
CRA AI Cloud

MAXIMÁLNÍ VÝKON A BEZPEČNOST S LOKÁLNÍM GPU CLOUDEM A OPENSOURCE
MODELY

Róbert Mitka, Head of AI Development, České Radiokomunikace

The other side of AI evolution



V1, February 2022



V2, April 2022



V3, July 2022



V4, November 2022



V5, March 2023



V6, December 2023



V6.1, July 2024

Global Arena

The race is on

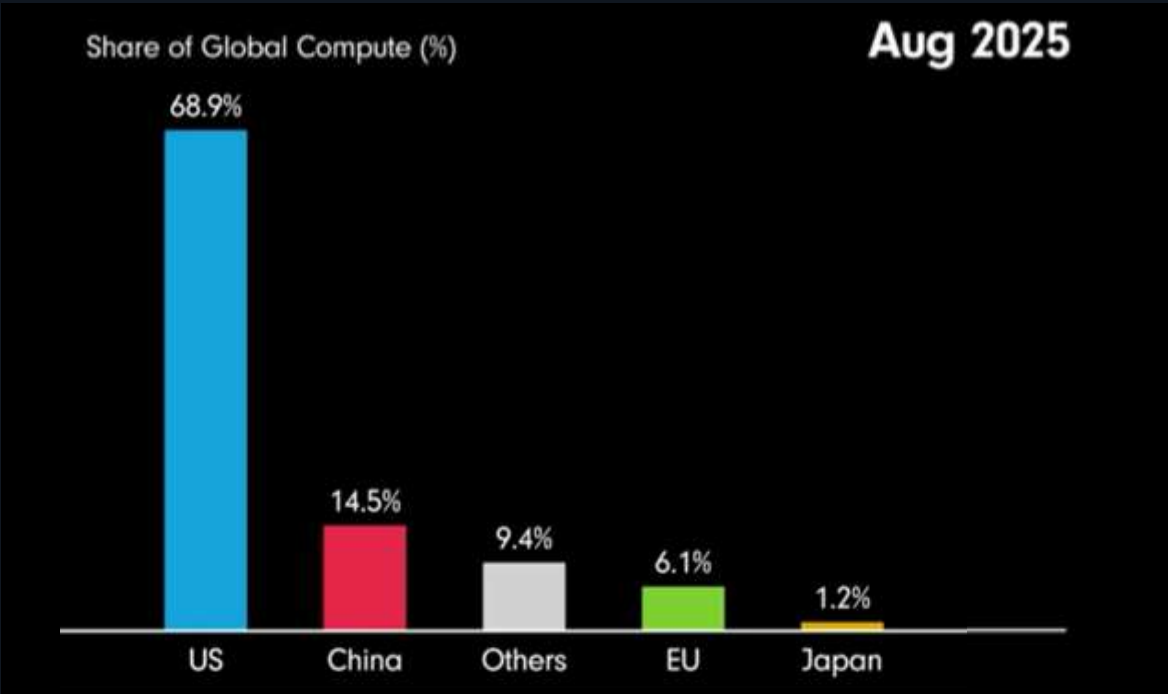
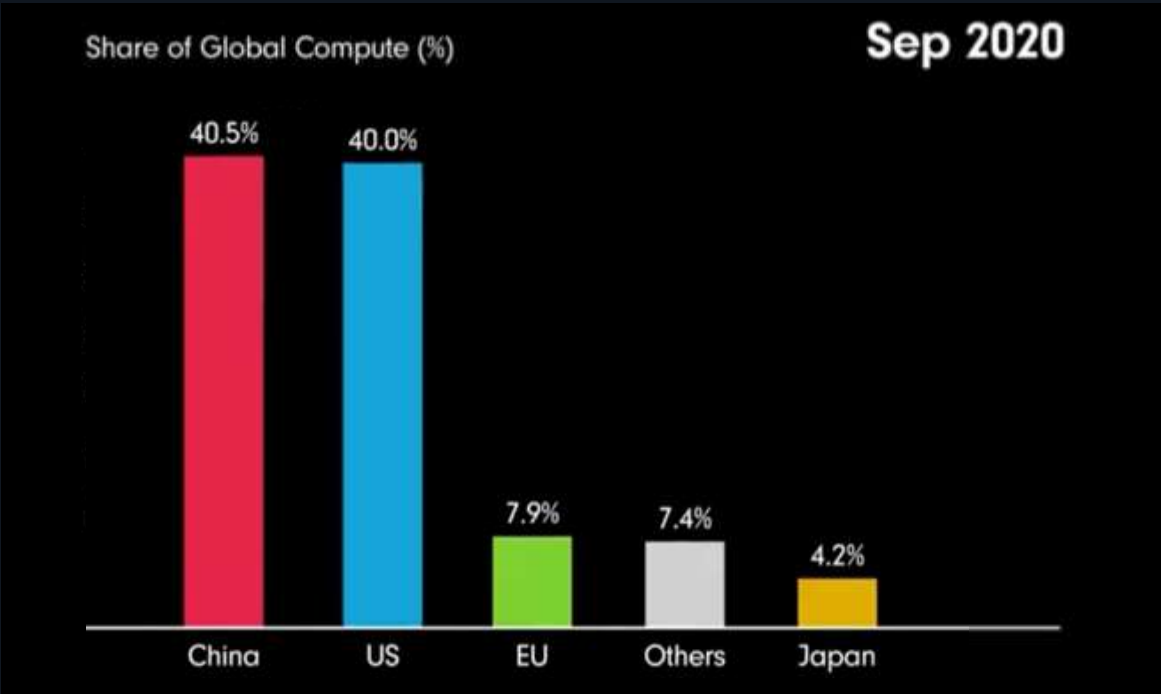


A photograph of an underwater data center. The scene is dimly lit with a blueish-green hue. In the foreground, a large, cylindrical metal structure, likely a server rack or part of the cooling system, is visible. It has several circular ports and a glowing blue ring. To the left, a vertical panel displays various digital readouts and lights. The background shows a long, narrow corridor of similar structures extending into the distance. The water surface is visible at the top, with light filtering through. The overall atmosphere is futuristic and industrial.

China

Underwater data center off the coast of Shanghai

AI Compute by Region



Number of newly funded AI companies by geographic area, 2013–24

Source: Quid, 2024 | Chart: 2025 AI Index report

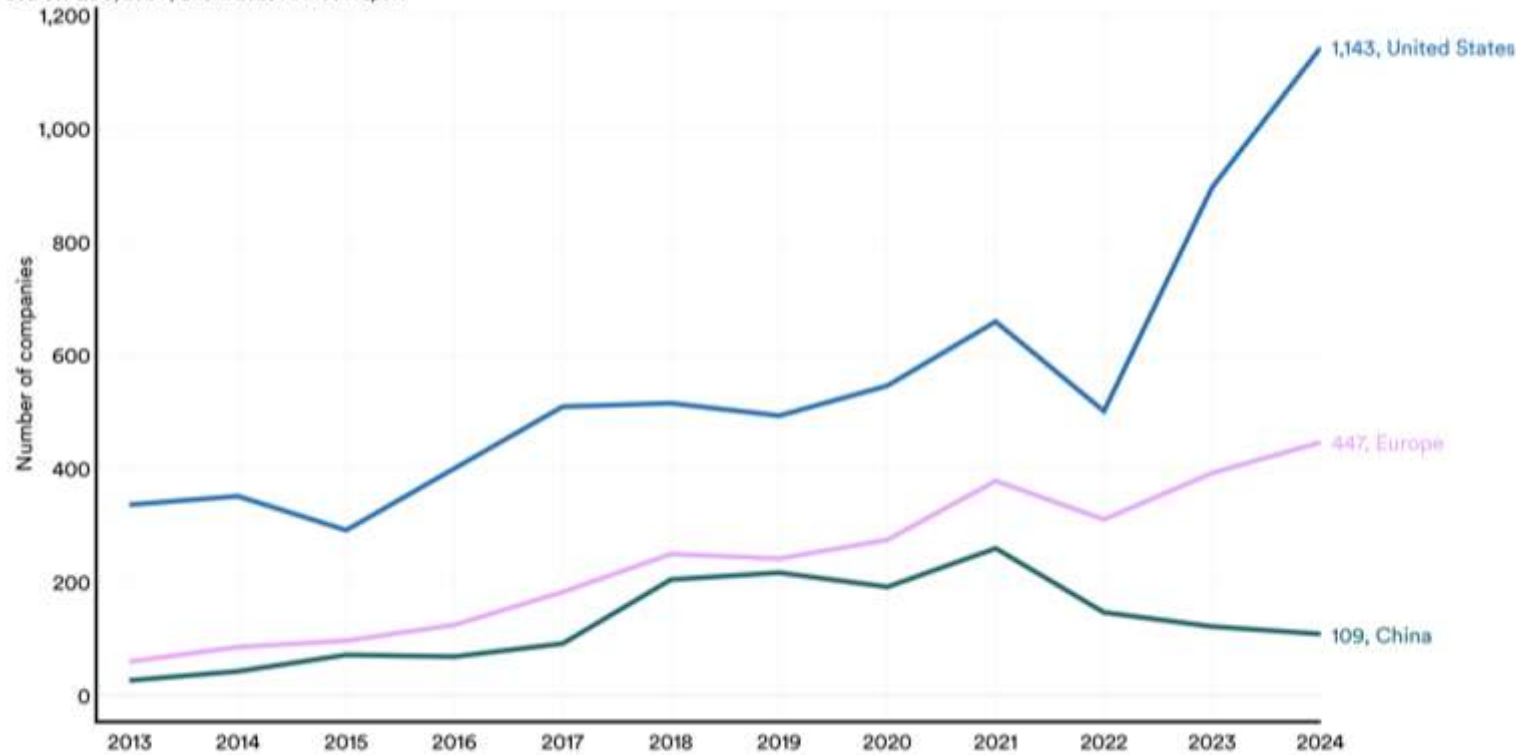


Figure 4.3.14

“When it comes to AI adoption, we don’t see a difference between European or U.S. companies. Whether they will be winners, yes or no, will be determined by who drives adoption faster.” MATTHIAS TAUBER, HEAD OF BCG

EUROPE, MIDDLE EAST, SOUTH AMERICA AND AFRICA

Czech Republic AI Landscape



Czech AI Global Startups



ROSSUM



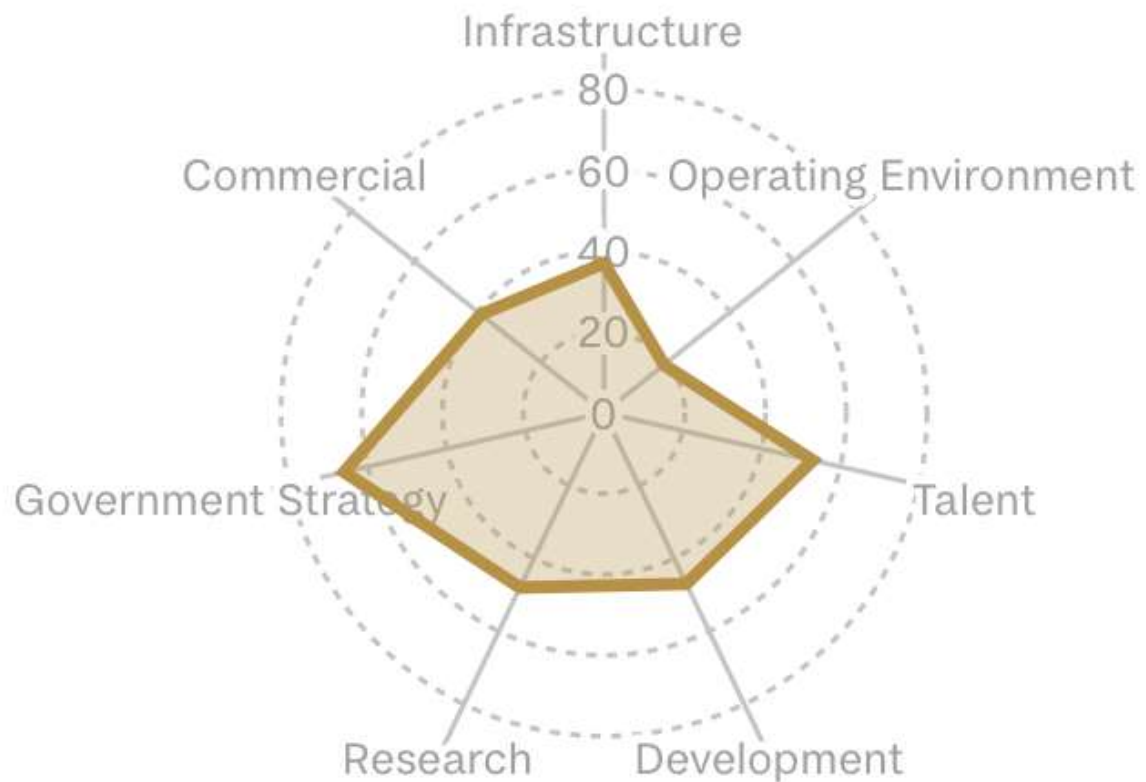
EQUILIBRE
TECHNOLOGIES

BottleCap AI

resistant_xai



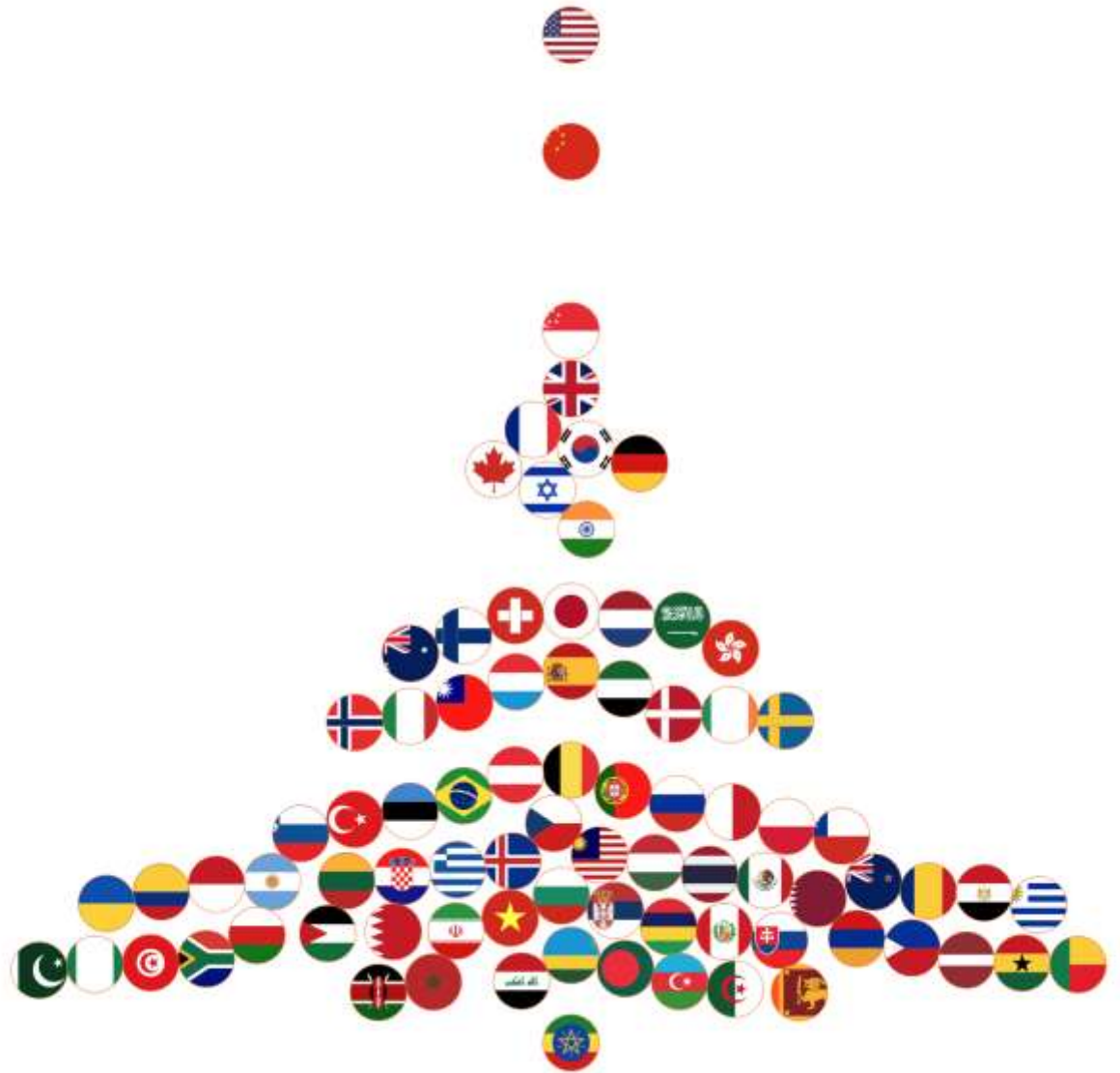
Czech AI Profile



Global AI Index

The Global AI Index aims to compare AI adoption in 83 countries

Implementation
Innovation
Investment



Global AI Index

The Global AI Index aims to compare AI adoption in 83 countries

Implementation
Innovation
Investment



Czech Republic New Opportunity



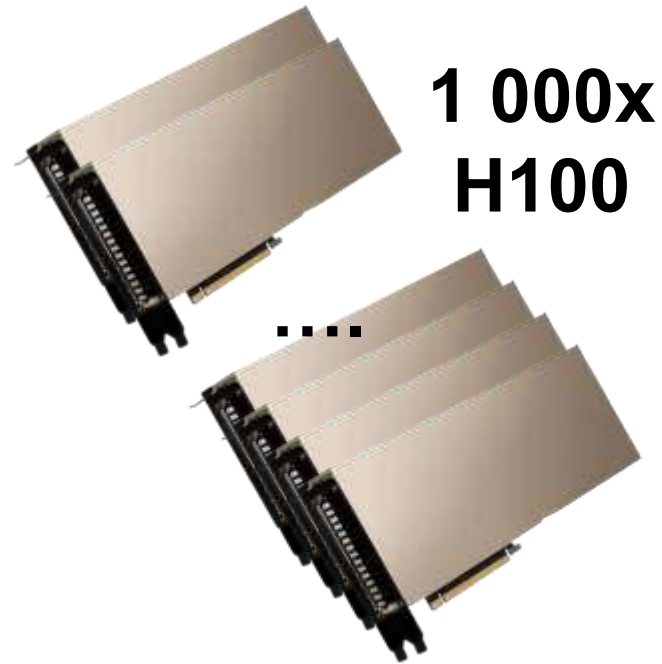
InvestAI

EU Initiative

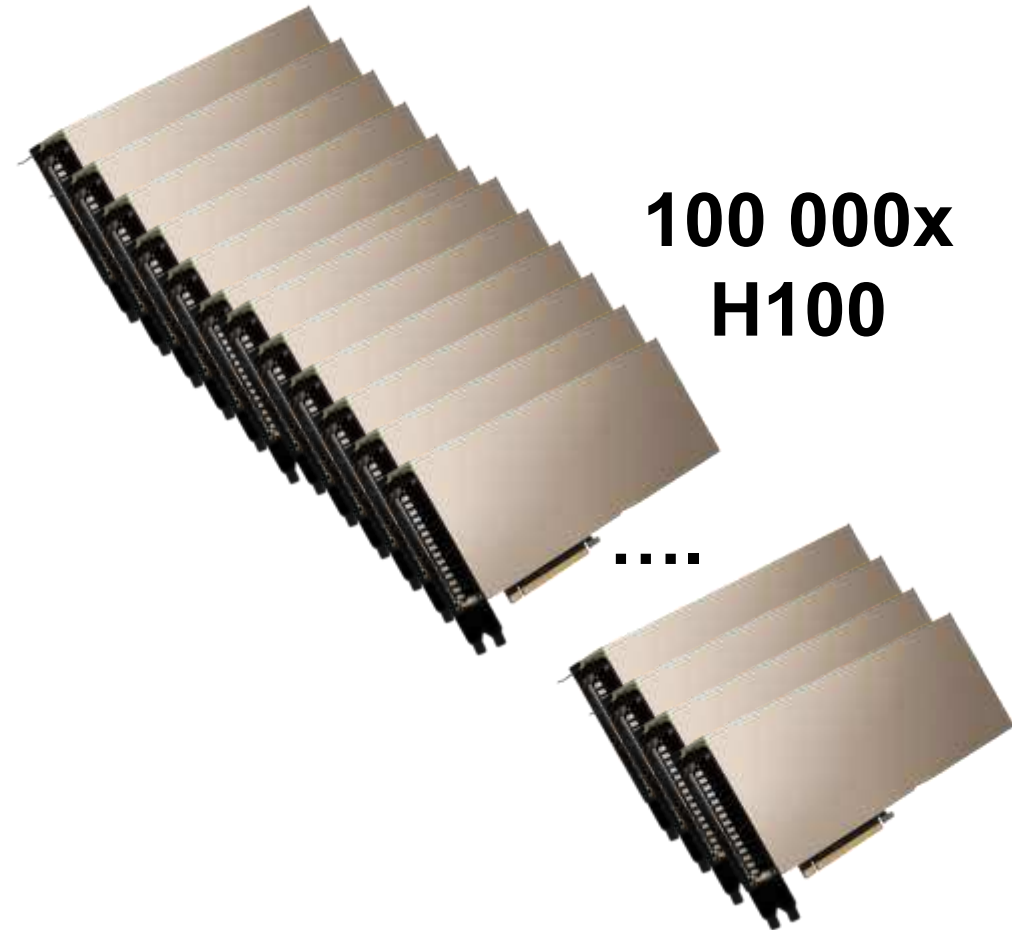
EUR 200 bil



AI Factory, AI GigaFactory

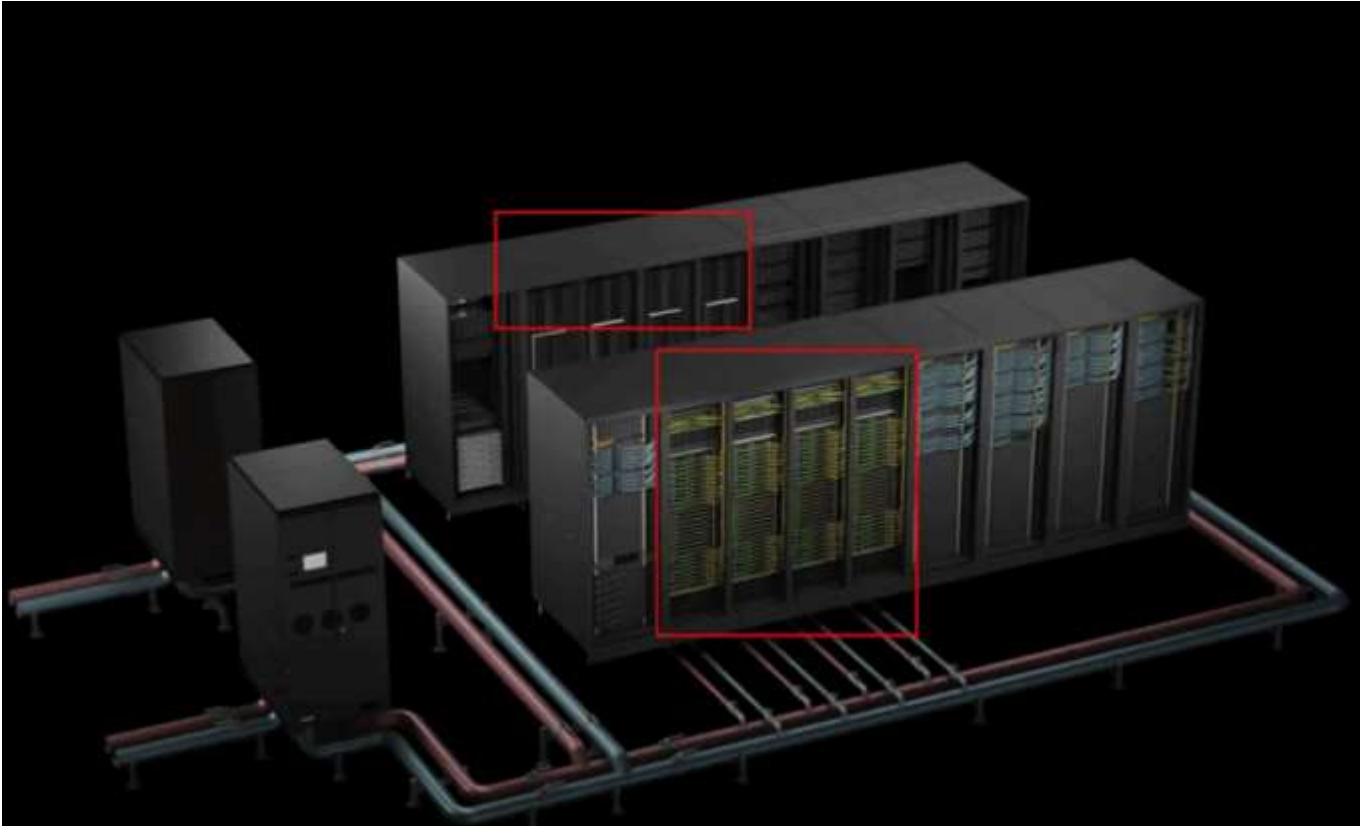


AI Factory









AI GigaFactory

Preliminary Characteristics



← KEY CHARACTERISTICS →

 Design GPU performance	~ 198 EFLOPS FP16
 Target GPU amount	~ 40 000 B200/B300 ~ 69 SuperPods
 Total Power	~80 MW
 Modularity	552 compute racks 1242 total racks
 Availability	99.9%
 Storage	~ 330 PB

Prague Gateway DC



Under Construction

Prague Gateway DC
construction site 11-08-2025

First capacity H2 2027



CRA AI CLOUD

DEDIKOVANÁ
GPU KARTA

DEDIKOVANÝ
AI SERVER

DEDIKOVANÝ
AI CLUSTER

Certified Security

Inside EU

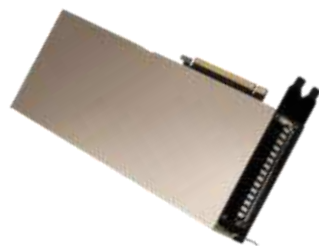
Lower Costs

No hidden fees

Czech Support

Low Latency

VYHRAZENÝ GPU
VÝKON VE
SDÍLENÉM AI
CLOUDU



VYHRAZENÝ
SERVER S VÍCE
GPU

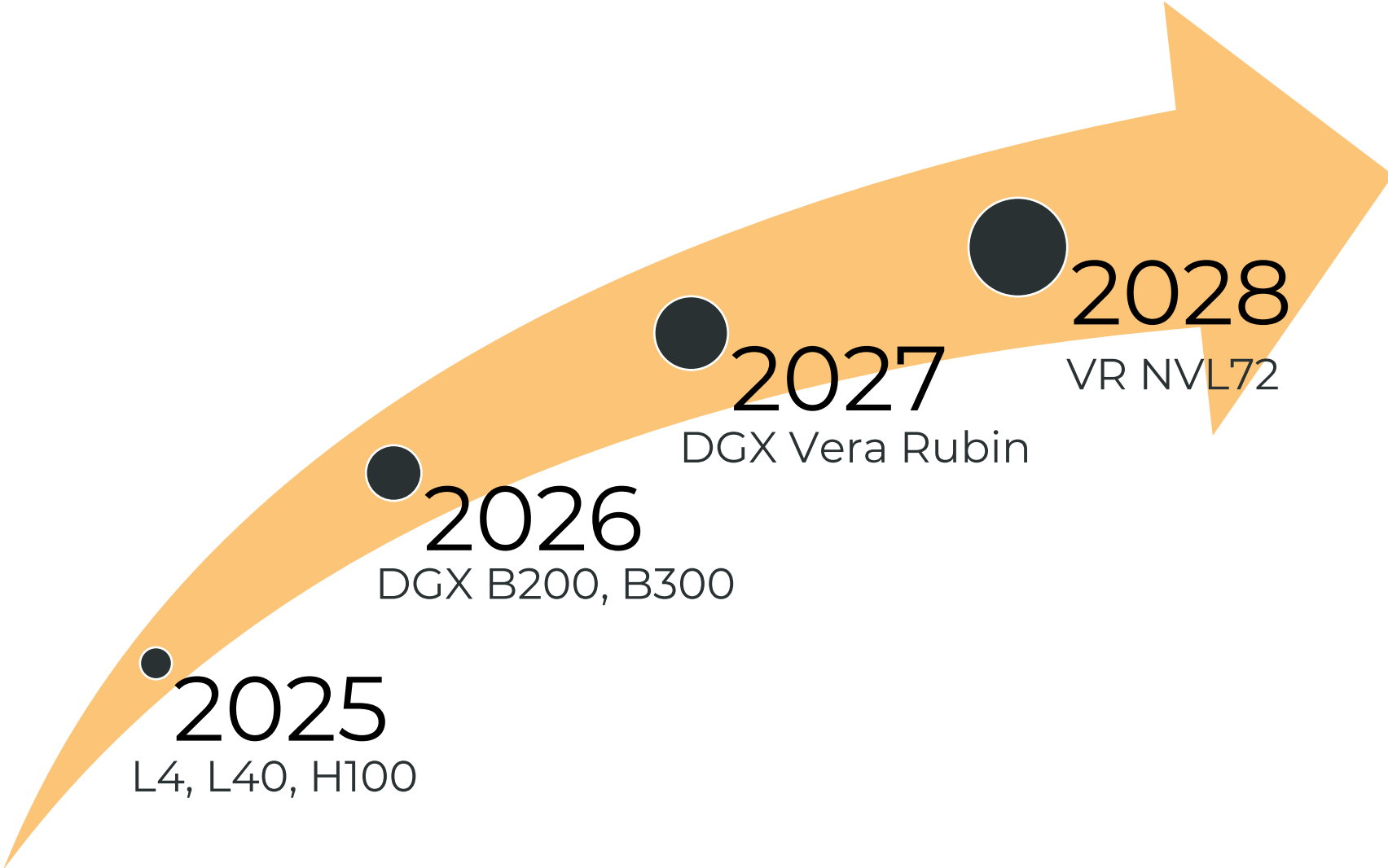


VYHRAZENÝ
CLUSTER GPU
SERVERŮ



CRA AI CLOUD

- Certified Security
- Inside EU
- Lower Costs
- No hidden fees
- Czech Support
- Low Latency



CRA AI CLOUD

Benefits

Hyperscaler

- Data jsou u Hyperscalerů, ale nemáte představu kde
- Flexibilní výkon
- Cena závislá od počtu tokenů, přenesených dat,..
- O administraci serverů se stará Hyperscaler
- Podpora v angličtině za příplatek

On-Premise

- Přísné požadavky na bezpečnost nad daty
- Pouze zakoupený HW
- Víte, kolik budete měsíčně platit za provoz AI
- Bez vlastních AI specialistů se neobejdu
- Mohu se obrátit na vlastní AI specialisty

CRA GPU Rental Services

- Data v ČR u poskytovatele kritické infrastruktury dodržující lokální regulace a zákony
- Flexibilní výkon/modely
- Fixní cena za model včetně konektivity a supportu
- Administraci serverů zajišťuje ČRA
- Lokální support v českém jazyce

Open Models vs Foundation Models



[Nvidia CES 2026](#)

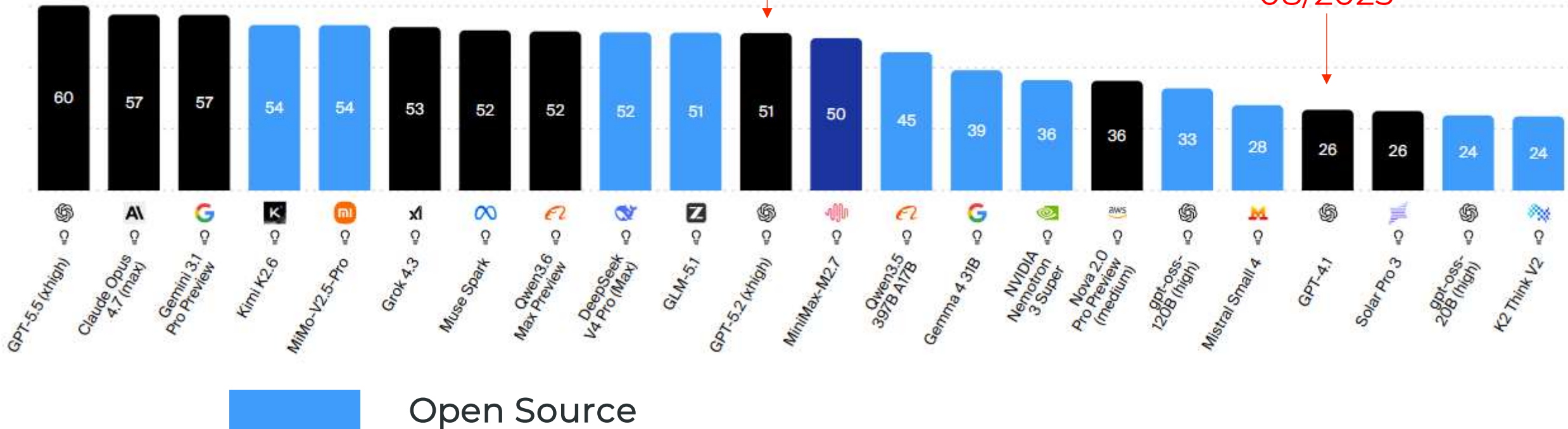
Open source vs. Foundation LLM

Artificial Analysis

04/2025

12/2025

08/2025



Artificial Analysis

Model Rental Services

- LLM modely (GPT OSS, Llama, Gemma, Qwen, Mistral, ...)
- Speech2Text modely (Whisper, Canary, ...)
- Multimodální modely (Qwen 3 VL, Deepseek OCR, ...)
- Modely pro generování obrázků a videí (SD, WAN, ...)
- Embedding modely (E5, dist-mpnet, Qwen Embed, ...)
- Další modely z Hugging Face

Certified Security

Inside EU

Lower Costs

No hidden fees

Czech Support

Low Latency

**Zabezpečený přístup k AI modelům s
garantovanou ochranou dat zákazníků**

Virtix Self-service

The screenshot displays the Virtix self-service portal interface. A modal window titled 'CANARY SPEECH2TEXT' is open, showing configuration options for a new model. The configuration includes:

- NÁZEV: Canary Speech2Text
- MODEL: CANARY
- CPU: LAOS
- GPU RAM: 48 GB
- CPU: 16 vCPU cores
- RAM: 96 GB
- SSD: 200 GB
- PARAMETRY: (empty field)
- POPS: (empty field)

At the bottom of the modal, the price is listed as 'CENA ZA MĚSÍC BEZ DPH' with a value of '+24 318 Kč'. A prominent orange button labeled 'ULOŽIT' (Save) is visible at the bottom center of the modal.

In the background, the main dashboard shows a 'PŘIDAT NOVÝ MODEL' button and a list of existing models, including 'SPEECH2TEXT' with specifications: 1x LAOS, 16 vCPU cores, 96 GB RAM, and 200 GB Local SSD.

Fakturace po dnech!

Q&A

Robert Mitka

r.mitka@cra.cz



Vyzkoušejte si naše GPUaaS a
MODELaaS zdarma na 5 dní!