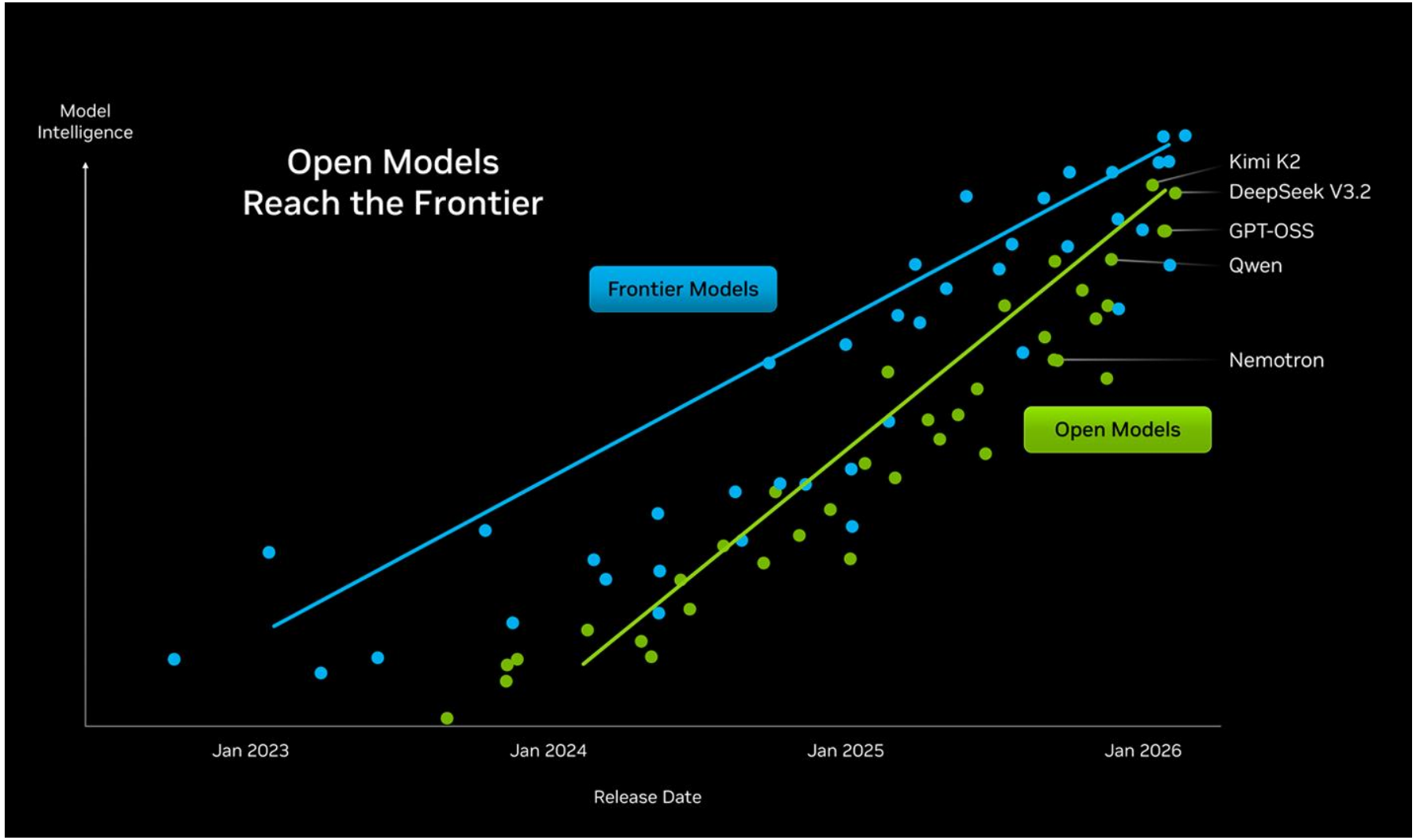
AI Confidence Workshop

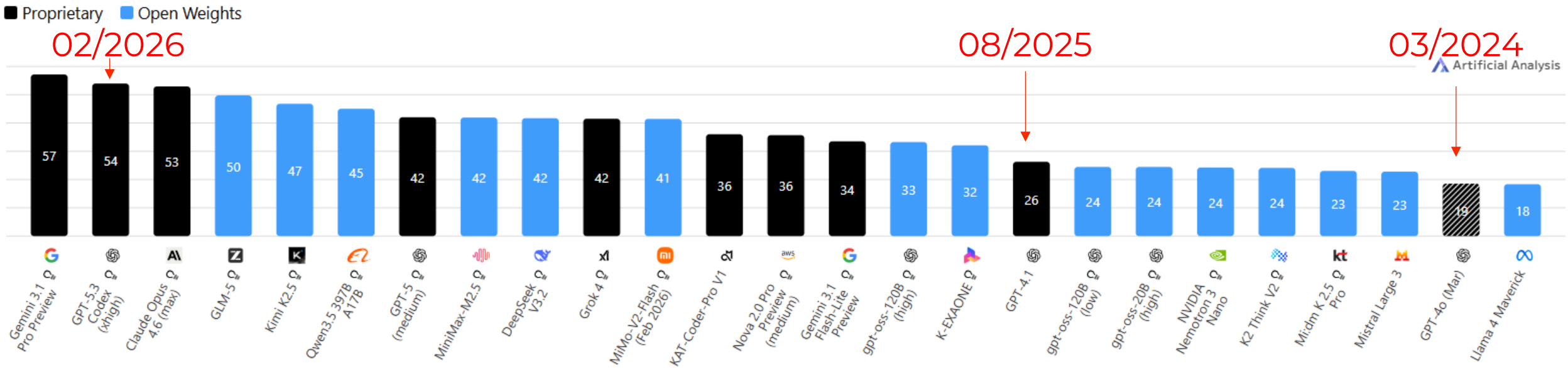
Robert Mitka, Head of AI Development, České Radiokomunikace

Open Models vs Foundation Models



Nvidia CES 2026

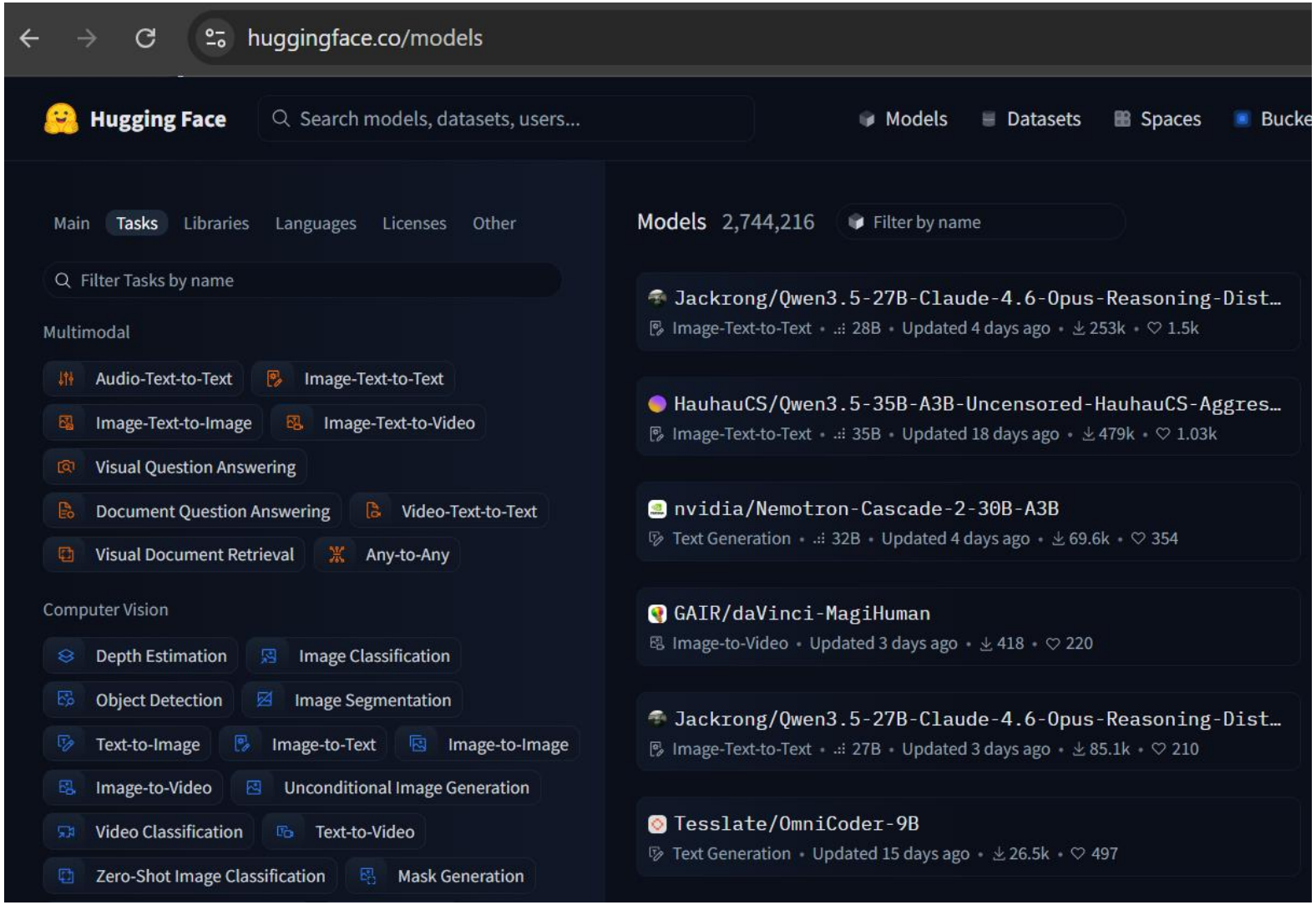
Opensource vs. Foundation LLM



Open Source

Artificial Analysis

Hugging Face



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 Hyperscalarů, ale nemáte představu kde
- Flexibilní výkon
- Cena závislá od počtu tokenů, přenesených dat,..
- O administraci serverů se stará Hyperscalar
- Podpora v angličtině za příplatek

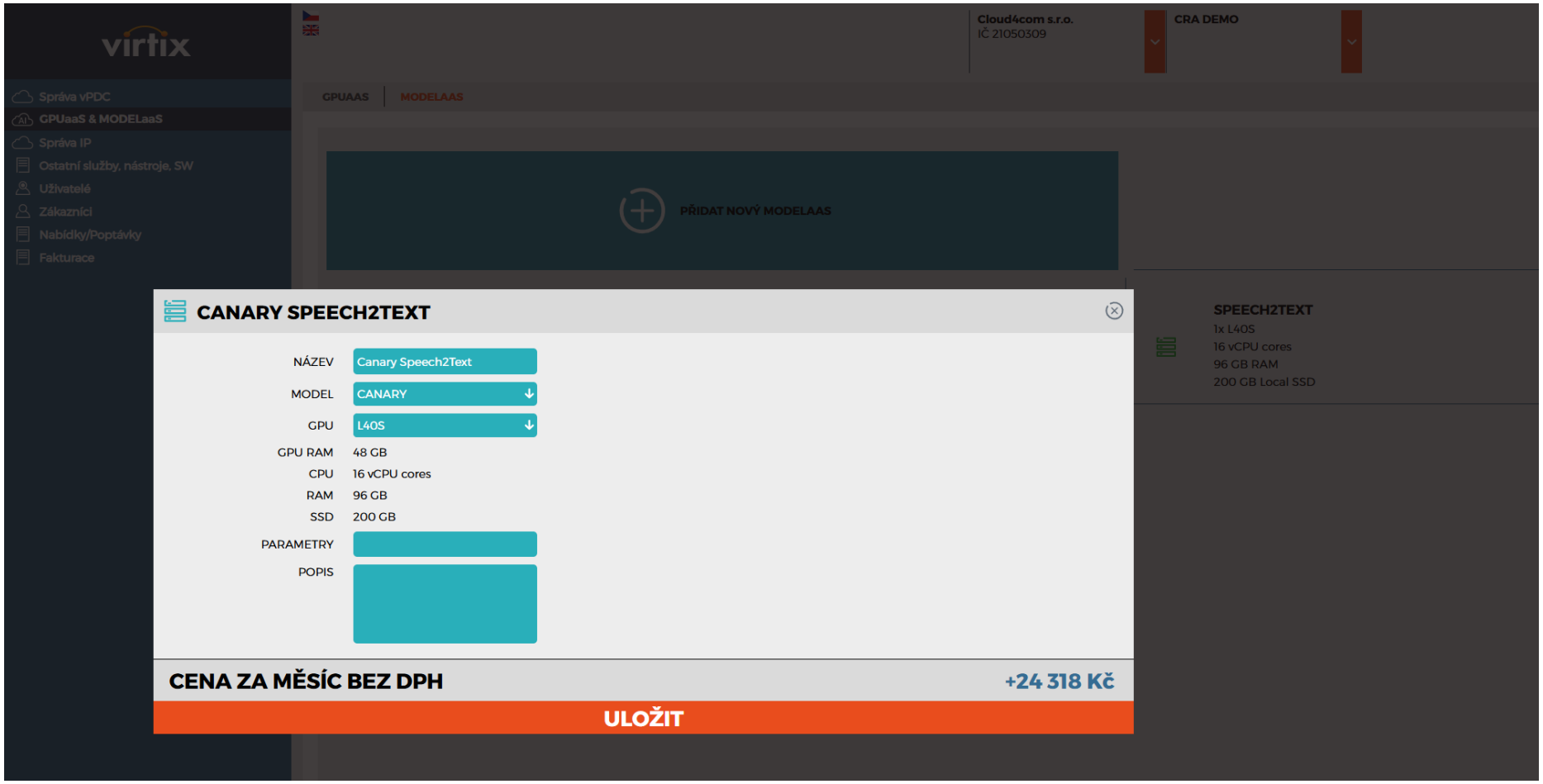
On-Premise

- 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



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

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

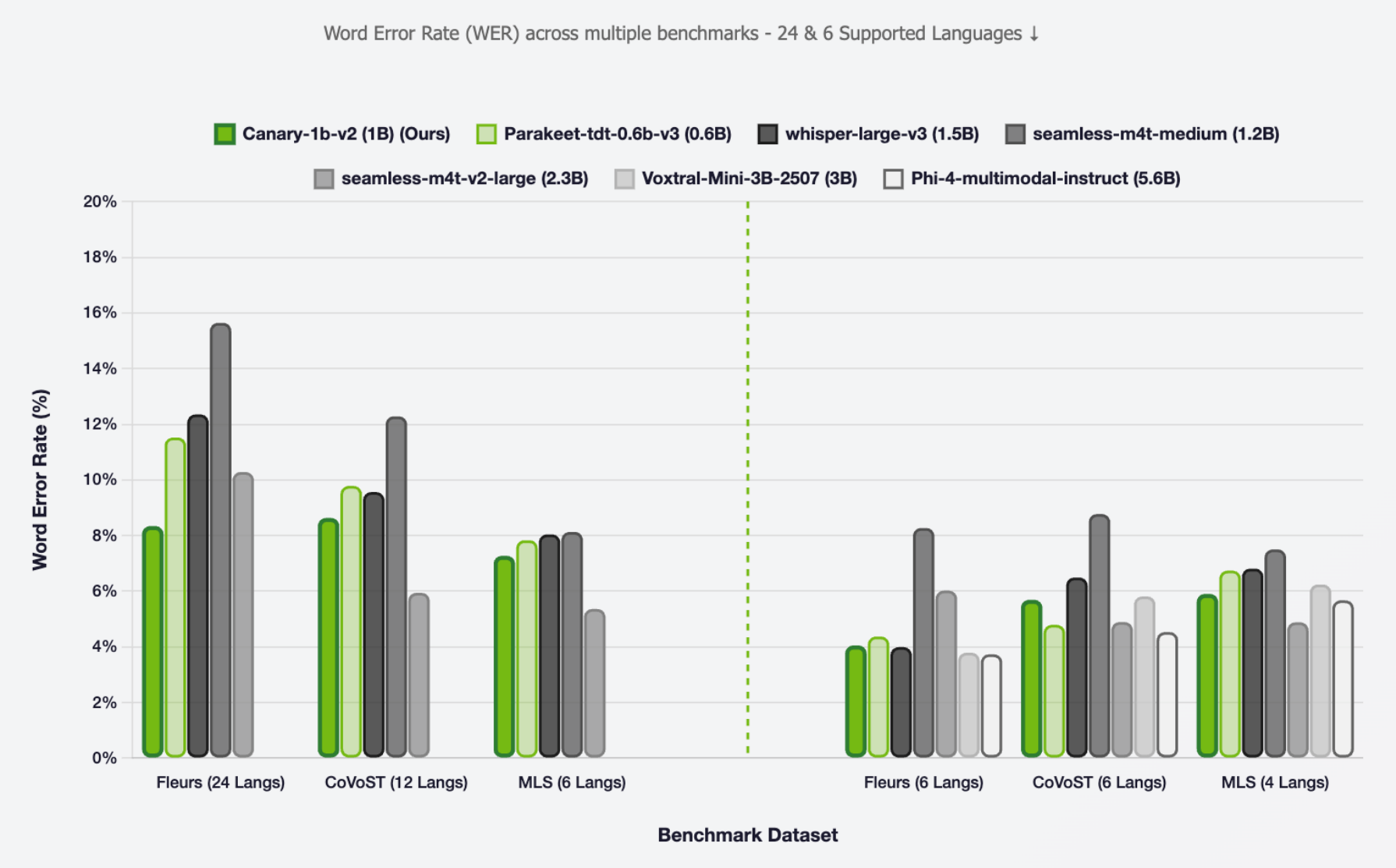
Multimodal Language Model Demo

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



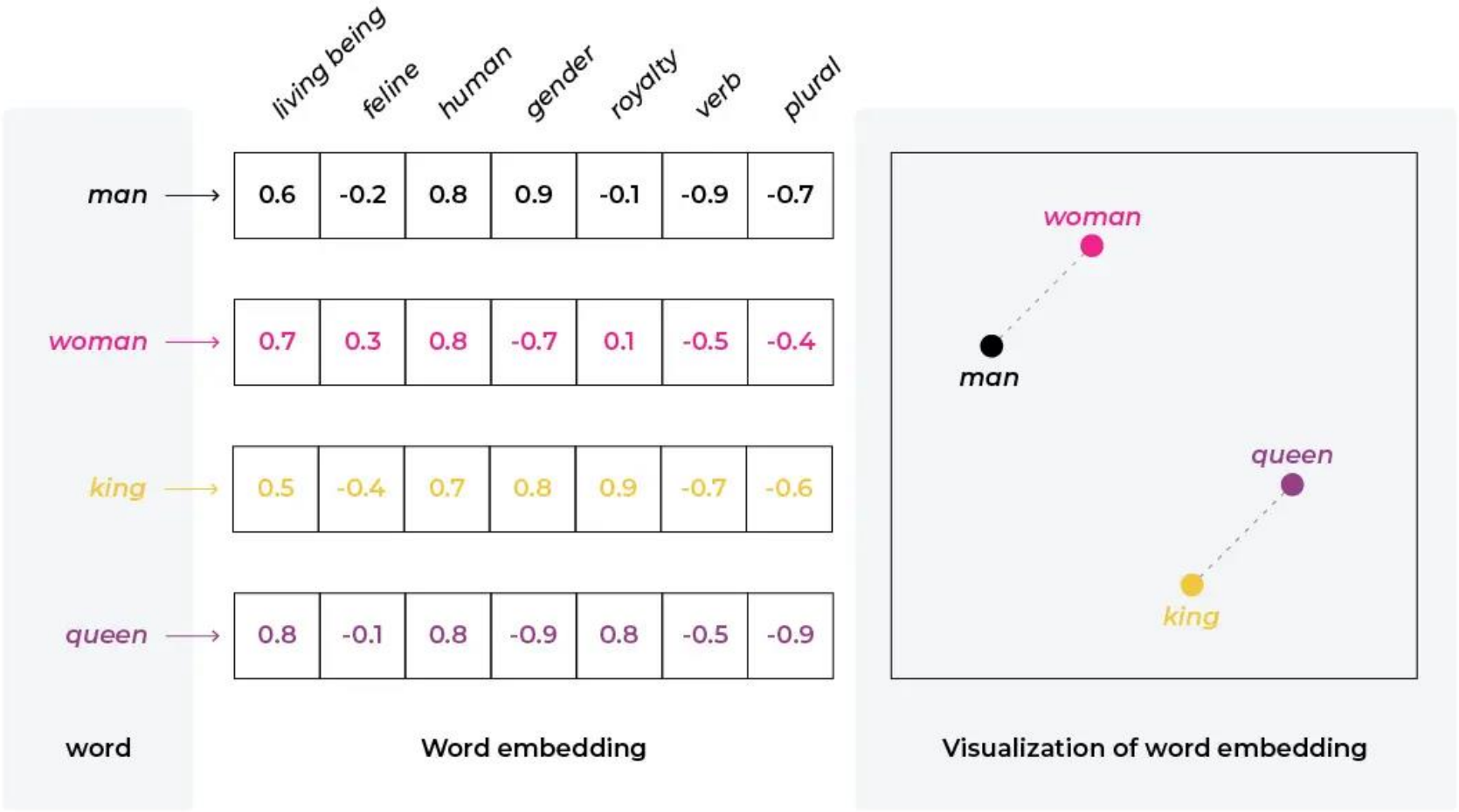
Speech to Text Model Demo

https://colab.research.google.com/drive/1nXLZySKd8Us9iEKp77hKYglneMsMMXyC?usp=sharing

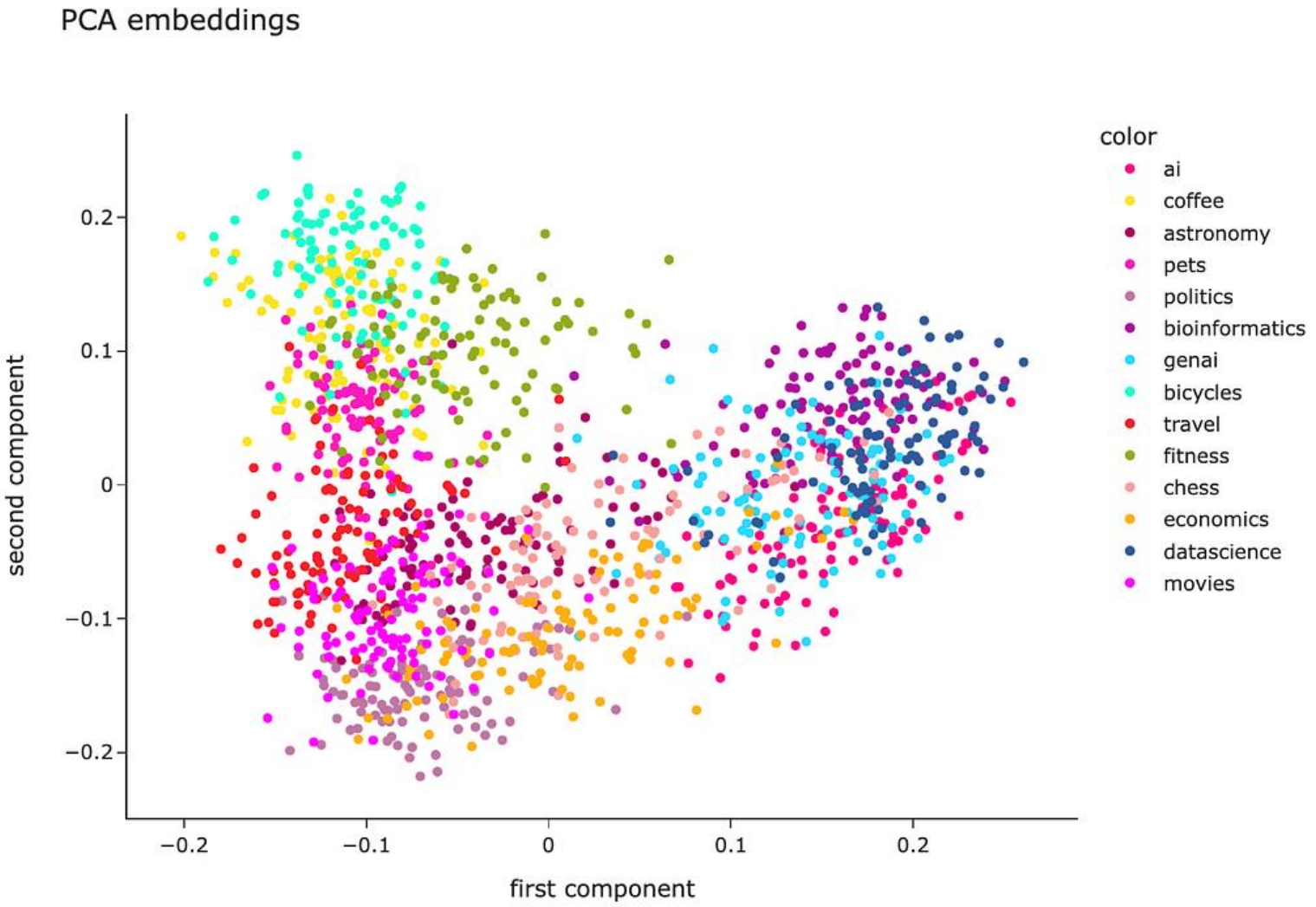
Embedding Models

- 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

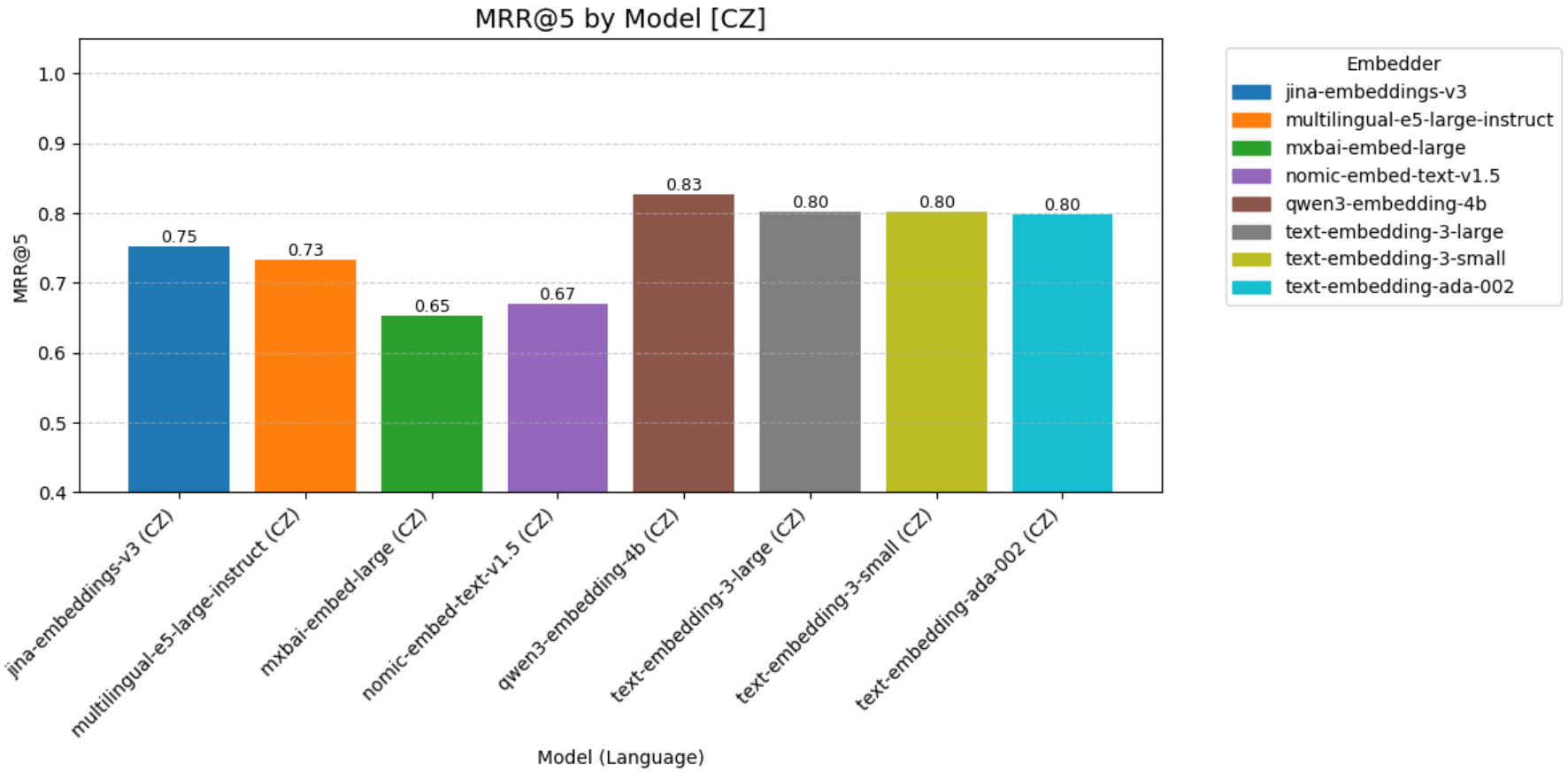
Embeddings


living being
feline
human
gender
royalty
verb
plural
man
0.6
-0.2
0.8
0.9
-0.1
-0.9
-0.7
woman
0.7
0.3
0.8
-0.7
0.1
-0.5
-0.4
king
0.5
-0.4
0.7
0.8
0.9
-0.7
-0.6
queen
0.8
-0.1
0.8
-0.9
0.8
-0.5
-0.9
word
Word embedding
woman
man
queen
king
Visualization of word embedding

Embeddings Visualization



Model Qwen 3 Embedding



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

Embedding Model Demo

https://colab.research.google.com/drive/1aR-1_aJ-zWd-xATZZKhbWpHoIJU6uFhv?usp=sharing