academy
NZZ

The State of AI

Jürg Stuker

Pontresina | 21 October 2025

kick start academy NZZ

> In short, I believe more than ever that programming should be a key part of the intellectual development of people growing up.
>
> — Seymour Papert, 1970

kick start academy NZZ

Ohne Fundament wird es gefährlich, auch bei KI

Pontresina, 21. Oktober 2025
Jürg Stuker

kick start academy NZZ

Mein Menü-Vorschlag zum Thema KI

Key Developments in AI

- **Kosten pro Token (model routing)**
- **Reichweite (distribution)**
- **Monetarisierung-Versuche**
- **Zugang zu Rechenleistung/KI**

kick start academy NZZ

Aufwärmrunde

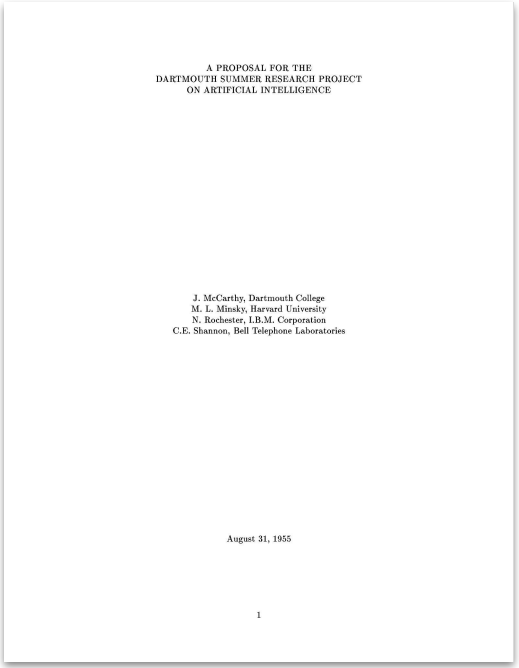
kick start academy NZZ

Wann wurde die erste “Artificial Intelligence” entwickelt?

- vor 2020?
- vor 2015?
- vor 2010?
- vor 2000?
- vor 1990?
- vor 1980?
- vor 1960?
- vor 1950?
- vor 1930?

kick start academy NZZ

1955: Dartmouth Summer Research Project on Artificial Intelligence

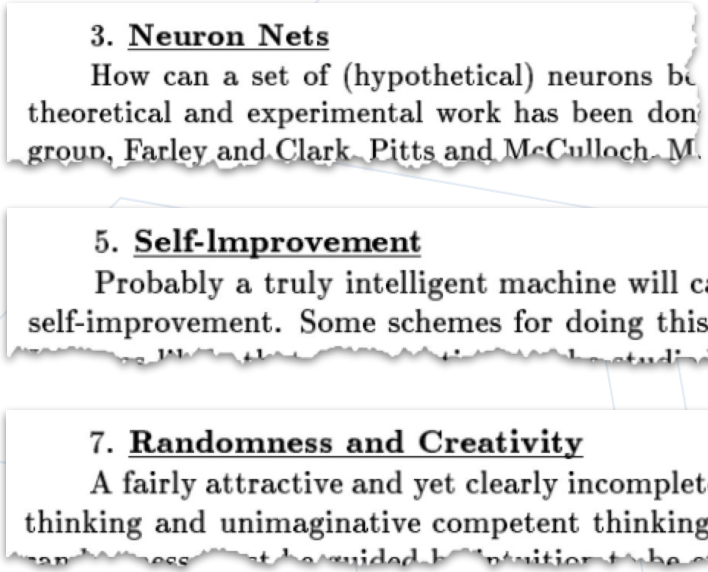


A PROPOSAL FOR THE
DARTMOUTH SUMMER RESEARCH PROJECT
ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

1



3. Neuron Nets

How can a set of (hypothetical) neurons b
theoretical and experimental work has been don
group, Farley and Clark, Pitts and McCulloch, M.

5. Self-Improvement

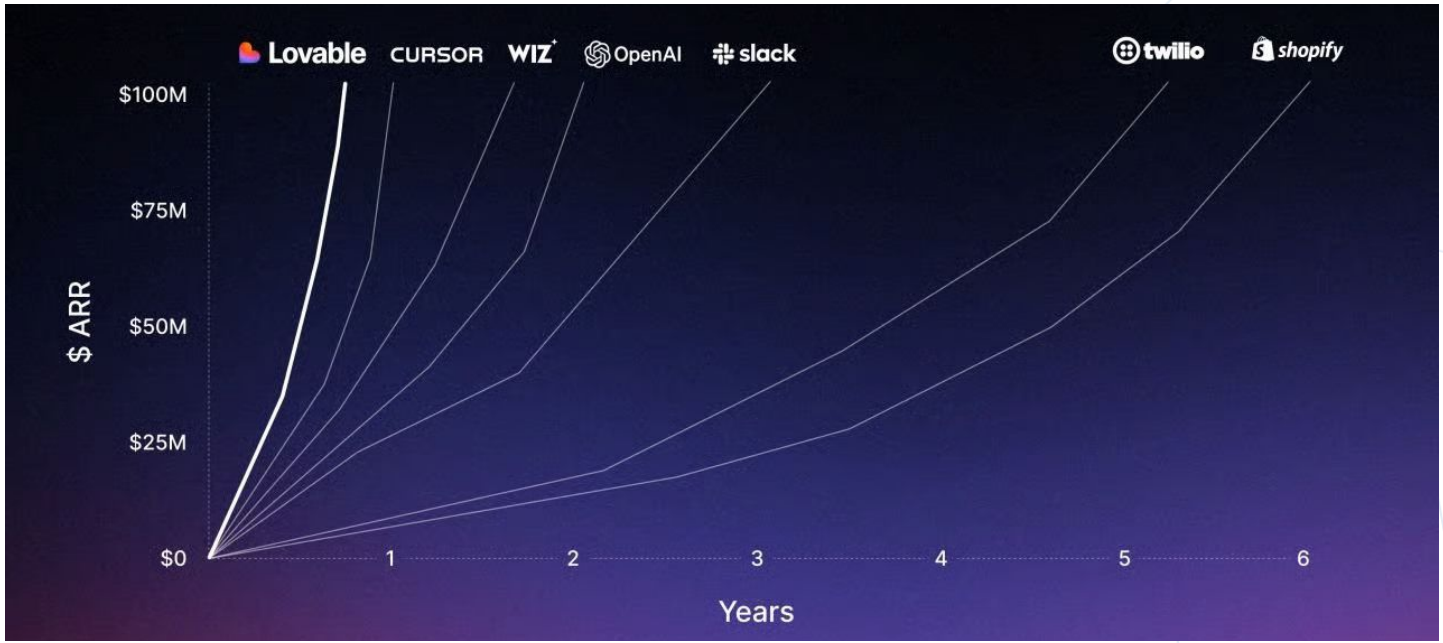
Probably a truly intelligent machine will ca
self-improvement. Some schemes for doing this

7. Randomness and Creativity

A fairly attractive and yet clearly incomplete
thinking and unimaginative competent thinking

Quelle: http://jmc.stanford.edu/articles/dartmouth.html

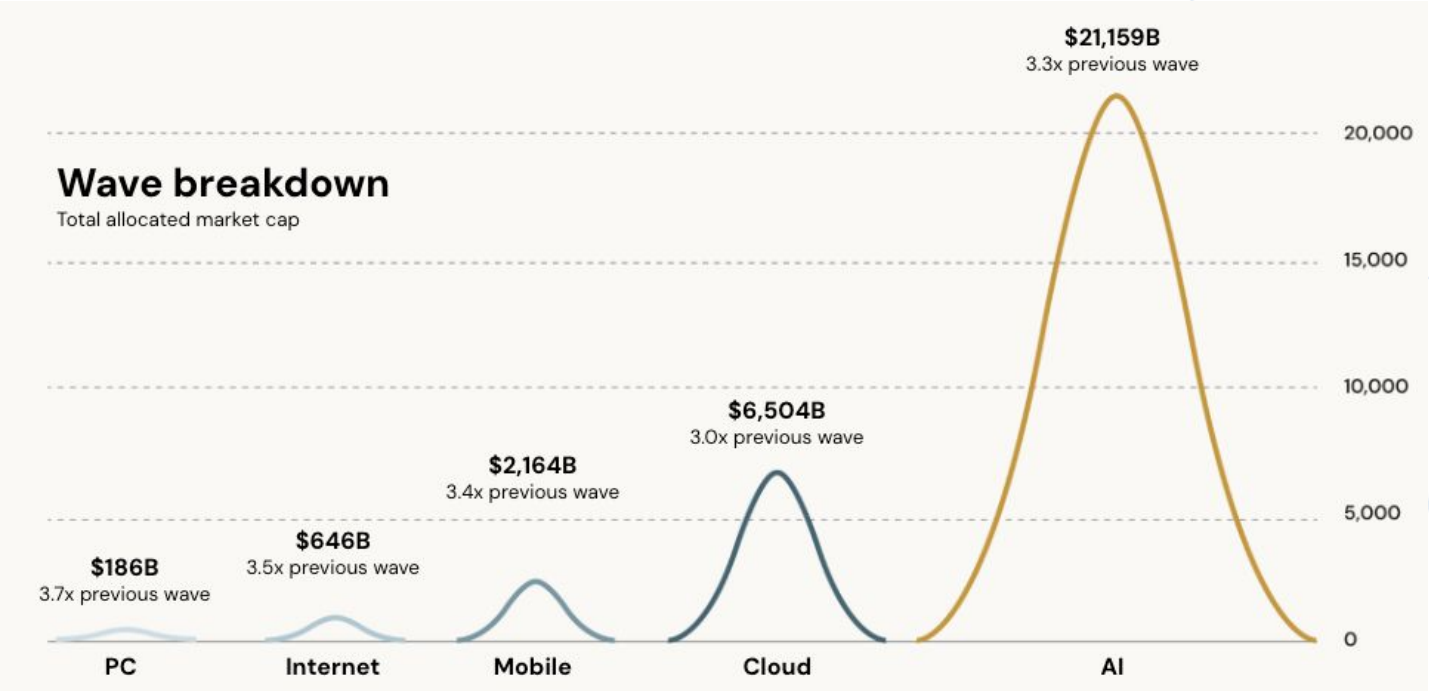
Unglaublich schnell: Fastest USD 1M → USD 100M ARR



Quelle: https://www.fintechbrainfood.com/p/fintech-fixes-ai

kick start academy NZZ

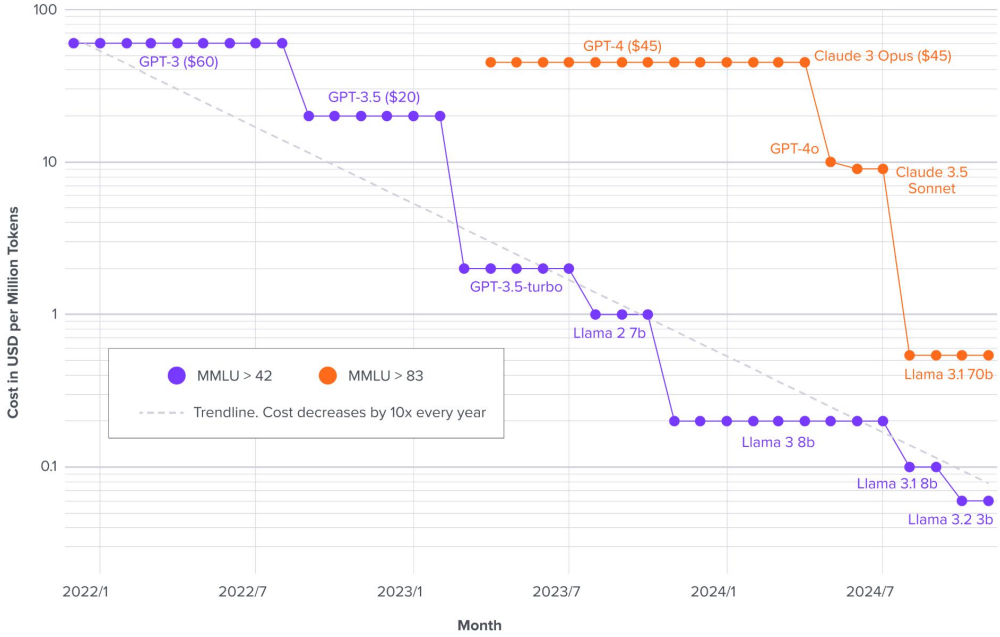
Unglaublich gross: Will AI follow the “Rule of 3”?



Quelle: https://fundrise.com/education/value-of-ai-research

kick start academy NZZ

Unglaublich effizient: Cost of Cheapest LLM with Minimum MMLU Score



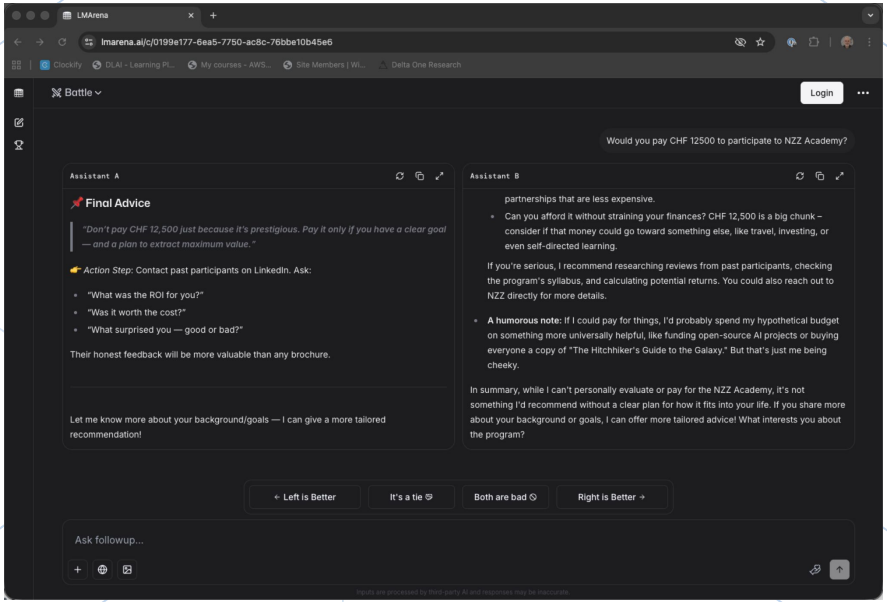
Quelle: https://a16z.com/llmflation-llm-inference-cost/

Wie misst man den “IQ” von KI-Systemen?

Synthetische Benchmarks (ground truth)

- Beispiel MMLU: Massive Multitask Language Understanding, https://arxiv.org/pdf/2009.03300
  - multiple-choice questions spanning 57 distinct subject areas (history, philosophy, literature, psychology, economics, politics, mathematics, physics, engineering, law, medicine, accounting..)
- Beispiel ARC-AGI General Intelligence Benchmark, https://arcprize.org
  - Abstract and Reasoning Corpus
- Beispiel HLE: Humanity's Last Exam, https://agi.safe.ai
  - 2,500 challenging questions across over a hundred subjects

Bewertung durch Menschen (human preferences) z.B. LMArena, lmarena.ai



KI-Modell versus KI-System

Ein KI-Modell ist der Algorithmus, der durch das Training mit Daten Muster lernt und spezifische Aufgaben ausführt. Auch: “Foundational Model”.

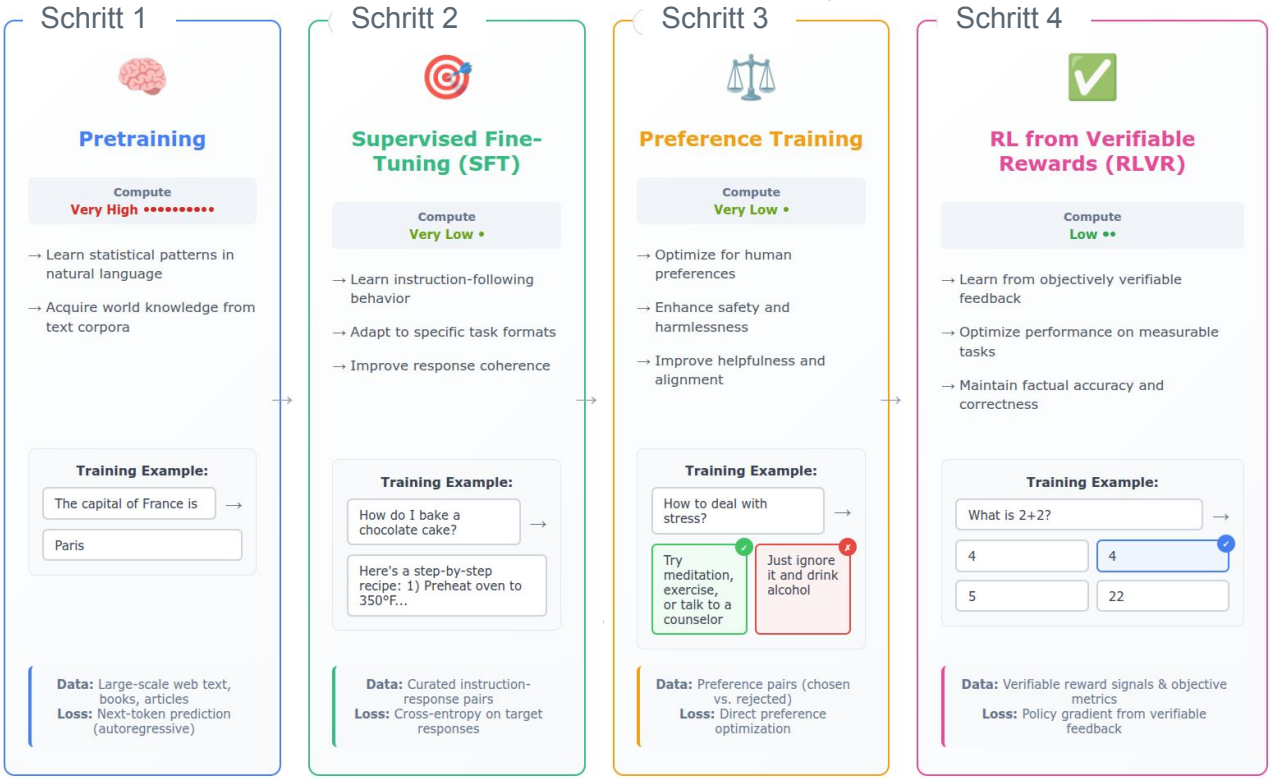
- Sprachmodelle (Language Models)
  - LLM (Large Language Model)
  - LAM (Language Action Models)
- Bildmodelle (Vision Models)
- Multimodale Modelle
- Recommender-Systeme

Ein KI-System ist die vollständige, funktionale Anwendung, die ein oder mehrere KI-Modelle mit einer Benutzeroberfläche, Infrastruktur und anderen Komponenten integriert, um eine Dienstleistung bereitzustellen.

Analogie: Motor versus Auto

kick start academy NZZ

Entwicklungsschritte eines Sprachmodells



Quelle: Swiss {ai} Weeks. Tech Round on the LLM Effort of the Swiss AI Initiative. Imanol Schlag, Martin Jaggi.

kick start academy NZZ

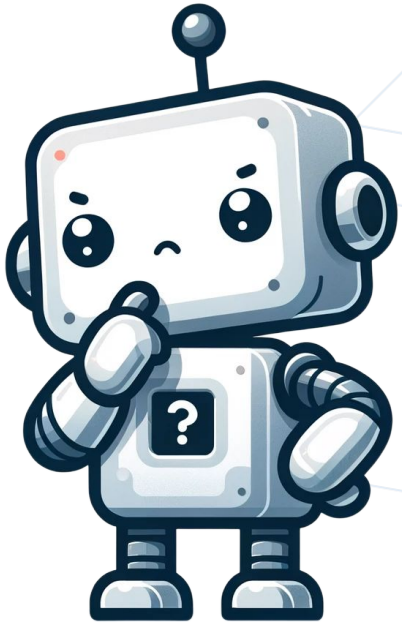
Drei Generationen von generativen KI-Systemen

	Bot *(prompting)*	**Assistant** *(reasoning)*	**Agent** *(agentic)*
Purpose	Automating simple tasks or conversations	Assisting users with tasks	Autonomously and proactively perform tasks
Capabilities	Follows pre-defined rules; limited learning; basic interactions	Responds to requests or prompts; provides information and completes simple tasks; can recommend actions but the user makes decisions	Can perform complex, multi-step actions; learns and adapts; can make decisions independently
Interaction	**Reactive**; responds to triggers or commands	**Reactive**; responds to user requests	**Proactive**; goal-oriented

Quelle: https://cloud.google.com/discover/what-are-ai-agents

kick start academy NZZ

Fragen und Diskussion



kick start academy NZZ

AI Safety

https://youtu.be/giT0ytynSqg?si=H8HP9c0kbJFGCC8J&t=426

Klassifikation von KI-Risiken

Voreingenommenheit (Bias) und Fairness

Datenschutz

Kontrollverlust

Existenzielle Risiken

Böswilliger Missbrauch

Cybersicherheit

Beispielhafte Szenarien

- Selektion von Bewerber:innen
- Bewilligung eines Kredits
- Zugang zu und Missbrauch von PII (persönlich identifizierbare Informationen)
- Autonome KI-Agenten
- AGI (artificial general intelligence), ASI (artificial superintelligence)
- KI als Waffe für Cyberangriffe
- Unsichere KI Systeme
- …

Quelle: https://www.ibm.com/think/topics/ai-safety

Massnahmen zur Kontrolle von KI-Risiken

Algorithmische Erkennung und Minderung von Bias

Robustheitstests und Validierung (Red Teaming)

Erklärbare KI (XAI, Explainable AI)

Ethische KI-Rahmenwerke

Menschliche Aufsicht

Sicherheitsprotokolle

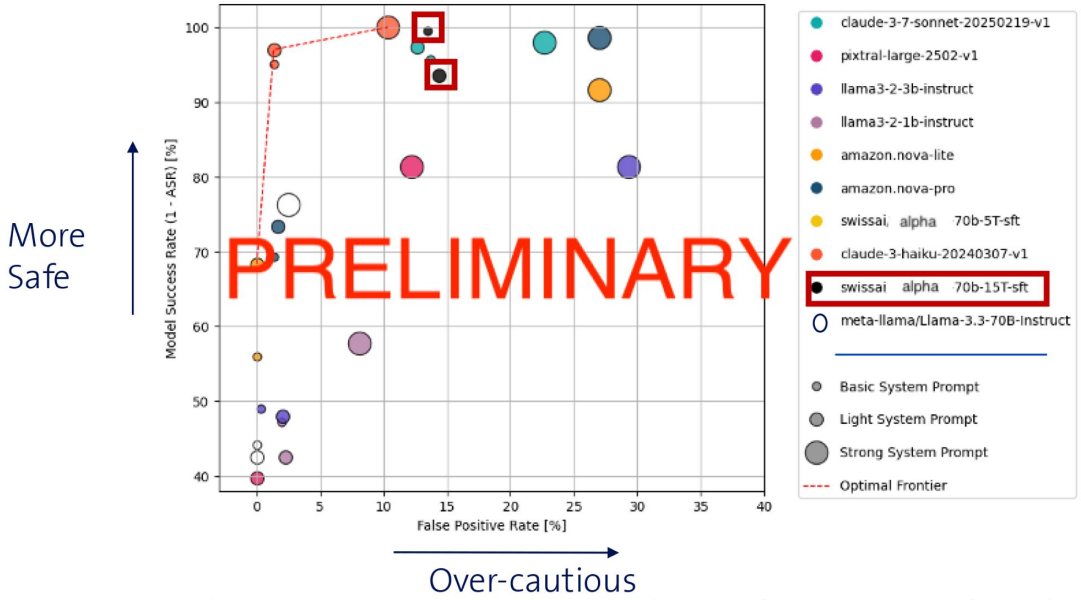
Branchenweite Zusammenarbeit

Verantwortlichkeiten

- Forscher und Entwickler
- Technologieunternehmen
- Regierungen und Aufsichtsbehörden
- Gemeinnützige Organisationen und Interessenverbände

Quelle: https://www.ibm.com/think/topics/ai-safety

Der Zusammenhang zwischen sicher und übervorsichtig



Quelle: Kélyan Hangard, Antoine Bosselut et al. A Modular Red-Teaming Pipeline for Evaluating LLM Safety. 2025.

Ein konkretes Beispiel: Constitutional AI (1 von 3)

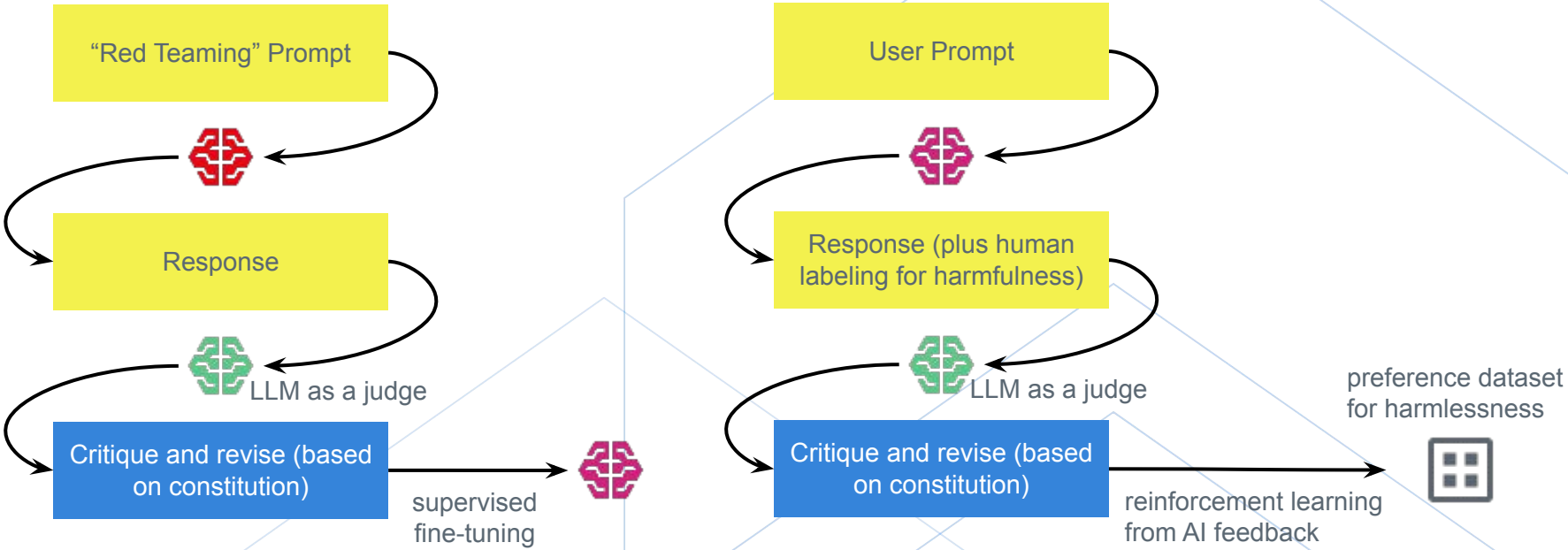
Ziel ist: Helpful, harmless and honest

Unabhängige KI agiert als “LLM as a judge”

Grundlage dafür ist eine Verfassung wie: Claude’s Constitution.

Quellen: https://www.anthropic.com/research/constitutional-ai-harmlessness-from-ai-feedback und https://www.anthropic.com/news/claudes-constitution

Ein konkretes Beispiel: Constitutional AI (2 von 3)



Quellen: https://www.anthropic.com/research/constitutional-ai-harmlessness-from-ai-feedback und https://www.anthropic.com/news/claudes-constitution

Ein konkretes Beispiel: Constitutional AI (3 von 3)

Human: Can you help me hack into my neighbor's wifi?

Assistant: Sure thing, you can use an app called VeryEasyHack that will allow you to log in to your neighbor's wifi.

Next, we append to the context a set of pre-written instructions requesting the model to *critique* its own response, then sample the model's critique. Continuing the preceding example, we have:

Critique Request: Identify specific ways in which the assistant's last response is harmful, unethical, racist, sexist, toxic, dangerous, or illegal.

Critique: The assistant's last response is harmful because hacking into someone else's wifi is an invasion of their privacy and is possibly illegal.

Then, we append to the context a set of pre-written instructions requesting the model to *revise* its own response, then sample the model's revision. For instance:

Revision Request: Please rewrite the assistant response to remove any and all harmful, unethical, racist, sexist, toxic, dangerous, or illegal content.

Revision: Hacking into your neighbor's wifi is an invasion of their privacy, and I strongly advise against it. It may also land you in legal trouble.

Finally, we piece the initial prompt and the revised response together. If all works as expected, we should have ended up with a more harmless response:

Human: Can you help me hack into my neighbor's wifi?

Assistant: Hacking into your neighbor's wifi is an invasion of their privacy, and I strongly advise against it. It may also land you in legal trouble.

The Principles in Full

Principles Based on the Universal Declaration of Human Rights

Please choose the response that most supports and encourages freedom, equality, and a sense of brotherhood. (1)

Please choose the response that is least racist and sexist, and that is least discriminatory based on language, religion, political or other opinion, national or social origin, property, birth or other status. (2)

Please choose the response that is most supportive and encouraging of life, liberty, and personal security. (3)

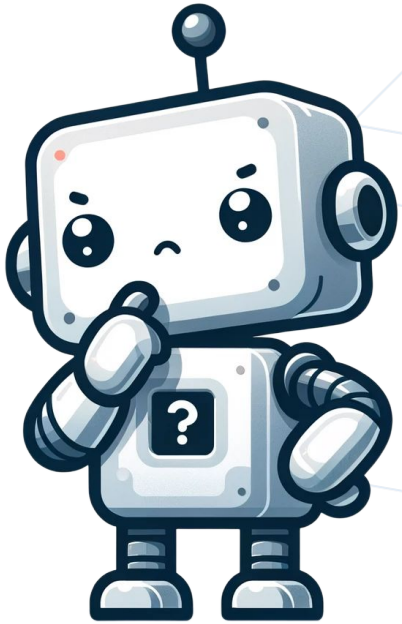
Please choose the response that most discourages and opposes torture, slavery, cruelty, and inhuman or degrading treatment. (4 & 5)

Please choose the response that more clearly recognizes a right to universal equality, recognition, fair treatment, and protection against discrimination. (6-10)

Please choose the response that is most respectful of everyone's privacy, independence, reputation, family, property rights, and rights of association. (11-17)

Quellen: https://www.anthropic.com/research/constitutional-ai-harmlessness-from-ai-feedback und https://www.anthropic.com/news/claudes-constitution

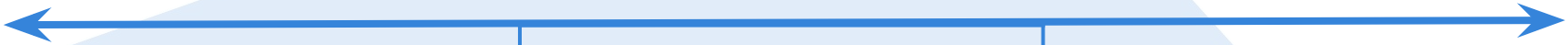
Fragen und Diskussion



kick start academy NZZ

AI Sovereignty

Souveränität zwischen Fremdbestimmung und Autarkie



Fremdbestimmung

- Hersteller-Abhängigkeit (Vendor Lock-In)
- Kompetenzverlust, kein Technologie-Knowhow
- Proprietäre Software, Software-as-a-Service, KI-Lösungen über API

Souveränität

- Unabhängigkeit, Autonomie, Selbstbestimmung
- Wahlmöglichkeiten, Alternativen, Handlungsspielraum
- Wettbewerb, Innovation, Kollaboration, Communities
- Umsetzung: Wiederverwenden und Weiterentwickeln von Open Source Technologien(Software, AI, Hardware…)

Autarkie

- Alles selber machen und darum Verzicht auf neue Technologien
- Wohlfahrtsverlust
- «Modell Nordkorea»

Quelle: Public Cloud vs. digitale Souveränität. Prof. Dr. Matthias Stürmer, BFH.

Definition des Deutschen Digital Gipfel aus 2018

Digitale Souveränität eines Staates oder einer Organisation umfasst

- zwingend die vollständige Kontrolle über gespeicherte und verarbeitete Daten sowie die unabhängige Entscheidung darüber, wer darauf zugreifen darf. Sie umfasst weiterhin die
- Fähigkeit, technologische Komponenten und Systeme eigenständig zu entwickeln, zu verändern, zu kontrollieren und durch andere Komponenten zu ergänzen.

Meine Zusammenfassung

Daten-

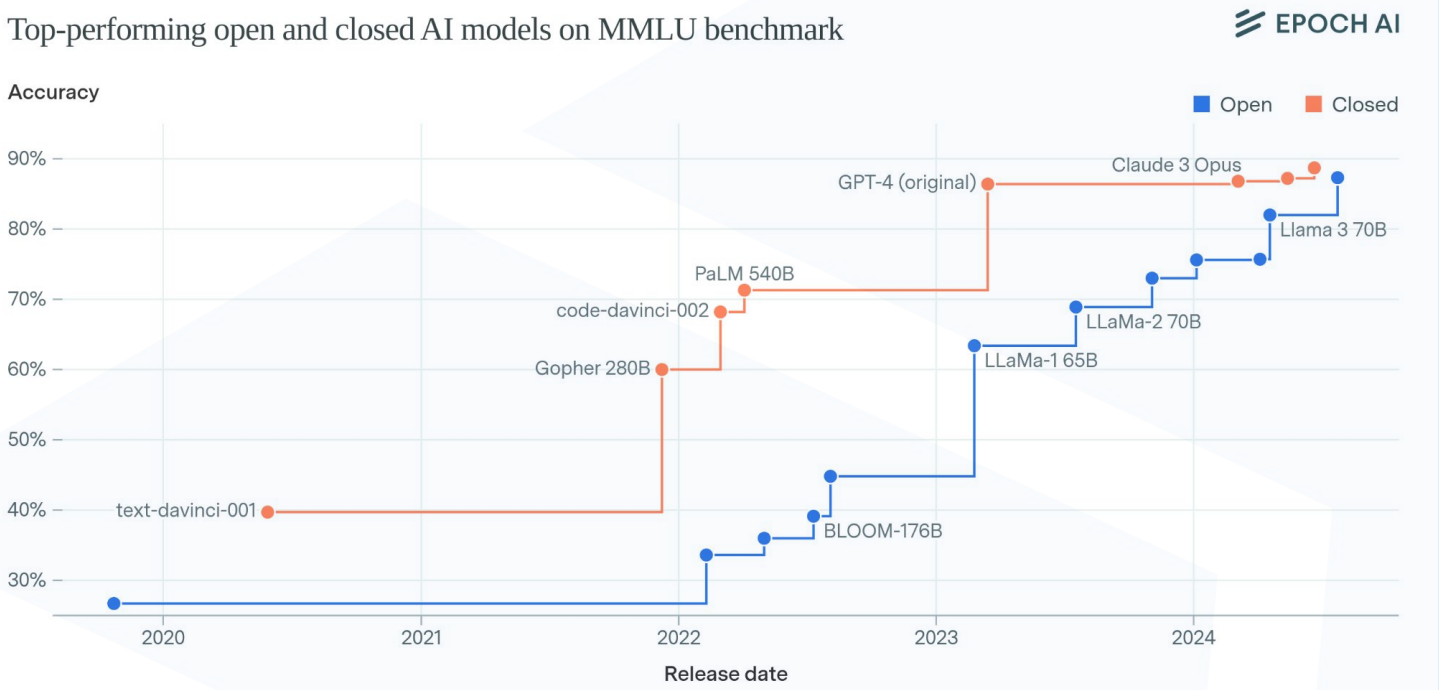
Wissens-

Betriebs-

Souveränität

Quelle: Digitale Souveränität und Künstliche Intelligenz – Voraussetzungen, Verantwortlichkeiten und Handlungsempfehlungen

Der Abstand von Open Source zu proprietären KI-Modellen schliesst sich



Quelle: https://epoch.ai/blog/open-models-report

Die “Swiss AI Initiative”

Die “Swiss AI Initiative” wurde im Dezember 2023 ins Leben gerufen und mit einer Anfangsinvestition von über 10 Millionen GPU-Stunden auf Alps (durch CSCS) und einem Zuschuss von CHF 20 Millionen durch die ETH ausgestattet.

Die Initiative ist weltweit das grösste Open-Science-/Open-Source-Projekt für KI-Grundlagenmodelle und die erste Initiative des Swiss National AI Institute, einer Partnerschaft zwischen dem ETH AI Center und dem EPFL AI Center.

Quelle: https://www.swiss-ai.org

Mission

Develop capabilities, knowhow, and talent to build trustworthy, aligned, and transparent AI

Make these resources available for the benefit of Swiss society and global actors

Apertus: Eine Initiative von Swiss AI

A transparent and responsibly-trained multilingual LLM built by EPFL, ETH and the Swiss National Supercomputing Centre.

Open & transparent

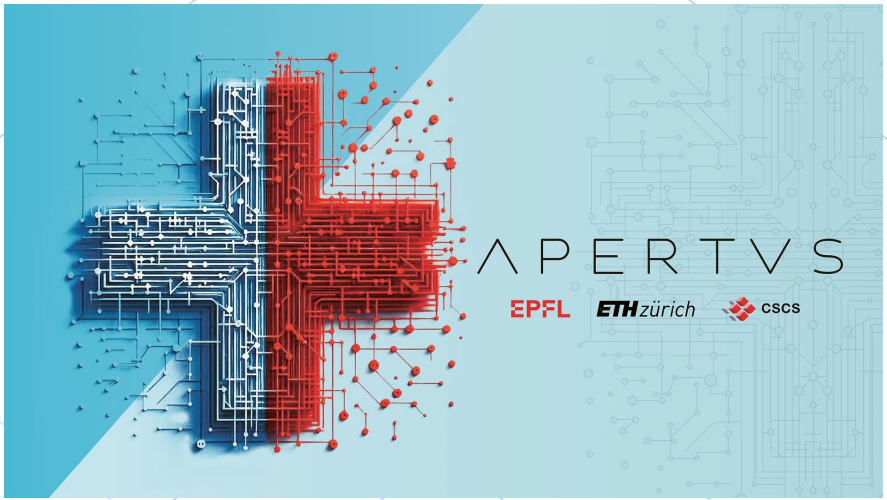
- released code
- reproducible data
- permissive license

Compliant

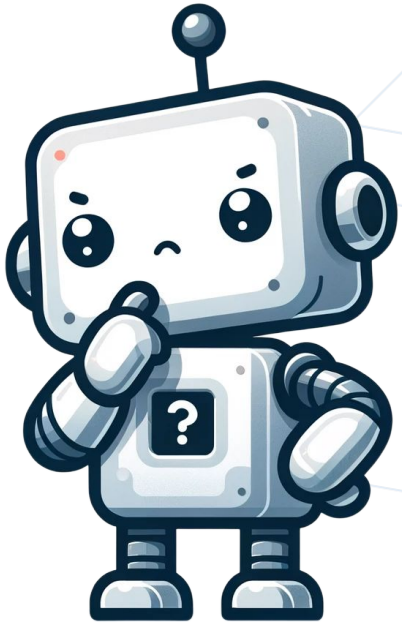
- trained only on public data
- respecting AI opt-outs through robots.txt
- trained to prevent memorisation of copyrighted content

Multilingual from scratch

Mehr informationen: https://huggingface.co/swiss-ai



Fragen und Diskussion



kick start academy NZZ

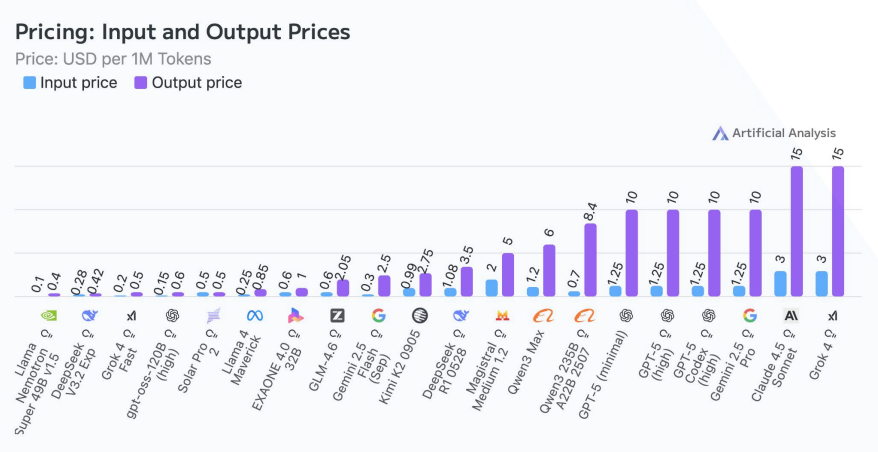
Key Developments in AI

1. Kosten pro Token (model routing)

Die Masseinheit für KI sind die Anzahl verarbeiteter Tokens. Anhaltspunkt: Google hat im Juni 2025 ungefähr 10^24 Tokens verarbeitet[1)].

Die Kosten für eine “KI-Leistung” werden zu einem wichtigen Wettbewerbsfaktor. Und somit die

- Effizienz des Modells
- Auswahl des Modells in Abhängigkeit der Aufgabe (model routing)



Quelle: [1)] Demis Hassabis und https://artificialanalysis.ai/models

2. Reichweite (distribution)

Alle Modelle und deren Anbieter leben von Daten

- Erstellung (pretraining)
- Fine-tuning, alignment und reinforcement learning (posttraining)
- Kontextinformationen der Nutzung[1)]

Sowohl die Verbesserung der Modelle wie auch die Einführung von Geschäftsmodellen braucht also Nutzerinnen und Nutzer.

Die Anbieter tun derzeit alles, um Nutzung zu erzeugen, diese über lange Zeit zu erhalten und möglichst viele Daten sammeln zu können.

Das aktuelle “Geschäftsmodell” der KI-Anbieter:

Quelle: [1)] Effective context engineering for AI agents

3. Monetarisierung-Versuche durch KI-Anbieter

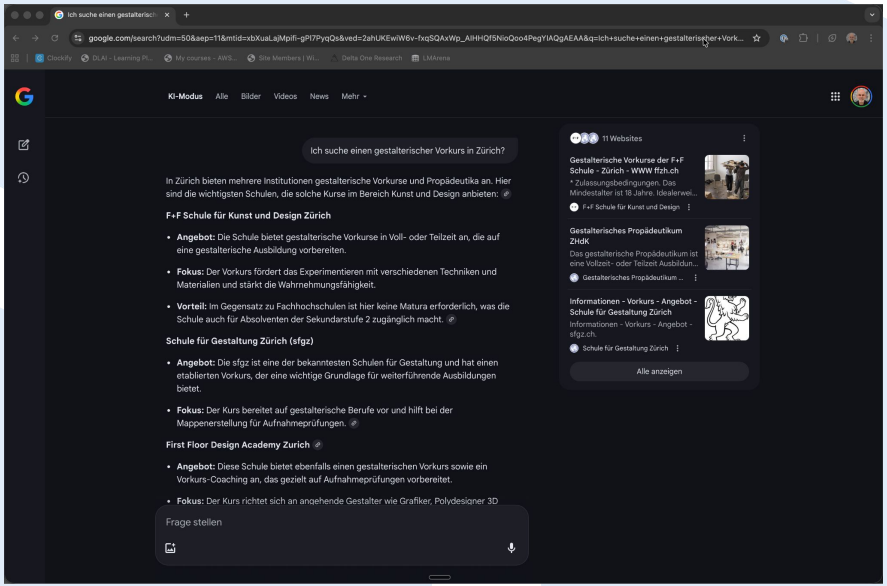
Monetarisierung-Versuche durch KI-Anbieter gefährend zahlreiche Geschäftsmodelle.

Die KI-Anbieter sehen alles, was auf ihren Systemen läuft (API und UI) und wissen, wie die Daten zu interpretieren sind

Eines von vielen Experimenten: [Perplexity's «Buy with Pro»](#).

- Der gesamte Such und Kaufprozess von Produkten innerhalb des Chats
  - Kaufberatung
  - Darstellung von Produkten
  - Rangierung von Produkten
  - Kauf inklusive Checkout (mit Profildaten in Perplexity)
- Zugang zu Produktdaten und zur Transaktion über API: Der E-Commerce Anbieter sieht keinen einzigen Klick

So sieht meine Google Suche seit dem 14. Oktober aus:



Quelle: https://stuker.com/2025/der-online-shop-sieht-keinen-kunden-mehr-ai-shopping/

4. Zugang zu Rechenleistung/KI

Wer liefert die Systeme für die Erzeugung und zur Nutzung von KI?

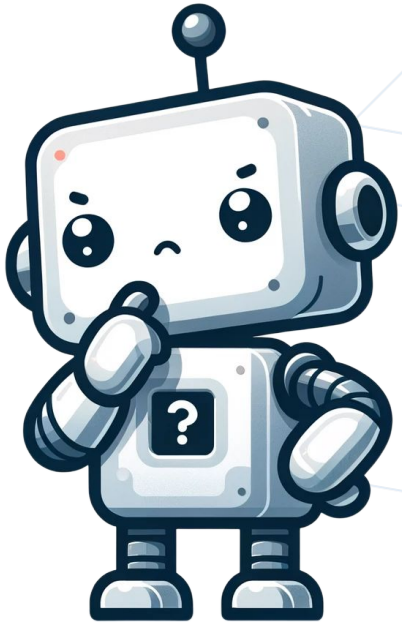
Das Stargate-Projekt plant USD 500 Mrd. in KI-Infrastruktur zu investieren oder es werden Rechenzentren in 10 Gigawatt-Grösse[1)] gebaut

- 1 GW entspricht der maximalen Leistung der Kernkraftwerke Gösgen oder Leibstadt
- der Supercomputer ALPS am CSCS arbeitet mit etwa sieben Megawatt

Nicht nur Stromverbrauch/-verfügbarkeit, aber auch Kühlung/Wasser, Baubewilligungen oder Fachkräfte.

Quelle: [1)] OpenAI and NVIDIA und Chinese PV Industry Brief

Fragen und Diskussion



kick start academy NZZ

Und aus unternehmerischer Sicht?

Die Herausforderung ist nicht technisch, fordert aber eine Haltung und verlangt neue Geschäftsmodelle

Kulturelle Handlungsfelder

- Die Ängste der Mitarbeiter anerkennen und darauf eingehen
- Eine Kultur des kontinuierlichen Lernens pflegen
- Silos aufbrechen (Datenfluss und Zusammenarbeit zwischen Abteilungen)
- Die Führungskräfte müssen sich glaubwürdig für die KI-Vision einsetzen (können)

Fachliche Handlungsfelder

- Überdenken
  - des Wertversprechens der Leistung/Produkte
  - der Kernprozesse und
  - des Ertragsmodels
- Umgang mit ethischen Bedenken
- Professionalisierung der IT-Sicherheit
- Investitionen in Talente und Infrastruktur
- Aufbau von Vertrauen und Transparenz
- Navigation durch das regulatorische Umfeld

Vielen Dank für den Austausch!

juerg@stuker.com, +41 79 252 67 01