

Herzlich Willkommen

# Digital souveräne KI in der Schweiz: Was ist heute möglich und sinnvoll?



ETH AI CENTER



# Das heutige Programm

Souveräne KI-Plattformen auf Kubernetes  
ipt

Apertus und der Schweizer Weg zu Open Source LLMs  
EPF AI Center / Swiss National AI Institute

**Kaffeepause & Austausch**

Souveräne KI im BIT: Private LLMs als Service –  
sicher, nachvollziehbar, betreibbar  
BIT (Bundesamt für Informatik und Telekommunikation)

Die Zukunft der Souveränen KI am ETH AI Center  
ETH AI Center

Battle: provokative Thesen und überspitzte Aussagen –  
was ist eure Meinung?

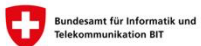
**Apero und Networking**



**Valentin Verschinin**  
Principal Architect, Director



**Oleg Lavrovsky**  
Technical Community Manager, EPFL AI Center / SNAI



**Bernhard von Allmen**  
Product Manager AI



**Gian Derungs**  
AI Engineer



**Daniel Naeff**  
Head of Innovation and Entrepreneurship

# Meine Reise: Warum souveräne KI-Plattformen?

## Ausgangslage

Reguliertes Projekt mit sensiblen Daten und hohen Nachweispflichten.

## Anfangsthese

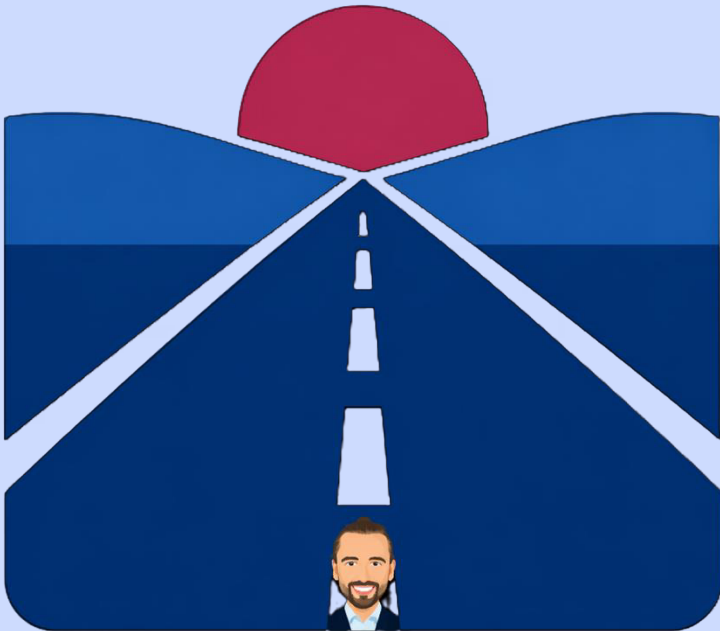
„Compliance wird das blockieren.“

## Tatsächliche Erkenntnis

Das Problem war nicht Compliance, sondern **fehlende technische Kontrolle**.

## Die entscheidende Frage

Wie baut man KI, damit Vertrauen Innovation ermöglicht?

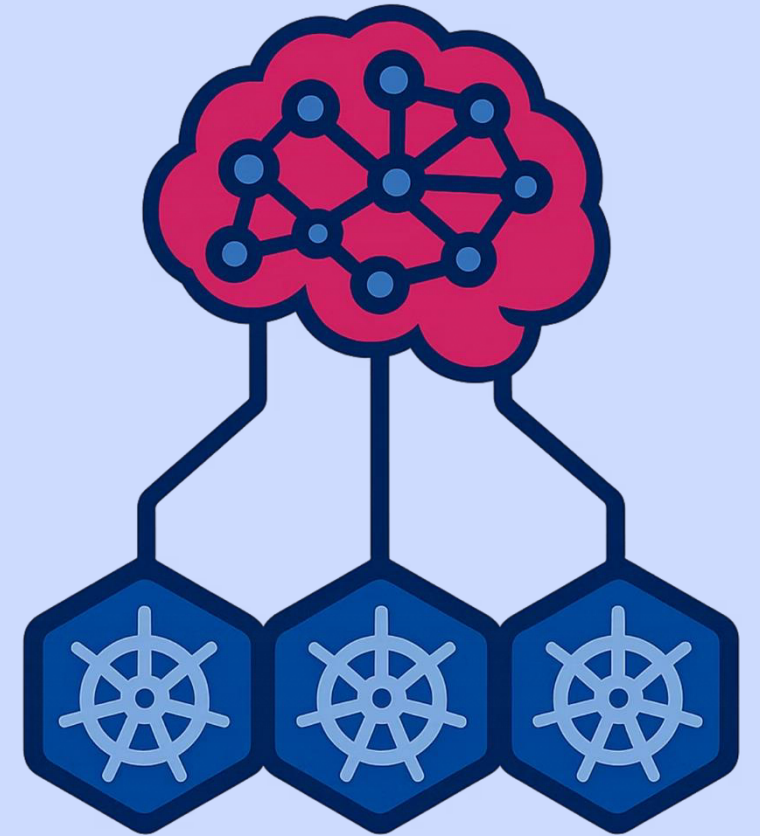




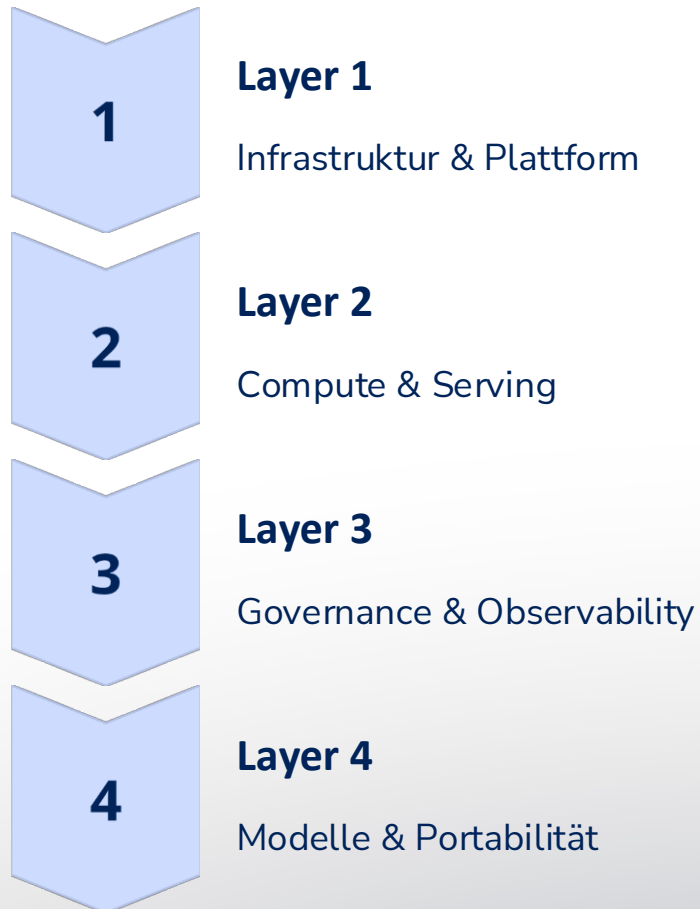
# Souveräne KI- Plattformen auf Kubernetes

Kontrolle, Governance und Portabilität

*Valentin Verschinin · ipt*



# Die Layer einer souveränen KI-Plattform



## Infrastruktur & Plattform

# Wo beginnt echte Souveränität?

Bei der Infrastruktur oder beim Betriebsmodell?



# Kontrolle ist eine Architekturentscheidung

## Kontrollpunkte

Identity, Datenpfad, Policy, Audit, Gateway, Exit

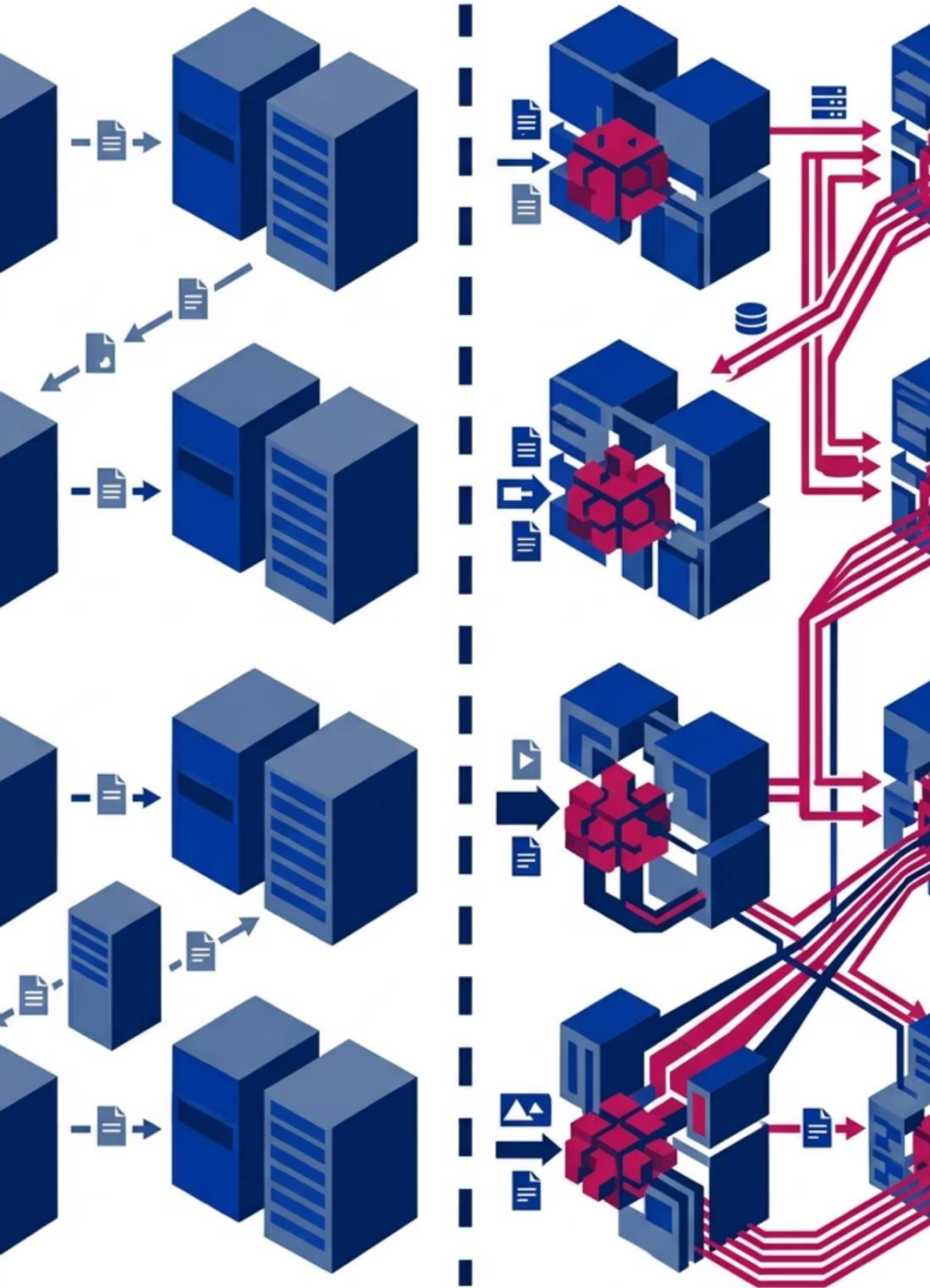
## Hybride Modelle

Können souverän sein, wenn Kontrollpunkte stimmen

## TEEs & Confidential Computing

Kryptografische Isolation für Sicherheit bei Data-in-Use

☐ Nicht Besitz allein schafft Kontrolle, sondern **belastbare Plattformmechanismen** und kryptografische Isolation.

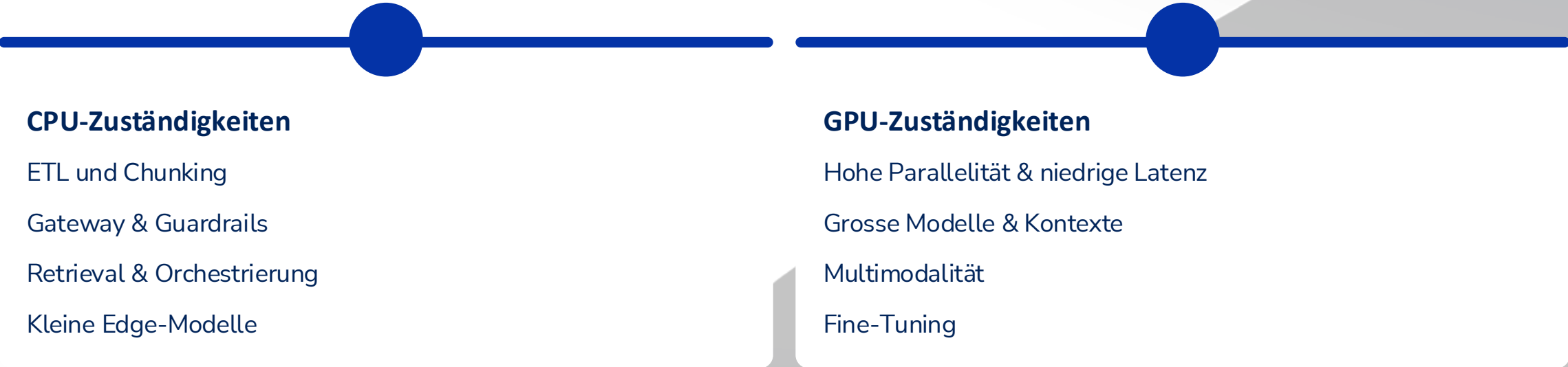


# Erweitern Standard- oder spezialisiert?

Wann reicht Standard-Infrastruktur und wann wird spezialisierte KI-Infrastruktur zwingend?

# Hardware-Realismus

Das dominante Muster, nicht GPU überall.



## CPU-Zuständigkeiten

- ETL und Chunking
- Gateway & Guardrails
- Retrieval & Orchestrierung
- Kleine Edge-Modelle

## GPU-Zuständigkeiten

- Hohe Parallelität & niedrige Latenz
- Grosse Modelle & Kontexte
- Multimodalität
- Fine-Tuning

- Public LLMs: schnelles Prototyping
- Private Inferenz: sensible Daten, SLA, Kostenkontrolle
- CPU-Beschleuniger für Edge & Air-gapped

# Kubernetes ist die Betriebsplattform

## Kern-Stack

### vLLM

Offener Default im Serving

### KServe

Stärkste offene Kubernetes-Schicht

## Spezialisierung

### SGLang

Structured Output, Multimodalität, Caching

### TensorRT-LLM

NVIDIA-maximierte Pfade

### llm-d

Prefill/Decode-Disaggregation, KV-aware Serving, CNCF Sandbox seit März 2026

- TGI: Maintenance-Mode, für Bestandssysteme, kein Neupfad.
- Der Stack bewegt sich noch; boring defaults schlagen Tool-Hype.

## Governance & Observability

# Risiko wird Freigabefähigkeit

Wie wird regulatorisches Risiko zur Freigabefähigkeit?



# Guardrails sind Startbahn, nicht Handschellen

## Gateway als erster Enforcement-Punkt

Policy, Budget, Auth und Routing gehören auf **Plattformebene**, nicht in den App-Code.

Nicht fehlende Regulierung bremst Innovation — sondern **fehlende technische Nachweisbarkeit**.

📄 **Supply-Chain-Vorfall LiteLLM 2026:** Gateways und Dependencies auf IOCs prüfen, kritische Angriffsfläche.

PII Detection

Abuse Monitoring

Prompt-Schutz

# Für produktive KI reicht Monitoring allein nicht mehr

1

## Traces

OpenTelemetry-  
Semantik

2

## Prompt- & Version- Tracking

Jede Interaktion  
nachvollziehbar

3

## Token- & Kostensicht

Ressourcenverbrauch  
transparent

4

## Eval-Datasets & Policy Policy Checks

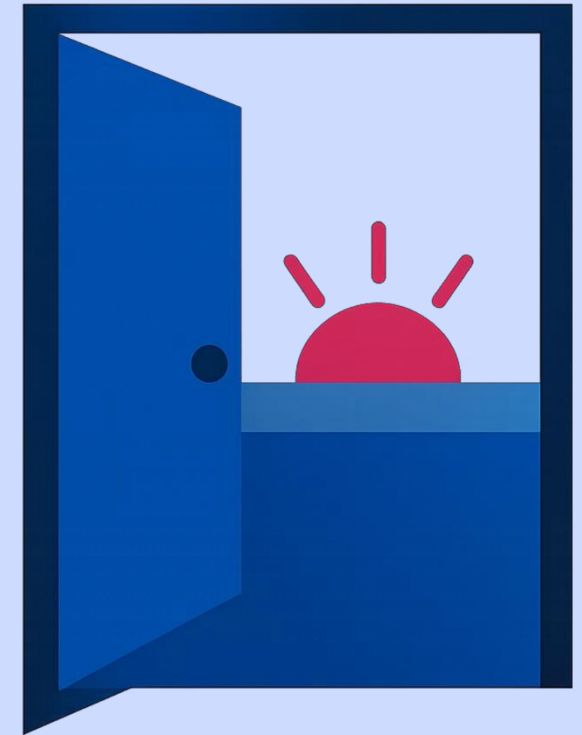
Vergleich und  
Compliance-Nachweis

Self-hosted Observability und Evals gehören zur **Plattform**, nicht nur in den App-Code.

# Modelle & Portabilität

# Freiheit, Verantwortung, Exit

Wann wird Modellwahl zur Frage von Freiheit, Verantwortung und Exit?



# Modelle sind nicht nur Benchmark-Entscheidungen

📄 **Llama 4:** Meta Custom License mit Schwellenwerten. **Gemma:** Googles Prohibited Use Policies. **Open-weight** ≠ **Open Source**.

## Saubere Beschaffungspfade (Apache 2.0 / MIT)

**Qwen**

**Mistral**

**Granite**

**Phi**

**DeepSeek**

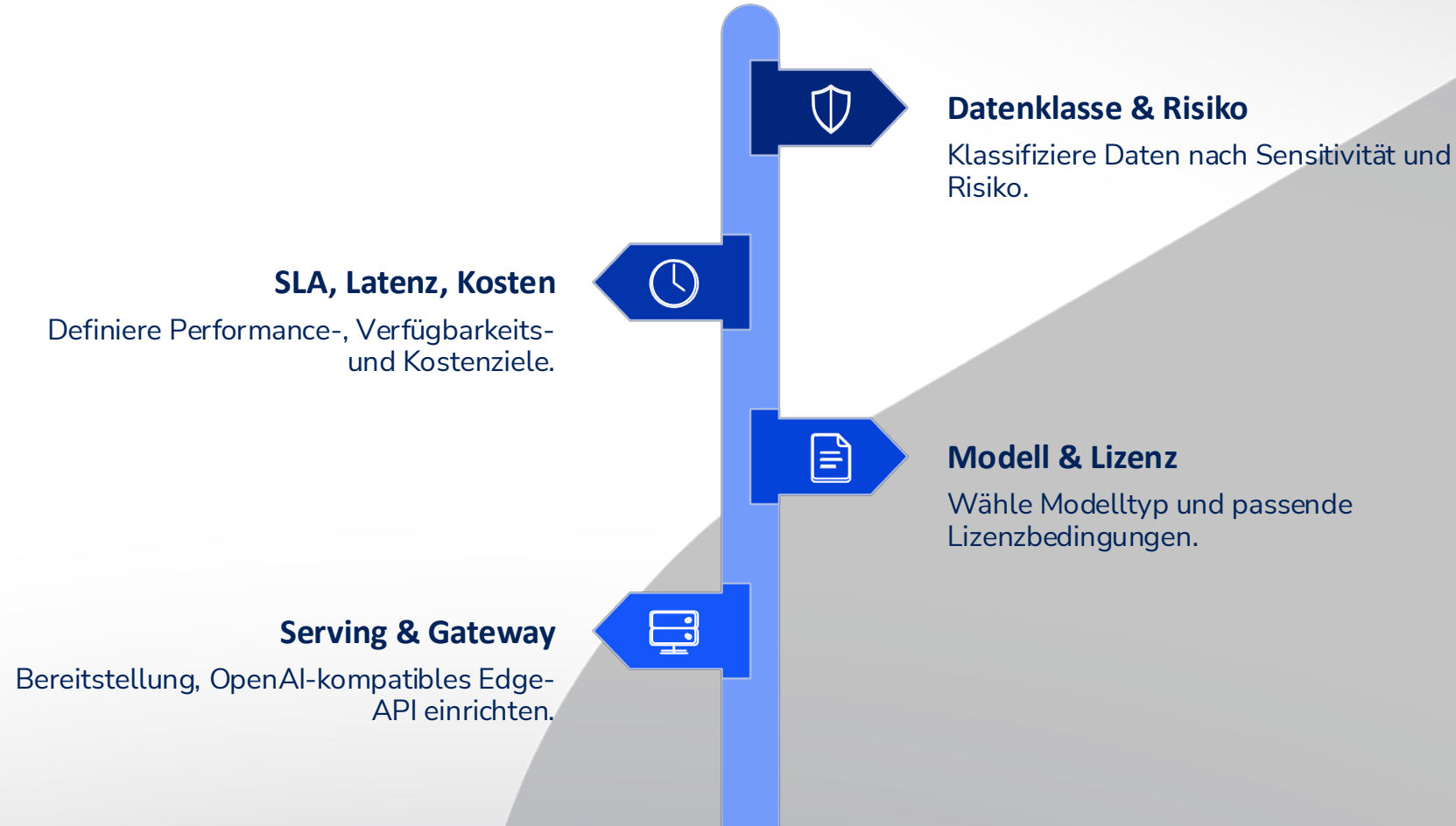
**gpt-oss**

Benchmarks kauft man einmal — Architektur- und Lizenzfehler zahlt man **jahrelang**.

Echte Freiheit entsteht durch:

- Austauschbare Architektur
- Gateway-Abstraktion
- Exit-Optionen

# Der praktikable Blueprint



☐ Souveränität ist kein Ort — sondern ein **Architekturprinzip**. Technische Guardrails sind nicht die Bremse deiner KI-Strategie, sie sind die **Startbahn**.

# Ausblick: Wohin die Reise geht

## Heutiges Verständnis

Souveränität ist die Fähigkeit, Kontrolle bewusst zu gestalten.

## Von Modellen zu Plattformen

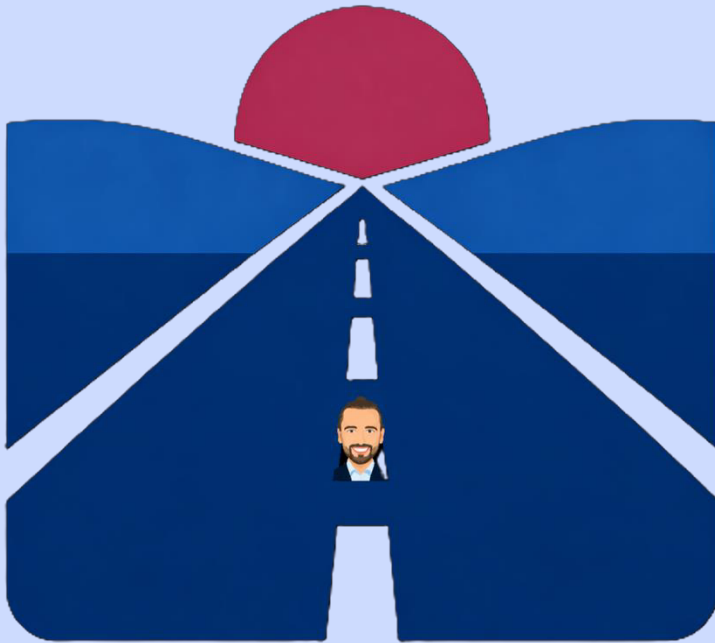
Entscheidend wird nicht das Modell, sondern die **Plattform: Serving, Gateway, Governance und Datenzugriff**.

## Von Experimenten zu belastbarem Betrieb

Reife entsteht, wenn KI **nachvollziehbar, steuerbar und produktionsfähig** wird.

## Von Tool-Hype zu Architekturprinzipien

**Offene Schnittstellen, modulare Layer und Exit-Fähigkeit** zählen mehr als kurzfristige Tool-Vorteile.





 -lichen Dank

-  **Zürich**  
Innovation Process Technology AG  
Löwenstrasse 28  
8001 Zürich  
**+41 41 727 25 25**  
**info@ipt.ch**
-  **Bern**  
Innovation Process Technology AG  
Marktgasse 28  
3011 Bern  
**+41 31 311 12 23**  
**info@ipt.ch**

# APERTVS

EPFL ETH zürich  CSCS

**Digital souveräne KI in der Schweiz:  
Was ist heute möglich und sinnvoll?**

Oleg Lavrovsky  
Community Manager

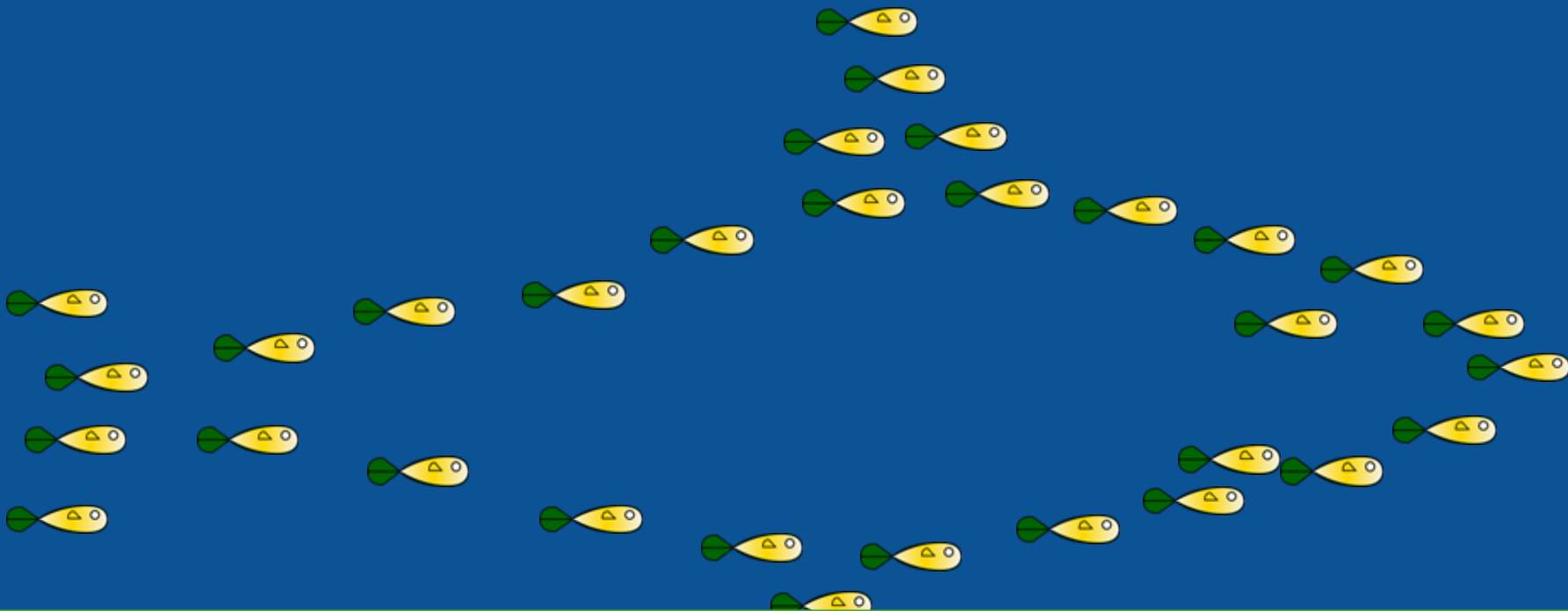
23. April 2026

The logo for 'aipt' is displayed in a large, white, sans-serif font. The letter 'a' is stylized with a hexagonal shape inside it. The background of the entire slide features a light blue gradient with a complex network of white lines and dots, resembling a circuit board or a data network.



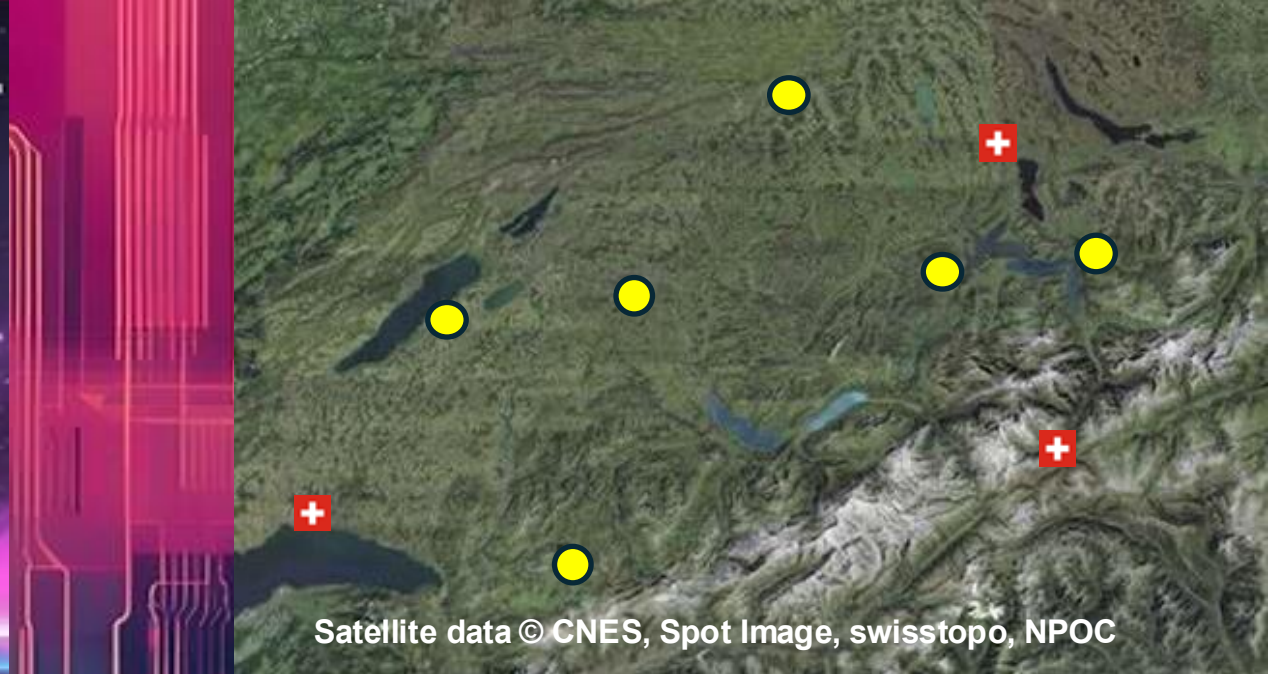
## Sovereign AI

AI systems that are developed, deployed, and governed by trusted national institutions, minimizing reliance on foreign entities.



## Public AI

AI as public infrastructure, provisioned like electricity, water, roads, libraries, or the Internet itself: ensuring everyone can participate, delivering accountability and value by the co-creation of public goods.



Satellite data © CNES, Spot Image, swisstopo, NPOC

# Swiss AI Initiative

**70+**  
Affiliated Professors

**800+**  
Researchers

**10+**  
Academic Institutions

Enabled by the ALPS  
Supercomputer, distributing 10-20  
Mio GPU hours in grants

Jointly led by



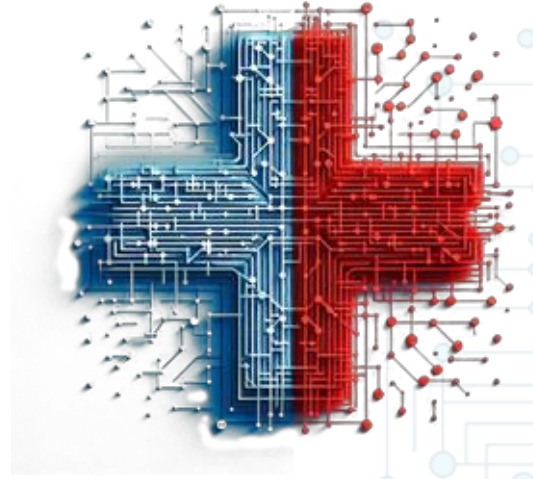
**CSCS**

More Information: [swiss-ai.org](https://swiss-ai.org)

# Apertus 2509

First release on September 2, 2025

- **Two** models with **8B** & **70B** parameters, that were trained on **15T tokens** of text.
- **Global representation:** Disambiguates data from over 1'000 languages, with 40% non-English content.
- **Open model:** Fully transparent platform to research and develop future AI systems. Apache 2.0 licensed.
- **Responsible Data Practices:** Satisfies the base provisions of EU AI Act and Swiss data protection laws.



# Why build Our own Models?

## 1. Legal necessity

Transparency obligations for responsible AI systems.

## 2. R&D autonomy

Focus on what matters (multilinguality, compliance, values)

## 3. Scientific integrity

Open data and reproducibility are foundational to science.

## 4. Deployment control

On-premise, no vendor lock, data stays internal (if you want)

# Why is Open our imperative?

## 1. Democracy

Governance of projects by actors with public accountability.

## 2. Representation

Address cultural bias, relevant metrics, inclusive oversight.

## 3. Trustworthy

Transparent practices, full reproducibility, local interests.

## 4. Regulatory Gap

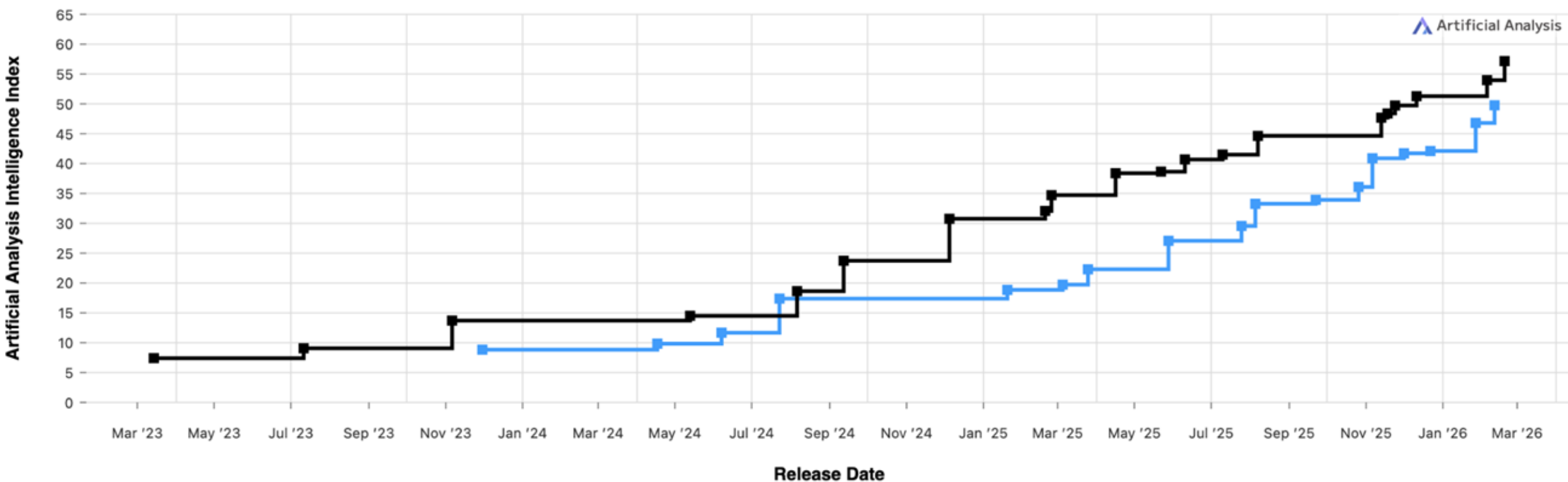
EU AI Act requires transparency and legal data sources.



## Progress in Open Weights vs. Proprietary Intelligence

Artificial Analysis Intelligence Index v4.0 incorporates 10 evaluations: GDPval-AA,  $\tau^2$ -Bench Telecom, Terminal-Bench Hard, SciCode, AA-LCR, AA-Omniscience, IFBench, Humanity's Last Exam, GPQA Diamond, CritPt

■ Open Weights ■ Proprietary



# Who's Who of Open Source AI

## Closed

US OpenAI  
US Anthropic  
US Perplexity  
US Google  
US xAI  
FR Mistral  
.....

Share outputs.

## Open-Weight

US Nemotron (NVIDIA)  
US Meta (Llama)  
CN Alibaba (Qwen)  
CN Moonshot (Kimi)  
CN Z.AI (GLM)  
FR Mistral (Devstral)  
.....

Share model weights.

## Fully Open

CH  Swiss AI (Apertus)  
EU  EuroLLM / OpenEuroLLM  
US  AI2 (Olmo)  
AE  MBZUAI IFM (K2-Think)  
SG  AI Singapore (SEA-LION)

...

**Share everything.**



= Non-profit institutions

# 15T

# Research Infrastructure as Key Enabler

Inaugurated in 2024, **ALPS** is one of Europe's largest (**11,000 GH200 GPUs**) and most efficient HPC infrastructures, maintained in collaboration w/ HPE & NVIDIA by the **Swiss National Supercomputing Centre (CSCS)**

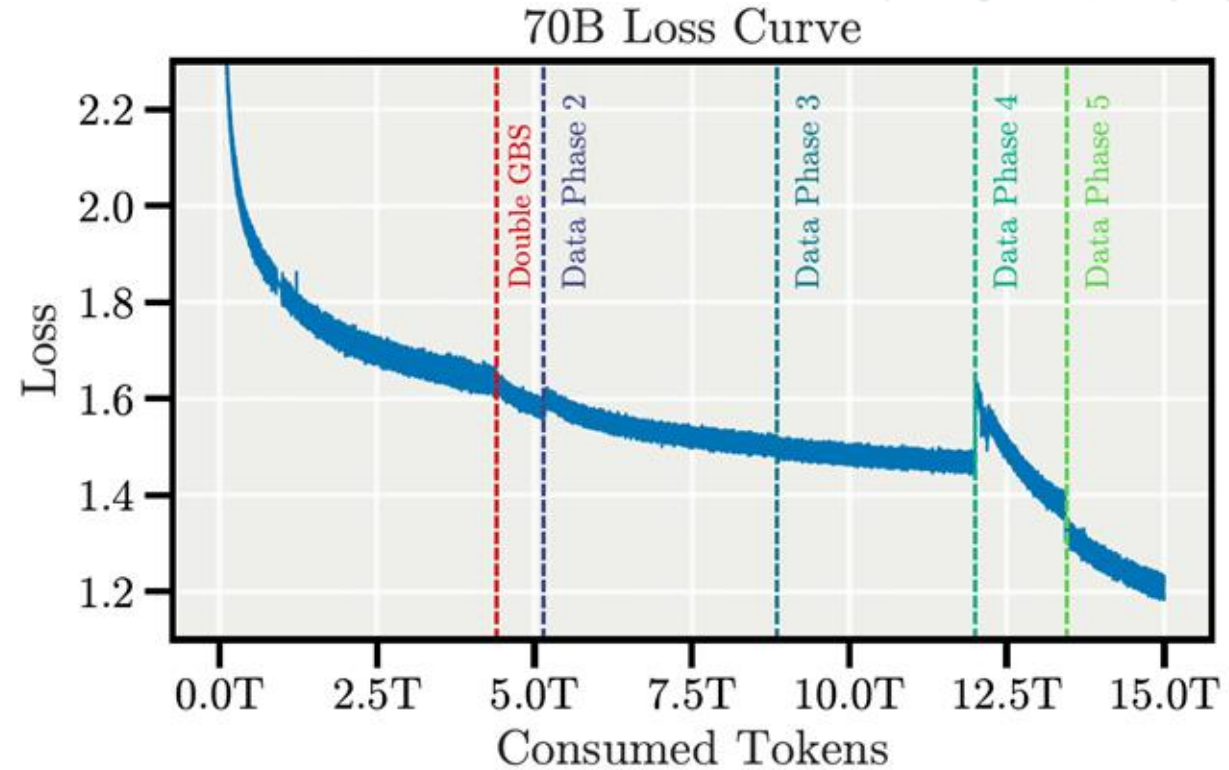
Hydro-powered & lake water-cooled, this is a green energy data science facility where training runs are tracked to optimize load.

 [cscs.ch/computers/alps](https://cscs.ch/computers/alps)



# Innovation in Architecture

- New optimizer:  
**AdEMAMix**
- New activation function:  
**XieLU**
- Memorization prevention:  
**Goldfish loss objective**



Learning **schedule** with data phases in continual training (warmup-stable-decay)

# Unparalleled Openness

- Source code
- **Deployment scripts**
- **Training data pipelines**
- **Evaluation harnesses**
- **Model checkpoints**
- **Technical report**

## APERTVS

DEMOCRATIZING OPEN AND COMPLIANT LLMs  
FOR GLOBAL LANGUAGE ENVIRONMENTS

APERTUS V1 TECHNICAL REPORT

### Project Apertus\*

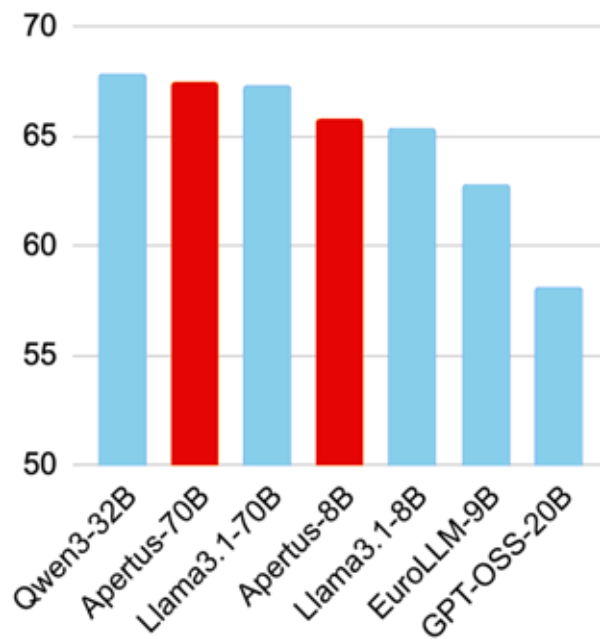
**Core Team:** Alejandro Hernández-Cano<sup>1</sup>, Alexander Hägele<sup>1</sup>, Allen Hao Huang<sup>1</sup>, Angelika Romanou<sup>1</sup>, Antoni-Joan Solergibert<sup>1,2</sup>, Barna Pasztor<sup>2</sup>, Bettina Messmer<sup>1</sup>, Dhia Garbaya<sup>1</sup>, Eduard Frank Āurech<sup>1,2</sup>, Ido Hakimi<sup>2</sup>, Juan García Giraldo<sup>1</sup>, Mete Ismayilzada<sup>1</sup>, Negar Foroutan<sup>1</sup>, Skander Moalla<sup>1</sup>, Tiancheng Chien<sup>2</sup>, Vinko Sabolčec<sup>1</sup>, Yixuan Xu<sup>1,2</sup>

**Contributors:** Michael Aerni<sup>2</sup>, Badr AlKhamissi<sup>1</sup>, Ines Altemir Marinas<sup>1</sup>, Mohammad Hossein Amani<sup>1</sup>, Matin Ansari-pour<sup>1</sup>, Ilia Badanin<sup>1,2</sup>, Harold Benoit<sup>1</sup>, Emanuela Boros<sup>1</sup>, Nicholas Browning<sup>3</sup>, Fabian Bösch<sup>3</sup>, Maximilian Bötter<sup>2</sup>, Niklas Canova<sup>2</sup>, Camille Challier<sup>1</sup>, Clement Charmillot<sup>1</sup>, Jonathan Coles<sup>3</sup>, Jan Deriu<sup>7</sup>, Arnout Devos<sup>2</sup>, Lukas Drescher<sup>3</sup>, Daniil Dzenhaliou<sup>1</sup>, Maud Ehrmann<sup>1</sup>, Dongyang Fan<sup>1</sup>, Simin Fan<sup>1</sup>, Silin Gao<sup>1</sup>, Miguel Gila<sup>3</sup>, Maria Grandury<sup>1</sup>, Diba Hashemi<sup>1</sup>, Alexander Hoyle<sup>2</sup>, Jiaming Jiang<sup>1</sup>, Mark Klein<sup>3</sup>, Andrei Kucharavy<sup>4</sup>, Anastasiia Kucherenko<sup>4</sup>, Frederike Lübeck<sup>2</sup>, Roman Machacek<sup>9</sup>, Theofilos Manitaras<sup>3</sup>, Andreas Marfurt<sup>5</sup>, Kyle Matoba<sup>1</sup>, Simon Matrenok<sup>1</sup>, Henrique Mendonça<sup>3</sup>, Fawzi Roberto Mohamed<sup>3</sup>, Syrielle Montariol<sup>1</sup>, Luca Mouchel<sup>1</sup>, Sven Najem-Meyer<sup>1</sup>, Jingwei Ni<sup>2</sup>, Gennaro Oliva<sup>3</sup>, Matteo Pagliardini<sup>1</sup>, Elia Palme<sup>3</sup>, Andrei Panferov<sup>6</sup>, Léo Paoletti<sup>1</sup>, Marco Passerini<sup>3</sup>, Ivan Pavlov<sup>1</sup>, Auguste Poiroux<sup>1</sup>, Kaustubh Ponskhe<sup>1</sup>, Nathan Ranchin<sup>1</sup>, Javi Rando<sup>2</sup>, Mathieu Sauser<sup>1</sup>, Jakhongir Saydaliev<sup>1</sup>, Muhammad Ali Sayfiddinov<sup>2</sup>, Marian Schneider<sup>2</sup>, Stefano Schuppli<sup>3</sup>, Marco Scialanga<sup>1</sup>, Andrei Semenov<sup>1</sup>, Kumar Shridhar<sup>2</sup>, Raghav Singhal<sup>1</sup>, Anna Sotnikova<sup>1</sup>, Alexander Sternfeld<sup>4</sup>, Ayush Kumar Tarun<sup>1</sup>, Paul Teiletche<sup>1</sup>, Jannis Vamvas<sup>8</sup>, Xiaozhe Yao<sup>2</sup>, Hao Zhao<sup>1</sup>

**Full 119-page guide on how  
to train an LLM from scratch**

# Independent Benchmarks

Apertus 1.0 has a strong foundation, with state-of-the-art results among fully open models on multilingual benchmarks, rivalling or surpassing open-weight counterparts in the same class.



Above: [3rd-party evaluation](#) of Base LLMs by Leibniz Supercomputing Center and Cerebras. Adapted from Fig. 4. All Tasks. Accuracy 0.0-0.5  
Left: Pre-training evaluation, Apertus 1.0 Technical Report

# Evaluating Representation



See Table 20, Post-training Evaluation: Performance (%) of open weight models on cultural knowledge. Focus on < 10B.  
[Apertus Technical Report, 2025](#)

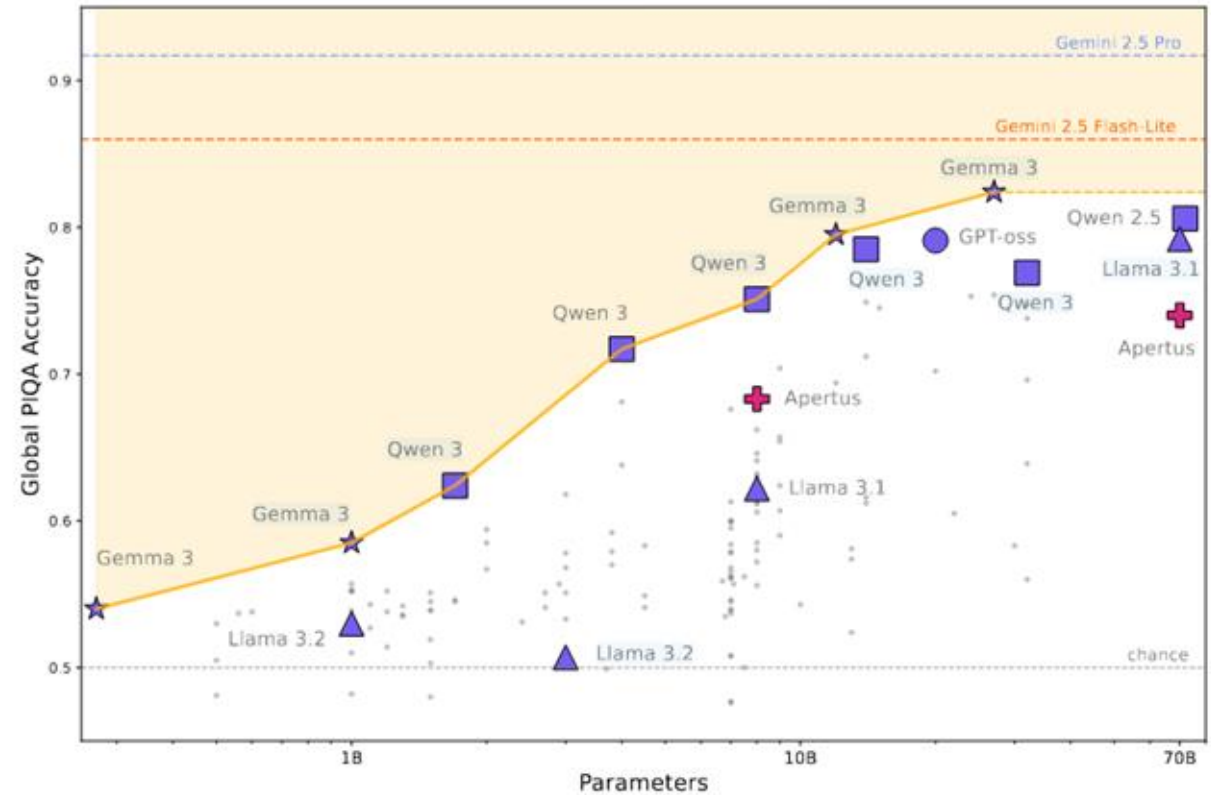


Figure 3: Accuracy averaged across all languages vs. parameter count for open-weight models.  
[Chang et al., 2025](#)

# Kindling international Collaboration

From the community, we have already seen:

- **2 Million+** downloads on Hugging Face
- **1000+** ❤️ likes and ⭐ stars (thank you!)
- **50** Quantizations, including MLX, GGUF, 1.7B, and more
- **30** Finetunes with curated datasets and capabilities



# Supporting the Public Sector

Pilot studies in Swiss public institutions:

- Federal Tribunal / GovTech
- Canton of Tessin (Artificialy)
- Clinical assistants (Moove)
- Legal assistants (PolyLex)
- Teaching assistants (Ethel)
- 10+ companies already offer support
- Many open source initiatives ...



**ZüriCityGPT OSS Version**

Ich bin ZüriCityGPT OSS Version und ich weiss (fast) alles, was auf stadt-zuerich.ch publiziert ist.

Was sind Argumente der Stadt Zürich zur Parkkartenverordnung Abstimmung vom 28. September? [Frage stellen](#)

Suchresultate	Antwort
<p>Für die Antwort berücksichtigte Links:</p> <p>Mittwoch, 5. Februar 2025 Medienmitteilungen</p> <p><b>Volksreferendum gegen die Parkkartenverordnung zustande gekommen</b> Das Volksreferendum ist mit der nötigen Anzahl gültiger Unterschriften zustande gekommen.</p> <p>Ein Volksreferendum gegen die Parkkartenverordnung wurde mit 4486 gültigen Unterschriften eingereicht, wobei 2057 Unterschriften als gültig anerkannt wurden, und die Abstimmung ist für den 28. ...</p> <p>Montag, 4. August 2025 <b>Abstimmungspublikation 28. September 2025 (PDF) (p. 16-17)</b></p> <p>Die Abstimmung am 28. September 2025 umfasst sechs Vorlagen, darunter eine Volksinitiative für ein günstigeres Jahresabo der VBZ, eine neue Parkkartenverordnung mit veränderten Gebühren und Regelungen...</p>	<p>Das Referendum gegen die Parkkartenverordnung wurde am 22. Januar 2025 mit 4486 Unterschriften eingereicht, wobei 2057 Unterschriften gültig waren. Die Volksabstimmung findet am 28. September 2025 statt. <a href="#">Link 1</a></p> <p>Der Stadtrat und der Gemeinderat empfehlen die Annahme der neuen Parkkartenverordnung. Sie argumentieren, dass die Verordnung die bisherige, oft unübersichtliche und unterschiedliche Regelung von über 20 Park- und Zufahrtbewilligungen vereinheitliche und modernisiert. Die neue Verordnung soll eine einheitliche Rechtsgrundlage schaffen und die Parkierungs- und Zufahrtbewilligungen an die Bedürfnisse der heutigen Stadt anpassen, insbesondere unter Berücksichtigung des neuen kommunalen Richtplans Verkehr, der die Reduktion von Parkplätzen in der Blauen Zone und die Förderung von umweltfreundlichen Verkehrsmitteln vorsieht. <a href="#">Link 2</a></p> <p>Das Referendumskomitee argumentiert dagegen, dass die Verordnung eine leere, komplizierte und unsoziale Gebühr einführen würde, insbesondere für Personen mit schmalen Budget, Familien und Besitzer von Elektrofahrzeugen. Sie sehen die neue Preisgestaltung, die sich nach dem Fahrzeuggewicht richtet, als unfair an. <a href="#">Link 3</a></p> <p>Die Minderheit im Gemeinderat sieht in der Verordnung eine Behördenwillkür und Wucherpreise und lehnt sie ab. <a href="#">Link 3</a></p> <p>Der Stadtrat und der Gemeinderat betonen jedoch, dass die Verordnung notwendig sei, um die Parkplatzplanung zu optimieren, den öffentlichen Verkehr zu fördern und den Klimaschutz zu unterstützen. Sie sehen die erweiterte Gewerbebewilligung als Vorteil für die lokale Wirtschaft und die Anpassung der Preise als fair und nachhaltig. <a href="#">Link 2</a></p> <p>Zusammengefasst: Die Stadt Zürich verteidigt die neue Parkkartenverordnung als notwendige Modernisierung und Vereinfachung der bisherigen Regelungen, während das Referendumskomitee und die Minderheit im Gemeinderat B. ...</p>

[oss.zuericitygpt.ch](https://oss.zuericitygpt.ch)

# Catalyzing Private Sector Innovation

154 Partners

33 Cities



AI TALK

15 Hackathons

240 Events

SRF news  
tagesschau

SRF news  
IO VOR IO

140  
Prototypes

50   
Apertus prototypes

RSI

on {ai}r 4'150 Sign-ups  
87% very satisfied/satisfied participants

4'400 followers  
LinkedIn

SWISS  
{ai} Weeks

2'820  
Luma subscribers

750 members  
Discord

4'100  
Newsletter subscribers

# Catalyzing Private Sector Innovation

>>venture>>

Join the Waiting List



## 2026 Spotlight: Apertus LLM

For the inaugural 2026 award, we're spotlighting Apertus – Switzerland's new open-source LLM launched by ETH Zurich, EPFL, and CSCS. It marks a milestone in sovereign, transparent, and multilingual AI, supporting underrepresented languages and developers worldwide.

We're looking for the most creative, high-impact application of Apertus within a Swiss startup, recognizing innovation, technical excellence, and real-world impact.

Learn More

# Enabling Industry Collaborations



Swisscom is a Strategic Partner of the Swiss AI Initiative



## Swisscom Innovator Stage at AI+X Summit 2025

Panel w/ Daniel Dobos (Research Director at Swisscom), Katharina Frey (Executive Director ICAIN), Imanol Schlag (Co-Lead Apertus) and David Froehlicher (Research Manager at Swisscom)

# Platform for Swiss AI Industry



## SME CIRCLE

Join the community, leverage the freely usable frontier models like Apertus and benefit from resources and updates.



## SNAI MEMBERSHIP

Access deeper AI expertise, events, and visibility across the entire Swiss AI ecosystem.



## STRATEGIC PARTNERSHIPS

Contribute to research through long-term, multi-year collaboration with customized research wallet or donation plans.

Are you building, or planning to build on Apertus? → [industry@swiss-ai.org](mailto:industry@swiss-ai.org)



**ETH AI CENTER**

**June 17, 2026**



**SME  
CIRCLE**

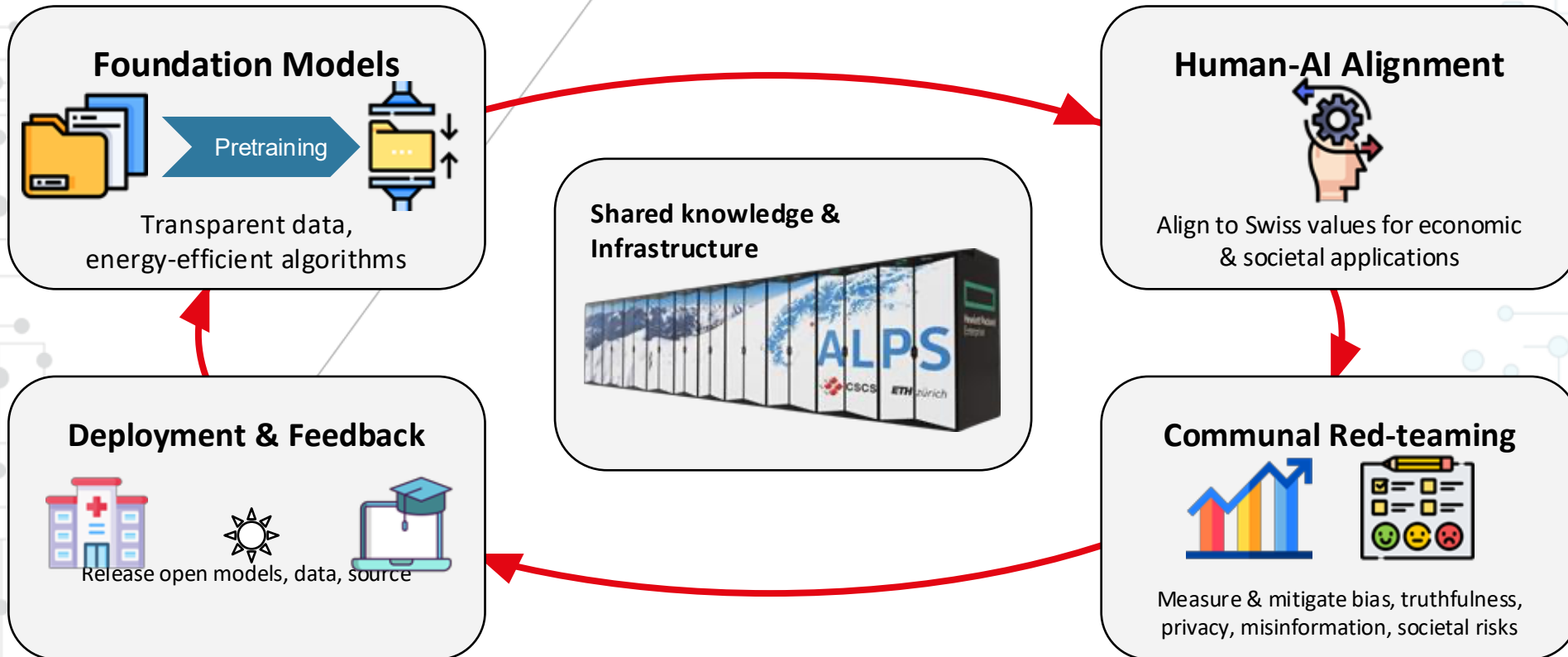
Join the community, leverage the freely usable frontier models like Apertus and benefit from resources and updates.

*INSIDE*

APERTVS

Subscribe to our newsletter at **APERTVS.ai**

# Apertus Release Cycle



# Apertus 1.5 Release

## **Continued Pretraining**

Additional high-quality data. Adding 4T tokens to previous training base. Data upsampling for strategic areas, e.g. health, education, justice.

**Multimodality:** Enable native audio and image input.

**Agentic Skills:** Improved instruction-following, tool-use, and advanced reasoning capabilities with our efficient RL codebase at scale.

>> Summer 2026

# Where can you find Apertus?

**Weights available on Hugging Face:** For individual deployment anywhere under an Apache 2.0 license, plus inference APIs.

**Available from cloud providers:** Amazon AWS, Microsoft Azure, Swisscom, Exoscale, Infomaniak, Phoenix Technologies, ...

**Supported for local inference in leading LLM software:**  
vLLM, SGLang, llama.cpp, MLX, Transformers

**Hosted by third-party non-profit organisations:**  
try Apertus chat on Hugging Face or at [publicai.co](https://publicai.co)



APERTVS

**EPFL**

**ETH** zürich



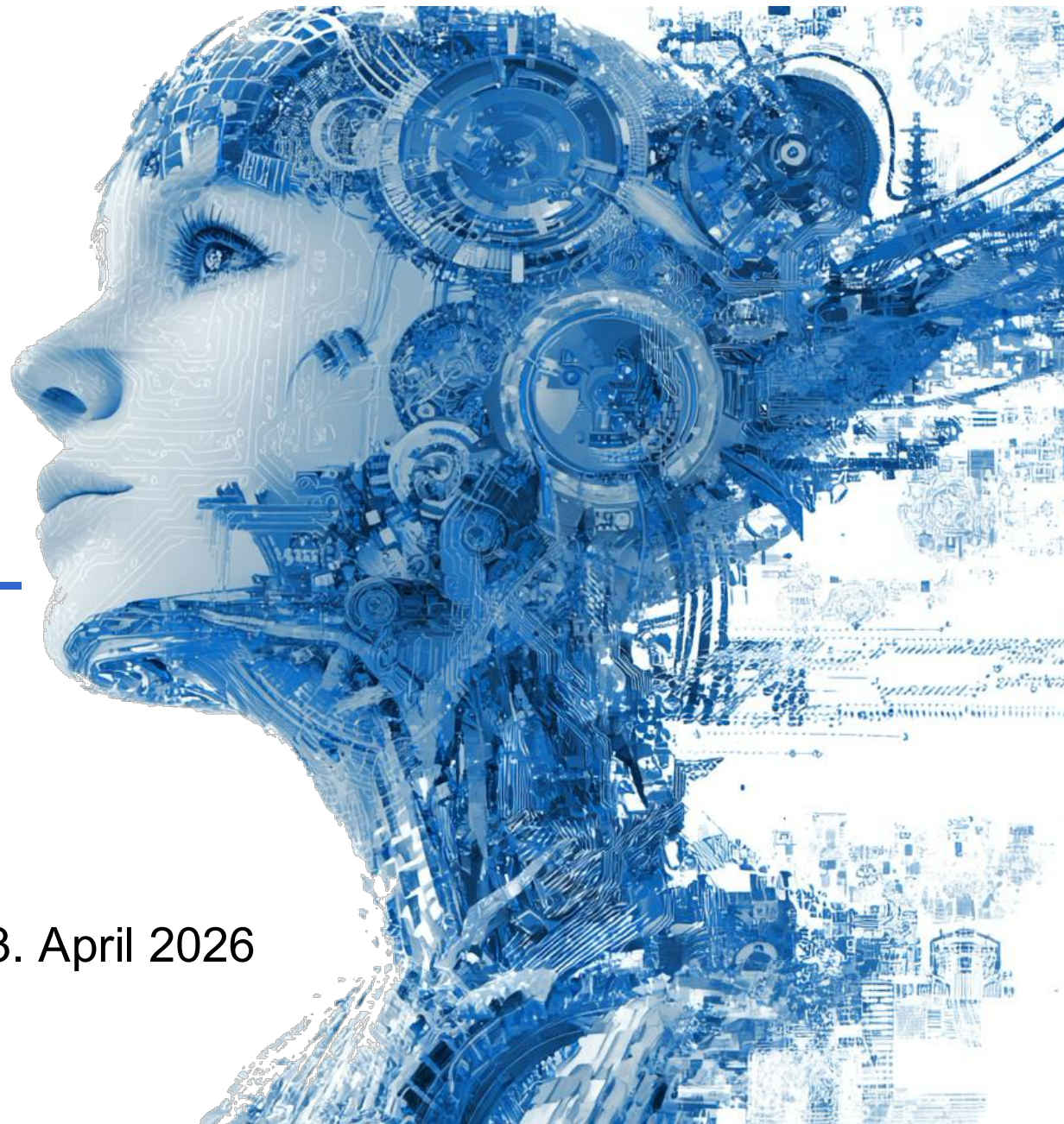
**CSCS**

**APERTVS.ai**

# Souveräne KI im BIT:

**Private LLMs als Service –  
sicher, nachvollziehbar,  
betreibbar**

Gian Derungs & Bernhard von Allmen – 23. April 2026





# AI Vision Statement BIT



*Wir machen künstliche Intelligenz zu einem **integralen Bestandteil** der **Geschäftsprozesse des BIT** und stellen diese neuen Erfahrungen in den Dienst der **öffentlichen Verwaltung**, um deren Entscheidungsfindungs-, Automatisierungs- und Lernfähigkeiten **nachhaltig** zu erweitern.*

AI Accelerator (Januar 2025)



# Warum souveräne KI wichtig ist

1

## Schatten-KI ist längst hier

54%

der Beschäftigten in der Schweiz nutzen nicht genehmigte GenAI-Tools.

*Patientendaten, interne Sitzungsprotokolle usw. alles bei ChatGPT hochgeladen, oft ohne Wissen der IT.*

*Im Juli 2025 waren tausende geteilte ChatGPT-Konversationen über Google indexiert.*

Quellen: [NZZ](#), [Salesforce-Studie CH](#)

2

## Dein Schreibstil verrät dich

1'000

Wörter genügen, um einen einzigartigen Schreib-Fingerabdruck zu bilden.

*ETH Zürich zeigte: LLMs können Reddit-Kommentare anonymer Nutzer mit hoher Genauigkeit mit echten Identitäten verknüpfen — allein aus Syntax, Wortwahl, Satzlänge.*

*Risiko: hoch personalisiertes Social Engineering, Phishing im persönlichen Stil, Deanonymisierung.*

Quelle: [ETH Zürich Stylometrie-Studie 2024/2025](#)

3

## Geopolitische Abhängigkeit

5-10%

der weltweiten KI-Rechenkapazität liegen in Europa.

*Was passiert, wenn der Zugang zu führenden Modellen aus politischen Gründen zurückgehalten wird?*

*Schweizer Geschäftsentscheide stützen sich heute auf KI, die ausserhalb des Landes trainiert, reguliert und betrieben wird.*

Quellen: [EY Schweiz KI-Trends 2026](#)

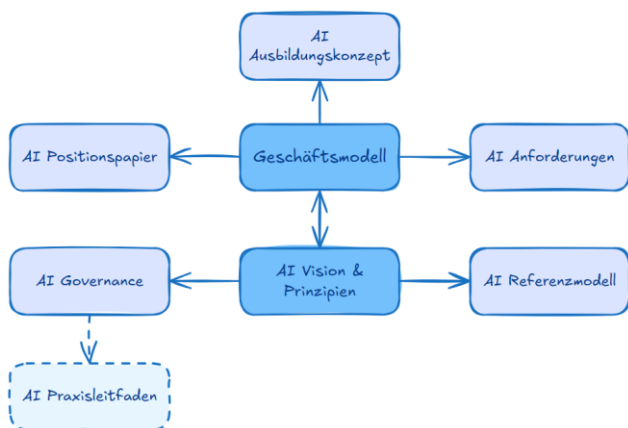


# Vom Experiment zur KI-befähigten Organisation

## AI Incubator

2024

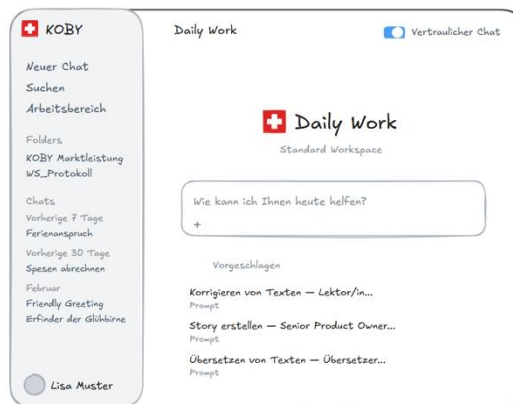
Erarbeitung der Grundlagen und Umsetzung eines PoC «GovGPT»



## AI Accelerator

2025

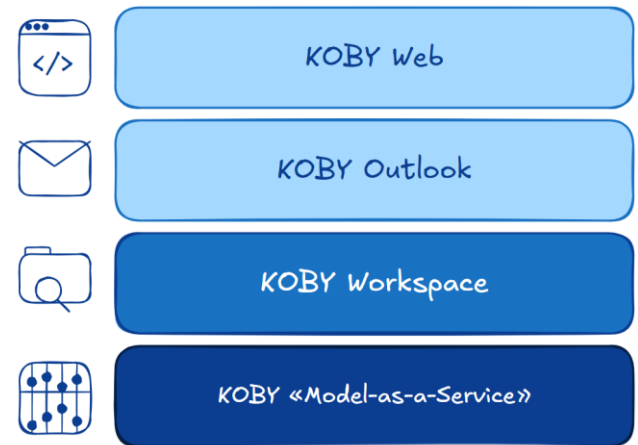
Aufbau der KI-Plattform, Bildung des AI-Teams und Umsetzung eines MVP für das BIT



## KI-Marktleistung

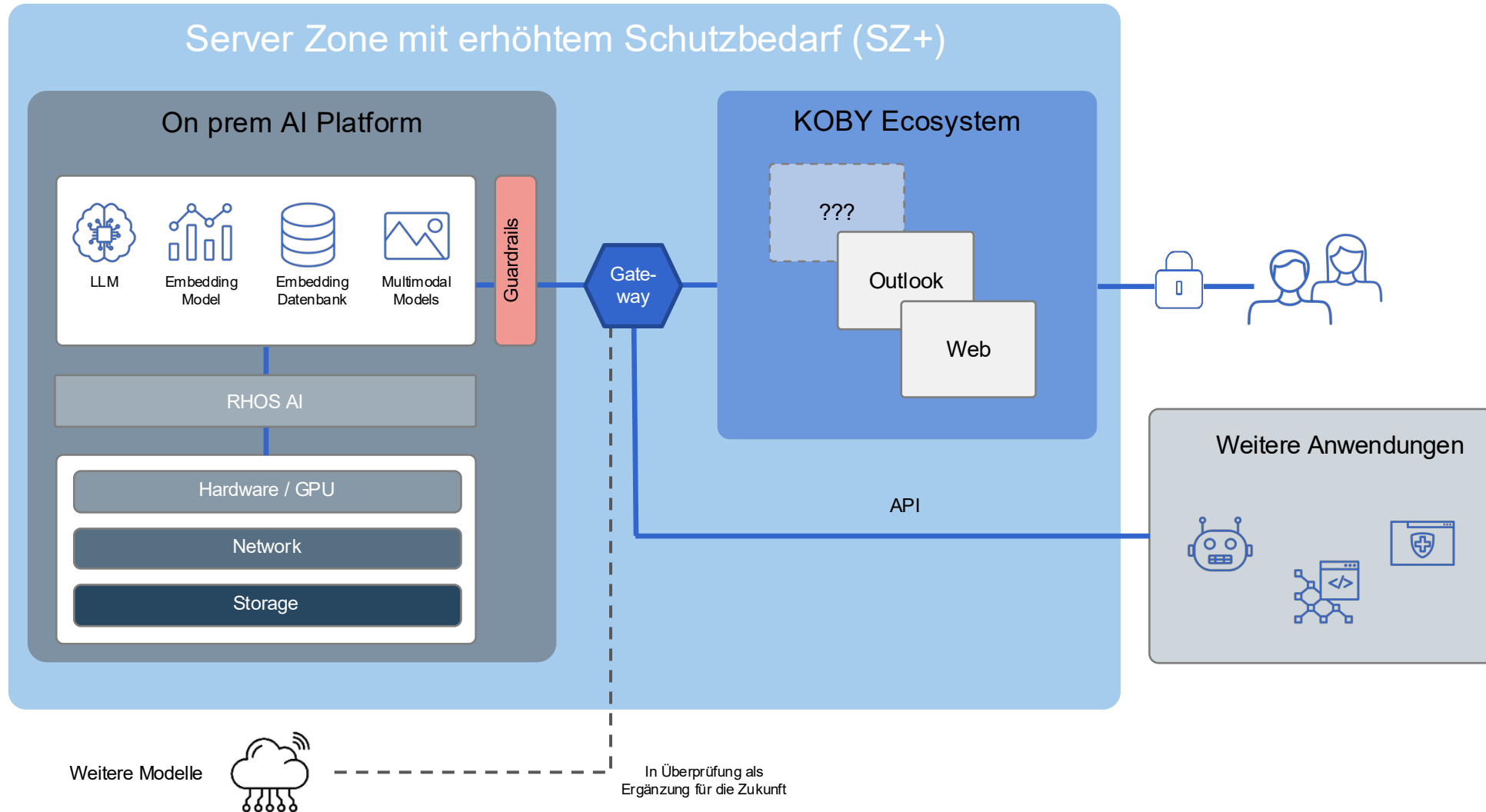
2026

Einführung der Marktleistung namens «KOBY» für unsere Kunden





# Lösungsansatz



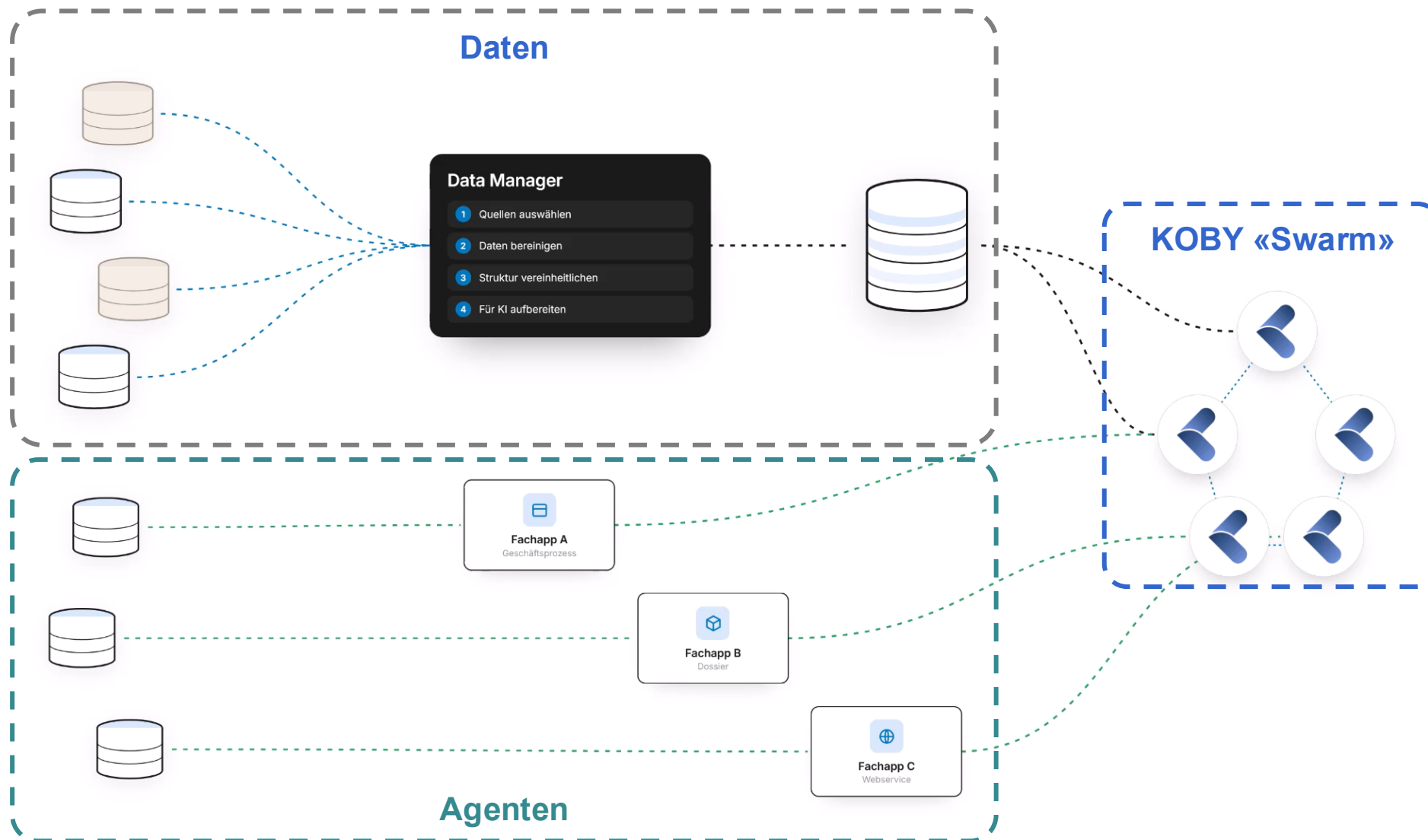


**Wie geht es nun weiter?**

**data is key**



# Daten im Mittelpunkt des enablement





# Danke für Ihre Aufmerksamkeit

Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

KOBY

Daily Work +  
Als Standard festlegen

Vertraulicher Chat 🔒

KOBY

Neuer Chat

Suchen

Arbeitsbereich

Folders

- KOBY Marktleistung
- WS\_Protokoll (Beispiel)

Chats

Vorherige 7 Tage

Dein Ferienanspruch Überbl...

Vorherige 30 Tage

Spesen abrechnen

Februar

Friendly Chat Greeting

Erfinder der Glühbirne

Lisa Muster

Daily Work  
Standard Workspace (GPT-OSS Medium)

Wie kann ich Ihnen heute helfen?

Vorgeschlagen

- Korrigieren von Texten - Agiere als professionelle/r Lektor/in inklusive Korrektor/in  
Prompt
- Story erstellen - Agiere als erfahrener Senior Product Owner. Erstelle eine auf das...  
Prompt
- Übersetzen von Texten - Agiere als professioneller Übersetzer. Aufgabe: Übersetze  
Prompt

## KOBY Web

## KOBY Outlook

Reporting Personalmanagement 2025

EPA-InfoPers <infopers@epa.admin.ch>  
To: von Allmen Bernhard BIT

Fr 06.03.2026 10:24

direkt an mich  
Follow up. Completed on Freitag, 6. März 2026.

InfoPers | Die Schweiz mitgestalten

**Reporting Personalmanagement 2025**

Liebe Mitarbeiterin, lieber Mitarbeiter

Der Bundesrat hat an seiner Sitzung vom 6. März 2026 das Reporting Personalmanagement 2025 gutgeheissen und an die Geschäftsprüfungs- und Finanzkommissionen der eidgenössischen Räte weitergeleitet.

Um den Fortschritt der Personalpolitik der Bundesverwaltung messen zu können, legt der Bundesrat für jede Legislatur eine Reihe strategischer Ziele in Form von Sollwerten fest. Der Bundesrat hat am 1. Mai 2024 beschlossen, sich für die Legislatur 2024–2027 auf Sollwerte in sechs personalpolitisch bedeutenden Bereichen zu fokussieren: die Geschlechterverteilung, die Sprachenanteile, den Anteil beschäftigter Menschen mit Behinderungen, den Anteil Lernende, den Anteil Hochschulpraktikantinnen und -praktikanten sowie neu die Verbleibquoten der Lernenden und Hochschulpraktikantinnen und -praktikanten. Die Sollwerte sind als Zielbänder ausgestaltet und bezeichnen den Stand, der bis Ende 2027 erreicht werden sollte.

Im zweiten Berichtsjahr der Legislaturperiode 2024–2027 wurden fünf der elf strategischen Sollwerte bereits erreicht. Es sind dies der Frauenanteil in den Lohnklassen 30–38 (von 28.7 auf 30 Prozent erhöht) und der Frauenanteil in der Verteidigung (von 15 auf 16.3 Prozent erhöht). Auch liegen die Anteile aller vier Landessprachen innerhalb der Sollwerte (69.1 Prozent Deutsch, 23.5 Prozent Französisch, 7 Prozent Italienisch, 0.5 Prozent Rätoromanisch). Der Anteil Lernende (4.1 Prozent) sowie der Anteil Hochschulpraktikantinnen und -praktikanten (2.3 Prozent) befinden sich weiterhin innerhalb des angestrebten Zielbandes. Bei der Geschlechterverteilung insgesamt, beim Frauenanteil in den Lohnklassen 24–29, beim Frauenanteil im Grenzwachtkorps und beim Anteil beschäftigter Menschen mit Behinderungen liegt der Wert nur noch ganz knapp unter den Sollwerten. Die

Chat

Bitte fasse diese E-Mail in 2 Sätzen zusammen.

Der Bundesrat hat das Reporting Personalmanagement 2025 genehmigt und die strategischen Sollwerte für die Legislatur 2024–2027 (Geschlechterverteilung, Sprachenanteile, Behinderte, Lernende, Hochschulpraktikant\*innen und Verbleibquoten) veröffentlicht. Im zweiten Berichtsjahr wurden bereits fünf der elf Sollwerte erreicht, während bei einigen Zielgrössen (z. B. Gesamt-Geschlechterverteilung, Frauenanteil in Lohnklassen 24–29, Behinderte, Verbleibquoten) noch leichte Nachholbedarfe bestehen.

Stelle irgendeine Frage

# ETH AI Center

Amplifying human potential through  
interdisciplinary AI research and  
enabling the path from science to society

**ETH** zürich

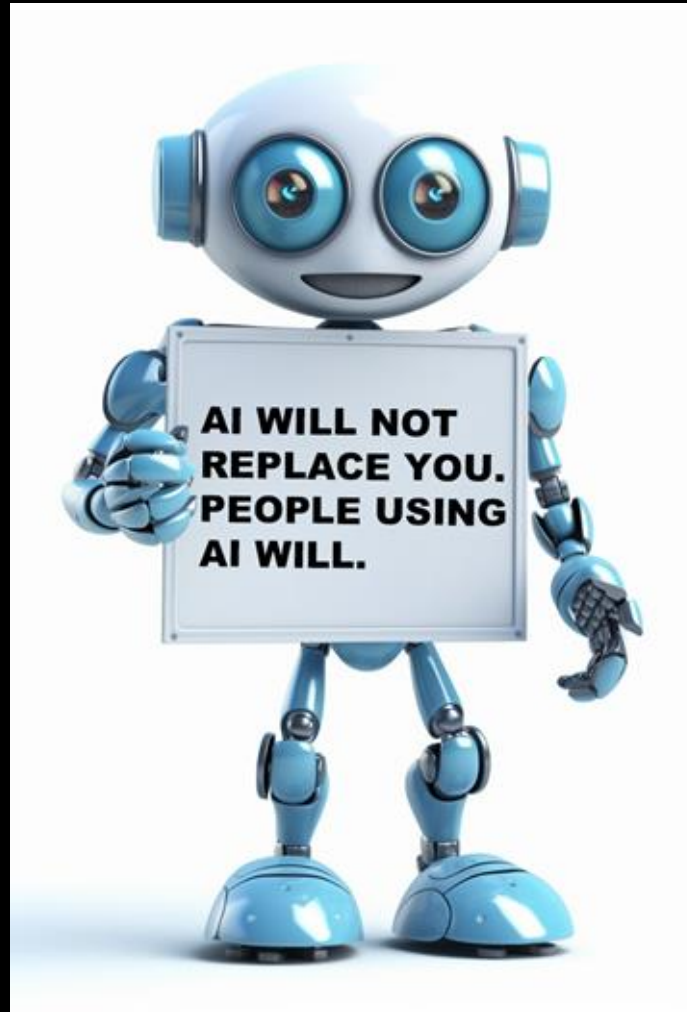


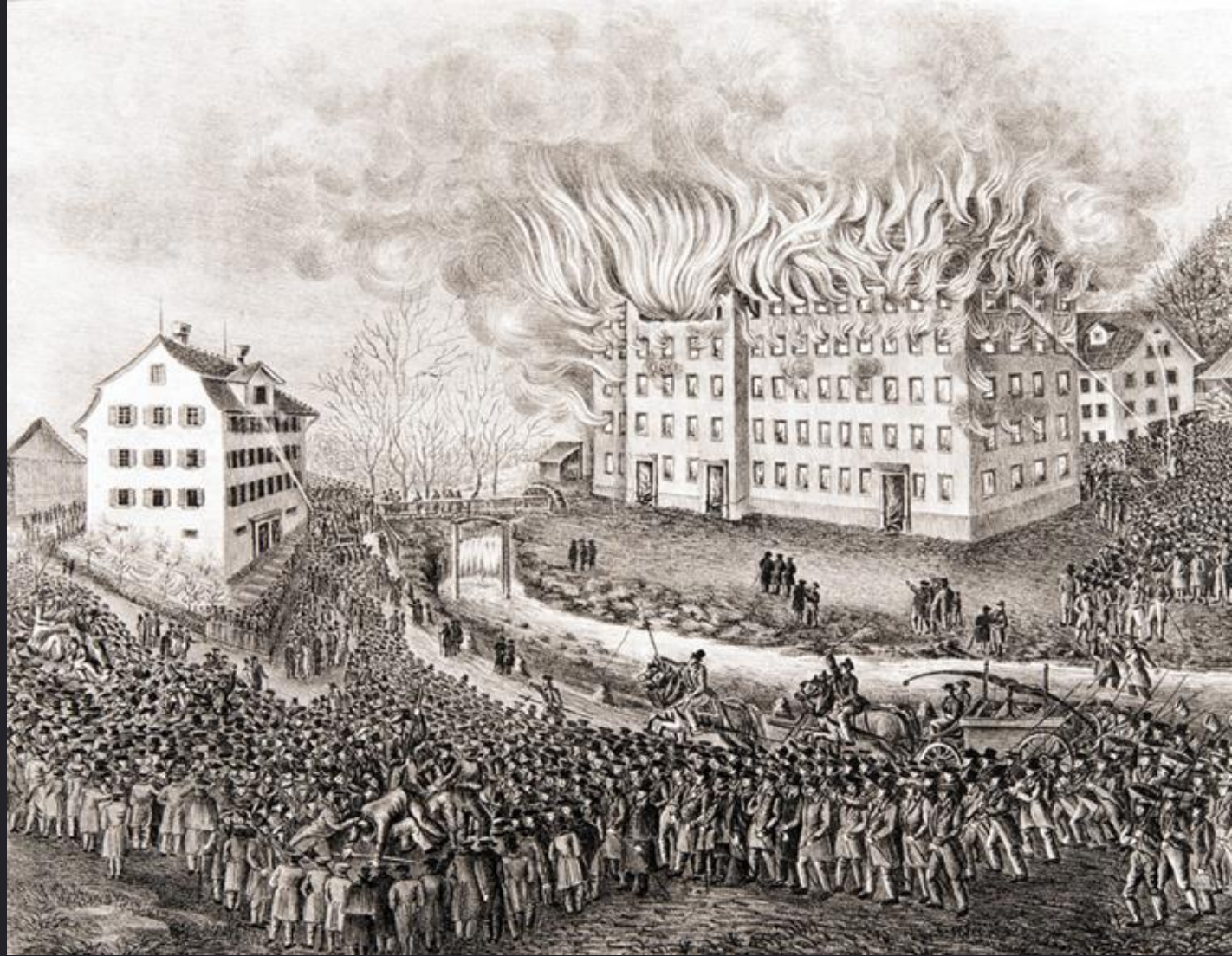
**ETH AI CENTER**

ipt Event

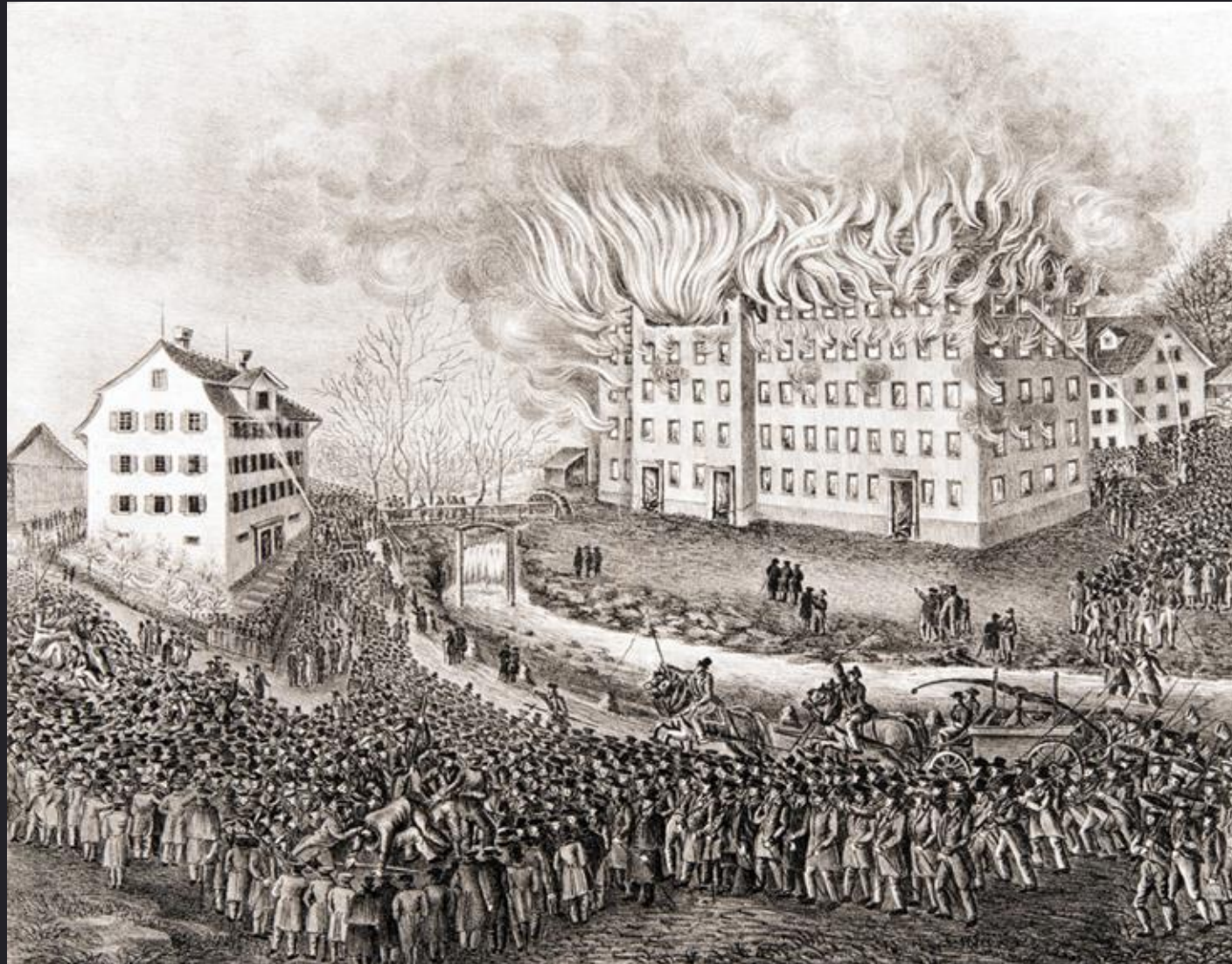
April 23, 2026

Daniel Naeff





# Brand von Uster 1832



Source: Wikipedia / Usterbrand

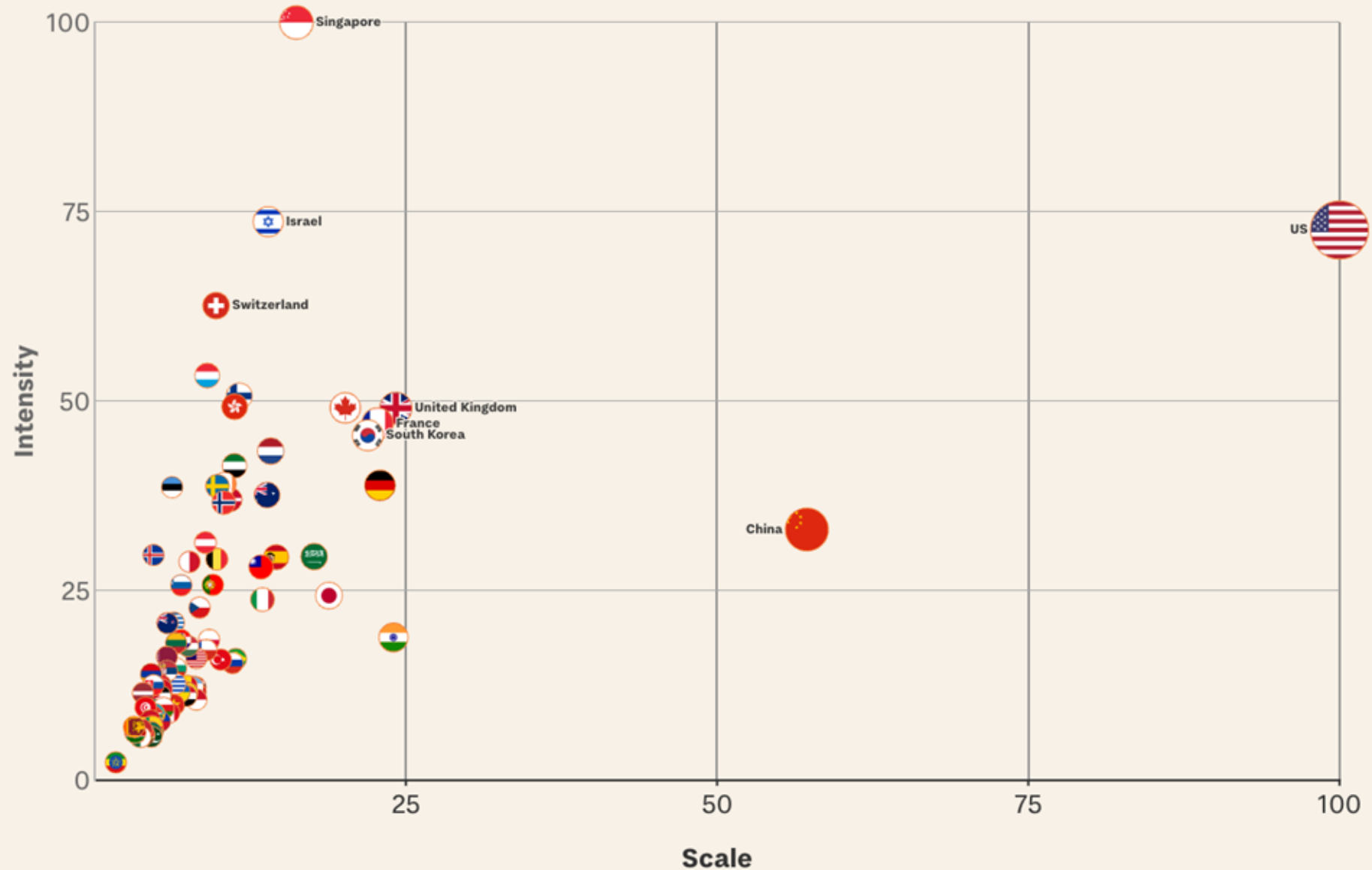


# Bündner Autoverbot 1900-1925



Source: Wikipedia / Stadtarchiv Chur

# National AI Capacity



# Position of **Switzerland**

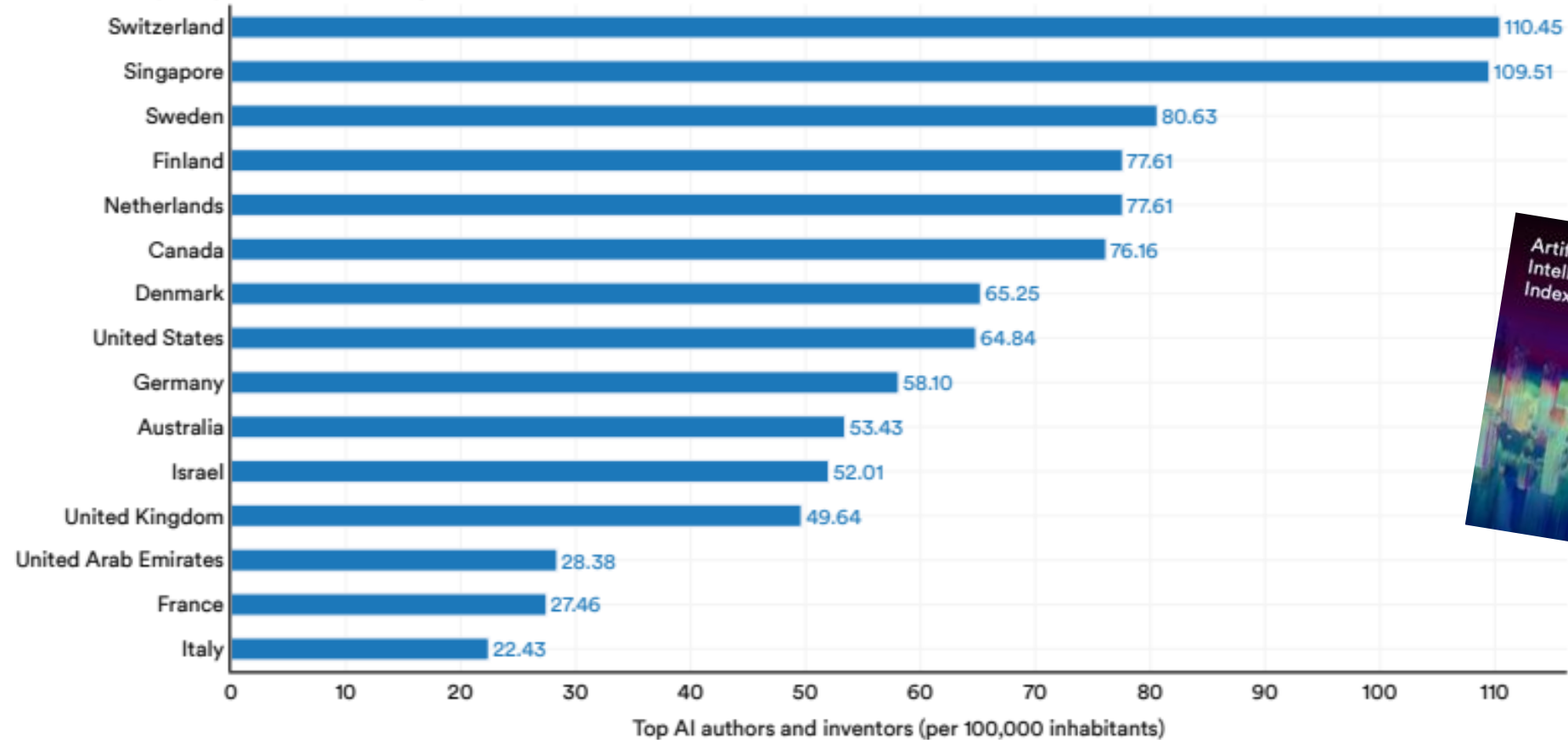


Source: Tortoisemedia (2024)

# Top AI Authors

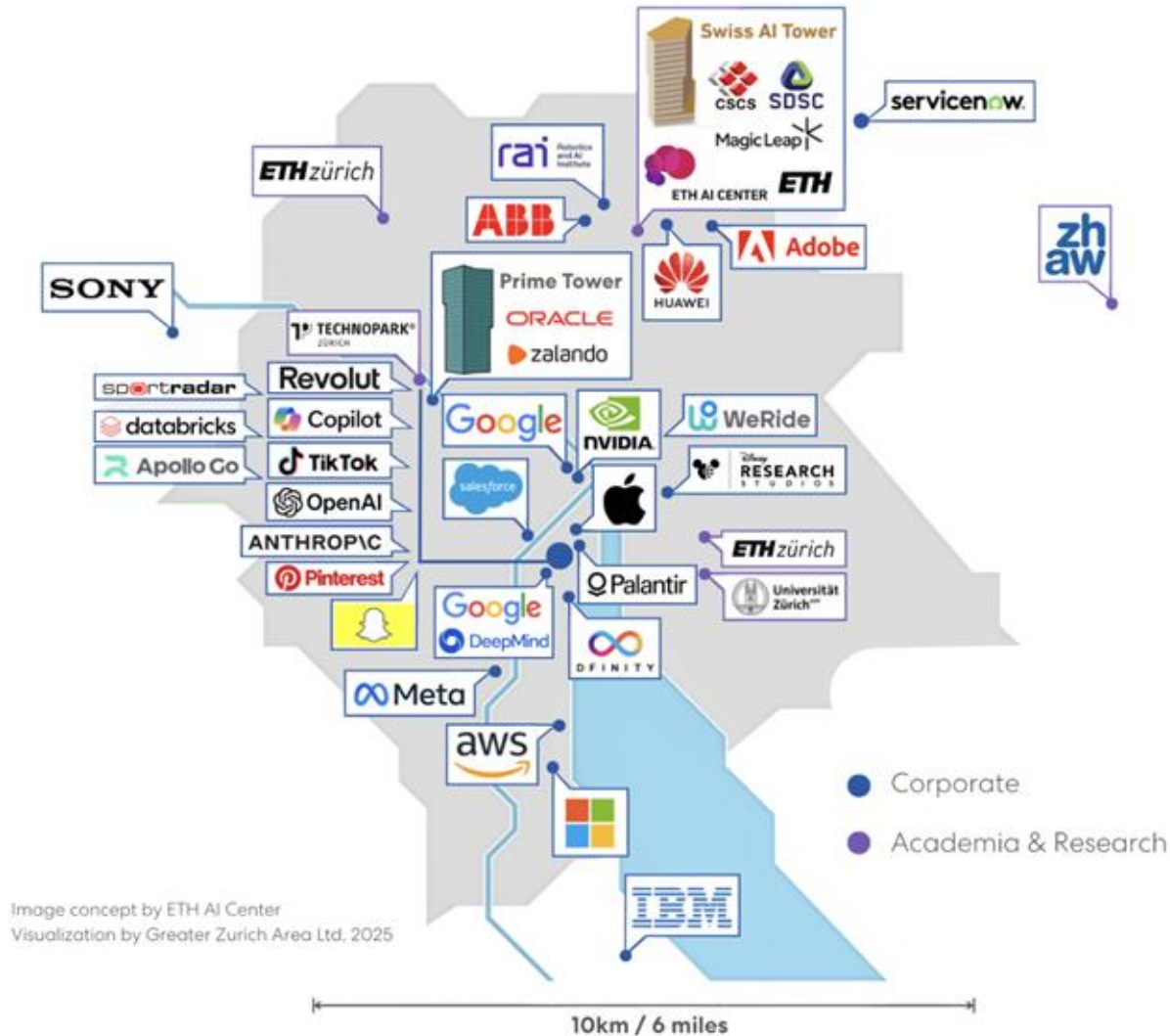
## Top AI authors and inventors per 100,000 inhabitants by country, 2025

Source: Zeki Data, 2025 | Chart: 2026 AI Index report



Source: [AI Index Report 2026](#)

# One of the most dense AI ecosystems in the world



# 1

Switzerland is a global leader in attracting AI talent  
(IBM Talent Ranking 2025)

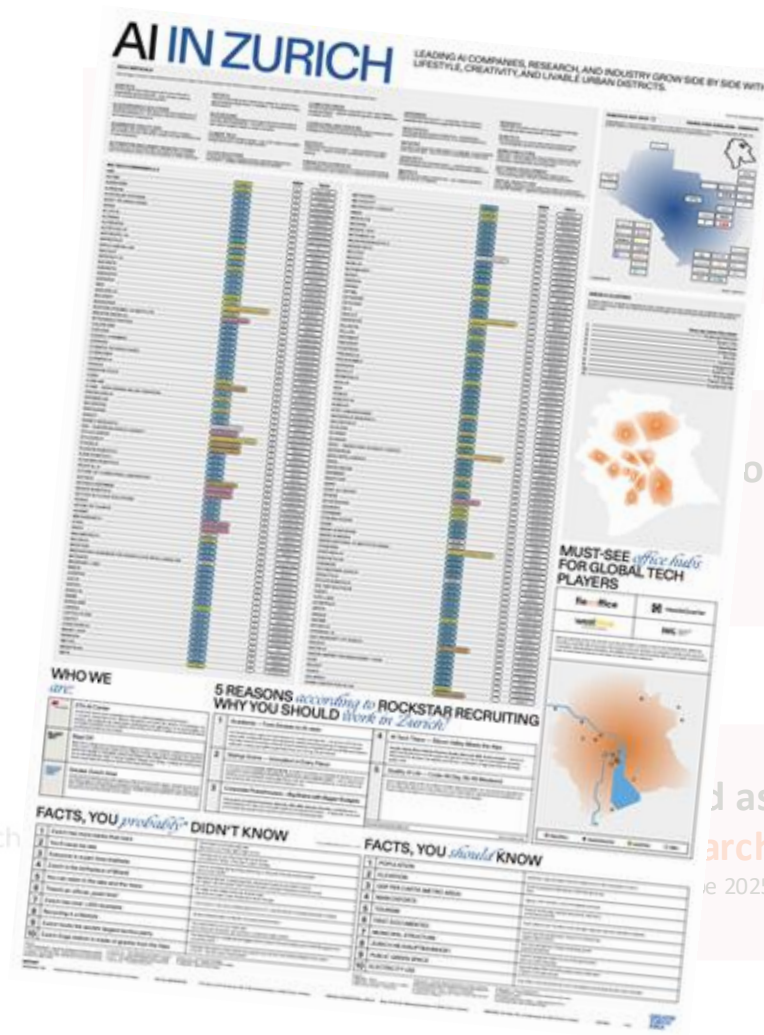
# 2

Zurich ranks 2nd in Europe for AI engineer density.  
(Greater Zurich Area)

# 1

ETH Zurich recognized as the top university in Europe for AI research  
(QS World University Ranking: Europe 2025)

# One of the most dense AI ecosystems in the world



ader in

ope for AI

l as the top university

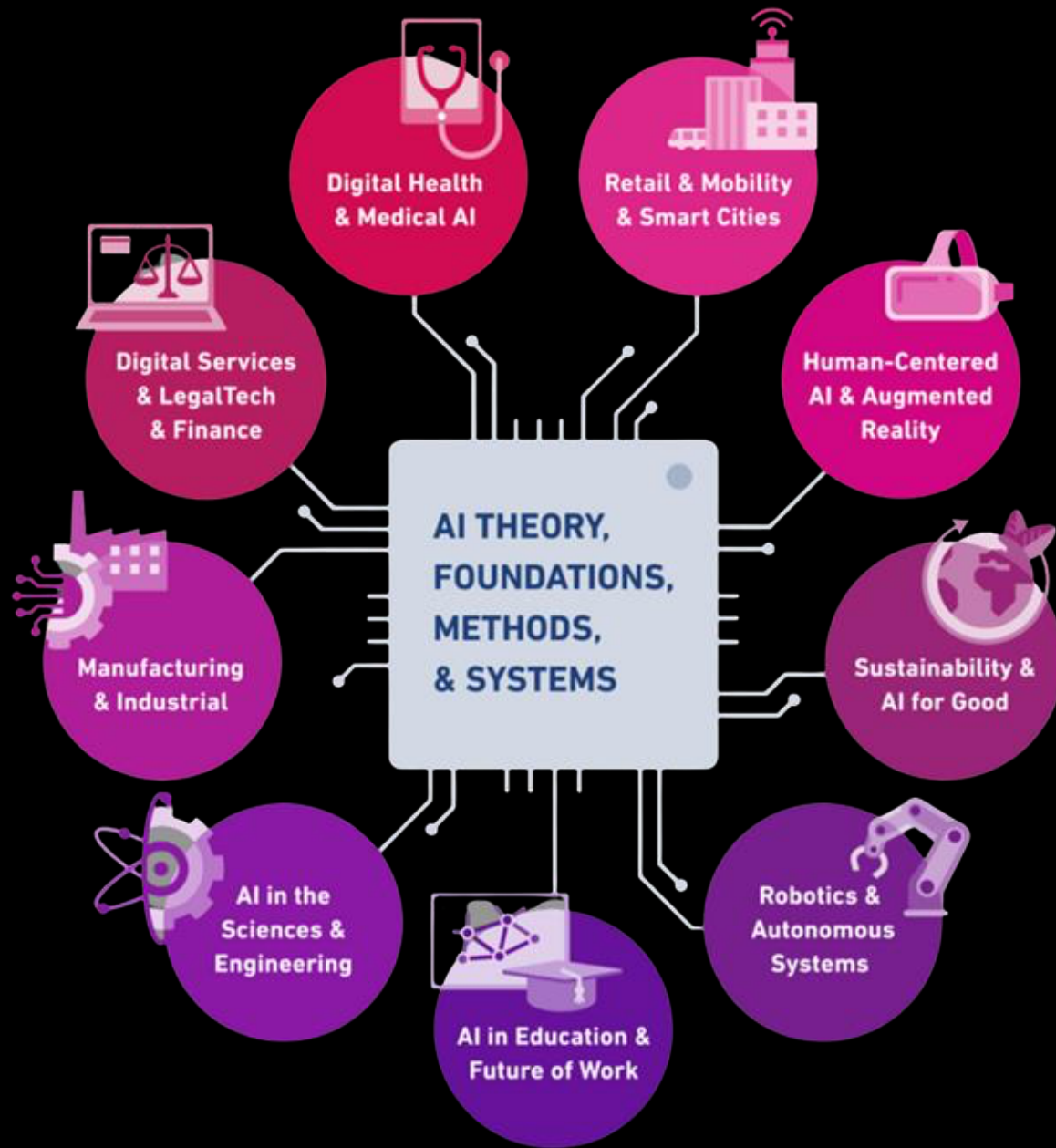
arch

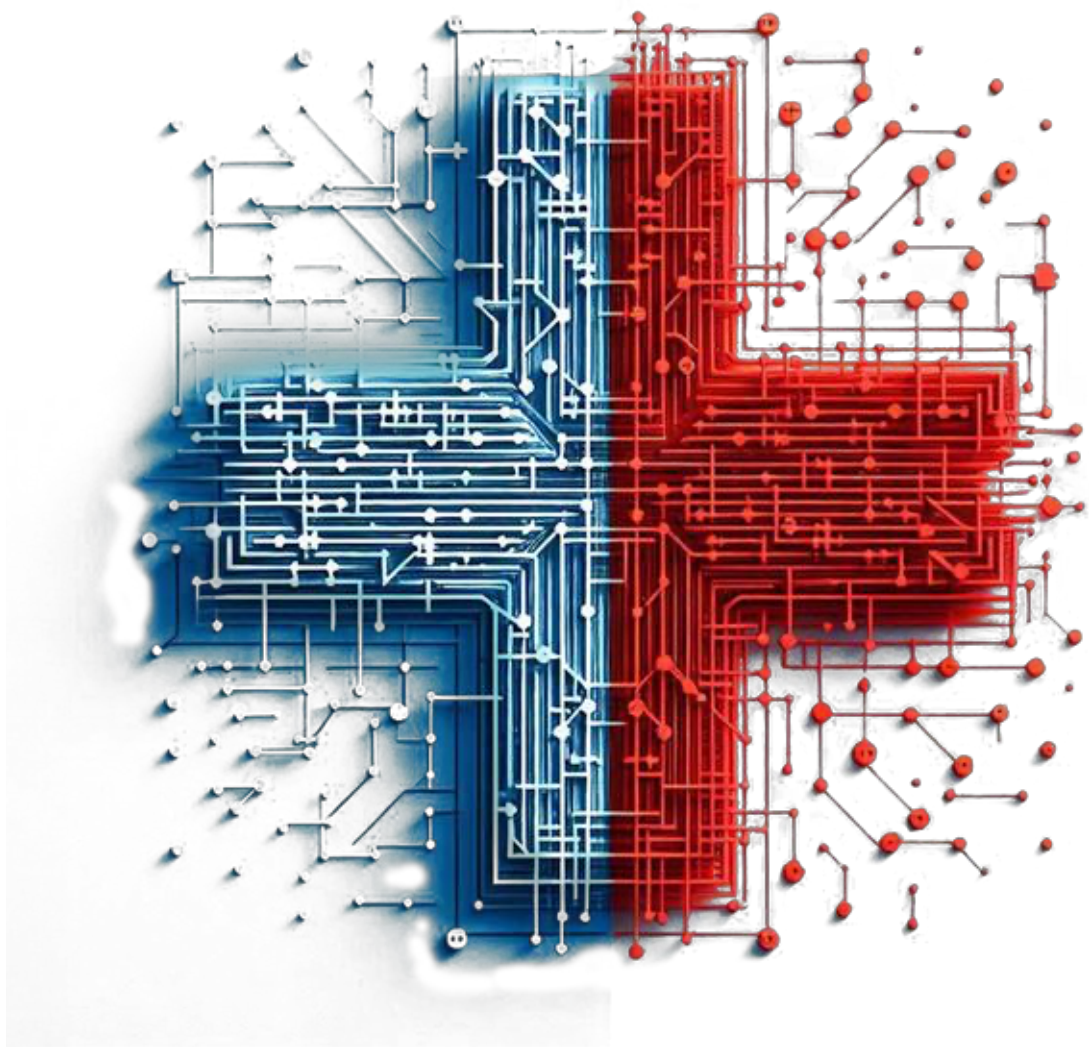
(e 2025)



ETH AI CENTER

Amplifying the human potential  
through **interdisciplinary**  
**AI research** and enabling the path  
from **science to society**





APERTVS

**EPFL**

**ETH** zürich



**CSCS**



# Profound Reshaping of Financial Industry

**No killer model available yet** to reliably take over financial related tasks effectively (while also compliant, secure & low risk)

**Challenges:**

Explainability/  
Correctness

Time series,  
tabular data

No available Data  
(non-public)

No available  
Benchmarks

# Profound Reshaping of Financial Industry

**No killer model available yet** to reliably take over financial related tasks effectively (while also compliant, secure & low risk)

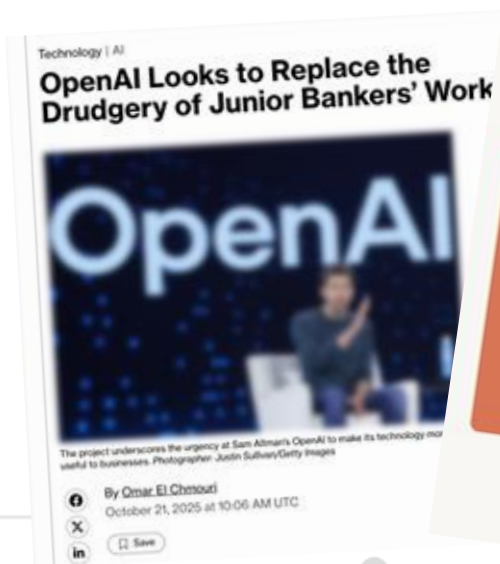
**Challenges:**

Explainability/  
Correctness

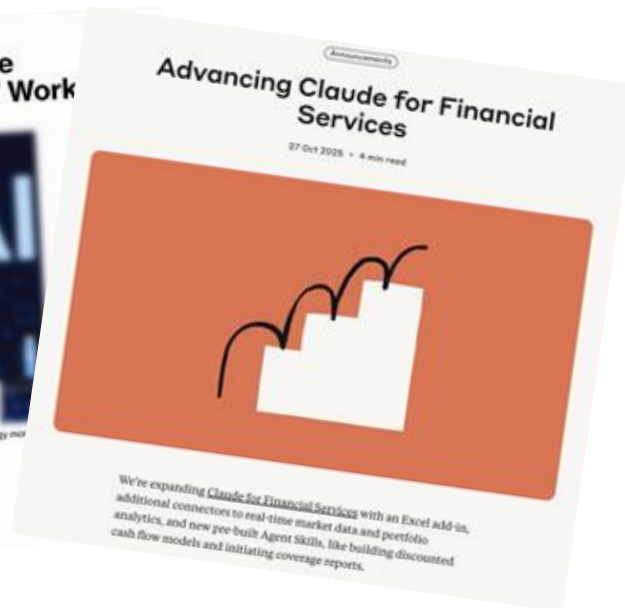
Time series,  
tabular data

No available Data  
(non-public)

No available  
Benchmarks



Source: [Bloomberg](#)



Source: [Anthropic](#)

# Profound Reshaping of Financial Industry

**No killer model available yet** to reliably take over financial related tasks effectively (while also compliant, secure & low risk)

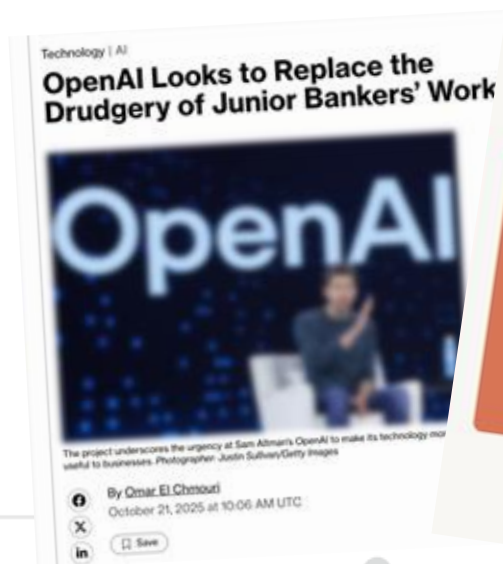
**Challenges:**

Explainability/  
Correctness

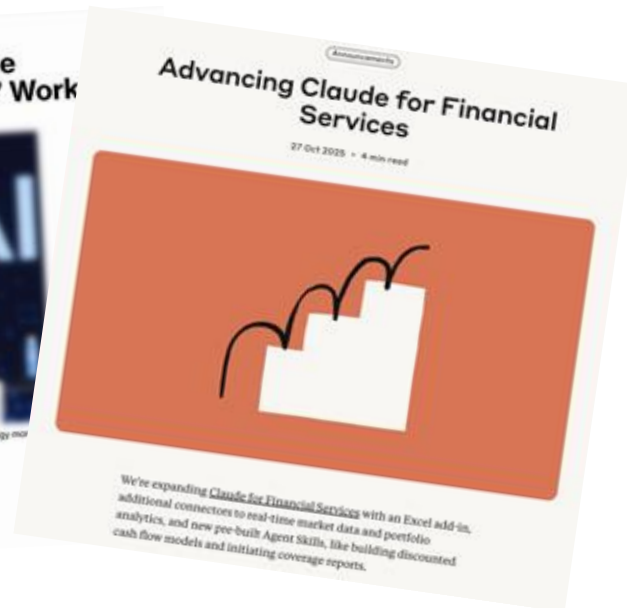
Time series,  
tabular data

No available Data  
(non-public)

No available  
Benchmarks



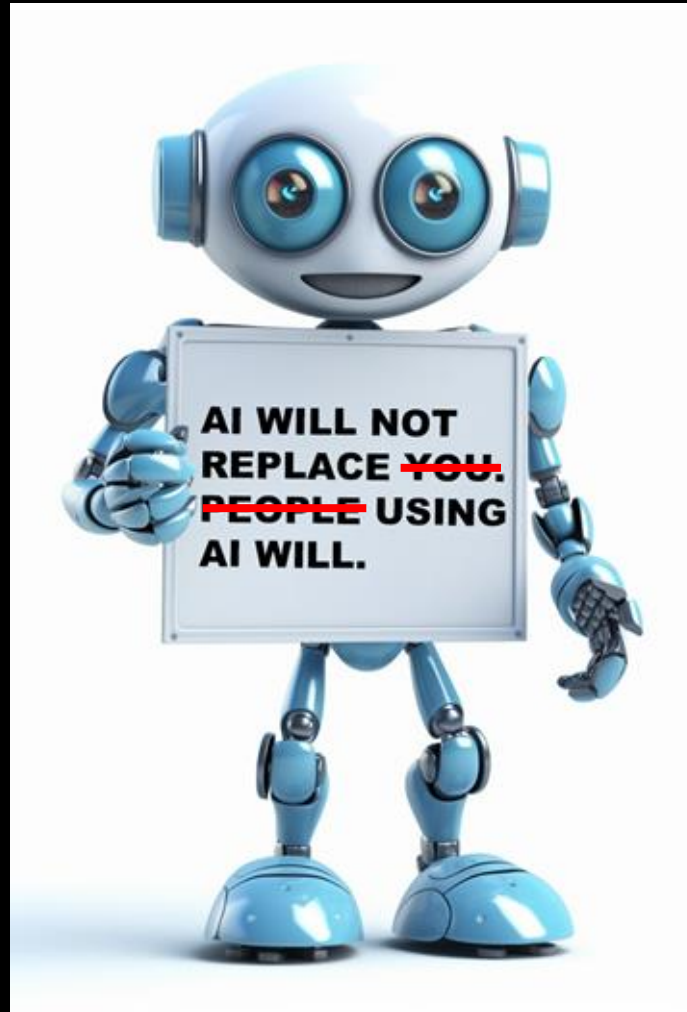
Source: [Bloomberg](#)

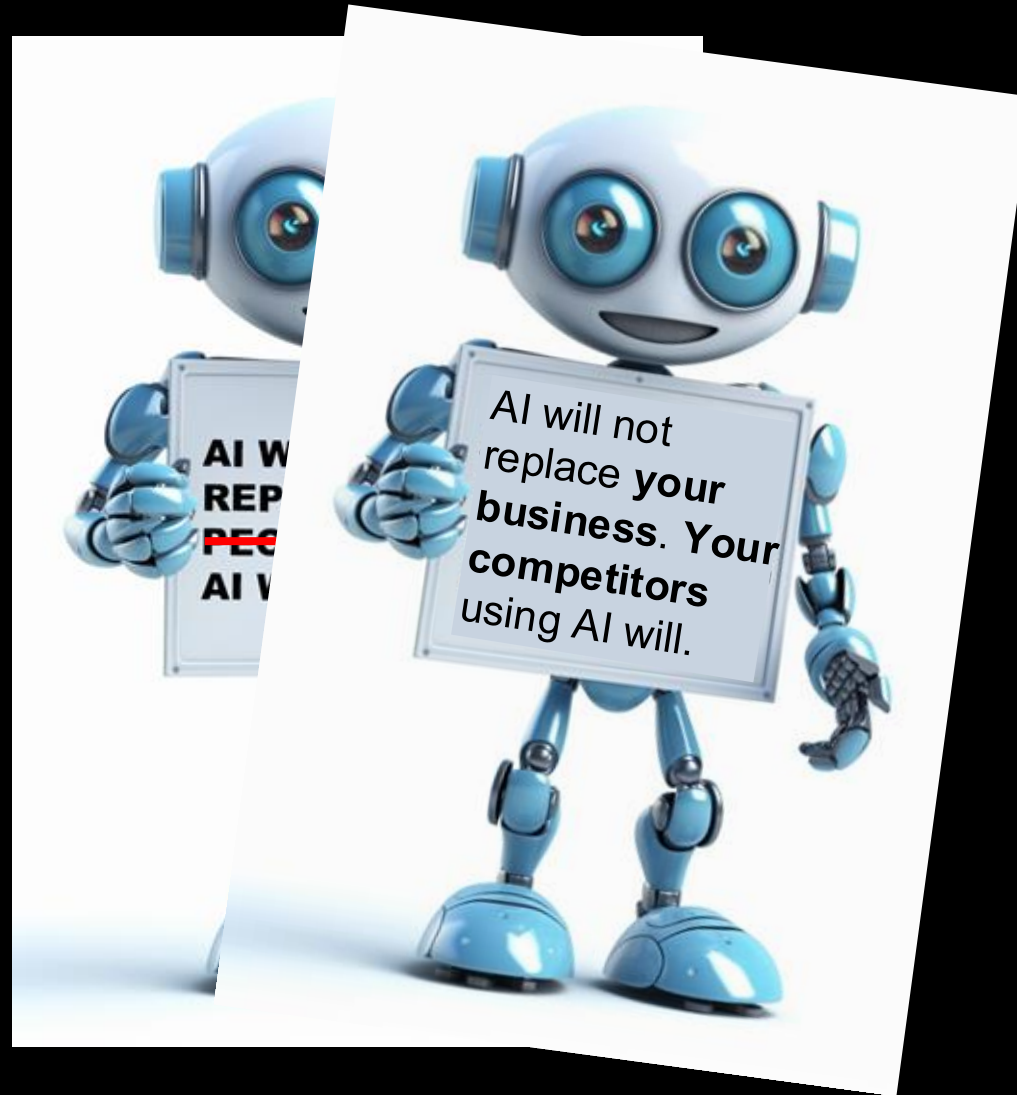



Source: [Anthropic](#)

**Options left to stay ahead:**

1. Adopt non-compliant, risky model
2. Build your own (CHF 100m+)
3. Join consortium to build a compliant, open Foundation Model for Finance (e.g. Apertus)







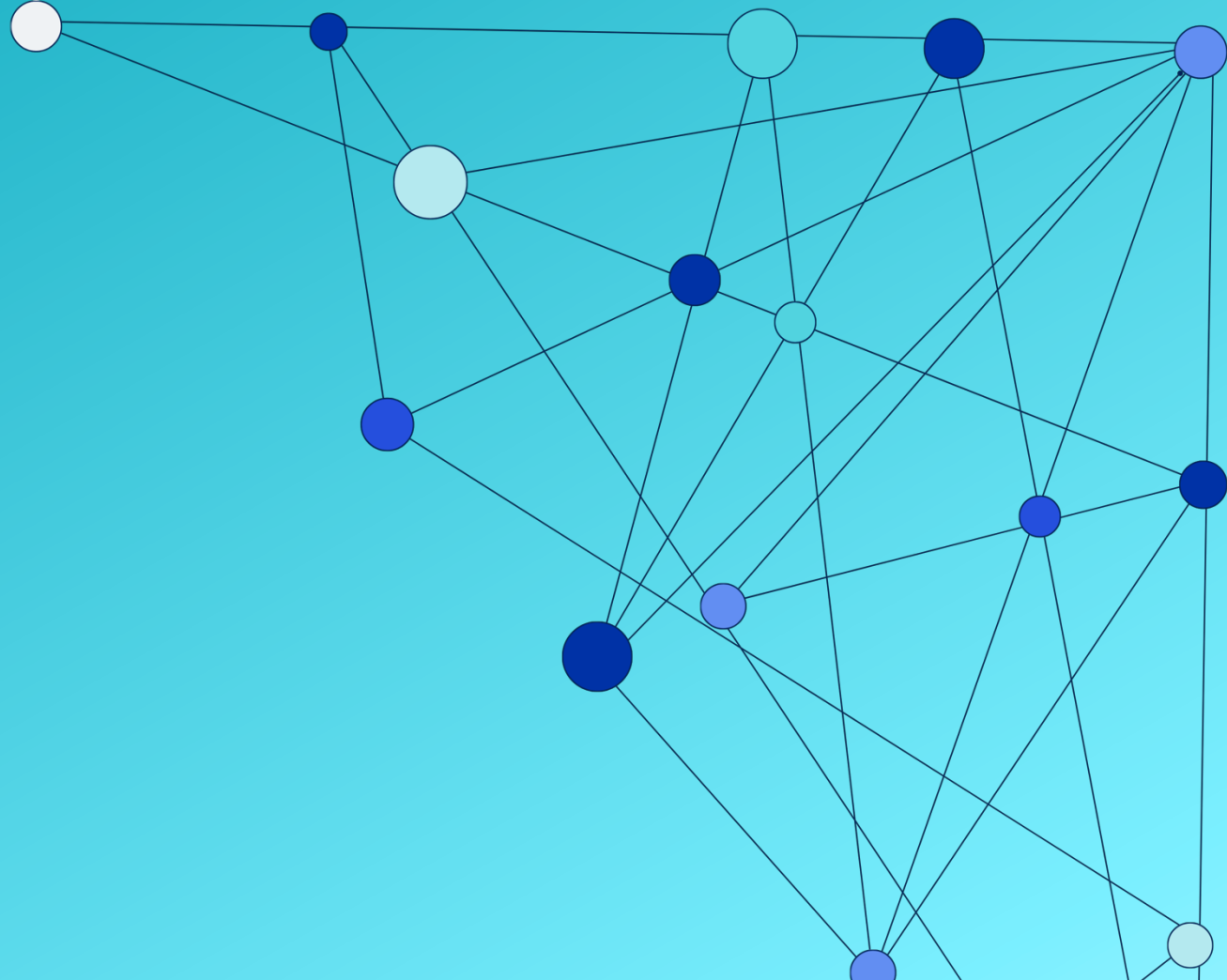
this doesn't  
roll at all!

AI Sovereignty  
Engineering at work

technically correct, but missing the point



 -lichen Dank



 **Zürich**  
Innovation Process Technology AG  
Löwenstrasse 28  
8001 Zürich  
**+41 41 727 25 25**  
**info@ipt.ch**

 **Bern**  
Innovation Process Technology AG  
Marktgasse 28  
3011 Bern  
**+41 31 311 12 23**  
**info@ipt.ch**