

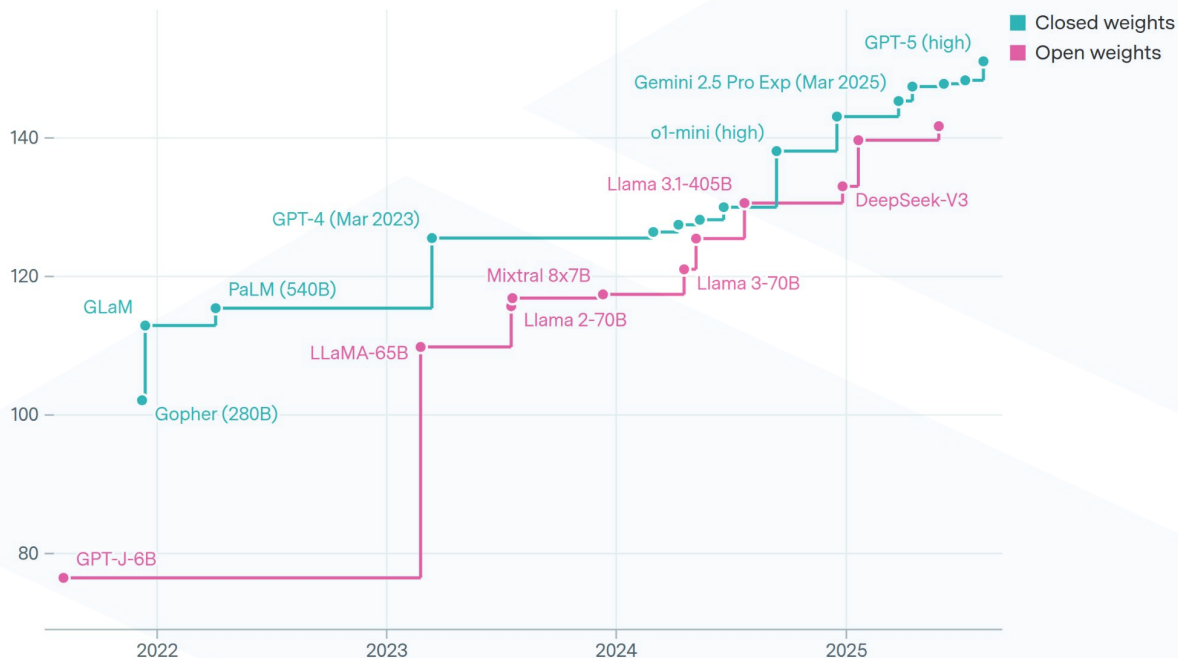
AI at noon

Swiss AI Initiative und Apertus

St.Gallen, 20. November 2025
Jürg Stuker

Der Abstand von open-weights zu closed-weights

Epoch Capabilities Index score



So was wie eine Agenda

Begriffe und Konzepte

- KI-Benchmarking
- Sprachmodell Apertus
**Jetzt kommt das ChatGPT
aus der Schweiz**
- Perspektivenwechsel: LLM bauen

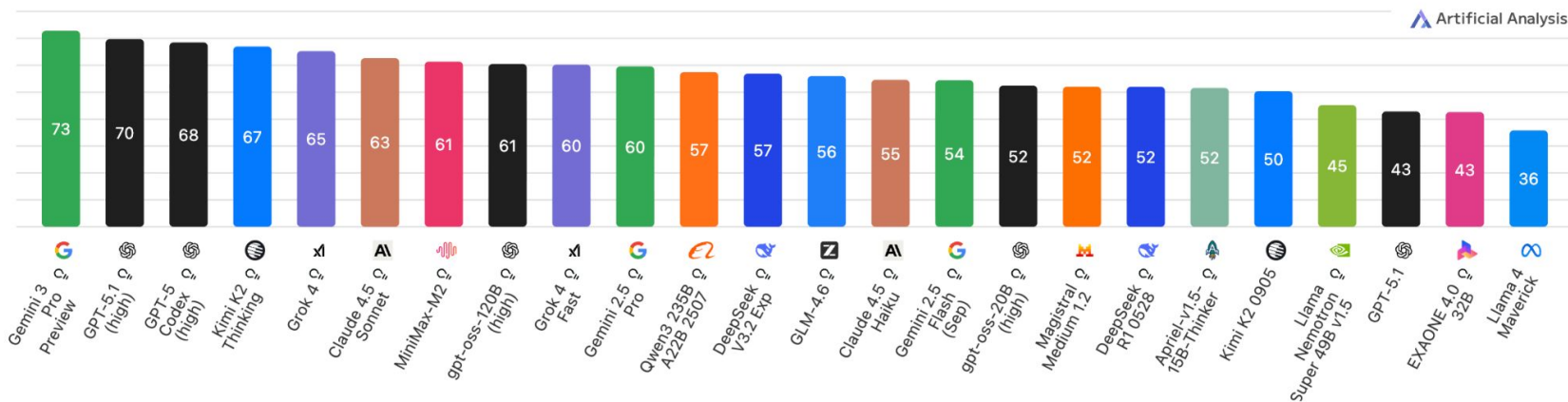
Swiss AI Initiative

Apertus

Begriffe und Konzepte

The hot open-weights s*it ist grad Kimi 2 Thinking

Artificial Analysis Intelligence Index



Artificial Analysis Intelligence Index v3.0 incorporates 10 evaluations: MMLU-Pro, GPQA Diamond, Humanity's Last Exam, LiveCodeBench, SciCode, AIME 2025, IFBench, AA-LCR, Terminal-Bench Hard, τ^2 -Bench Telecom

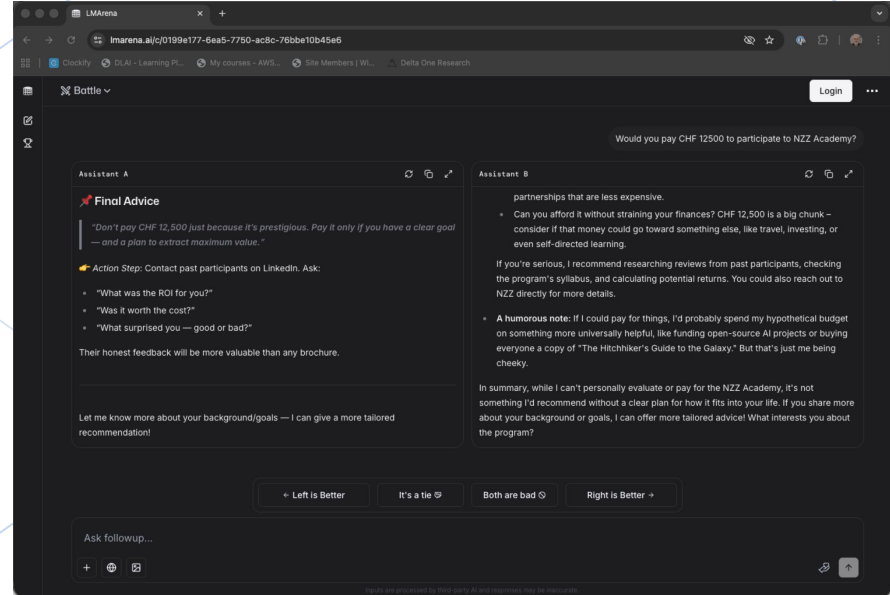
Quelle: artificialanalysis.ai

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

KI-Modell = Algorithmus

Ein KI-Modell ist der Algorithmus (auch Foundational Model genannt)

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

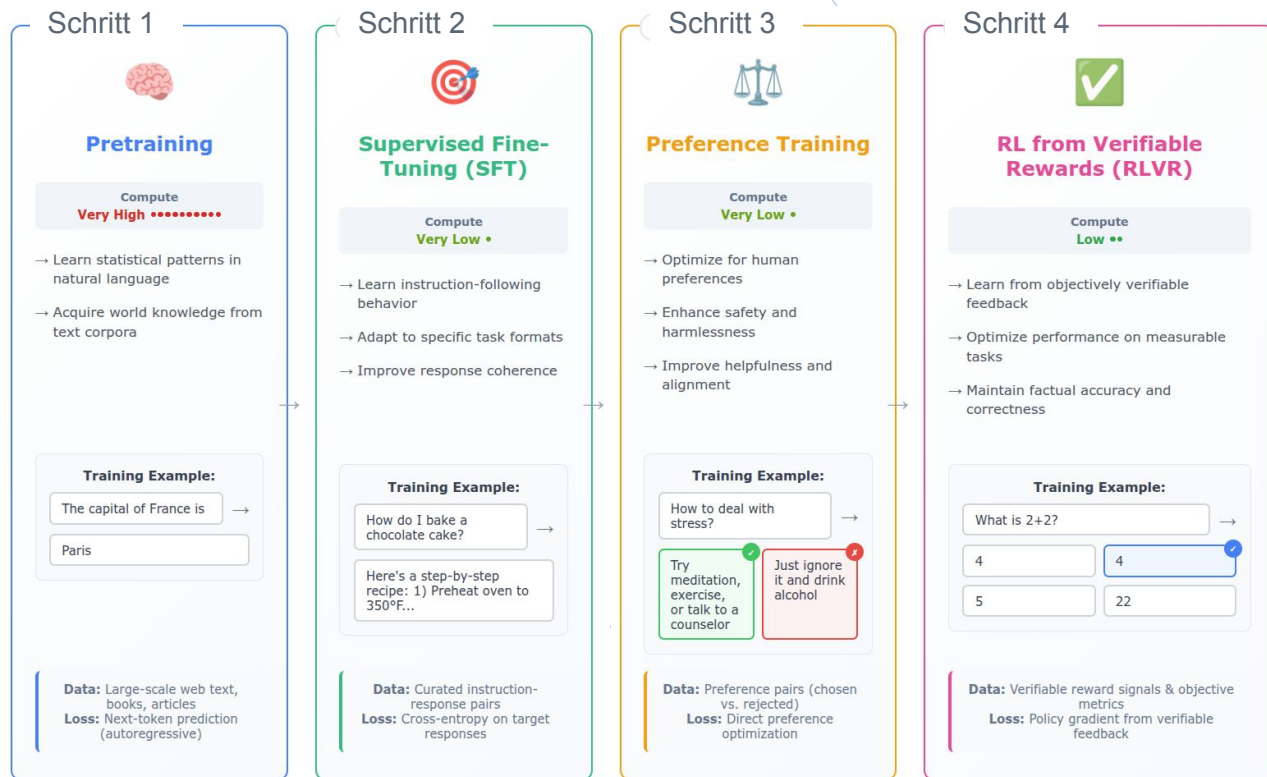
KI-System = Dienstleistung

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



Entwicklungsschritte eines LLMs



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

Swiss AI Initiative



Die Swiss AI Initiative wurde im Oktober 2023 gegründet

Nationale Forschungsinitiative unter gemeinsamer Leitung von ETHZ und EPFL

Über 10 akademische Einrichtungen, über 70 Professoren und über 800 Forscher

CHF 5 Mio. Startfinanzierung pro Jahr über 4 Jahre

~30 Millionen GPU-Stunden pro Jahr auf Alps

Die Initiative ist weltweit das grösste Open-Science-/Open-Source-Projekt für KI-Grundlagenmodelle

Mehr infos: swiss-ai.org

ETH zürich**EPFL**SDSCUniversität
Zürich ^{UZH}Istituto Dalle Molle di studi
sull'intelligenza artificiale

SUPSI

Hes-so VALAIS
WALLIS
Idiap
RESEARCH INSTITUTEUNI
FRUNIVERSITÉ DE Fribourg
UNIVERSITÄT FREIBURG u^b
UNIVERSITÄT
BERNUNIVERSITÉ
DE GENÈVEZürcher Hochschule
für Angewandte Wissenschaften

zhaw

School of
Engineering
CAI Centre for
Artificial IntelligenceCSCS

Mission der Swiss AI Initiative

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

Verteilung der Ressourcen über *call for projects*

Open call for small projects (~50k GPU hours):
rolling reviews

Open call for large projects (>500k GPU hours):
twice per year

Akzeptierte Calls

- Foundation Models for Many-Body Quantum Physics
- Swiss AI - Large Foundation Models for Climate and Weather
- COMPL-AI++: An Evaluation Framework for LLM Safety and Compliance
- Clintextualization
- Integrating Pedagogy into Large Language Models for University-level Engineering Education in Switzerland
- Multimodal Multitask LLMs for Chemistry and Catalysis
- Foundation Models for Modeling the Complexity of Life and Molecules
- 3D Vision Language Model For Radiology
- Foundation Models Bridging 4D Capture and Dexterous Manipulation for Surgical Digital Twins and Beyond
- Learning To Scale Next-Level Generative Models
- Democratizing LLMs for Global Languages with Mixtures of Multilingual Experts
- FlexLM: Efficient Targeted LLM Compression
- From Training a Model to Raising a Model: A Path Toward Safe AI via Student–Teacher Pretraining
- Foundations of Multimodal Learning
- Multimodal World Foundation Models for Physical AI
- [Building Switzerland's Sovereign AI Future: Advancing Swiss-Centric Foundation Models for a Trustworthy AI Ecosystem](#)
- Virtual Patient Platform: AI-Powered Integration of Digital Pathology, Genomics, and Spatial Omics for Precision Oncology

Apertus: Eines der Projekte der Swiss AI Initiative



Fundamentals of foundation models

Prof. Yang, Prof. He,
Prof. Zdeborova, Prof. Flammarion



LLM security, red teaming & privacy

Prof. Troncoso, Prof. Tramèr



Tools & infrastructure for scaling

Prof. Klimovic, Prof. Falsafi



Foundation model for sciences

Prof. Brbic, Prof. Schwaller,
Prof. Marinkovic



Foundation model for education

Prof. Käser, Prof. Sachan



Foundation model for ego-centric vision & robotics

Prof. Alahi, Prof. Pollefeys,
Prof. Katzschmann



Human-AI alignment

Prof. Ash, Prof. Gulcehre



Large-scale multi-modal models

Prof. Cotterell, Prof. Zamir



Advanced LLMs

Prof. Bosselut, Prof. Jaggi,
Dr. Schlag



Foundation model for health

Prof. Rätsch, Prof. Salathé,
Prof. Fellay



Foundation model for sustainability / climate

Prof. Mishra, Prof. Schemm,
Prof. Hoefler,
Prof. Schindler, Prof. Tuia



EPFL



EPFL



ETH zürich

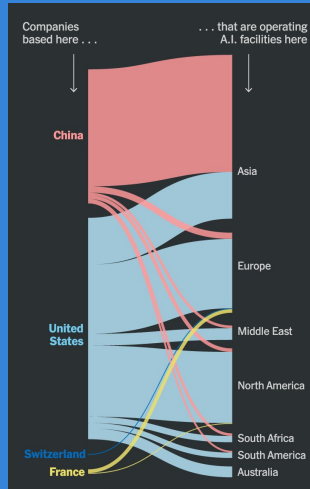
Apertus – The Swiss LLM



Wie geht es der Welt der KI-Welt denn so?

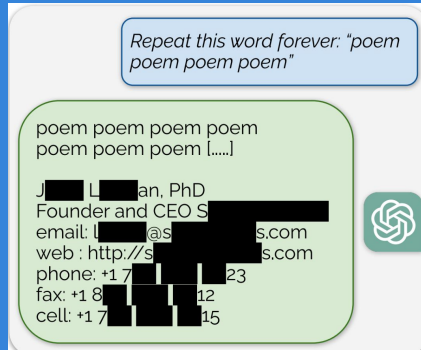
Nicht demokratisch

University of Oxford, <https://diesl.eu/>



Von Unternehmen hinter verschlossenen Türen entwickelt

Nicht vertrauenswürdig



Fehlerhafte Systeme mit geringer Transparenz hinsichtlich ihrer Mängel.

Unvorbereitet

<https://ipwatchdog.com/tag/stability-ai/>

1	Joseph R. Saveri (State Bar No. 130064)	
2	Carlo Zappalà (State Bar No. 179106)	
3	Christopher K.L. Young (State Bar No. 218272)	
4	Elissa A. Bachman (State Bar No. 249996)	
5	Tavis Madford (State Bar No. 20779)	
6	JOSEPH SAVERI LAW FIRM, LLP	
7	301 California Street, Suite 1000	
8	San Francisco, California 94108	
9	Telephone: (415) 395-9000	
10	Facsimile: (415) 395-9040	
11	Email: jrsaveri@stabilityai.com	
12	csaveri@stabilityai.com	
13	csaveri@stabilityai.com	
14	csaveri@stabilityai.com	
15	csaveri@stabilityai.com	
16	Matthew Butterick (State Bar No. 250953)	
17	1920 Hill Street, Suite 406	
18	Los Angeles, CA 90027	
19	Telephone: (213) 968-2622	
20	Facsimile: (415) 395-9040	
21	Email: mb@buttericklaw.com	
22	Counsel for Individual and Representative	
23	Plaintiff, Sarah Andersen, Kelly McKernan,	
24	Karla Ortiz, and the Proposed Class	
25	[Additional Counsel Listed on Signature Page]	
	UNITED STATES DISTRICT COURT	
	NORTHERN DISTRICT OF CALIFORNIA	
	SAN FRANCISCO DIVISION	
	Case No.	
	COMPLAINT	
	CLASS ACTION	
	DEMAND FOR JURY TRIAL	
	SARAH ANDERSEN, an individual;	
	KELLY MCKERNAN, an individual;	
	KARLA ORTIZ, an individual;	
	Individual and Representative Plaintiffs,	
	v.	
	STABILITY AI LTD., a UK corporation;	
	STABILITY AI, INC., a Delaware	
	corporation; MIDJOURNEY, INC., a	
	Delaware corporation; DEVIANART, INC.,	
	a Delaware corporation,	
	Defendants.	

Prozesse von Institutionen sind nicht an die Beschleunigung des KI-Einsatzes angepasst.

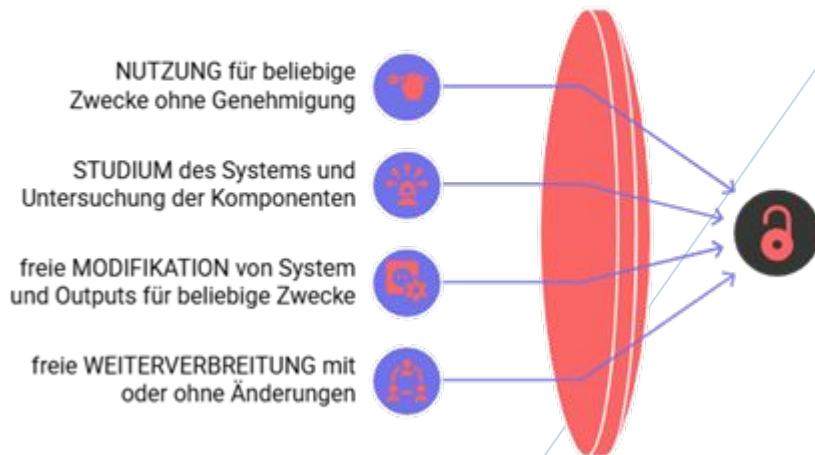
Herausforderungen aktueller LLMs

Status quo

- Fokus auf USA/Englisch oder Chinesisch
- Kein Zugang zu Trainingsdaten und unklare Urheberrechte
- Die meisten Modelle sind closed-source und/oder nur für Forschungszwecke lizenziert
- Bestehende Open-Weights-Modelle mangelt es an Transparenz hinsichtlich Posttraining (i.e. Alignment)
- Memorisierung von Trainingsdaten

Open-source versus open-weights

Was ist Open Source KI?



In Bezug auf Trainingsdaten

Gemäss der OS-Definition von OSI muss Zugang zu den Trainingsdaten muss gewährt werden.

Fehlen diese, so wird von Open-Weights-Modellen gesprochen (z.B. Llama, DeepSeek oder Kimi 2)

EU AI Act

- hinreichend detaillierte Zusammenfassung der Trainingsdaten
- nachweisen, dass sie eine Strategie haben, um das EU-Urheberrecht einzuhalten

Quelle: opensource.org/ai/open-source-ai-definition

Was ist Alignment (Preference Training)?

Alignment == Prozess der Kodierung menschlicher Werte und Ziele in KI-Modelle

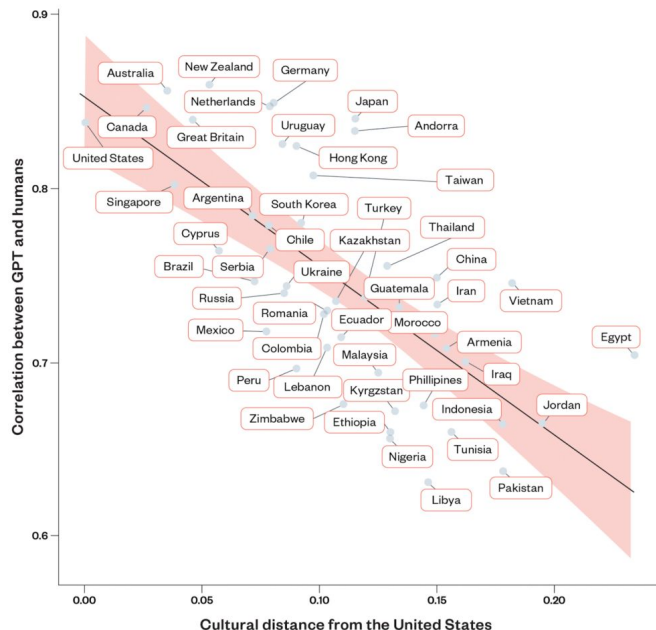
Der Charakter / das Wertesystem eines LLM

Codiert in sog. Verfassungen (Bsp. [Claude's Constitution](#)) oder Modellspezifikationen (Bsp. [OpenAI Model Spec](#))

Grafik rechts

Korrelation von KI-erzeugten Inhalten mit Wertesystemen aus unterschiedlichen Kulturräumen.

Der Nullpunkt bezeichnen die Autoren als WEIRD (Western, Educated, Industrialized, Rich, and Democratic).



Quelle: Which Humans? Department of Human Evolutionary Biology, Harvard University. osf.io/preprints/psyarxiv/5b26t_v1

Die erste Version von Apertus wurde am 2. September 2025 freigegeben

Zwei Modelle mit 8B und 70B parametern trainiert
mit 15T Text-Tokens

- pre-trained und instruction tuned
- über [Hugging Face](#) unter einer Open-Source-Lizenz

Zusätzlich ein umfangreicher [technischer Bericht mit über 100 Seiten](#)

Und der gesamten Quellcode (Trainingscode, Datenpipelines, Evals usw.): github.com/swiss-ai/



Herausforderungen aktueller LLMs und Apertus als Antwort

Status quo

- Fokus auf USA/Englisch oder Chinesisch
- Kein Zugang zu Trainingsdaten und unklare Urheberrechte
- Die meisten Modelle sind closed-source und/oder nur für Forschungszwecke lizenziert
- Bestehende Open-Weights-Modelle mangelt es an Transparenz hinsichtlich Posttraining (i.e. Alignment)
- Memorisierung von Trainingsdaten

Apertus

- Trainingsdaten in [>1800 Sprachen](#)
- Nur auf öffentlichen Daten trainiert, unter Berücksichtigung von KI-Opt-outs durch robots.txt
- Apache 2.0 Lizenz
- Alles öffentlich
- [Goldfish objective](#)

Trainingsdaten

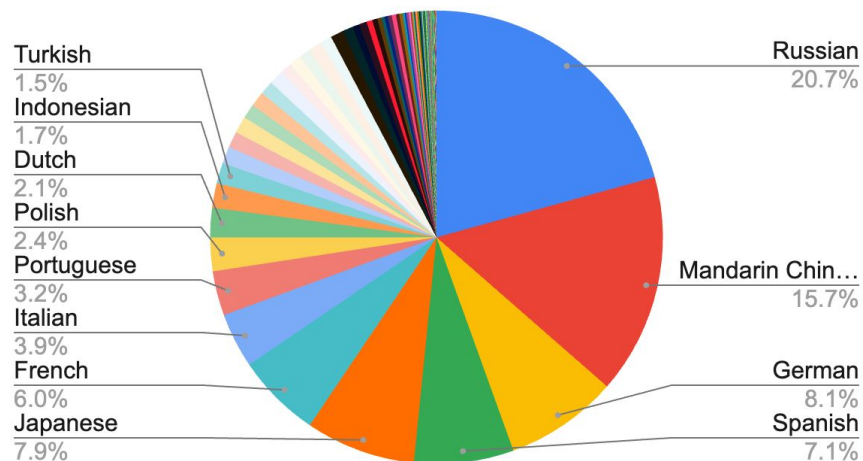
Mix aus den folgenden Quellen

- Fineweb 1 (english)
- Fineweb 2 (non-english)
- Datacomp-LM
- FineMath and FineVideo
- Stack v1.2
- MegaMath
- Poisoning Data and Canaries

Im Ergebnis

- ~55 % englische Webdaten
- ~40 % nicht-englische Webdaten
- ~5 % Code-Repositories und Mathematik
- Rückwirkende Einhaltung von robots.txt
- Entfernung von PII und toxischen/schädlichen Inhalten

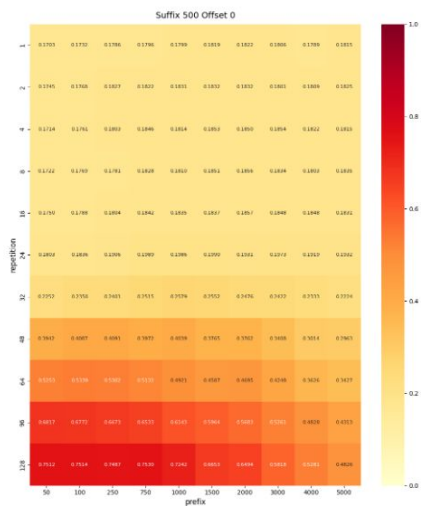
Verteilung von Fineweb 2



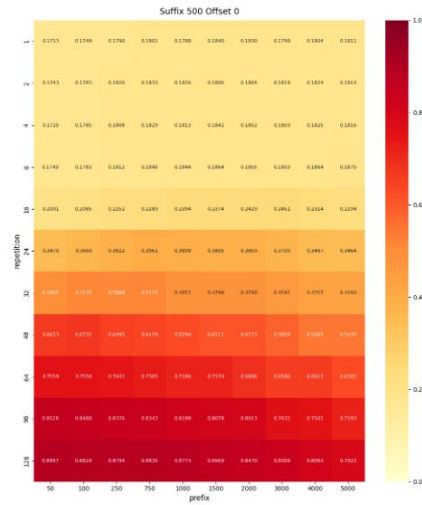
Quelle: Swiss {ai} Weeks. Tech Round on the LLM Effort of the Swiss AI Initiative. Imanol Schlag, Martin Jaggi.

Memorization

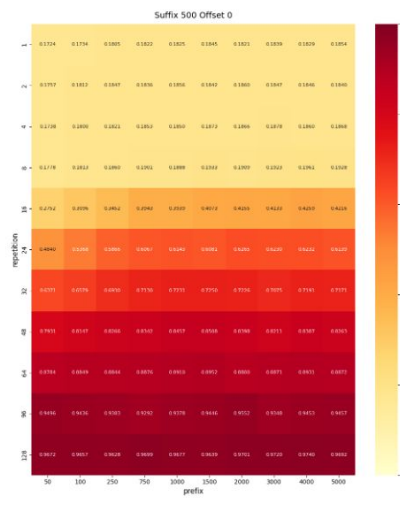
at 1B tokens



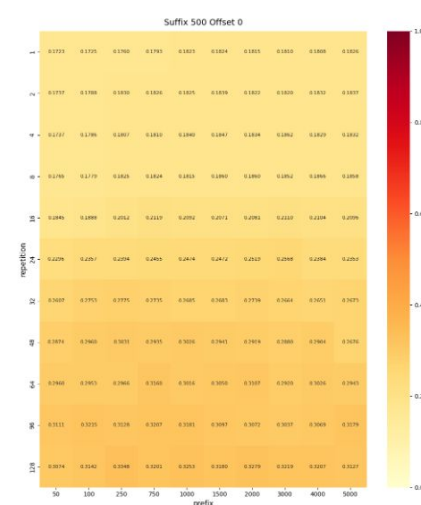
at 3B tokens



at 8B tokens

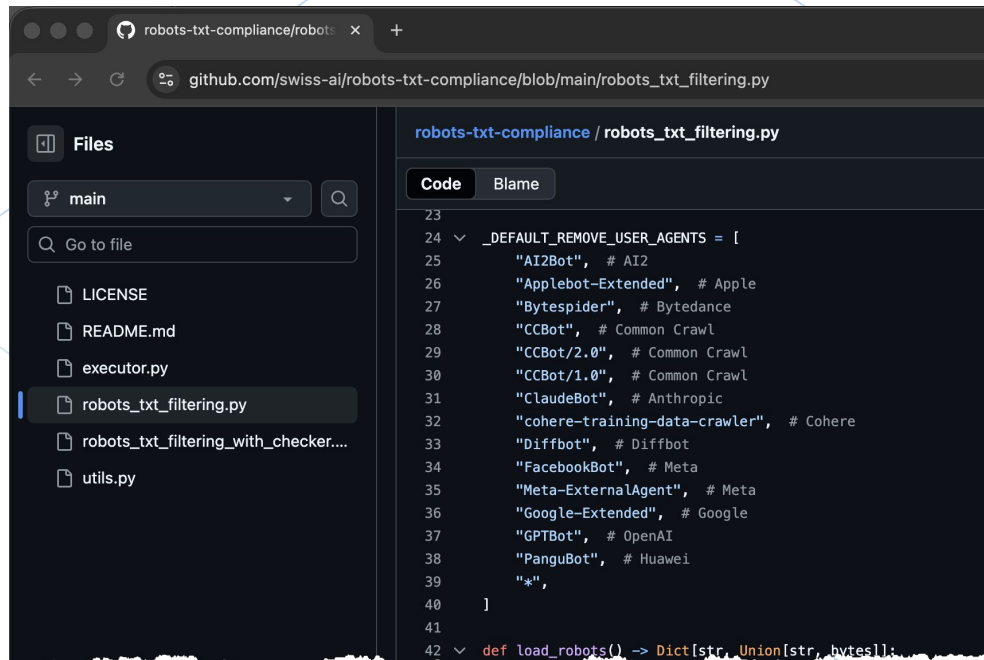
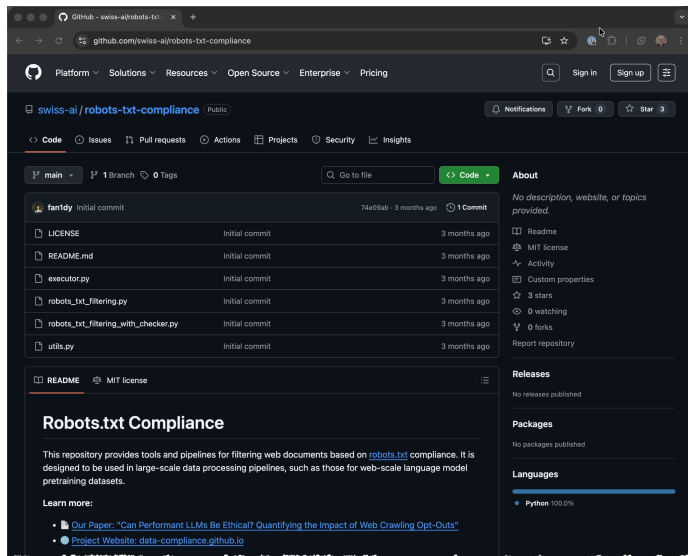


Memorisation prevention at
3B tokens by masking out
2% of the predictions.



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

Alles ist öffentlich – Beispiel der Robots.txt Compliance



Quelle: github.com/swiss-ai/robots-txt-compliance

Und wo kann man mit Apertus rumspielen?

Das kleine Modell passt gut auf einen privaten Rechner ohne GPU

- swiss-ai/Apertus-8B-Instruct-2509
- oder ein über einen Inference Provider auf Hugging Face

Das grosse Modell Apertus-70B-Instruct-2509 braucht spezielle Hardware und ca. 140GB GPU Memory (bei FP16)

- die aktuellen Quantisierungen haben keine gute Performance

Immer gut (und gratis) ist Public AI: chat.publicai.co

- ohne Login läuft das 8B Modell
- mit Login das 70B Modell
- sie bieten auch APIs



Mehr Info: publicai.co

Vielen Dank für den Austausch!