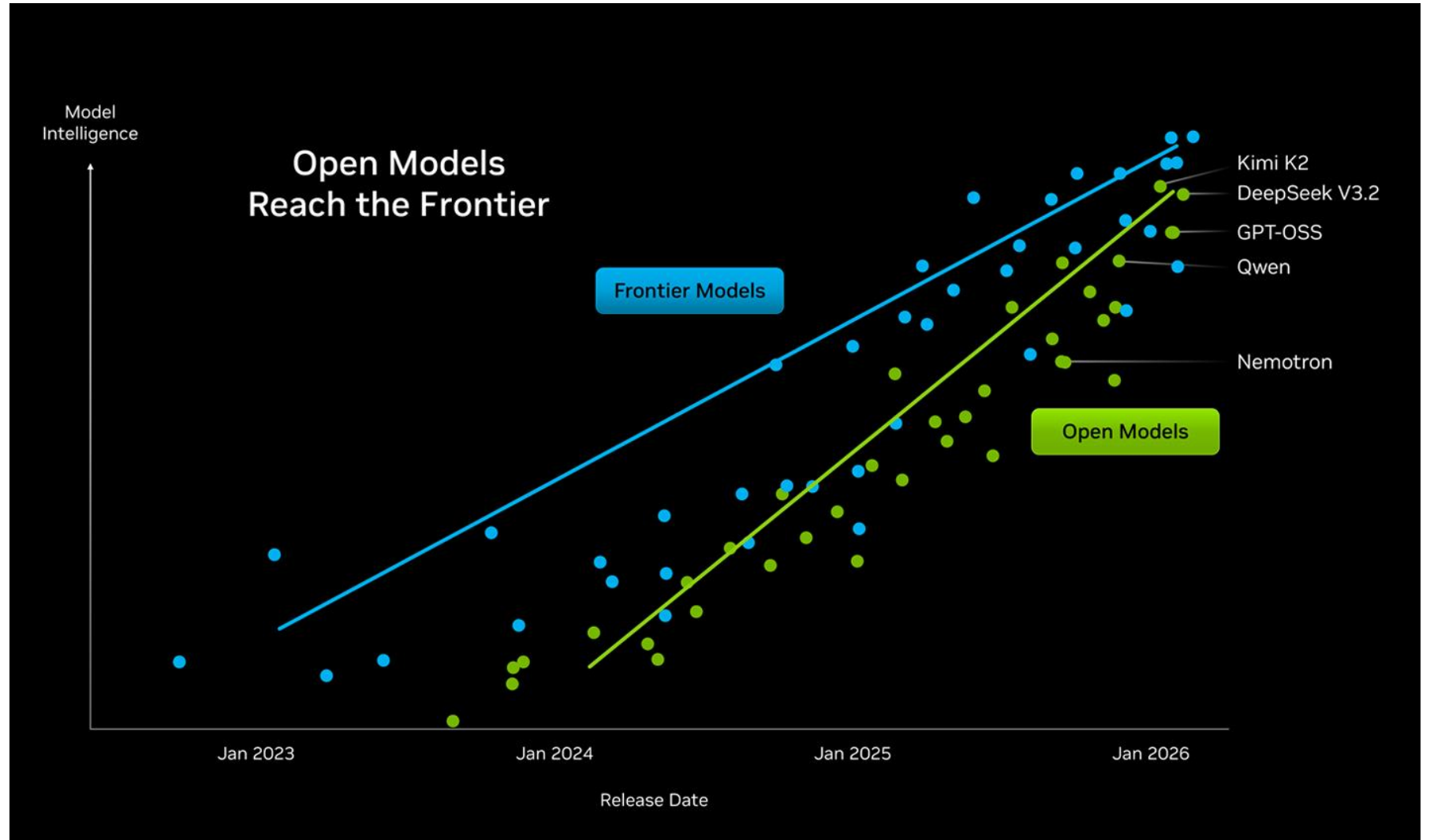

AI Confidence Workshop

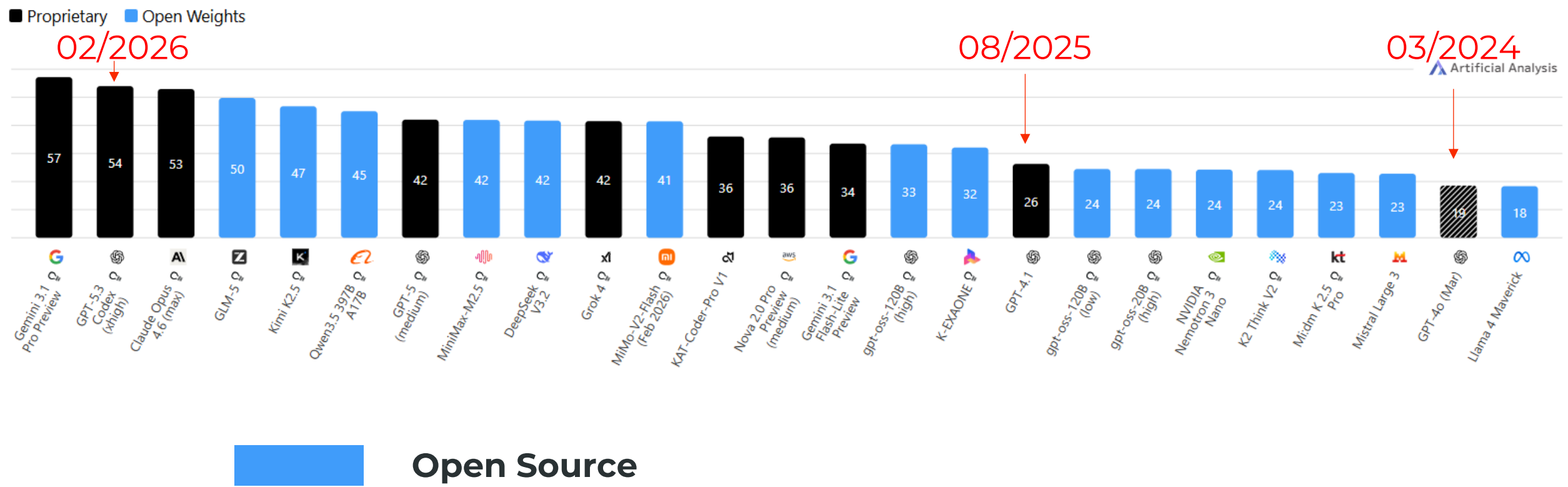
Robert Mitka, Head of AI Development, České Radiokomunikace

Open Models vs Foundation Models



[Nvidia CES 2026](#)

Open source vs. Foundation LLM



Artificial Analysis

Hugging Face

The screenshot shows the Hugging Face website interface. At the top, there is a navigation bar with the Hugging Face logo, a search bar, and links for Models, Datasets, Spaces, and Buckets. Below the navigation bar, there are tabs for Main, Tasks, Libraries, Languages, Licenses, and Other. The 'Tasks' tab is selected, and a search bar for filtering tasks by name is visible. The main content area is divided into two columns. The left column lists various task categories under 'Multimodal' and 'Computer Vision'. The right column displays a list of models, including 'Jackrong/Qwen3.5-27B-Claude-4.6-Opus-Reasoning-Dist...', 'HauhauCS/Qwen3.5-35B-A3B-Uncensored-HauhauCS-Aggres...', 'nvidia/Nemotron-Cascade-2-30B-A3B', 'GAIR/daVinci-MagiHuman', 'Jackrong/Qwen3.5-27B-Claude-4.6-Opus-Reasoning-Dist...', and 'Tesslate/OmniCoder-9B'. Each model entry includes its name, task type, size, update date, download count, and heart count.

huggingface.co/models

Hugging Face Search models, datasets, users...

Models Datasets Spaces Buckets

Main Tasks Libraries Languages Licenses Other

Filter Tasks by name

Multimodal

- Audio-Text-to-Text
- Image-Text-to-Text
- Image-Text-to-Image
- Image-Text-to-Video
- Visual Question Answering
- Document Question Answering
- Video-Text-to-Text
- Visual Document Retrieval
- Any-to-Any

Computer Vision

- Depth Estimation
- Image Classification
- Object Detection
- Image Segmentation
- Text-to-Image
- Image-to-Text
- Image-to-Image
- Image-to-Video
- Unconditional Image Generation
- Video Classification
- Text-to-Video
- Zero-Shot Image Classification
- Mask Generation

Models 2,744,216 Filter by name

- Jackrong/Qwen3.5-27B-Claude-4.6-Opus-Reasoning-Dist...
Image-Text-to-Text · 28B · Updated 4 days ago · 253k · 1.5k
- HauhauCS/Qwen3.5-35B-A3B-Uncensored-HauhauCS-Aggres...
Image-Text-to-Text · 35B · Updated 18 days ago · 479k · 1.03k
- nvidia/Nemotron-Cascade-2-30B-A3B
Text Generation · 32B · Updated 4 days ago · 69.6k · 354
- GAIR/daVinci-MagiHuman
Image-to-Video · Updated 3 days ago · 418 · 220
- Jackrong/Qwen3.5-27B-Claude-4.6-Opus-Reasoning-Dist...
Image-Text-to-Text · 27B · Updated 3 days ago · 85.1k · 210
- Tesslate/OmniCoder-9B
Text Generation · Updated 15 days ago · 26.5k · 497

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

- Přístup k modelům přes API kompatibilní s OpenAI
- Možnost omezení přístupu na vybrané IP adresy nebo rozsahy
- Náhodně generované adresy pro přístup k API

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

Virtix Self-service

The screenshot displays the Virtix self-service portal interface. The top navigation bar includes the 'virtix' logo, a language selector (CZ/EN), and company information for 'Cloud4com s.r.o.' (IČ 21050309) and 'CRA DEMO'. The left sidebar contains menu items: 'Správa vPDC', 'GPUaaS & MODELaaS', 'Správa IP', 'Ostatní služby, nástroje, SW', 'Uživatelé', 'Zákazníci', 'Nabídky/Poptávky', and 'Fakturace'. The main content area is titled 'GPUaaS | MODELaaS' and features a '+ PŘIDAT NOVÝ MODELaaS' button. A modal window titled 'CANARY SPEECH2TEXT' is open, showing the following configuration details:

- NÁZEV: Canary Speech2Text
- MODEL: CANARY
- GPU: L40S
- GPU RAM: 48 GB
- CPU: 16 vCPU cores
- RAM: 96 GB
- SSD: 200 GB
- PARAMETRY: (empty field)
- POPIS: (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 'ULOŽIT' button is located at the bottom center of the modal. On the right side of the main interface, a preview of the 'SPEECH2TEXT' model is visible, listing its specifications: '1x L40S', '16 vCPU cores', '96 GB RAM', and '200 GB Local SSD'.

Fakturace po dnech!

Large Language Models

- Generování textu
- Odpovídání na otázky
- Shrnování dlouhých dokumentů
- Extrakce informací z textu např. jmen, částek, datumů
- Oprava gramatiky a stylu
- Překlad textu
- Analýza a zpracování dokumentů
- Chatboti

Large Language Model Demo

https://colab.research.google.com/drive/17C_JzHrgGgKf34h8nIvYYUrQlvDviiI6?usp=sharing

Multimodal Language Models

- Porozumění obrázkům a fotografiím
- Popis obsahu obrázku slovy
- Odpovídání na otázky k obrázku
- Čtení textu z obrázků a dokumentů
- Analýza grafů, tabulek a diagramů
- Pomoc při zpracování faktur, formulářů

Standard OCR vs. Multimodal Language Models

Document 1



Document 2



Document 3



Vizuální modely na rozdíl od starého dobrého OCR umí zachytit pozici textu na obrázku a pochopit kontext (dotazník)

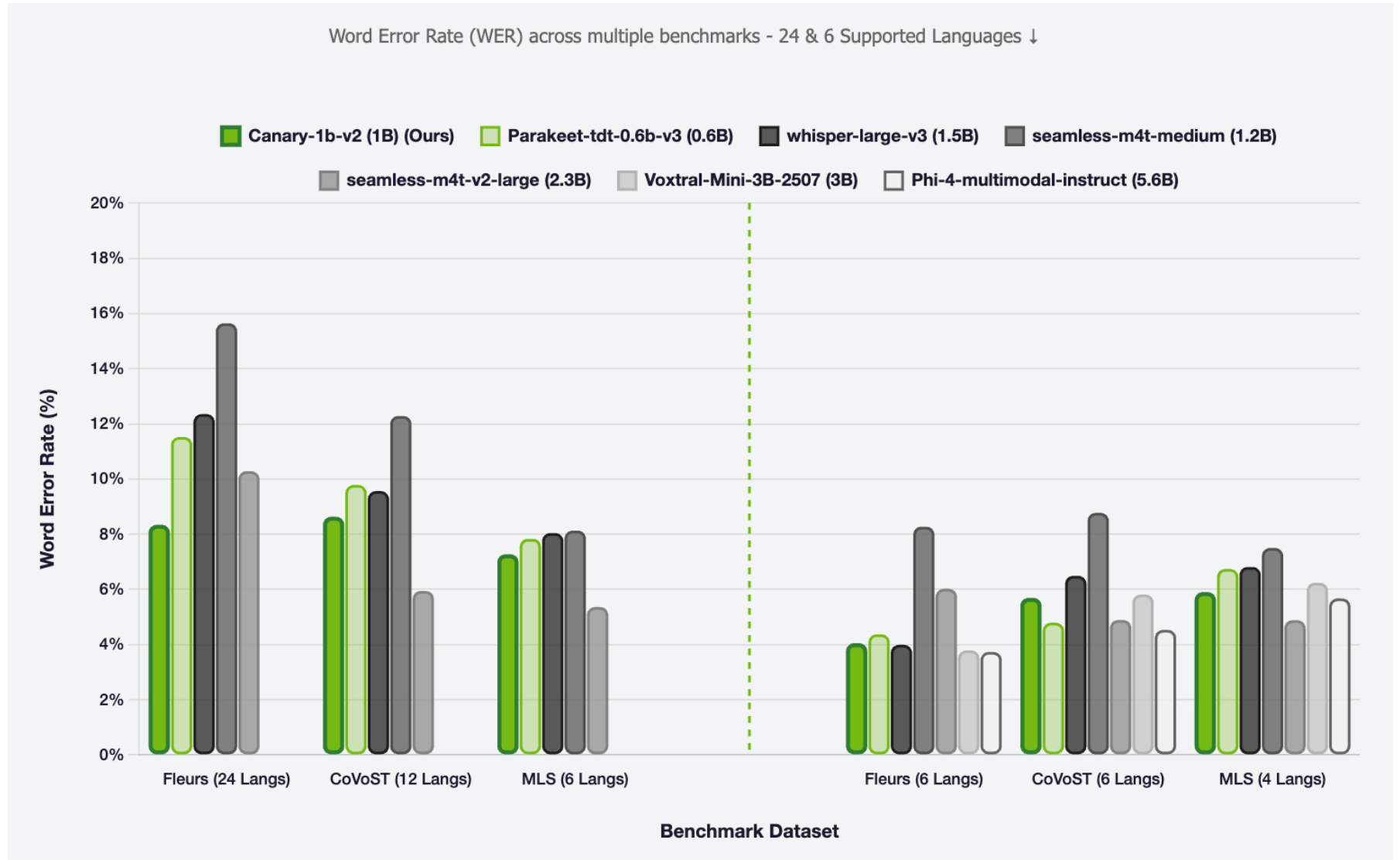
Multimodal Language Model Demo

https://colab.research.google.com/drive/12VtuwybTYzxk_pWDJnrkPQmMp_TMsVWc?usp=sharing

Speech to Text Models

- Přepis řeči do textu
- Překlad do jiného jazyka
- Generování titulků/timestampů
- Rozpoznávání mluvčí (diarizace)

Model Canary 1b v2



Speech to Text Model Demo

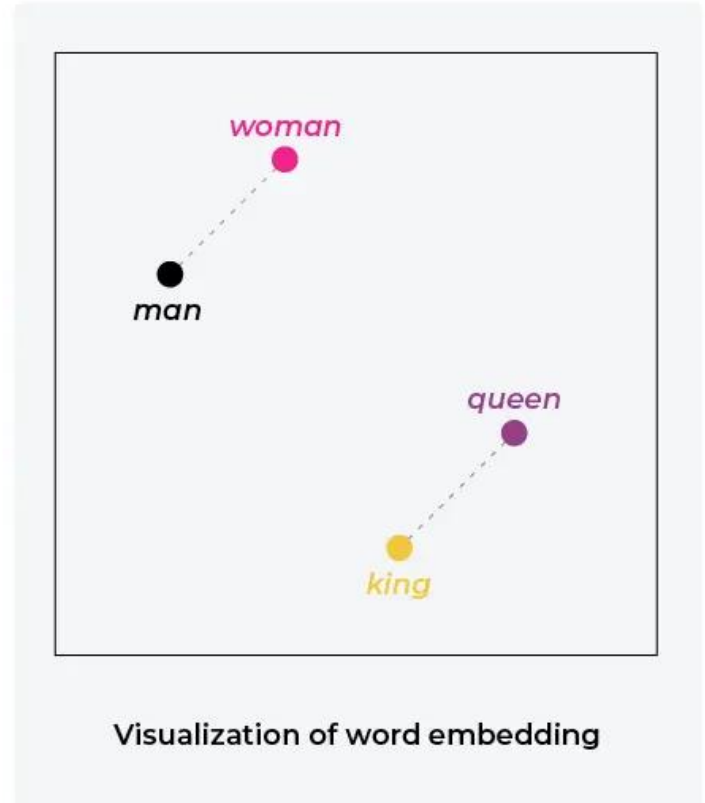
<https://colab.research.google.com/drive/1nXLZySKd8Us9iEKp77hKYgIneMsMMXyC?usp=sharing>

Embedding Models

- Převádějí text, obrázky do číselných vektorů
- Umožňují měřit podobnost mezi významy
- Používají se pro sémantické vyhledávání
- Pomáhají při doporučování
- Slouží ke shlukování podobných dokumentů nebo vět
- Jsou důležité pro RAG systémy, kde se vyhledávají relevantní informace pro LLM

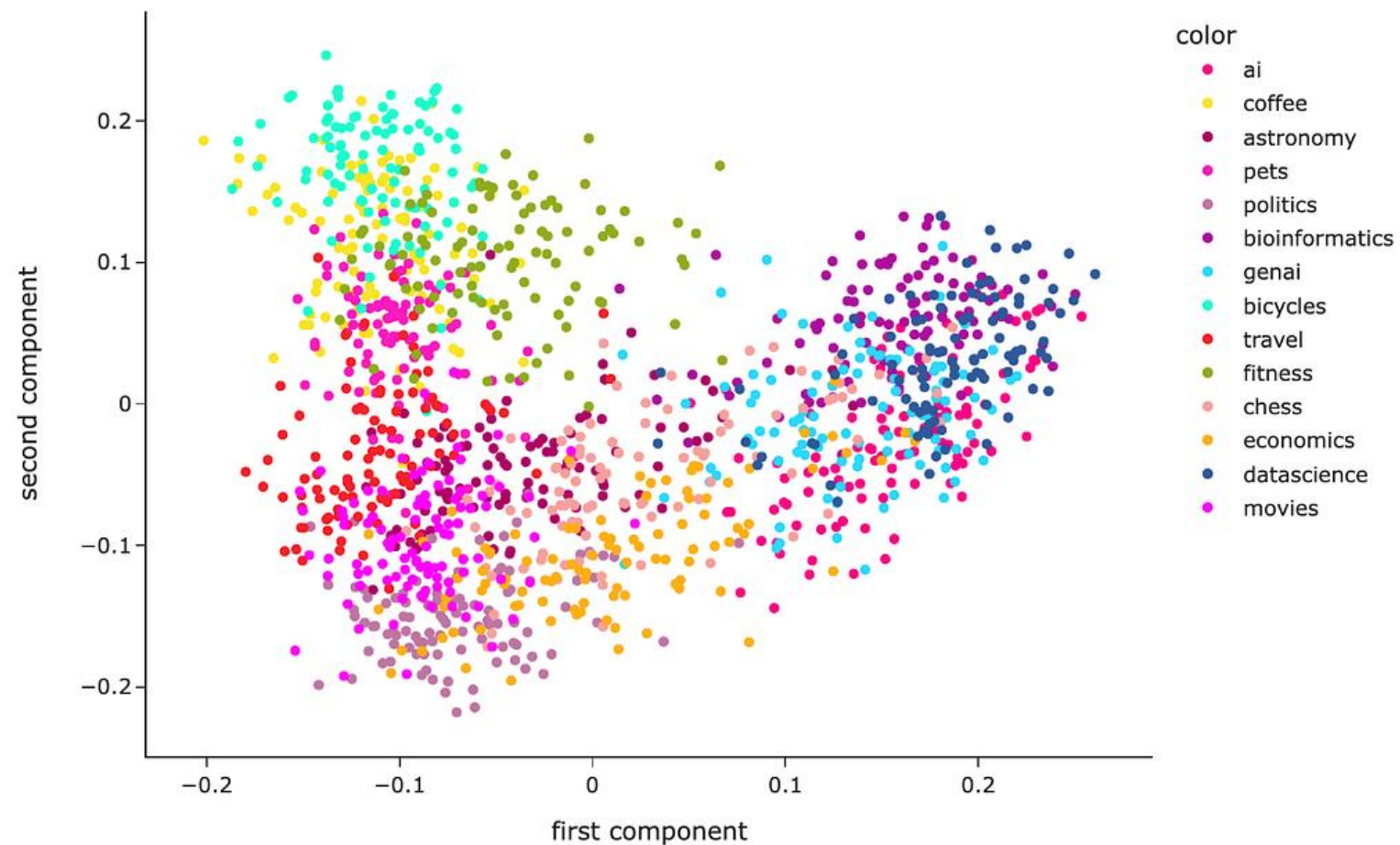
Embeddings

		living being	feline	human	gender	royalty	verb	plural
<i>man</i>	→	0.6	-0.2	0.8	0.9	-0.1	-0.9	-0.7
<i>woman</i>	→	0.7	0.3	0.8	-0.7	0.1	-0.5	-0.4
<i>king</i>	→	0.5	-0.4	0.7	0.8	0.9	-0.7	-0.6
<i>queen</i>	→	0.8	-0.1	0.8	-0.9	0.8	-0.5	-0.9
word		Word embedding						

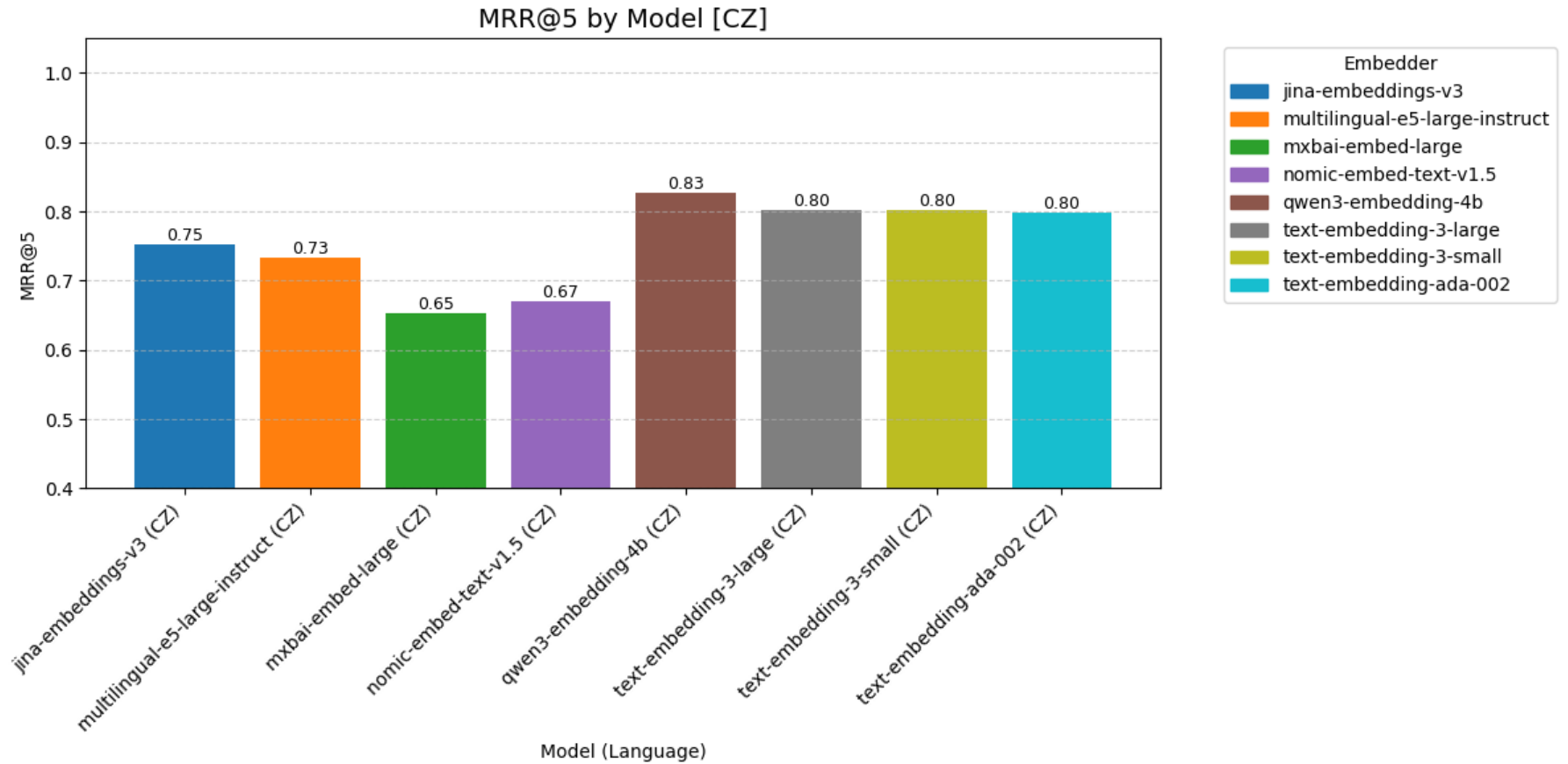


Embeddings Visualization

PCA embeddings



Model Qwen 3 Embedding



<https://blog.e-infra.cz/blog/embedders/>

Embedding Model Demo

https://colab.research.google.com/drive/1aR-1_aJ-zWd-xATZZKhbWpHolJU6uFhv?usp=sharing