

LLM Engineering

MASTER AI & LARGE LANGUAGE MODELS



Today equips you to make important decisions

What you can now do

- Code with Frontier Models including AI Assistants with Tools
- Build solutions with open-source LLMs with HuggingFace transformers
- Compare LLMs to identify the right one for the task at hand

After today you will be able to

- Navigate the most useful leaderboards and Arenas to evaluate LLMs
- Give real-world use cases of LLMs solving commercial problems
- Confidently choose LLMs for your projects

Six Leaderboards

A tour of the essential leaderboards for selecting LLMs

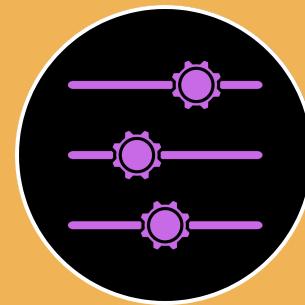


HuggingFace Open LLM

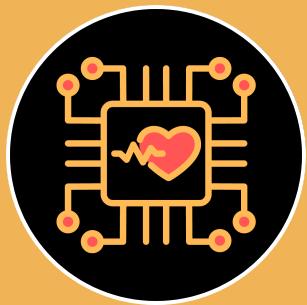
New and old version



HuggingFace BigCode



HuggingFace LLM Perf



HuggingFace Others

eg Medical, Language-specific



Vellum

includes API cost and context window



SEAL

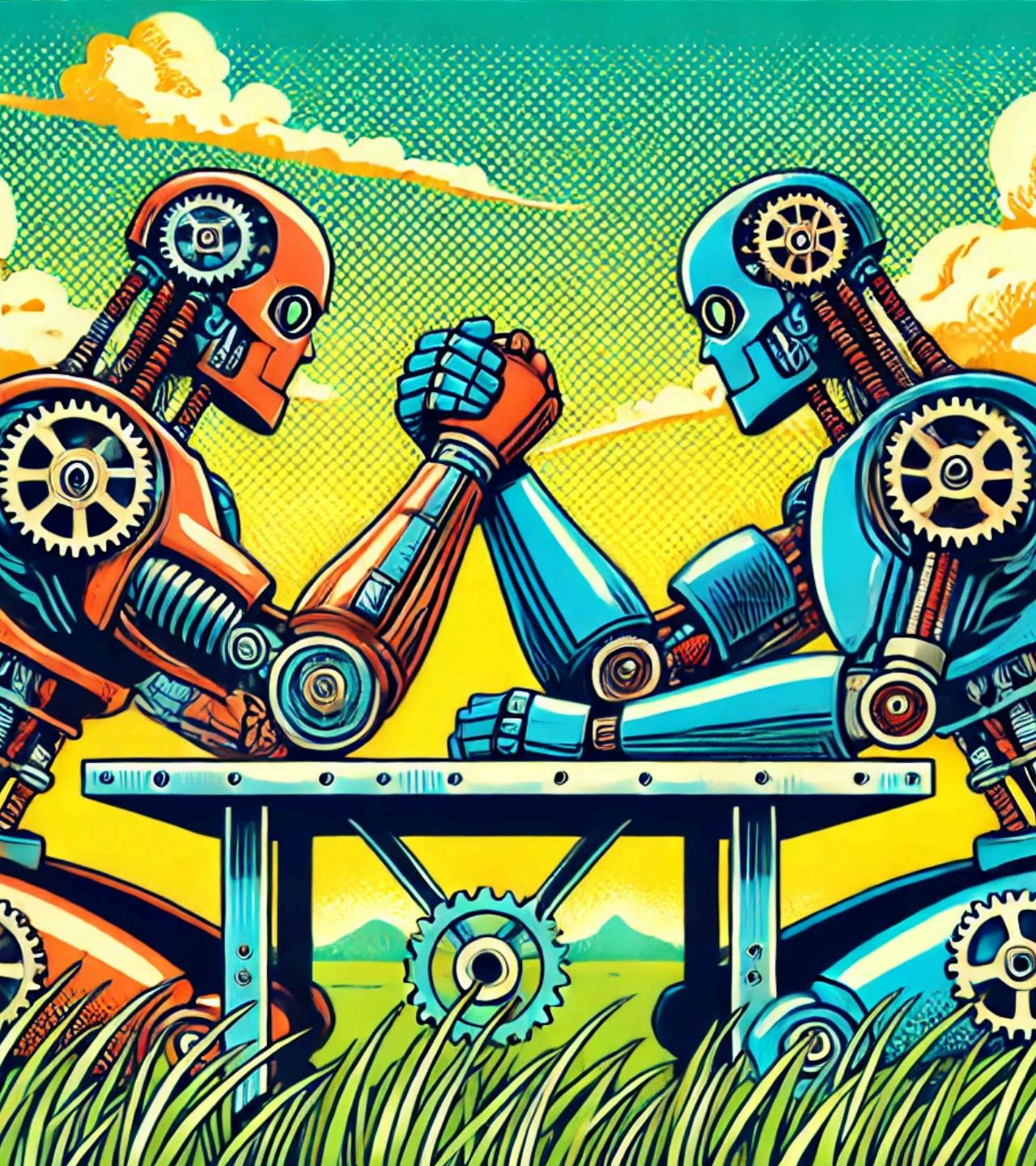
expert skills

The Arena

The LMSYS Chatbot Arena is an amazing resource

- Compare Frontier and Open-source models directly
- Blind human evals based on head-to-head comparison
- LLMs are measured with an ELO rating

Participate in the voting - it's a terrific way to learn about different models while adding to the ecosystem



Commercial Use Cases



Law

Harvey



Talent

nebula.io



Porting code

Bloop



Healthcare

Salesforce Health

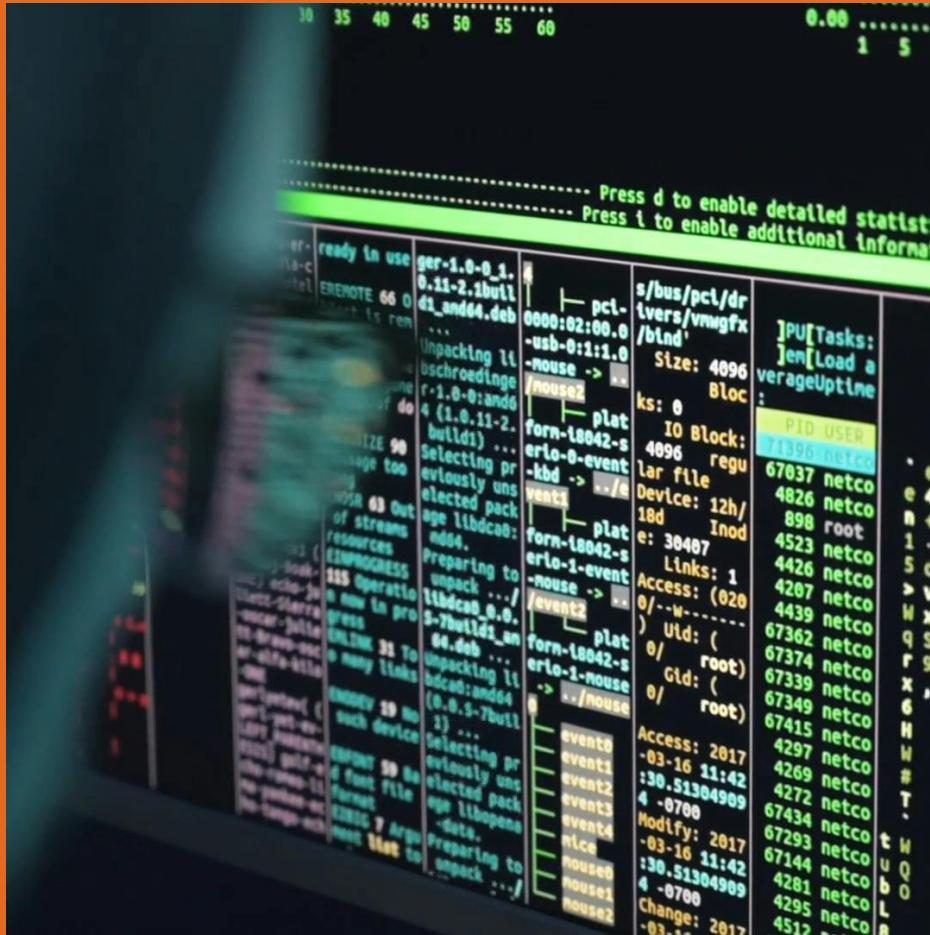


Education

Khanmigo

BUSINESS CHALLENGE

Introducing our commercial challenge this week



Build a product that converts Python code to C++ for performance

Solution with a Frontier model

Solution with an Open-Source model

Let's start by selecting the LLMs most suited for the task

Ready for our code challenge

What you can now do

- Code with Frontier Models including AI Assistants with Tools
- Build solutions with open-source LLMs with HuggingFace transformers
- Confidently choose the right LLM for your project, backed by metrics

After the next session you will be able to

- Assess models for coding ability
- Use Frontier models to generate code
- Build a solution that uses LLMs to generate code