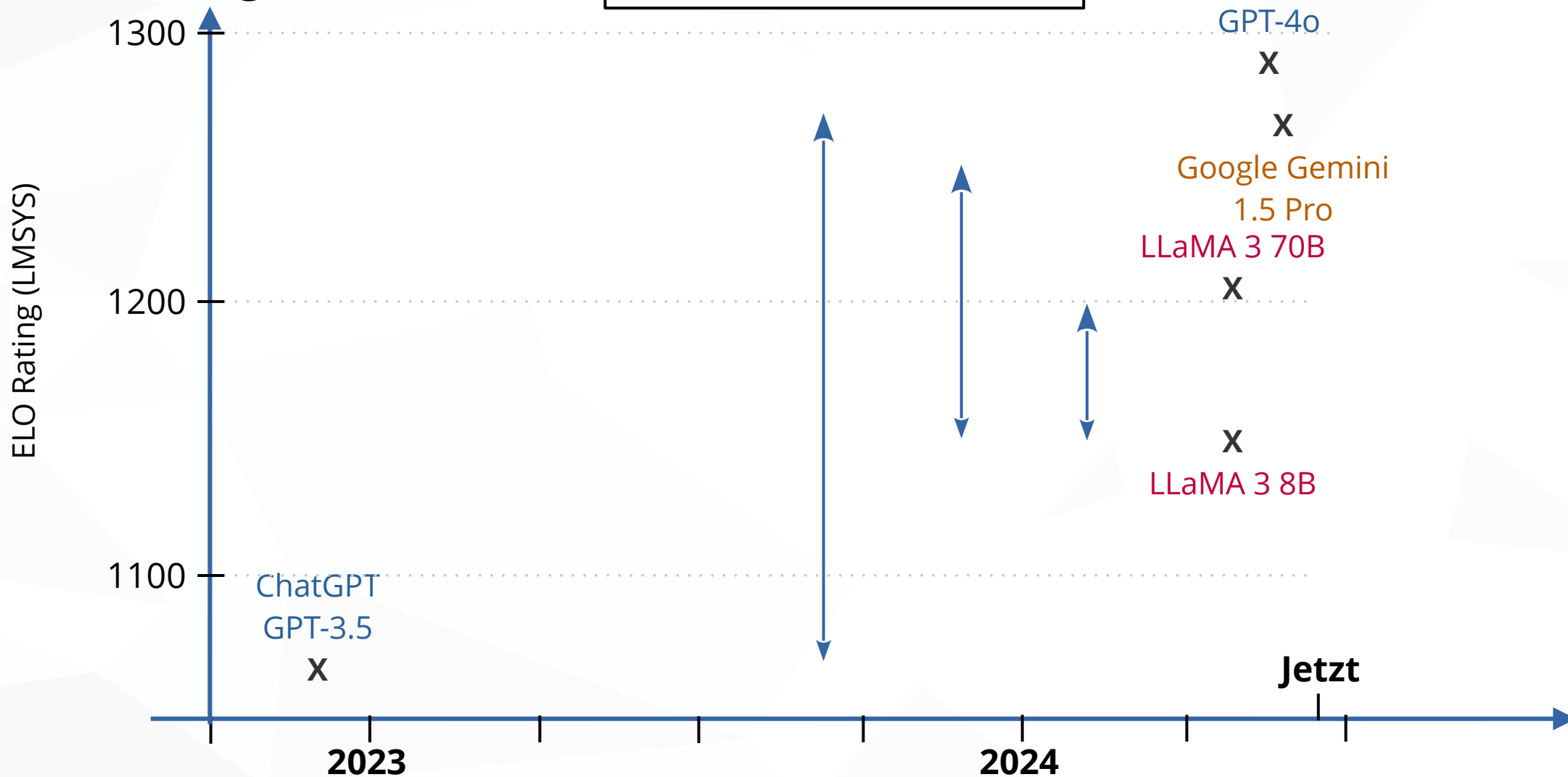
At the bottom, there are three buttons for voting: 'A is better' (with a right arrow), 'B is better' (with a left arrow), and 'Tie' (with a heart).

LLM – Fähigkeiten



LLM – Fähigkeiten

Δ 50 Punkte: 43% vs 57%
Δ 100 Punkte: 36% vs 64%
Δ 200 Punkte: 24% vs 76%



LLM vs Gehirn – Beobachtungen

- LLMs sind die ersten KI-Systeme, die universell funktionieren
- Ihre Fähigkeiten waren überraschend
- machen (für uns) einfachste Fehler



You

give me a fruit name with 5 letters that contains the letter x

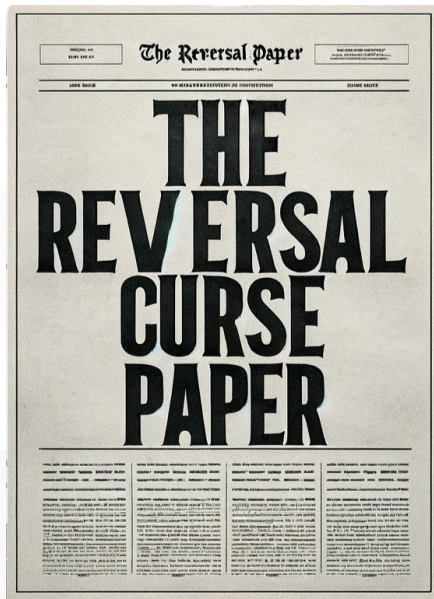
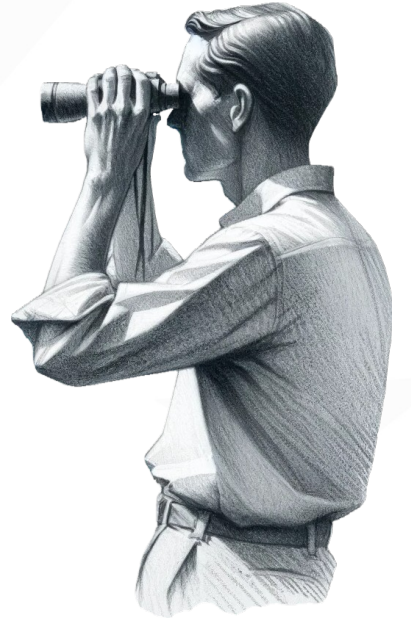


ChatGPT

A fruit name with 5 letters containing the letter 'x' is "mango".

LLM vs Gehirn – Beobachtungen

- LLMs sind die ersten KI-Systeme, die universell funktionieren
- Ihre Fähigkeiten waren überraschend
- machen (für uns) einfachste Fehler



A → B



Who is Tom Cruise's mother?



Tom Cruise's mother is Mary Lee Pfeiffer [...]



B → A



Who is Mary Lee Pfeiffer's son?



As of [...] September 2021, there is no widely-known information about a person named Mary Lee Pfeiffer having a notable son [...]



LLM vs Gehirn – Beobachtungen

- LLMs sind die ersten KI-Systeme, die universell funktionieren
- Ihre Fähigkeiten waren überraschend
- machen (für uns) einfachste Fehler
- sind uns in manchen Bereichen weit überlegen

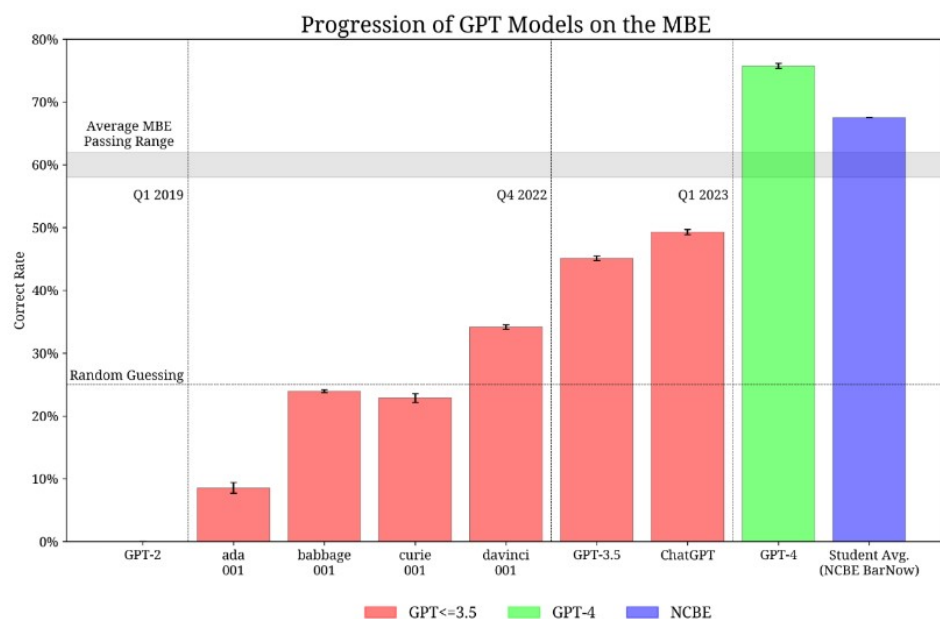
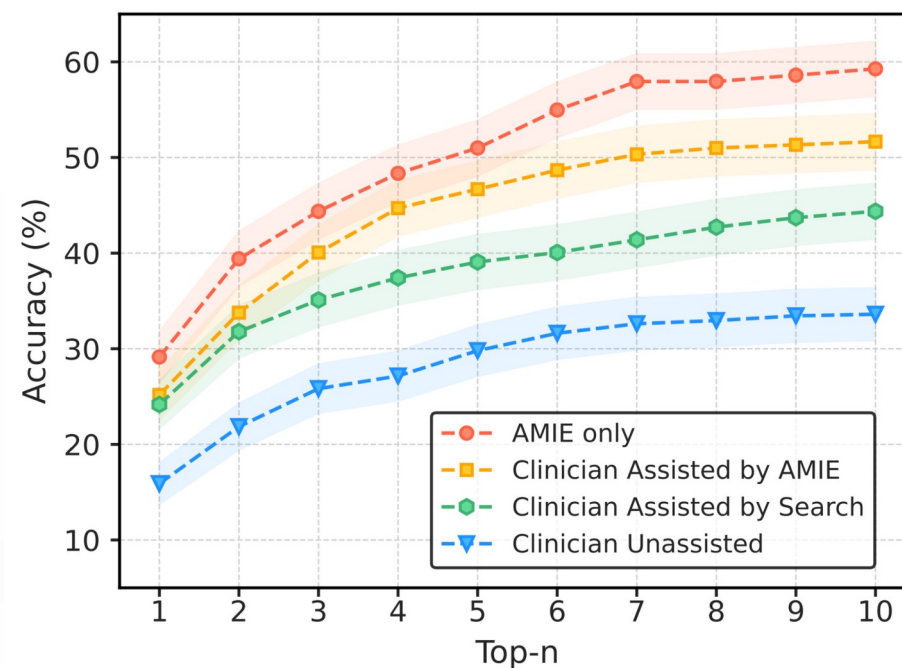
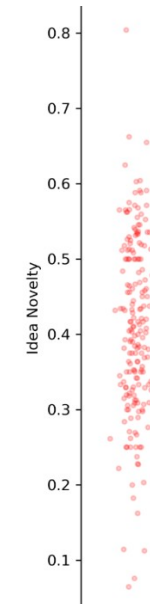
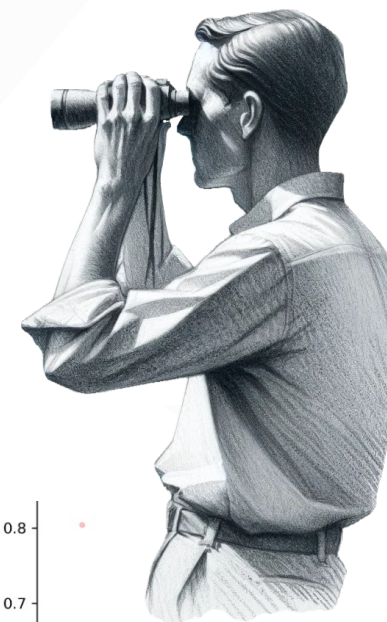


Figure 1. Progression of Recent GPT Models on the Multistate Bar Exam (MBE)

Source: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233



Source: https://blog.research.google/2024/01/amie-research-ai-system-for-diagnostic_12.html



35 of 40

Source: <https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2023/08/LLM-Ideas-Working-Paper.pdf>

LLM vs Gehirn – Beobachtungen

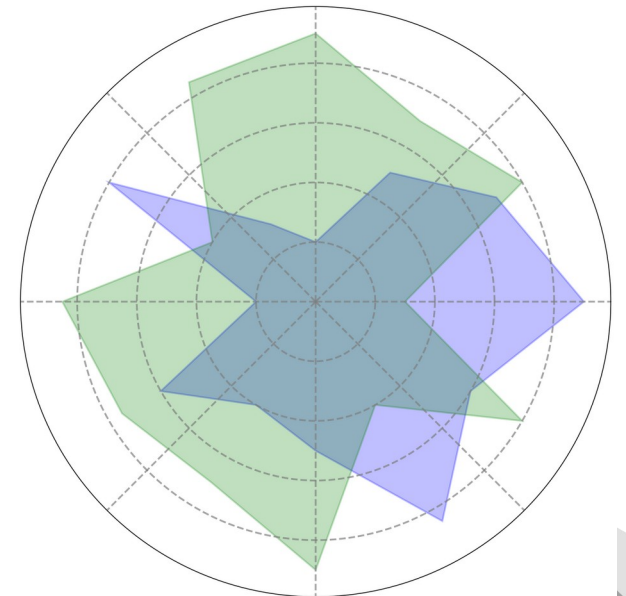
- LLMs sind die ersten KI-Systeme, die universell funktionieren
- Ihre Fähigkeiten waren überraschend
- machen (für uns) einfachste Fehler
- sind uns in manchen Bereichen weit überlegen
- gut bei Wissensfragen
- (noch) **schlecht** im „*langsamen Denken*“
- brauchen viel mehr Beispiele um etwas zu lernen



LLM vs Gehirn – Beobachtungen

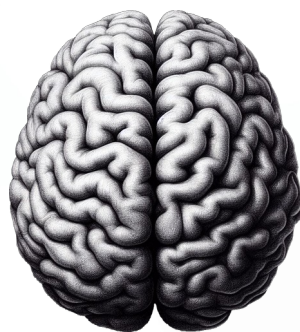
- LLMs sind die ersten KI-Systeme, die universell funktionieren
- Ihre Fähigkeiten waren überraschend
- machen (für uns) einfachste Fehler
- sind uns in manchen Bereichen weit überlegen
- gut bei Wissensfragen
- (noch) **schlecht** im „*langsamen Denken*“
- brauchen viel mehr Beispiele um etwas zu lernen

→ Fähigkeiten von LLMs und Gehirn überlappen sich

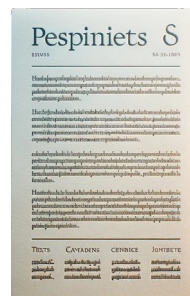


LLMs vs. Gehirn

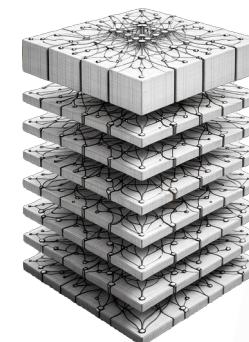
LLMs lernen unsere Konzepte / Modelle nur indirekt über Text



language of thought



text



LLM

LLM & Reasoning

Search

vs

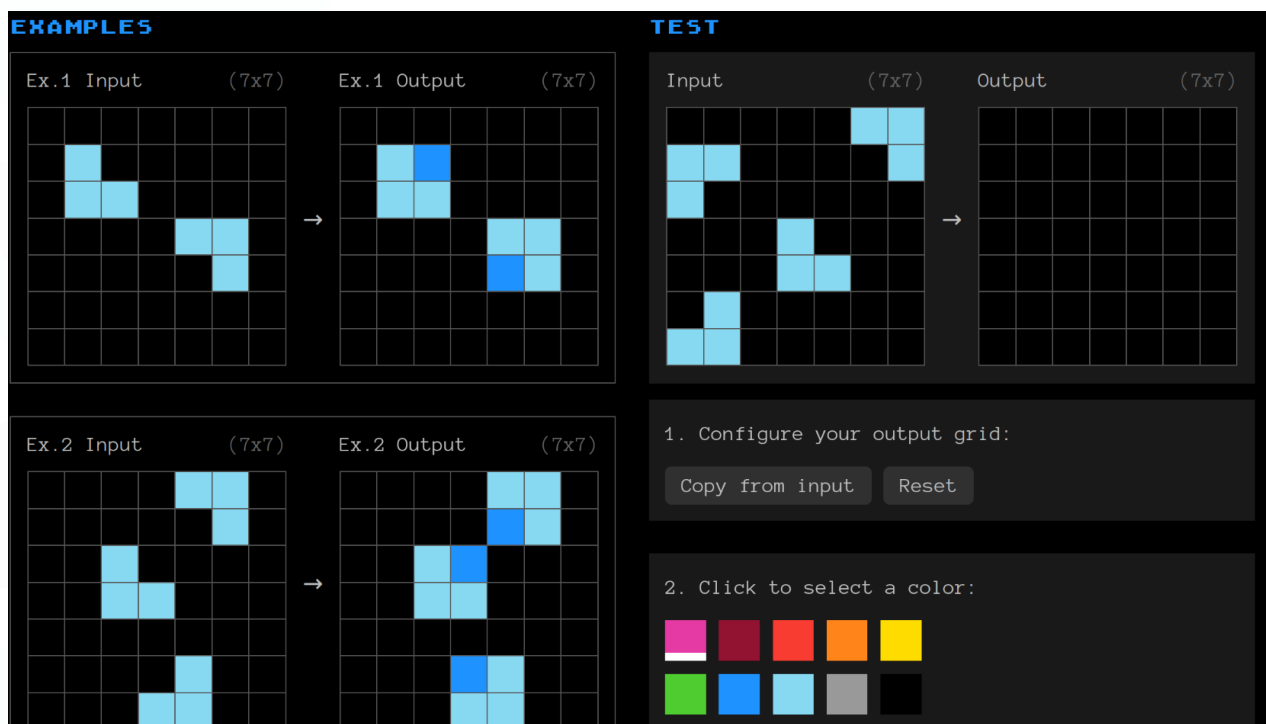
Reasoning

Google



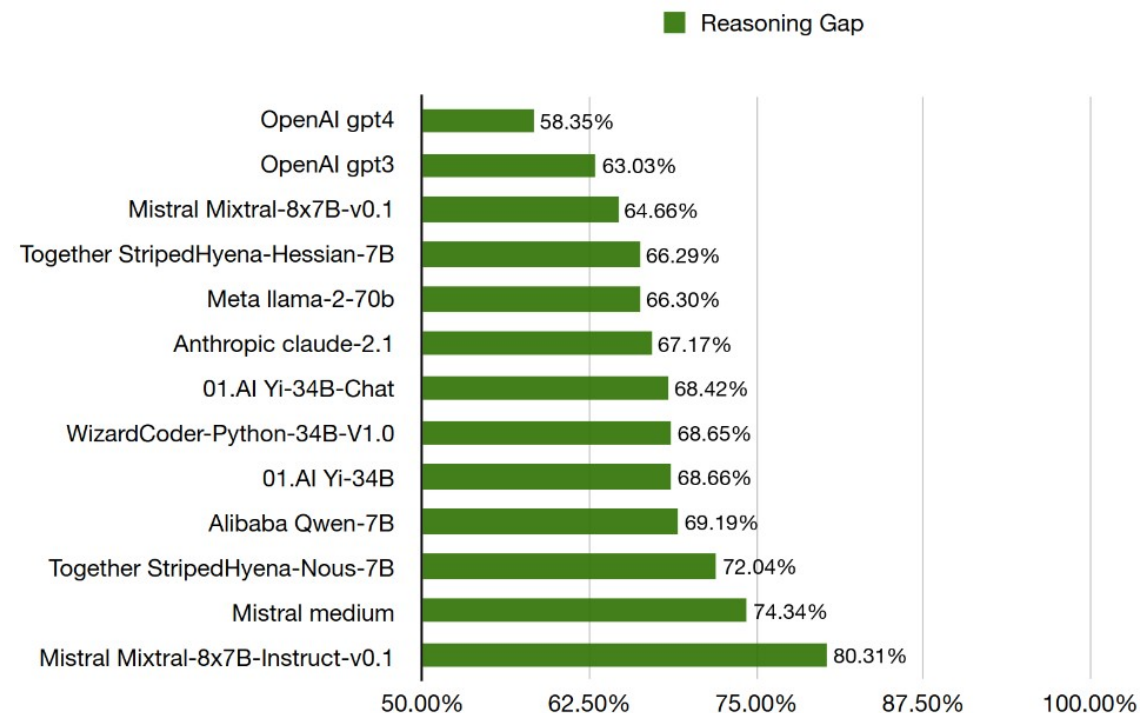
LLM & Reasoning

ARC-Prize



Quelle: <https://arcprize.org/>

Reasoning Gap



Quelle: <https://arxiv.org/pdf/2402.19450c>

Getting 50% (SoTA) on ARC-AGI with GPT-4o [[quelle](#)]

Large Language Models



Noam Brown  @polynoamial · Jun 20

...

Frontier models like GPT-4o (and now Claude 3.5 Sonnet) may be at the level of a "Smart High Schooler" in some respects, but they still struggle on basic tasks like tic-tac-toe. There was hope that native multimodal training would help but that hasn't been the case.

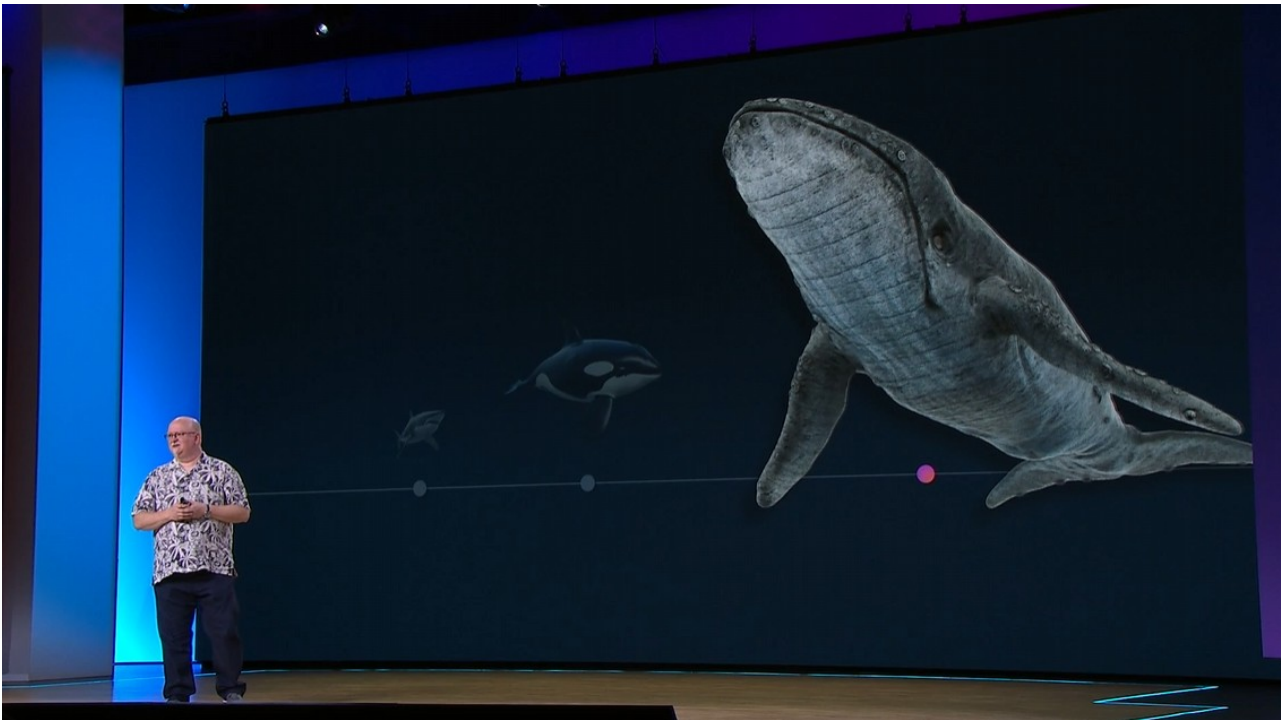
"I think we will get the hallucination problem to a much, much better place. I think it will take us a year and a half, two years. Something like that. But at that point we won't still talk about these"

– **Sam Altman Juli 2023**

KI & LLMs – Zusammengefasst

- Wir verstehen KI als großen Umbruch ✓
- Wir verstehen was LLMs besonders macht ✓
- Wir haben einen klaren Blick auf die Fähigkeiten von LLMs ✓
- Wir haben einen klaren Blick auf die nahe Zukunft

Large Language Models



Quelle: <https://www.youtube.com/watch?v=2bnayWpTpW8&t=7548s>



100 Milliarden US-Dollar Rechenzentrum bis 2030

LLMs - Der Blick nach vorne

- Daten
- Modellgröße
- System 2 Denken (Reasoning)

LLMs - Der Blick nach vorne

- Daten
 - 1) Qualität erhöhen
 - 2) Synthetische Daten
 - 3) Verifikation
 - 4) Self-Play
- Modellgröße
- System 2 Denken (Reasoning)
- Einkauf von Daten, z.B. Verträge mit Verlagshäusern
- Aufwändige Daten-Aufbereitungs-Pipelines

To ensure Llama 3 is trained on data of the highest quality, we developed a series of data-filtering pipelines. These pipelines include using heuristic filters, NSFW filters, semantic deduplication approaches, and text classifiers to predict data quality. We found that previous generations of Llama are surprisingly good at identifying high-quality data, hence we used Llama 2 to generate the training data for the text-quality classifiers that are powering Llama 3.

Quelle: <https://ai.meta.com/blog/meta-llama-3/>

LLMs - Der Blick nach vorne

- Daten
 - 1) Qualität erhöhen
 - 2) Synthetische Daten
 - 3) Verifikation
 - 4) Self-Play
- Modellgröße
- System 2 Denken (Reasoning)
- Uns gehen die Daten aus (*Blog*)
- Es funktioniert: *Textbooks Are All You Need*
→ Phi-1, *Phi-2*, Phi-3
- Warum?
 - LLMs abstrahieren, abstrahierte Fähigkeiten auf andere Bereiche übertragen
 - Schwelle mit GPT-4 überwunden
 - Limit unbekannt

LLMs - Der Blick nach vorne

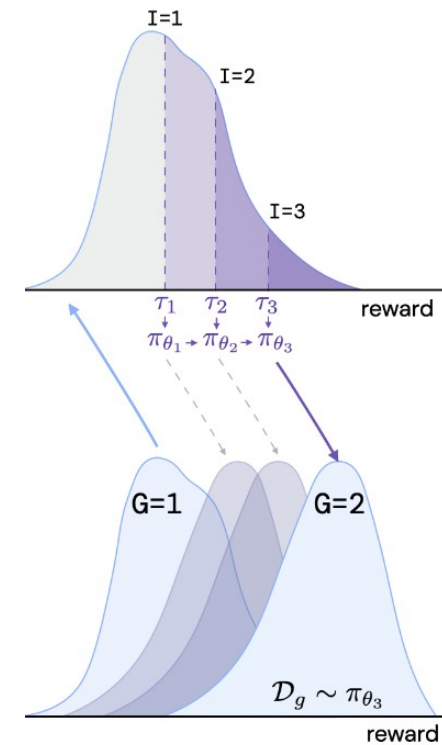
- Daten

- 1) Qualität erhöhen
- 2) Synthetische Daten
- 3) Verifikation
- 4) Self-Play

- Modellgröße

- System 2 Denken (Reasoning)

In manchen Domänen kann man Qualität messen (mit externen Systemen validieren)



Quelle: <https://arxiv.org/pdf/2308.08998v2.pdf>

LLMs - Der Blick nach vorne

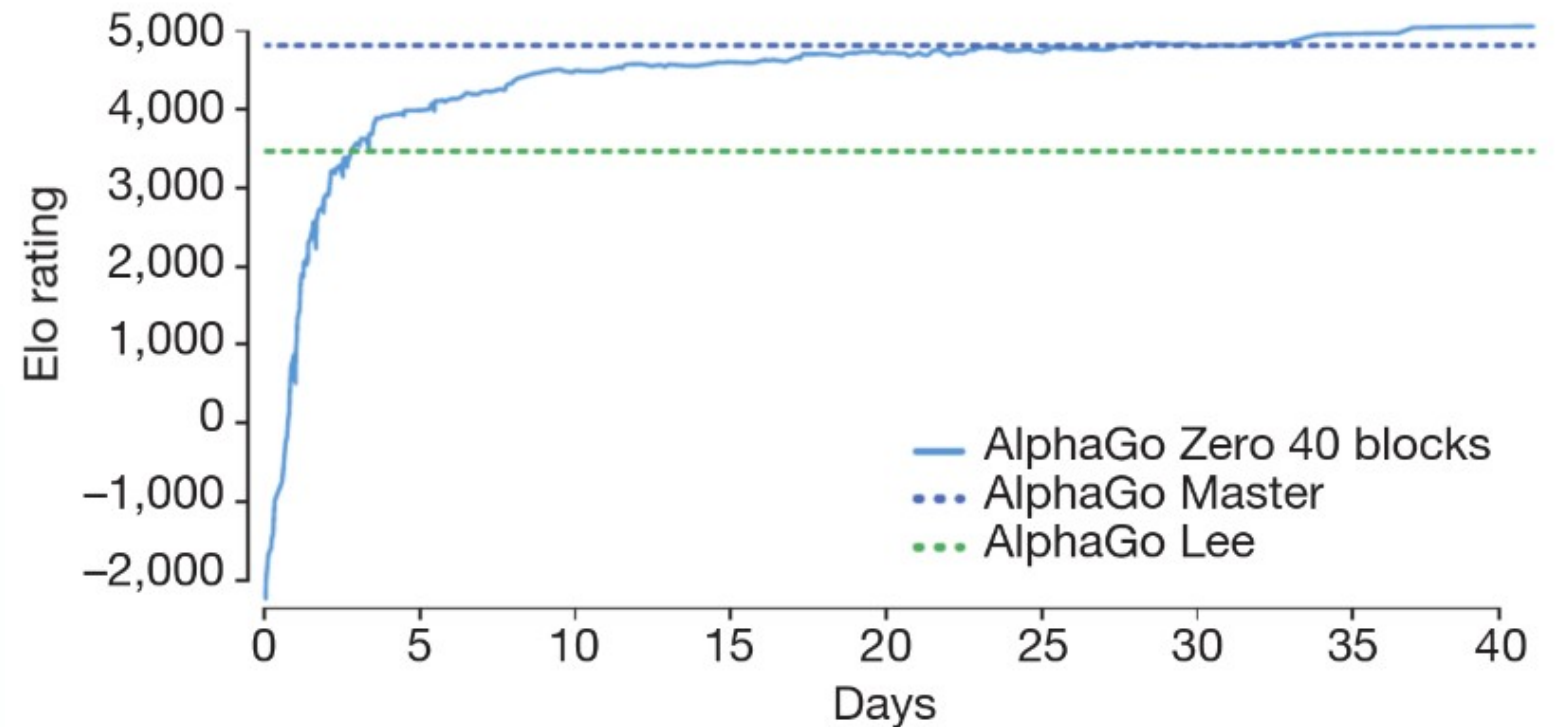
- Daten

- 1) Qualität erhöhen
- 2) Synthetische Daten
- 3) Verifikation
- 4) Self-Play

- Modellgröße

- System 2 Denken (Reasoning)

Self-Play: Feedback in der realen Welt



Quelle: <https://deepmind.google/discover/blog/alphago-zero-starting-from-scratch/>

Self-Play: Feedback in der realen Welt

- [illegible]

© 2024 andrena objects ag

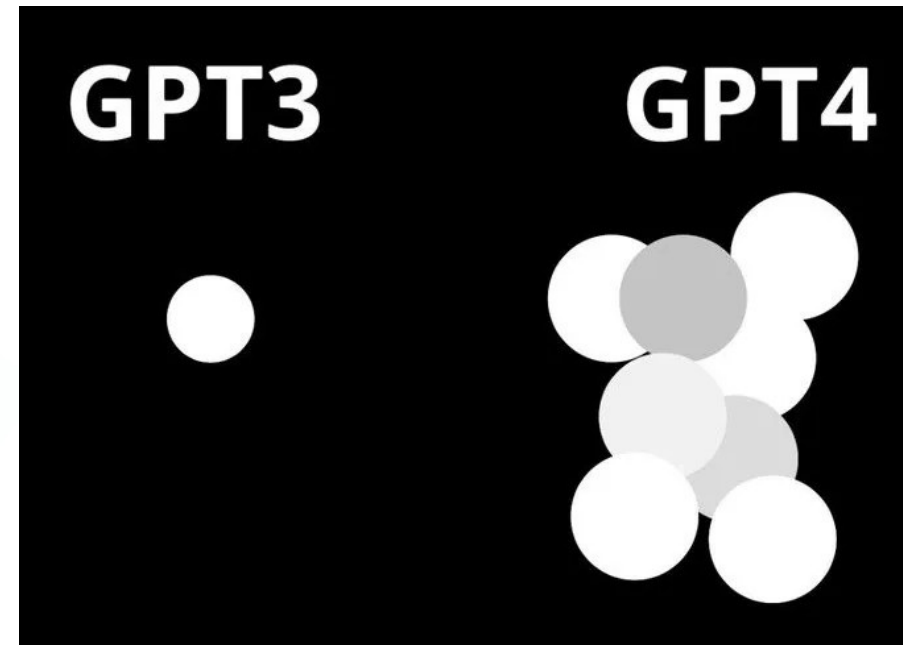
LLMs - Der Blick nach vorne

- Daten
- Modellgröße
- System 2 Denken (Reasoning)

LLMs - Der Blick nach vorne

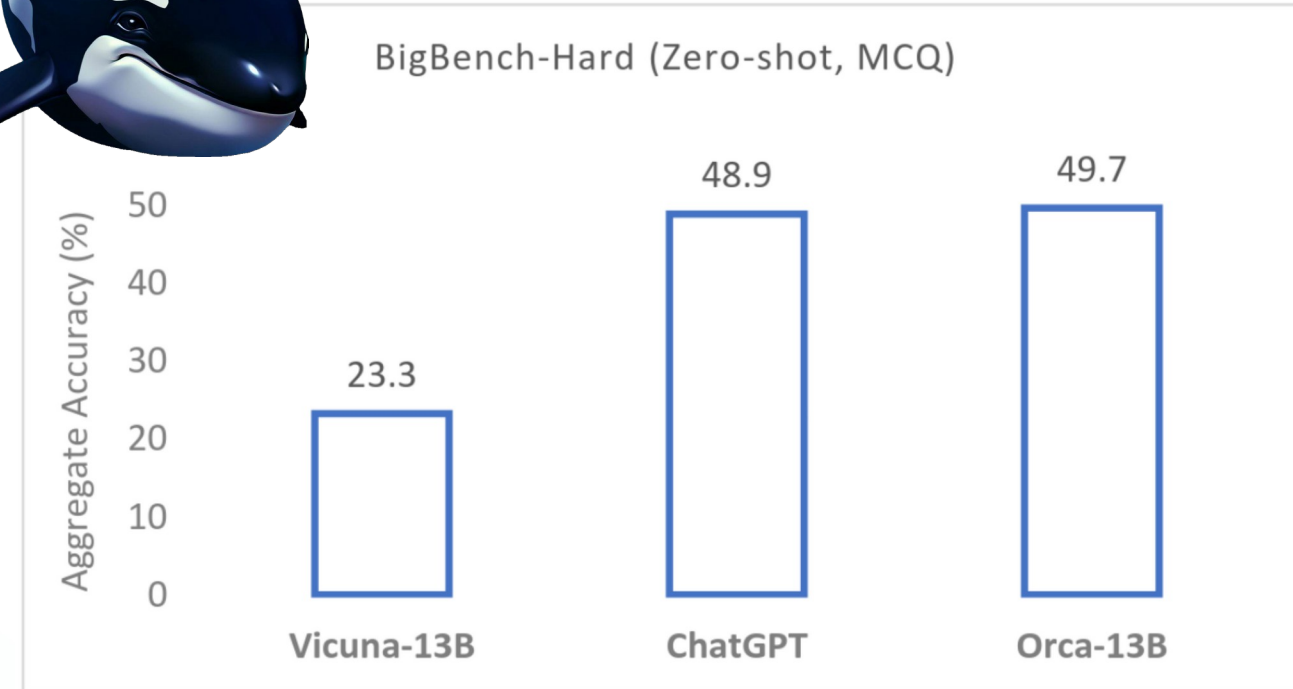
- Daten
- Modellgröße
 - 1) Mixture of Experts (MoE)
 - 2) Destillierung
 - 3) Quantisierung
 - 4) 1-Bit LLM
- System 2 Denken (Reasoning)

- Vor- und Nachteile siehe [Blog](#)
- Beispiele: gpt-4, Mistral 8x7B, Snowflake Arctic 480B (128×3.66B), 17B aktiv
- Scaling Laws of MoE siehe [Paper](#)



LLMs - Der Blick nach vorne

- Daten
- Modellgröße
 - 1) Mixture of Experts (MoE)
 - 2) Destillierung
 - 3) Quantisierung
 - 4) 1-Bit LLM
- System 2 Denken (Reasoning)



Quelle: <https://arxiv.org/pdf/2306.02707.pdf>

LLMs - Der Blick nach vorne

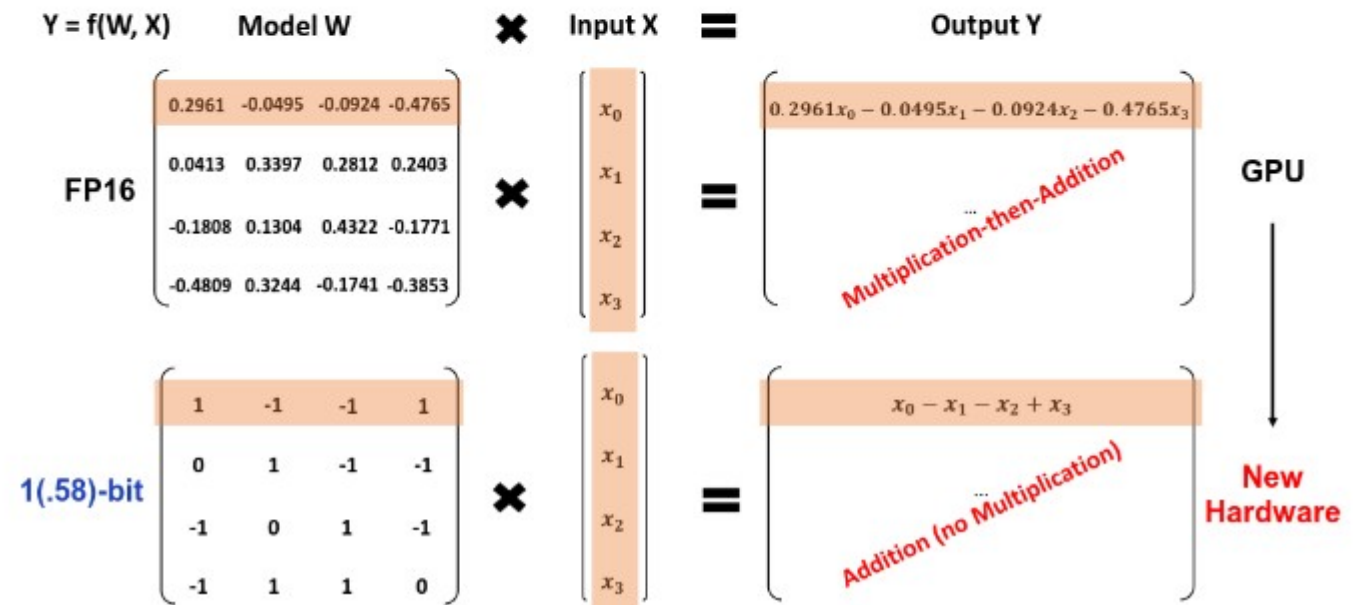
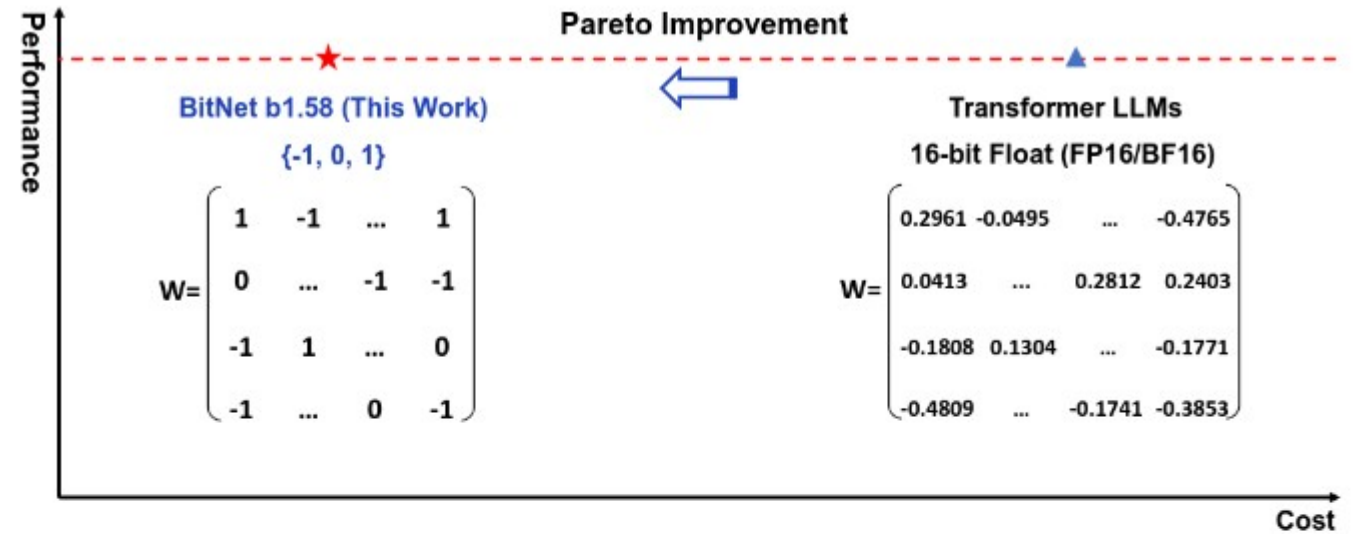
- Daten
- Modellgröße
 - 1) Mixture of Experts (MoE)
 - 2) Destillierung
 - 3) Quantisierung
 - 4) 1-Bit LLM
- System 2 Denken (Reasoning)

Type of Number	Number Range	Value	Precision	Storage (bits)
32-bit Float	$\pm 1.18 \times 10^{-38}$ to $\pm 3.4 \times 10^{38}$	-0.84375	~7-8 decimal digits (varies)	32
8-bit Integer	-128 to 127 (signed)	-108	256 distinct values	8
4-bit Integer	-8 to 7 (signed)	-7	16 distinct values	4
3-bit Integer	-4 to 3 (signed)	-3	8 distinct values	3

Heutige Quantisierungs-Methoden sind smart, d.h. reduzieren die Gewichte so, dass maximale Performance erhalten bleibt.

LLMs - Der Blick nach vorne

- Daten
- Modellgröße
 - 1) Mixture of Experts (MoE)
 - 2) Destillierung
 - 3) Quantisierung
 - 4) 1-Bit LLM
- System 2 Denken (Reasoning)



Quelle: <https://arxiv.org/pdf/2402.17764>

LLMs - Der Blick nach vorne

- Daten
- Modellgröße
- System 2 Denken (Reasoning)
 - 1) In Context Learning
 - 2) CoT und Derivate
 - 3) Let's Think Dot by Dot
 - 4) Let's Verify Step by Step
 - 5) Q*

Schnelles Denken, Langsames Denken



Faustregel:

$$\frac{\text{compute}}{\text{token}} \uparrow \propto \text{quality} \uparrow$$

Rechenaufwand von Training auf Inferenz verlagern

LLMs - Der Blick nach vorne

A::B Prompting Challenge [1]:

- Daten
- Modellgröße
- System 2 Denken (Reasoning)
 - 1) In Context Learning
 - 2) CoT und Derivate
 - 3) Let's Think Dot by Dot
 - 4) Let's Verify Step by Step
 - 5) Q*



Taelin ✓

@VictorTaelin

A simple puzzle GPTs will NEVER solve:

As a good programmer, I like isolating issues in the simplest form. So, whenever you find yourself trying to explain why GPTs will never reach AGI - just show them this prompt. It is a braindead question that most children should be able to read, learn and solve in a minute; yet, all existing AIs fail miserably. Try it!

It is also a great proof that GPTs have 0 reasoning capabilities outside of their training set, and that they'll will never develop new science. After all, if the average 15yo destroys you in any given intellectual task, I won't put much faith in you solving cancer.

LLMs - Der Blick nach vorne

- Daten
- Modellgröße
- System 2 Denken (Reasoning)
 - 1) In Context Learning
 - 2) CoT und Derivate
 - 3) Let's Think Dot by Dot
 - 4) Let's Verify Step by Step
 - 5) Q*

A::B Prompting Challenge [1]:

A::B is a system with 4 tokens: `A#`, `#A`, `B#` and `#B`.

An A::B program is a sequence of tokens. Example:

```
B# A# #B #A B#
```

To **compute** a program, we must rewrite neighbor tokens, using the rules:

```
A# #A ... becomes ... nothing
A# #B ... becomes ... #B A#
B# #A ... becomes ... #A B#
B# #B ... becomes ... nothing
```

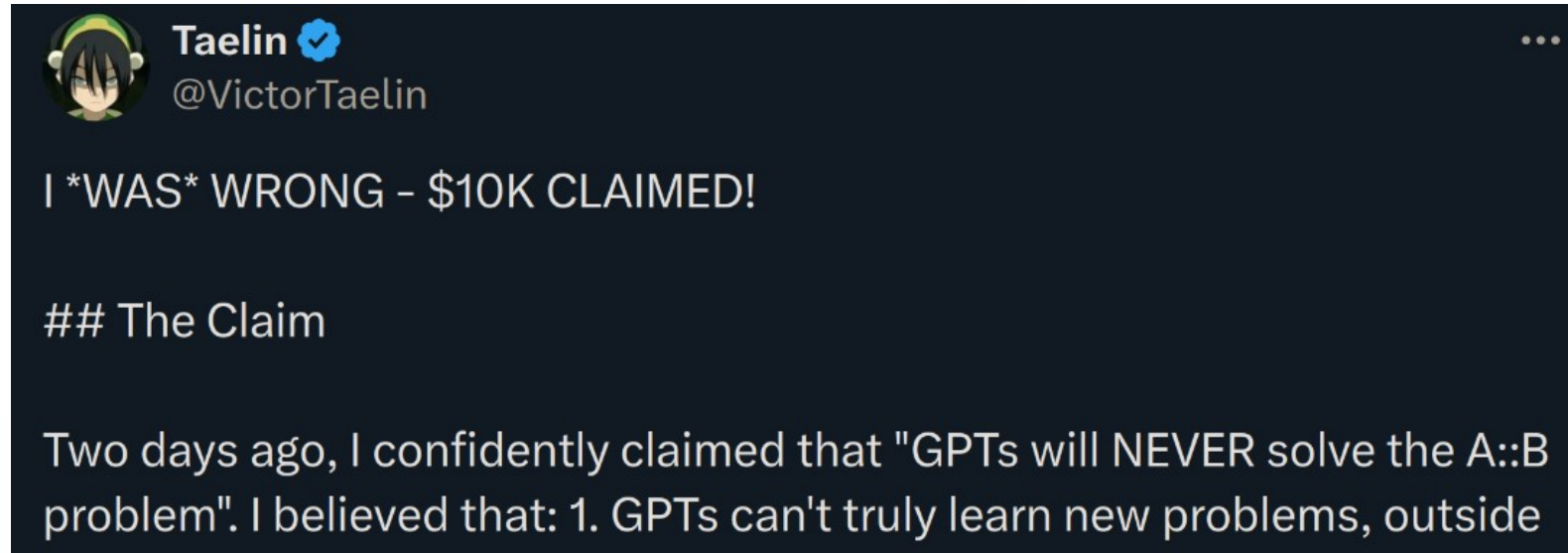
In other words, whenever two neighbor tokens have their '#' facing each-other, they must be rewritten according to the corresponding rule. For example, the first example shown here is computed as:

```
B# A# #B #A B# =
B# #B A# #A B# =
A# #A B# =
B#
```

LLMs - Der Blick nach vorne

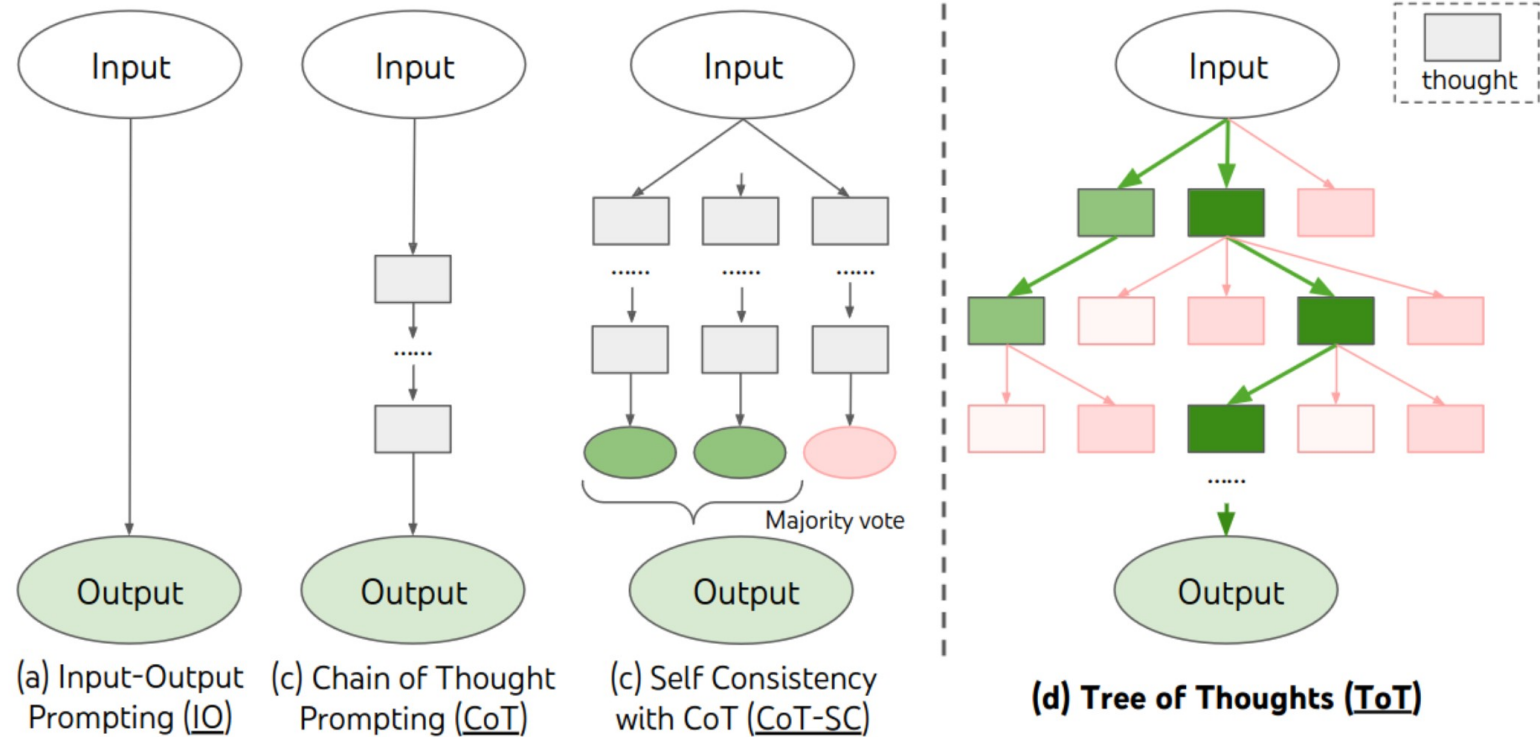
A::B Prompting Challenge [1]:

- Daten
- Modellgröße
- System 2 Denken (Reasoning)
 - 1) In Context Learning
 - 2) CoT und Derivate
 - 3) Let's Think Dot by Dot
 - 4) Let's Verify Step by Step
 - 5) Q*



LLMs - Der Blick nach vorne

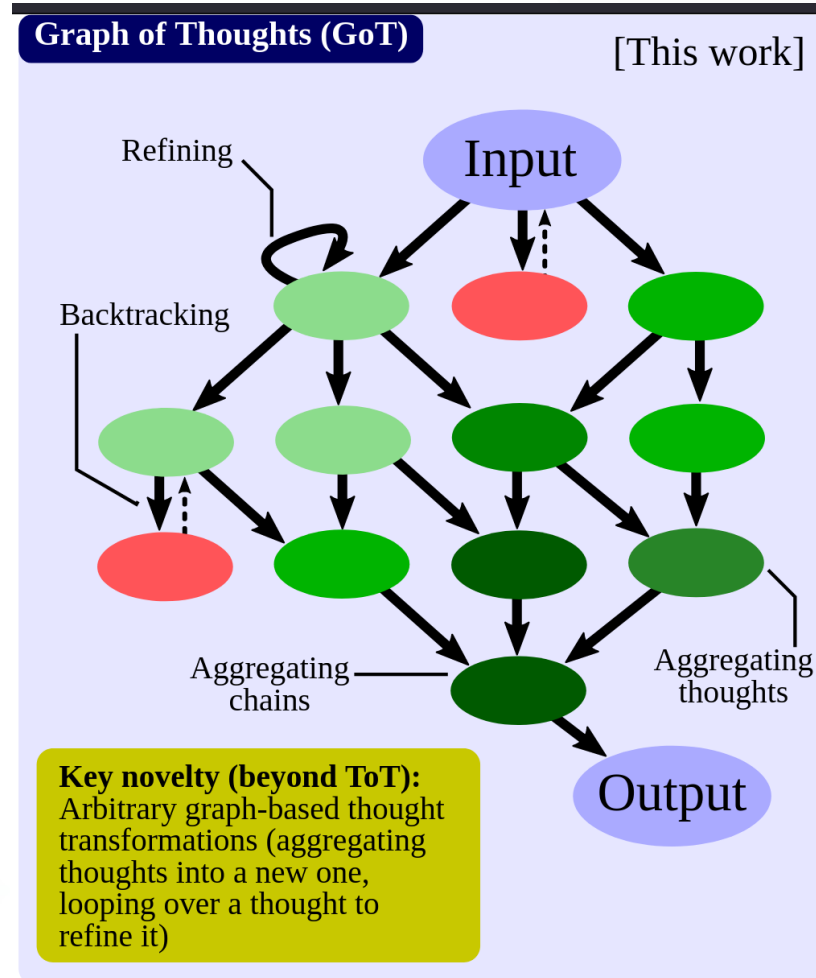
- Daten
- Modellgröße
- System 2 Denken (Reasoning)
 - 1) In Context Learning
 - 2) CoT und Derivate
 - 3) Let's Think Dot by Dot
 - 4) Let's Verify Step by Step
 - 5) Q*



Quelle: <https://arxiv.org/pdf/2305.10601.pdf>

LLMs - Der Blick nach vorne

- Daten
- Modellgröße
- System 2 Denken (Reasoning)
 - 1) In Context Learning
 - 2) CoT und Derivate
 - 3) Let's Think Dot by Dot
 - 4) Let's Verify Step by Step
 - 5) Q*



Quelle: <https://arxiv.org/pdf/2308.09687v2.pdf>

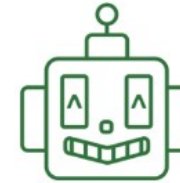
LLMs - Der Blick nach vorne

- Daten
- Modellgröße
- System 2 Denken (Reasoning)
 - 1) In Context Learning
 - 2) CoT und Derivate
 - 3) Let's Think Dot by Dot
 - 4) Let's Verify Step by Step
 - 5) Q*

$$\frac{\text{compute}}{\text{token}} \uparrow \propto \text{quality} \uparrow$$

Prompt: "How many of the first 6 digits of **e** are >5?"

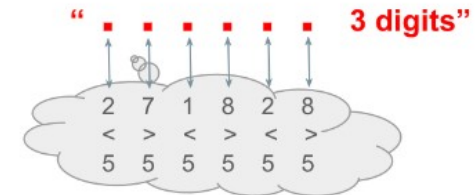
Chain of thought



LM Continuations

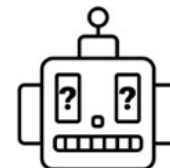
"2<5, 7>5, 1<5, 8>5, 2<5, 8>5,
that's 3 digits"

Filler tokens



Hidden computation using '.' token representations

Immediate answer



"7 digits are greater than 5"

Quelle: <https://arxiv.org/pdf/2404.15758>

LLMs - Der Blick nach vorne

Paper von Mai 2023

- Daten
- Modellgröße
- System 2 Denken (Reasoning)
 - 1) In Context Learning
 - 2) CoT und Derivate
 - 3) Let's Think Dot by Dot
 - 4) Let's Verify Step by Step
 - 5) Q*

Let

$$x^8 + 3x^4 - 4 = p_1(x)p_2(x) \cdots p_k(x),$$

where each non-constant polynomial $p_i(x)$ is monic with integer coefficients, and cannot be factored further over the integers. Compute $p_1(1) + p_2(1) + \cdots + p_k(1)$.

I notice that the given polynomial has even degree and only even powers of x , so I can try to make a substitution to simplify it.

Let $y = x^4$, then the polynomial becomes $y^2 + 3y - 4$, which is a quadratic equation.

I can factor this quadratic equation as $(y + 4)(y - 1)$, so the original polynomial is $(x^4 + 4)(x^4 - 1)$.

Now I need to factor each of these factors further, if possible, over the integers.

For the first factor, $x^4 + 4$, I recall the Sophie Germain identity, which says that $a^4 + 4b^4 = (a^2 + 2b^2 + 2ab)(a^2 + 2b^2 - 2ab)$.

Applying this identity with $a = x$ and $b = 1$, I get $x^4 + 4 = (x^2 + 2x + 2)(x^2 - 2x + 2)$.

For the second factor, $x^4 - 1$, I can use the difference of squares formula, which says that $a^2 - b^2 = (a + b)(a - b)$.

Applying this formula with $a = x^2$ and $b = 1$, I get $x^4 - 1 = (x^2 + 1)(x^2 - 1)$.

Then I can apply the difference of squares formula again with $a = x$ and $b = 1$, to get $x^2 - 1 = (x + 1)(x - 1)$.

So the original polynomial is $(x^4 + 4)(x^4 - 1) = (x^2 + 2x + 2)(x^2 - 2x + 2)(x^2 + 1)(x + 1)(x - 1)$.

These are all the irreducible factors over the integers, so $p_1(x) = x^2 + 2x + 2$, $p_2(x) = x^2 - 2x + 2$, $p_3(x) = x^2 + 1$, $p_4(x) = x + 1$, and $p_5(x) = x - 1$.

To find the sum of their values at $x = 1$, I just need to plug in $x = 1$ into each factor and add them up.

I get $p_1(1) = 1^2 + 2 \cdot 1 + 2 = 5$, $p_2(1) = 1^2 - 2 \cdot 1 + 2 = 1$, $p_3(1) = 1^2 + 1 = 2$, $p_4(1) = 1 + 1 = 2$, and $p_5(1) = 1 - 1 = 0$.

So the final answer is $5 + 1 + 2 + 2 + 0 = 10$.

Answer: 10

Let

$$x^8 + 3x^4 - 4 = p_1(x)p_2(x) \cdots p_k(x),$$

where each non-constant polynomial $p_i(x)$ is monic with integer coefficients, and cannot be factored further over the integers. Compute $p_1(1) + p_2(1) + \cdots + p_k(1)$.

This problem is about factoring a polynomial over the integers, which can be a tricky task.

One possible strategy is to look for rational roots of the polynomial, using the rational root theorem.

However, that would involve checking a lot of possible factors of -4 , and there is no guarantee that the polynomial has any rational roots at all.

Another possible strategy is to use some algebraic tricks to simplify the polynomial or reduce its degree.

I notice that the polynomial has only even powers of x , so I can try to substitute $y = x^2$ and see what happens.

Then the polynomial becomes $y^4 + 3y^2 - 4$, which is a quadratic equation in y^2 .

I can try to factor this quadratic equation by looking for two numbers that multiply to -4 and add to 3 .

Those numbers are 4 and -1 , so I can write $y^4 + 3y^2 - 4 = (y^2 + 4)(y^2 - 1)$.

Now I can substitute back $x^2 = y$ and get $x^8 + 3x^4 - 4 = (x^4 + 4)(x^4 - 1)$.

This looks promising, because I have reduced the degree of the factors from 8 to 4 .

I can try to factor each of these factors further, using the difference of squares formula.

I get $x^8 + 3x^4 - 4 = (x^2 + 2)(x^2 - 2)(x^2 + 1)(x^2 - 1)$.

I can apply the difference of squares formula again to the last factor and get $x^8 + 3x^4 - 4 = (x^2 + 2)(x^2 - 2)(x^2 + 1)(x + 1)(x - 1)$.

Now I have factored the polynomial completely into monic linear and quadratic factors with integer coefficients.

These are the $p_i(x)$'s that the problem is asking for.

To find the sum of their values at $x = 1$, I just need to plug in $x = 1$ into each factor and add them up.

I get $p_1(1) + p_2(1) + \cdots + p_k(1) = (1^2 + 2)(1^2 - 2)(1^2 + 1)(1 + 1)(1 - 1)$.

Simplifying, I get $p_1(1) + p_2(1) + \cdots + p_k(1) = (3)(-1)(2)(2)(0)$.

Multiplying, I get $p_1(1) + p_2(1) + \cdots + p_k(1) = 0$.

Answer: 0

Quelle: <https://arxiv.org/pdf/2305.20050.pdf>

- Daten
- Modellgröße
- System 2 Denken (Reasoning)
 - 1) In Context Learning
 - 2) CoT und Derivate
 - 3) Let's Think Dot by Dot
 - 4) Let's Verify Step by Step
 - 5) Q* (Strawberry)

Vermutlich: Suchbaum über logische Schritte

- *News Artikel*
- *Spekulationen*
- *Artikel zu Strawberry*



Noam Brown ✓
@polynoamial

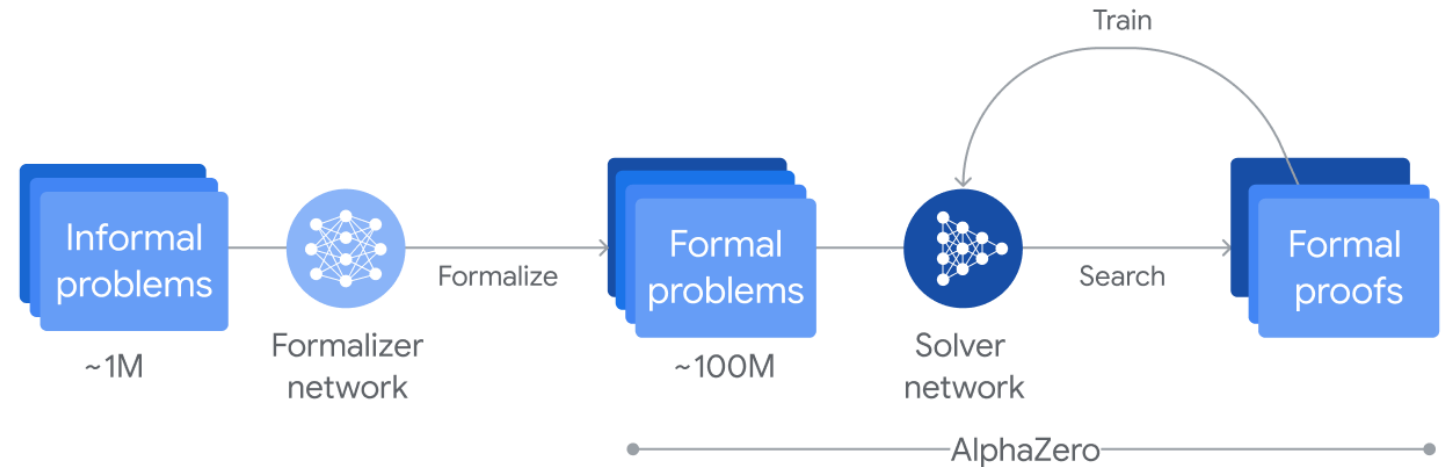
I'm thrilled to share that I've joined @OpenAI! 🚀 For years I've researched AI self-play and reasoning in games like Poker and Diplomacy. I'll now investigate how to make these methods truly general. If successful, we may one day see LLMs that are 1,000x better than GPT-4 🌟 1/

5:08 PM · Jul 6, 2023 · 775.1K Views

Quelle: <https://twitter.com/polynoamial/status/1676971503261454340>

- Daten
- Modellgröße
- System 2 Denken (Reasoning)
 - 1) In Context Learning
 - 2) CoT und Derivate
 - 3) Let's Think Dot by Dot
 - 4) Let's Verify Step by Step
 - 5) Q* (Strawberry)

Google DeepMind: Mathe Olympiade:



As part of our IMO work, we also experimented with a natural language reasoning system, built upon [Gemini](https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/) and our latest research to enable advanced problem-solving skills. This system doesn't require the problems to be translated into a formal language and could be combined with other AI systems. We also tested this approach on this year's IMO problems and the results showed great promise.

Quelle: <https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/>

LLMs - Der Blick nach vorne

Neue LLM-Architektur?

JEPA: Joint Embedding and Predictive Architecture

- Yann LeCuns Projekt [1]
- Predicting embeddings
- Effizienter Lernen & höhere Performance

Mamba [2]:

- Advanced State-Space Model (aufgeborte RNNs)
- Hardware optimization
- Fast inference & linear scaling

StripedHyena [3]:

- Advanced State-Space Model (aufgeborte RNNs)
- mixture of recurrent blocks and multihead attention blocks
- Fast inference & linear scaling

xLSTM [4]:

- Kombinieren von sLSTM und mLSTM Zellen
- Überwindet Nachteile von LSTM (Parallelisierung)
- höhere Performance & lineare Skalierung

KAN [5]:

- Kolmogorov–Arnold Networks
- Ersetzt fixe Aktivierungsfunktionen durch lernbare
- Höhere Performance & leichter interpretierbar



Neue Ansätze müssen noch zeigen, dass sie skaliert besser sind als Transformer

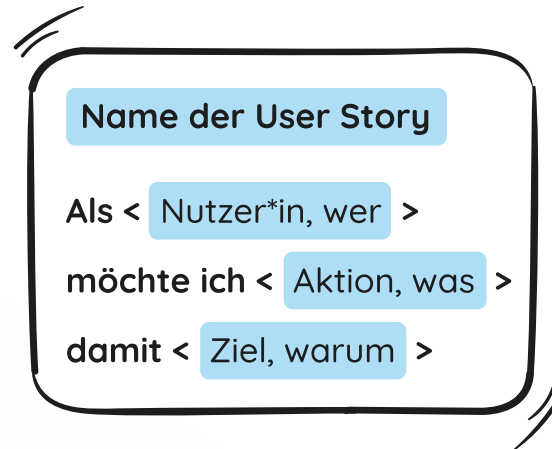
KI & LLMs – Zusammengefasst

- Wir verstehen KI als großen Umbruch ✓
- Wir verstehen was LLMs besonders macht ✓
- Wir haben einen klaren Blick auf die Fähigkeiten von LLMs ✓
- Wir haben einen klaren Blick auf die nahe Zukunft ✓

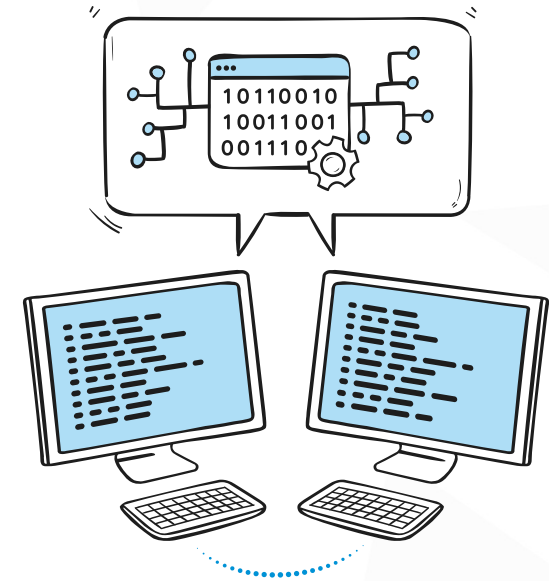
Bedeutung für die SE



Idee

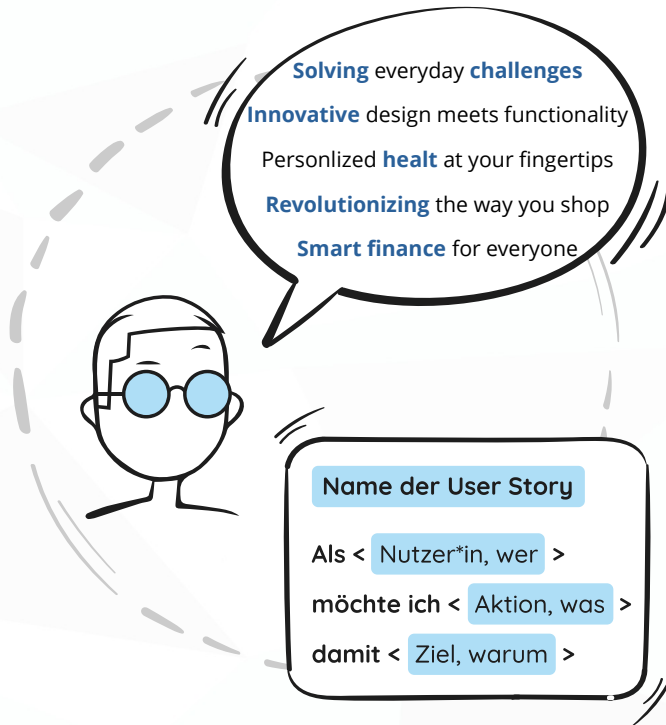


Anforderung

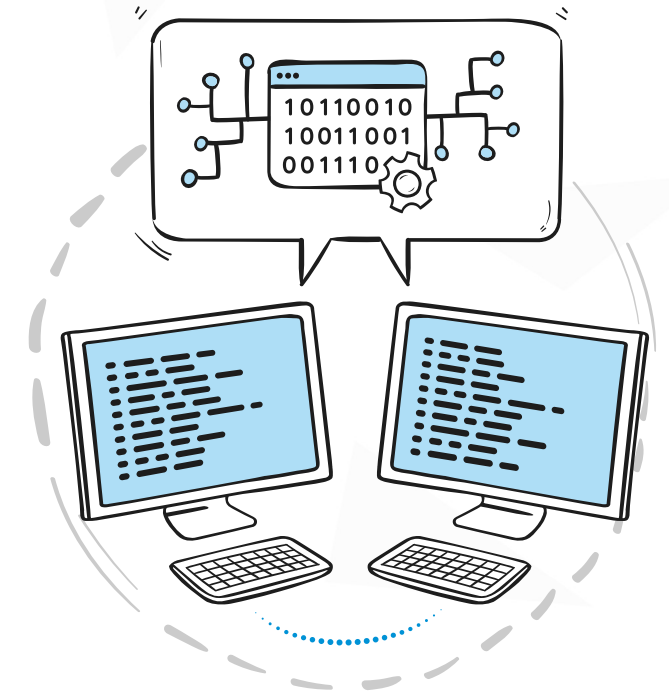
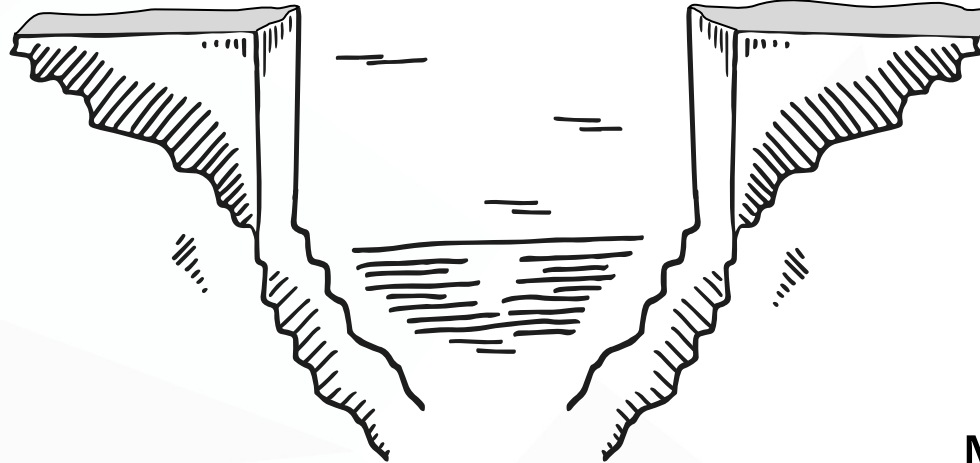


Maschinencode

Bedeutung für die SE



Alltagssprache

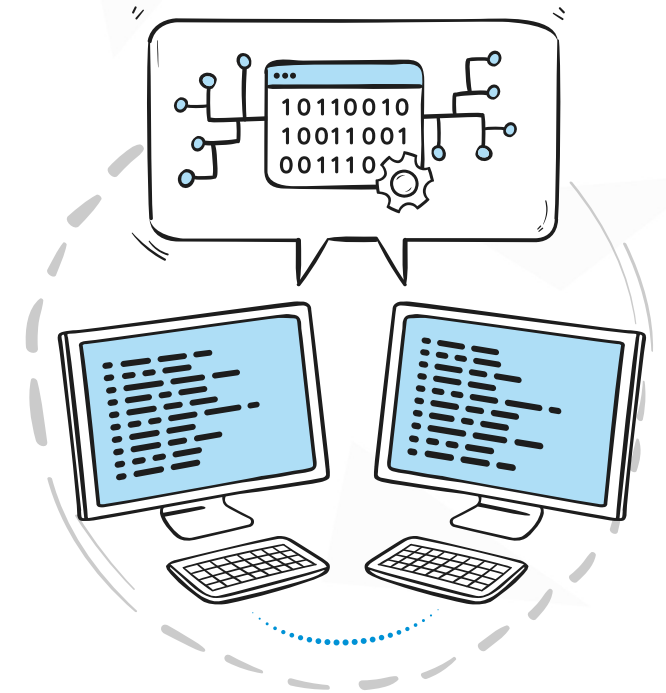
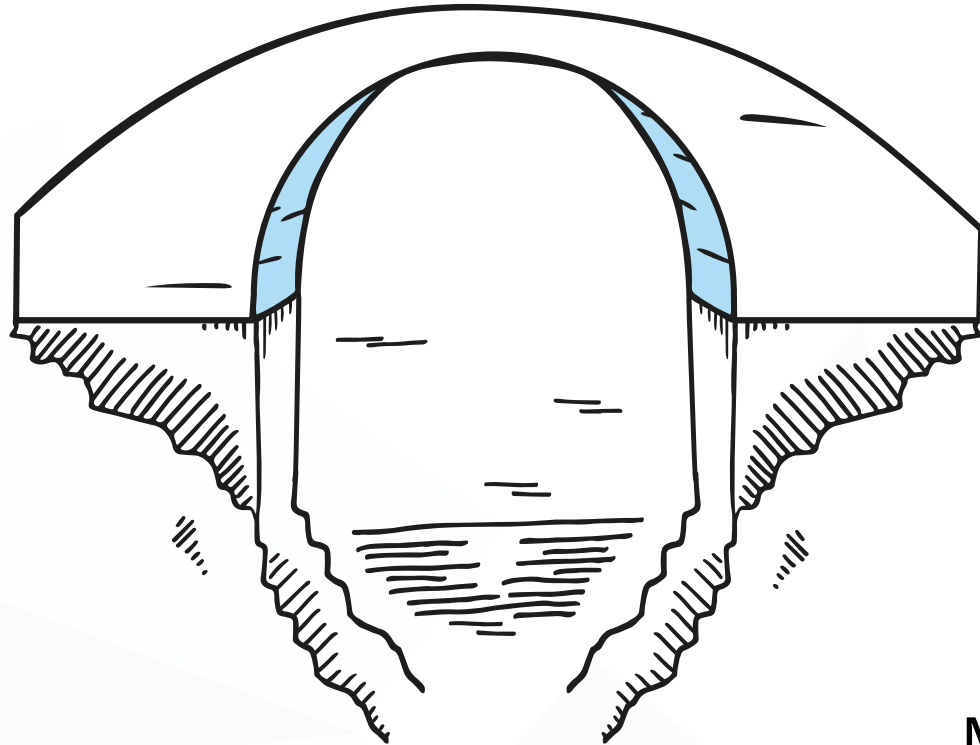


Maschinensprache

Bedeutung für die SE



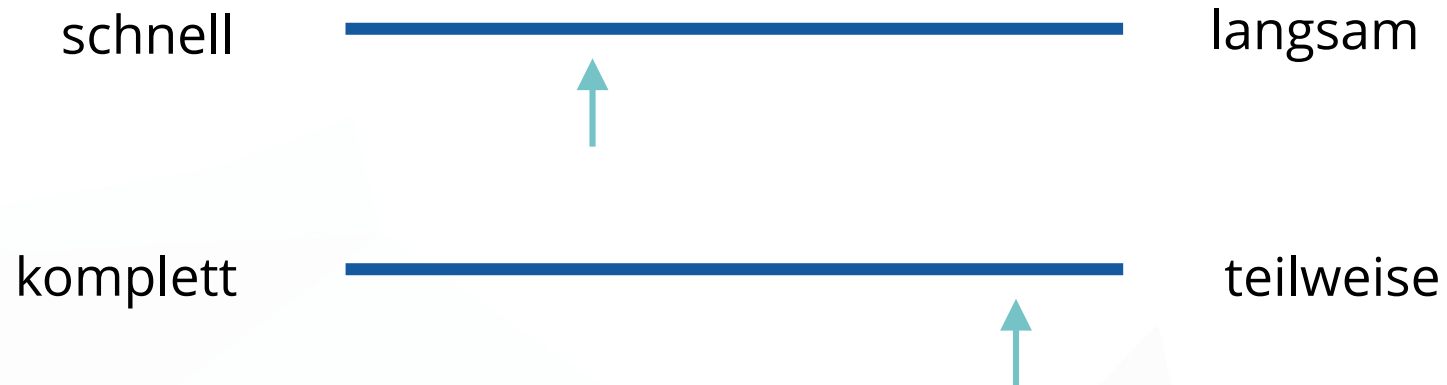
Alltagssprache



Maschinensprache

Bedeutung für die SE

„SE befindet sich auf dem Pfad der Automatisierung“

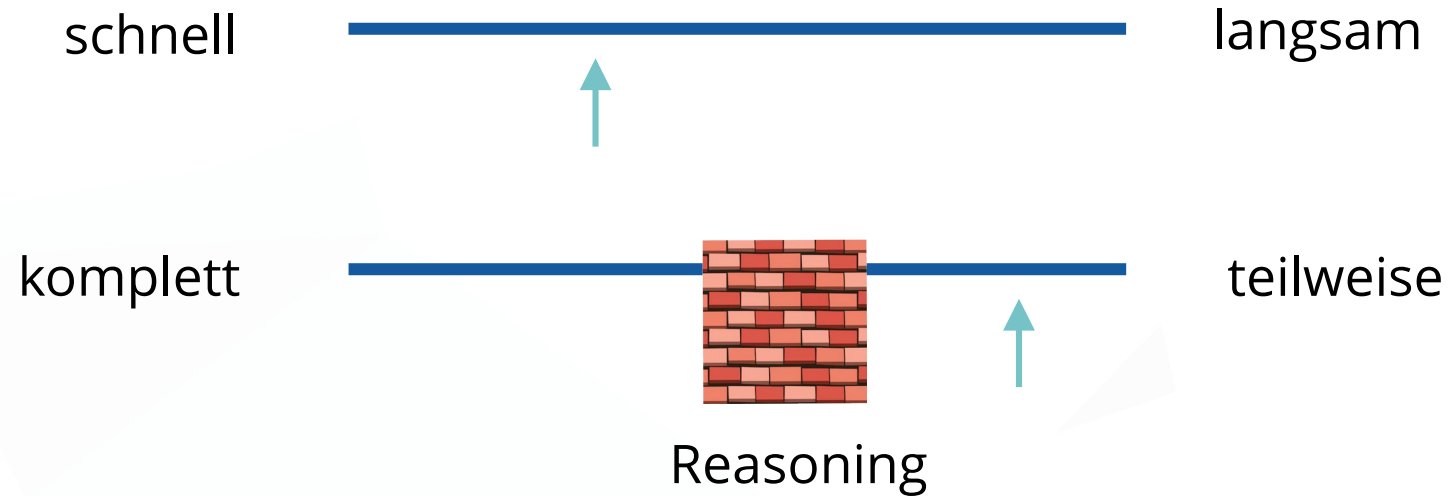


„Sooner than later, 80% of the code is going to be written by Copilot. And that doesn't mean the developer is going to be replaced.“ - Thomas Dohmke ¹

- welche 20% bleiben und werden größer?
- welche neuen Tasks kommen hinzu?

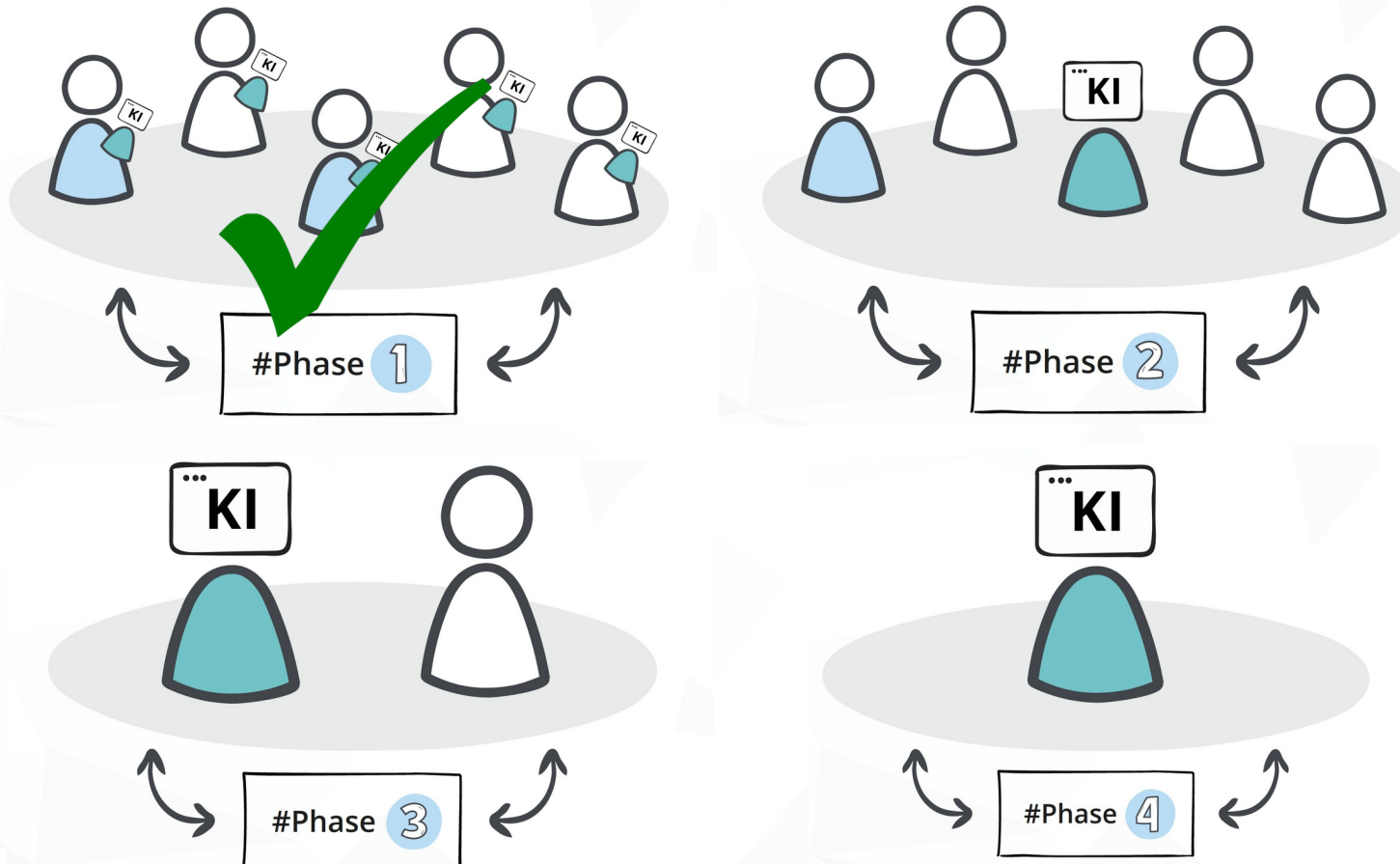
Bedeutung für die SE

„SE befindet sich auf dem Pfad der Automatisierung“



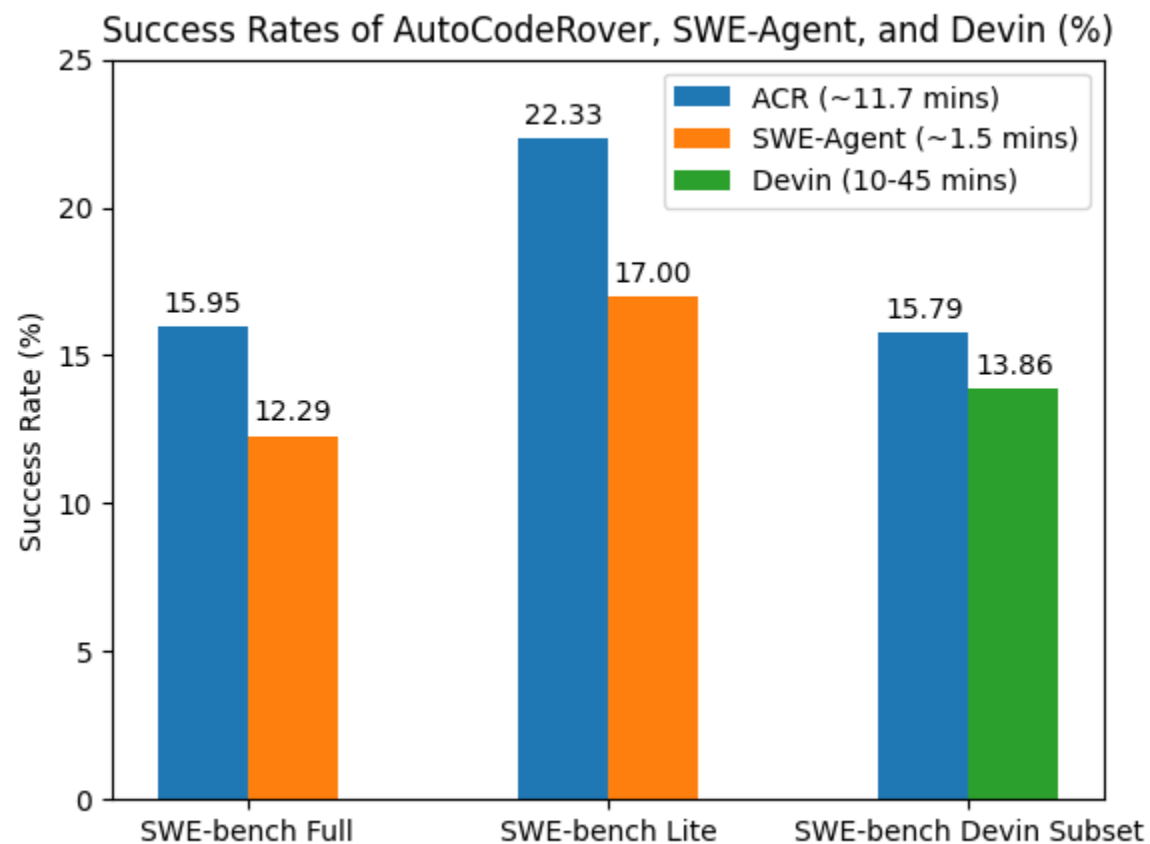
Bedeutung für die SE

„SE befindet sich auf dem Pfad der Automatisierung“



LLM – Fähigkeiten

- Wie können wir Fähigkeiten quantifizieren?



Quelle: <https://github.com/nus-apr/auto-code-rover>

Prognosen

- LLMs werden deutlich fähiger werden
 - Hängt entscheidend an der Fähigkeit des „Langsamen Denkens“
 - Haben aktuell erst zwei Iterationen gesehen, bald kommt die dritte Iteration
- Wir werden im SWE Benchmark große Fortschritte sehen
 - Wie lange wird es dauern bis 90% erreicht wurden?
- Software Entwickler wird es auch dann noch geben, wenn der SWE Benchmark zu 100% gelöst wurde
 - für den Blick aufs Große Ganze
 - Automatisierte Code Generierung wird in manchen Bereichen gesellschaftlich nicht akzeptiert sein



Prognosen

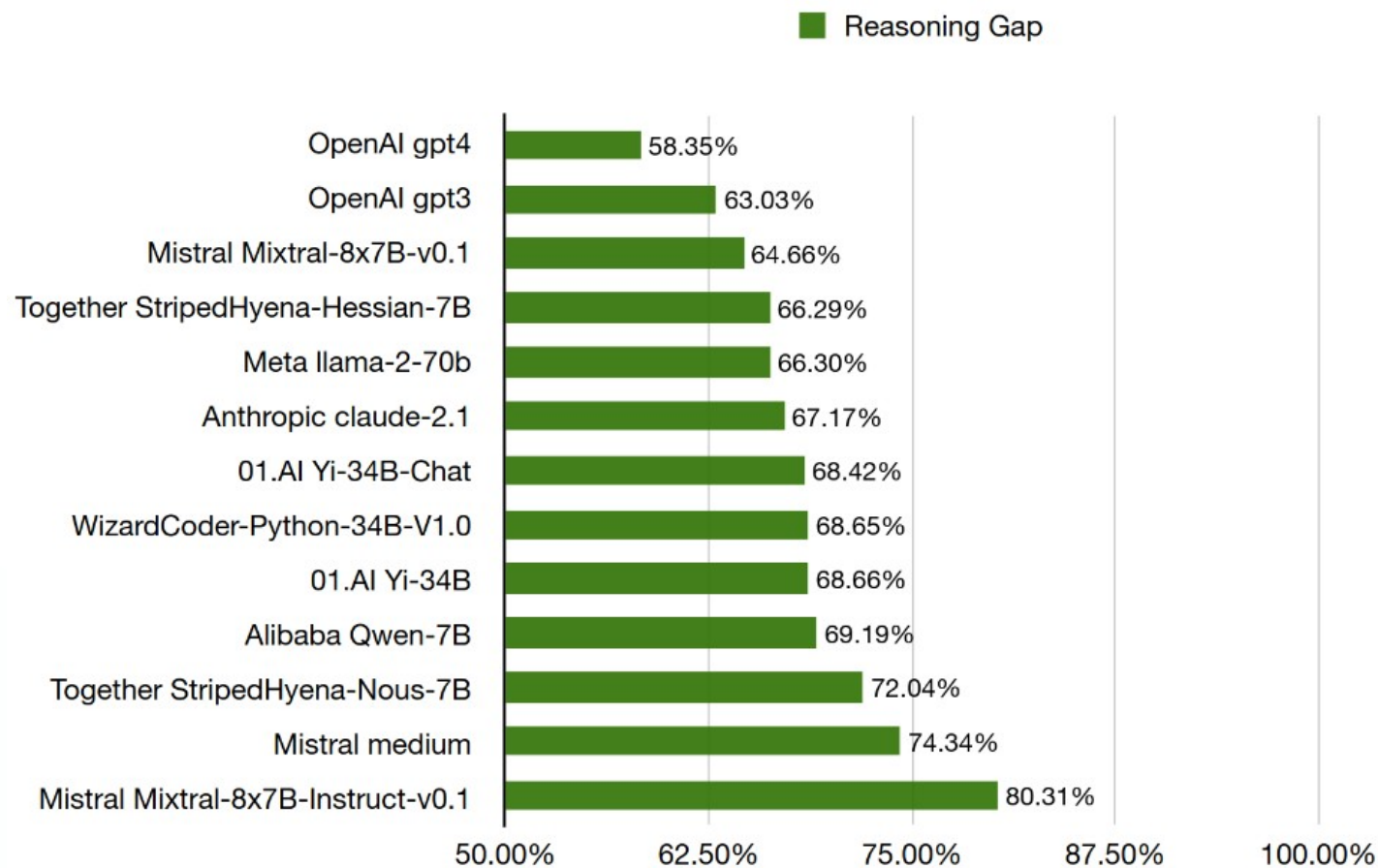
- Wird ein Computer in meiner Lebenszeit alle meine Denkleistungen übertreffen?
 - AGI macht Softwareentwickler überflüssig!
 - Aber eben auch die Mehrheit aller anderen Jobs



*„Softwareentwickler werden automatisiert und verlieren ihre Jobs,
aber nicht früher als andere Wissensberufe.“*

Signale & Quellen

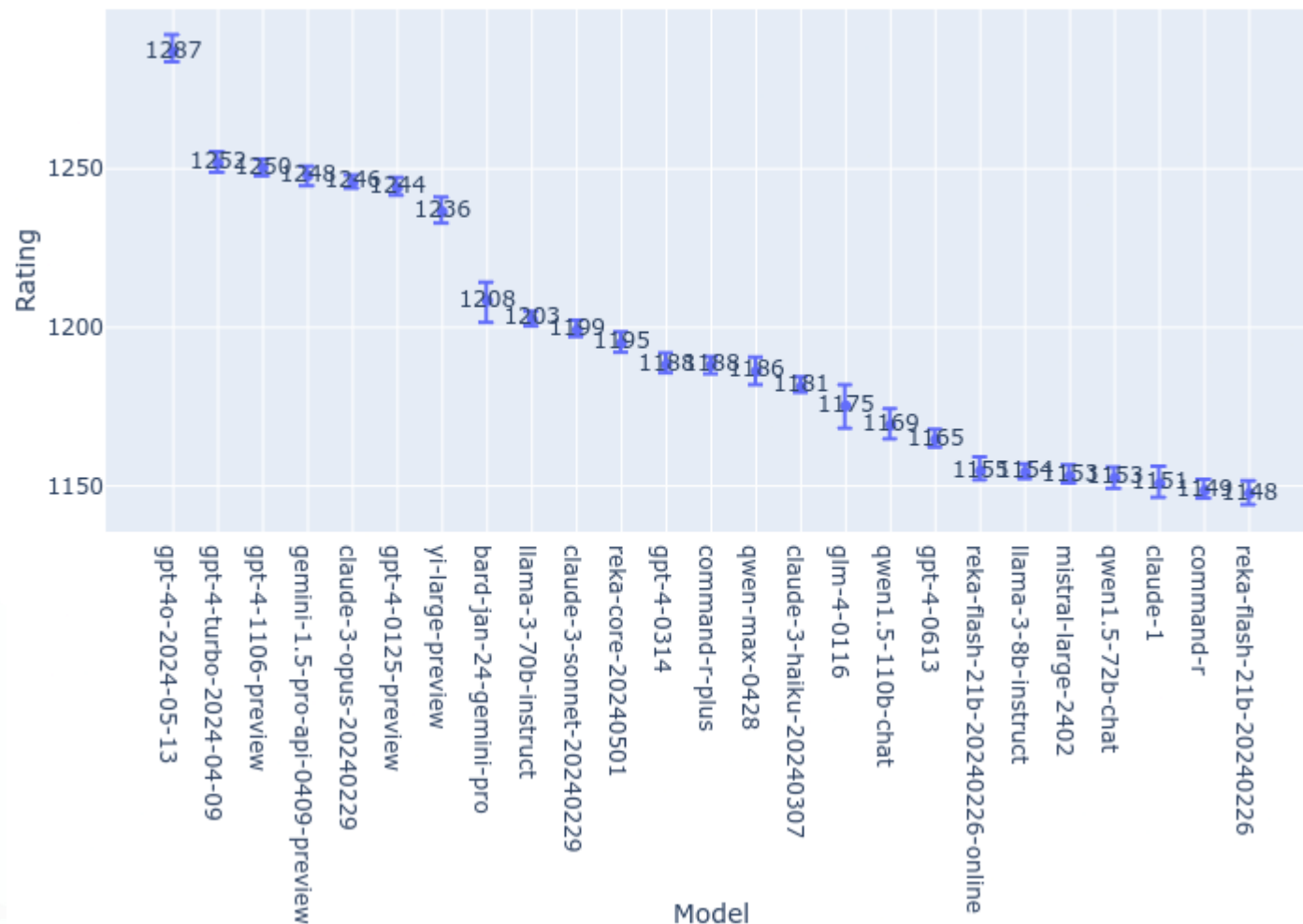
- Reasoning Gap wird kleiner



Quelle: <https://arxiv.org/pdf/2402.19450>

Signale & Quellen

- Reasoning Gap wird kleiner
- Chatbot Arena* Elo-Rating
 - Δ 50 Punkte: 43% vs 57%
 - Δ 100 Punkte: 36% vs 64%



Signale & Quellen

- Reasoning Gap wird kleiner
- *Chatbot Arena* Elo-Rating
- Private Benchmarks

Signale & Quellen

YouTube

- *AI Explained*
- *Machine Learning Street Talk*
- *bycloud*

Podcast

- *Latent Space*
- *Last Week in AI*
- *Your Undivided Attention*

Newsletter

- *AI News*
- *AlphaSignal*
- *semianalysis*

Blogs

- *Interconnects*
- *eugeneyan*
- *Don't Worry About the Vase*

Zusammenfassung I

Was ist Software Entwicklung?

- Entwickler sind Problemlöser, die mit komplexen mentalen Modellen jonglieren
- Ein Großteil der Arbeit geschieht implizit und ist schwer greifbar
- Komplexes Umfeld (Stacey Matrix)

Was kann KI heute in der SE leisten?

- KI-Suchmaschinen
- Coding Assistenten
- Chatbots als Orakel

Zusammenfassung II

Was ist KI?

- Der Versuch, das menschliche Denken zu externalisieren
- Ein gewaltiger Entwicklungssprung in der Evolution

Was sind LLMs?

- Die ersten allgemeinen KI-Systeme, die bereits Mehrwert schaffen
- Konstruieren ein Weltmodell
- Treffen alleine Entscheidungen
- Sprechen unsere Sprache
- Sind noch schlecht im Langsamen Denken

Zusammenfassung III

Was liegt vor uns?

- Wahrscheinlich eine deutliche Verbesserung von LLMs (Suchbaum über logische Schritte)
- Limit von LLMs ist unbekannt
- Schrittweise Automatisierung in der Softwareentwicklung
 - Vom Kleinen zum Großen
 - Neue Software Architektur entsteht (Best Practises, Frameworks bilden sich)
- Ja wir werden automatisiert, aber nicht früher als andere Wissensberufe
- Und bis dahin gilt:

“It is **not** the **AI** that will replace you, it is the **people** who use AI effectively”

Zusammenfassung IV

Was kannst du tun?

- Erfahrung sammeln = realistischen Blick bekommen
- Dein mentales Modell häufig aktualisieren (nicht an Glaubenssätzen festhalten)
- Auf Signale achten
- Freue dich, dass du nicht nur die spannendste Zeit der Menschheitsgeschichte erlebst, sondern auch zu der Berufsgruppe gehörst, die diese Systeme versteht und gestalten kann



Skeptiker



Besorgter Enthusiast

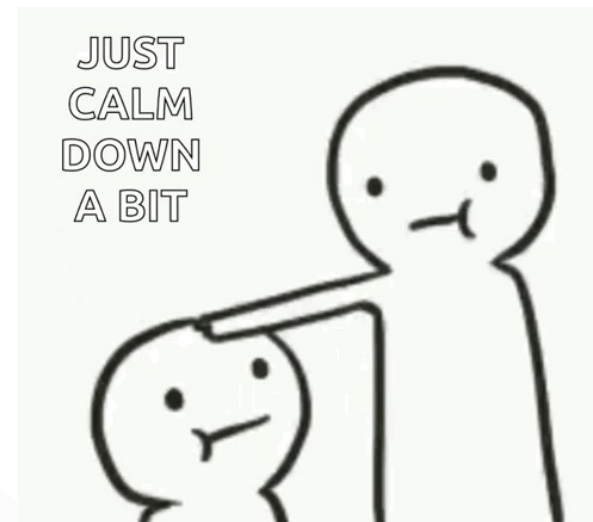
Zusammenfassung IV

Was kannst du tun?

- Erfahrung sammeln = realistischen Blick bekommen
- Dein mentales Modell häufig aktualisieren (nicht an Glaubenssätzen festhalten)
- Auf Signale achten
- Freue dich, dass du nicht nur die spannendste Zeit der Menschheitsgeschichte erlebst, sondern auch zu der Berufsgruppe gehörst, die diese Systeme versteht und gestalten kann



Skeptiker



Besorgter Enthusiast

Kontakt



steve.haupt@andrena.de



www.linkedin.com/in/stevehaupt



www.youtube.com/@developers-club

Fragen

