

Lokale LLMs für die Industrie

Michael Eichenberger
DINAcon 2025
18. November 2025

Übersicht



- Über uns
 - Digitale Souveränität
 - Wie erreichen wir die Digitale Souveränität?
 - Use Cases
 - Fazit
-

Über uns

Über die stepping stone AG und mich



- Cloud Provider
- Gegründet 2004
- Eigene Infrastruktur
- IaaS, CaaS, PaaS und SaaS
- Managed Services
- Michael Eichenberger
- Gründer und Mitinhaber
- CEO



Digitale Souveränität

Was ist digitale Souveränität?

Das «Netzwerk SDS» orientiert sich an der Definition von «digitaler Souveränität» des deutschen [Digital Gipfel 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.»

Wie erreichen wir die Digitale Souveränität?

Rechenzentren



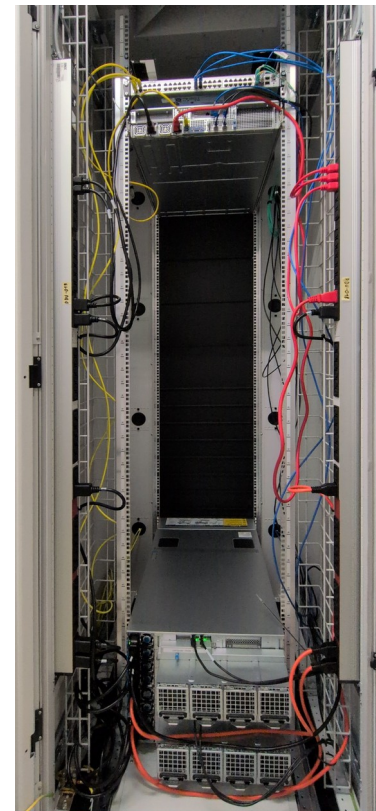
Rechenleistung



18. November 2025

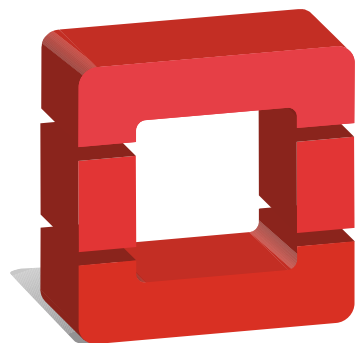


Lokale LLMs für die Industrie



9/27

Software



openstack



kubernetes

OpenTofu[📦]



puppet

AI ready Cloud Images



- Cloud Images
 - Mit den gängigsten Werkzeugen
 - Ollama
 - vLLM
 - Hugging Face CLI
 - Open WebUI
 - Und mit den neuesten Treibern
- Regelmässig aktualisiert
- Zusammenarbeit mit der ETHZ

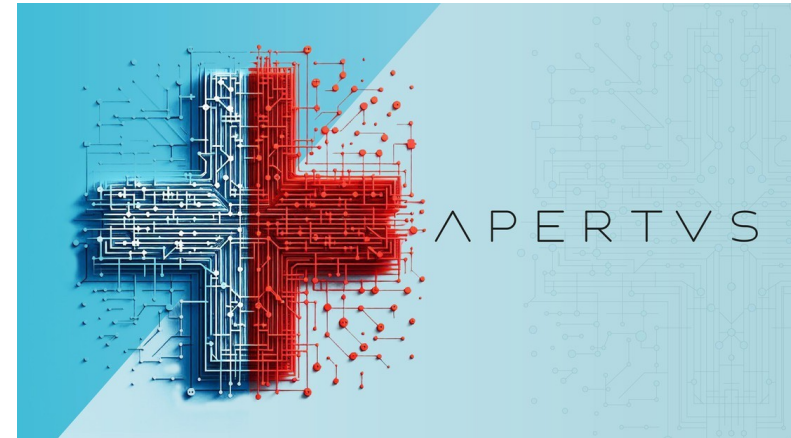


AI Cloud Images



Vorinstallierte Large Language Models:

- Apertus
- Mistral
- Llama
- Boltz-2
- BoltzGen
- OpenFold3



Unser Ansatz



Die zentralen Merkmale unseres Ansatzes sind

- Schweizer Ressourcen
- Isolierte LLM-Umgebungen
- Auditierbarkeit
- Datensparsamkeit
- Rechtliche Sicherheit

Use Cases

CADD Consulting GmbH

CADD Consulting bietet massgeschneiderte Rechendienstleistungen rund um wissenschaftliche Kernkompetenzen

- Structure Analysis and Structure Prediction
- Computer-Aided Drug Design
- Large-Scale Docking
- Cheminformatics Services
- Databases and Data Science
- Competitor Information

Arzneimittelforschung

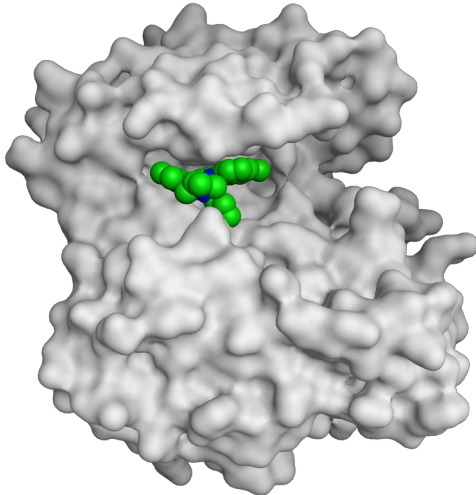
Anwendungen aus der Arzneimittelforschung

- Krebsmutationen modellieren und verstehen
- Proteinbindung und -faltung simulieren
- Der Heilige Gral: Gigantische Proteinfaltungen

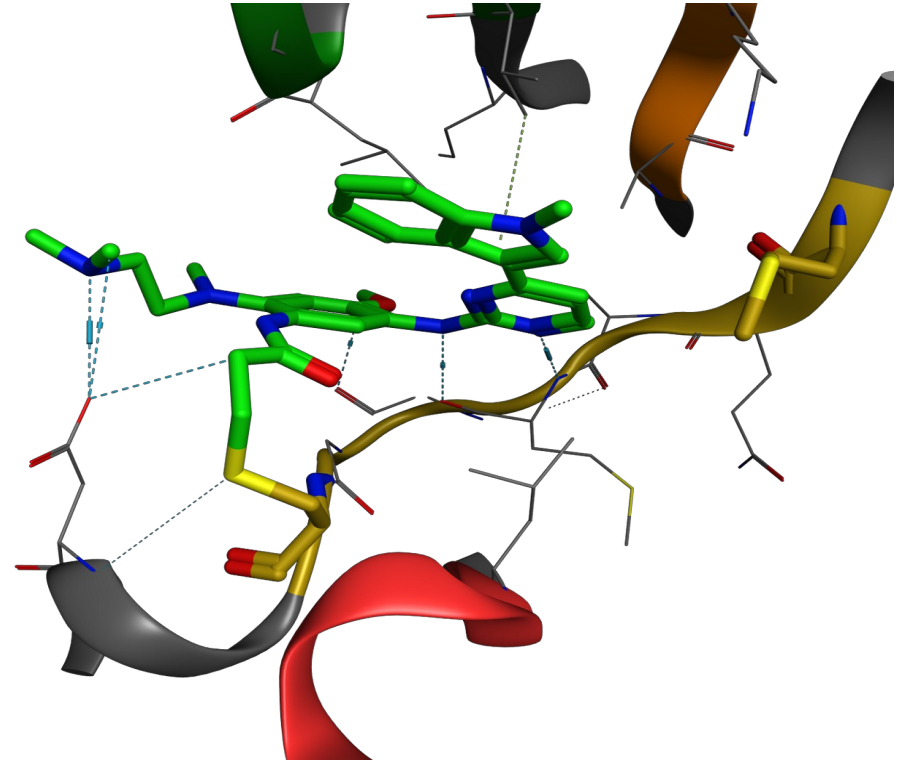
Krebsmutationen modellieren und verstehen

Cancer drug Tagrisso™ is used to treat EGFR-mutated lung cancer.

Using tools like Boltz-2 [1], new resistance mutations can be modeled in minutes to guide development of next-generation cancer drugs.



EGFR protein (white) with T790M mutation is inhibited by Tagrisso™ (green).



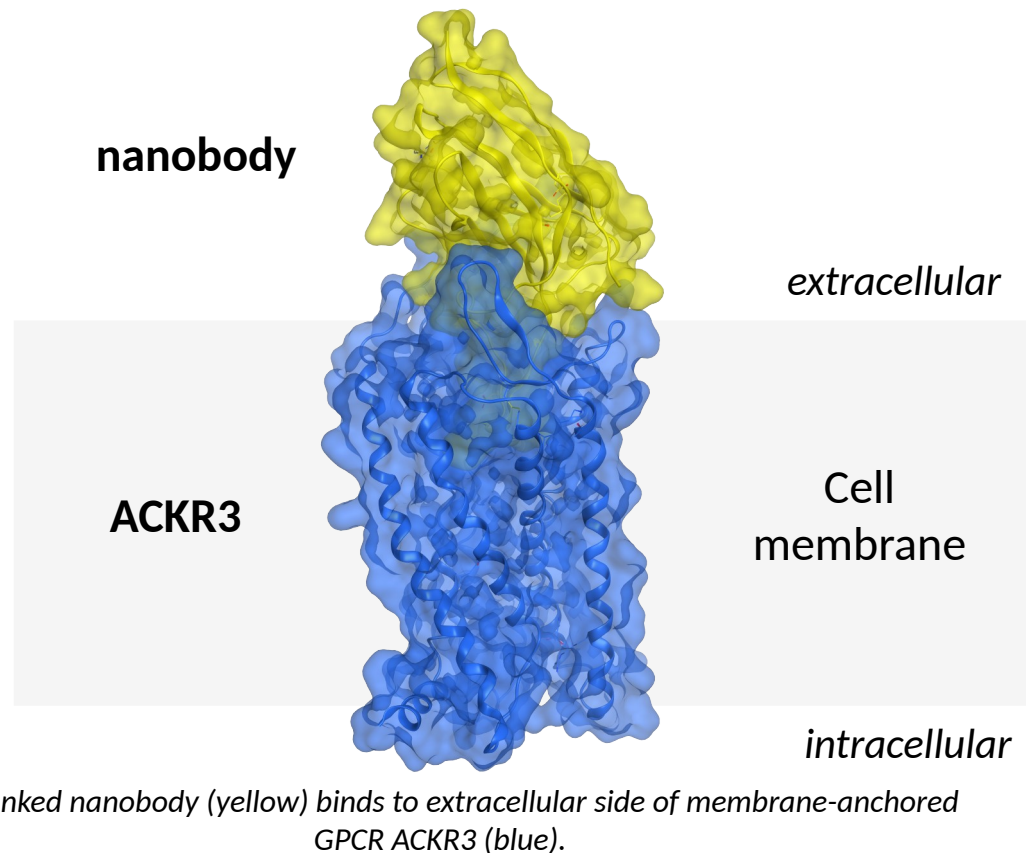
Close-up view shows how drug interacts with residues in binding site. EGFR mutants and covalent drugs can be calculated using Boltz-2^[1].

Proteinbindung und -faltung simulieren

Nanobody generation traditionally done by experiments.

Today, nanobodies can be generated in-silico which is much faster and more cost-efficient.

Example ACKR3, a novel prostate cancer target: Using 8 GPUs, new nanobody binders can be generated in 60 hours with BoltzGen [2].



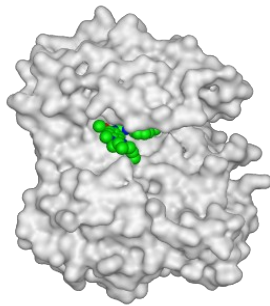
Der Heilige Gral: Gigantische Proteinfaltungen

Most proteins are less than 1500 amino acids in length.

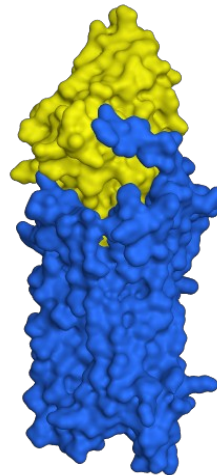
However, some proteins are huge. For example TRRAP: very long length of 3859 amino acids!

Calculation of big proteins requires a lot of VRAM.

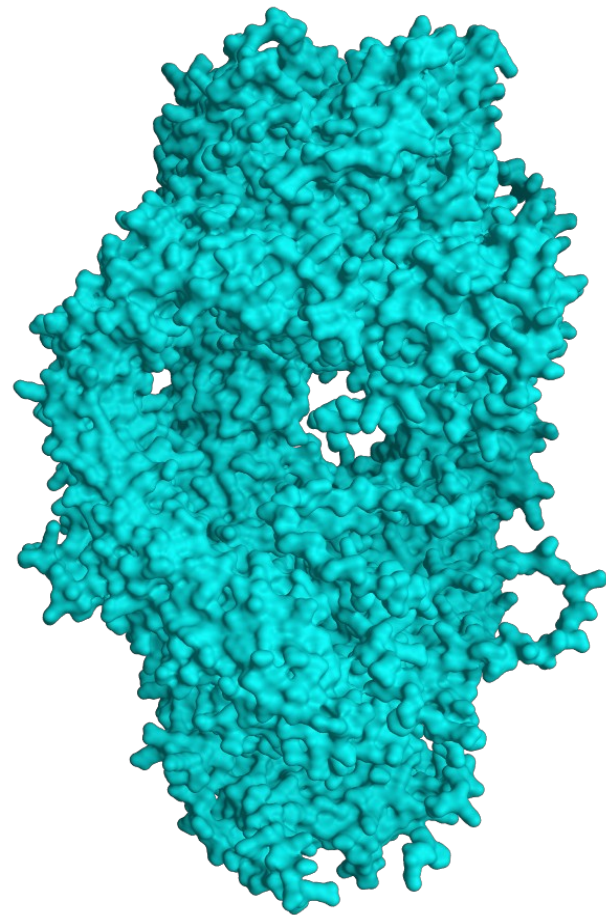
TRRAP calculated on 96 GB GPU with OpenFold-3 [3].



EGFR
323 aa



ACKR3
306 aa



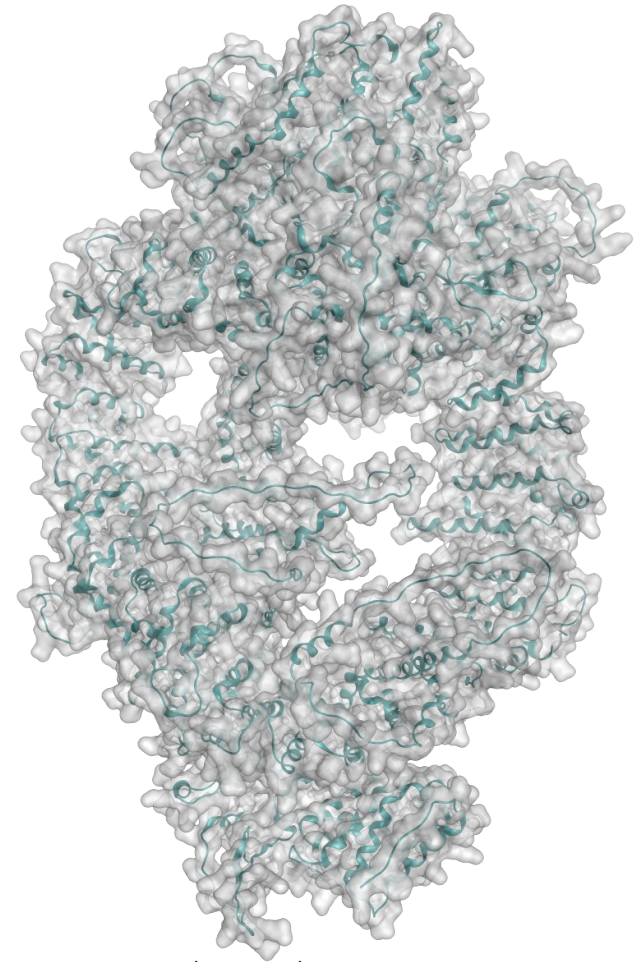
TRRAP
3859 amino acids

Gigantische Proteinstruktur durch massive GPU Setups

2023: höchstens 1700
Aminosäuren modelliert
(A100)

2025: komplettes Protein
(3859 Aminosäuren)
modelliert in 16 Minuten (!)

“I wish I had this before!”
Oliv Eidam, PhD
Director at CADD Consulting GmbH



TRRAP protein (3859aa) depicted with surface.
Calculated using OpenFold3-Preview.

Verwendete AI Modelle

- 1. Boltz-2:** <https://github.com/jwohlwend/boltz>
In silico screening for early-stage drug discovery from MIT.
- 2. BoltzGen:** <https://github.com/HannesStark/boltzgen>
Generative model for universal binder design from MIT.
Evaluated by leading academic labs and biotechnology companies.
- 3. OpenFold3:** <https://github.com/aqlaboratory/openfold-3>
Freely available AlphaFold3 reproduction developed by consortium.

Erkenntnisse aus der Verwendung von GPUs von der stepping stone AG

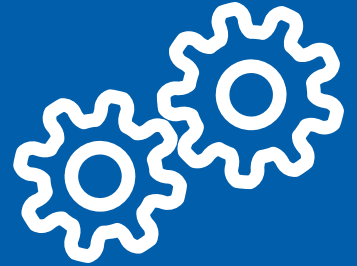
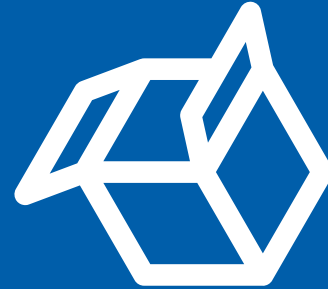
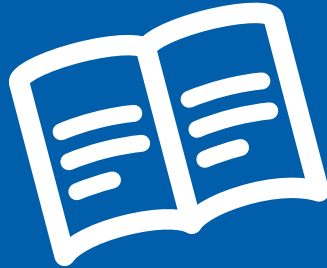
stepping stone AG bietet eine hervorragende Umgebung für anspruchsvolle Aufgaben im Bereich der computergestützten Wirkstoffforschung:

- Mithilfe paralleler GPU-Berechnungen konnte ich innerhalb weniger Tage neue Proteinbinder für ein Krebsziel generieren – statt wochenlanger Experimente.
- Dank der hohen VRAM-Kapazität der RTX 6000 Blackwell-Grafikkarten konnte ich die gigantische Proteinstruktur von TRRAP berechnen – etwas, von dem ich zuvor nur träumen konnte.

Fazit

Digitale Souveränität durch die Nutzung von lokalen LLMs bringt heute schon einen konkreten Mehrwert für die Industrie.

Fragen?



Links



- <https://www.stepping-stone.ch/>
- <https://www.stoney-backup.com/>
- <https://www.stoney-cloud.com/>
- <https://www.stoney-mail.com/>
- <https://www.stoney-meet.com/>
- <https://www.stoney-office.com/>
- <https://www.stoney-services.com/>
- <https://www.stoney-storage.com/>
- <https://www.stoney-wiki.com/>

stepping stone AG

Wasserwerksgasse 7
3011 Bern

Telefon: +41 31 332 53 63

www.stepping-stone.ch

info@stepping-stone.ch

CADD Consulting GmbH

Rigiweg 1
4800 Zofingen

www.cadd-consulting.com