Budapest ML Meetup #1

Arató Bence
BI Consulting

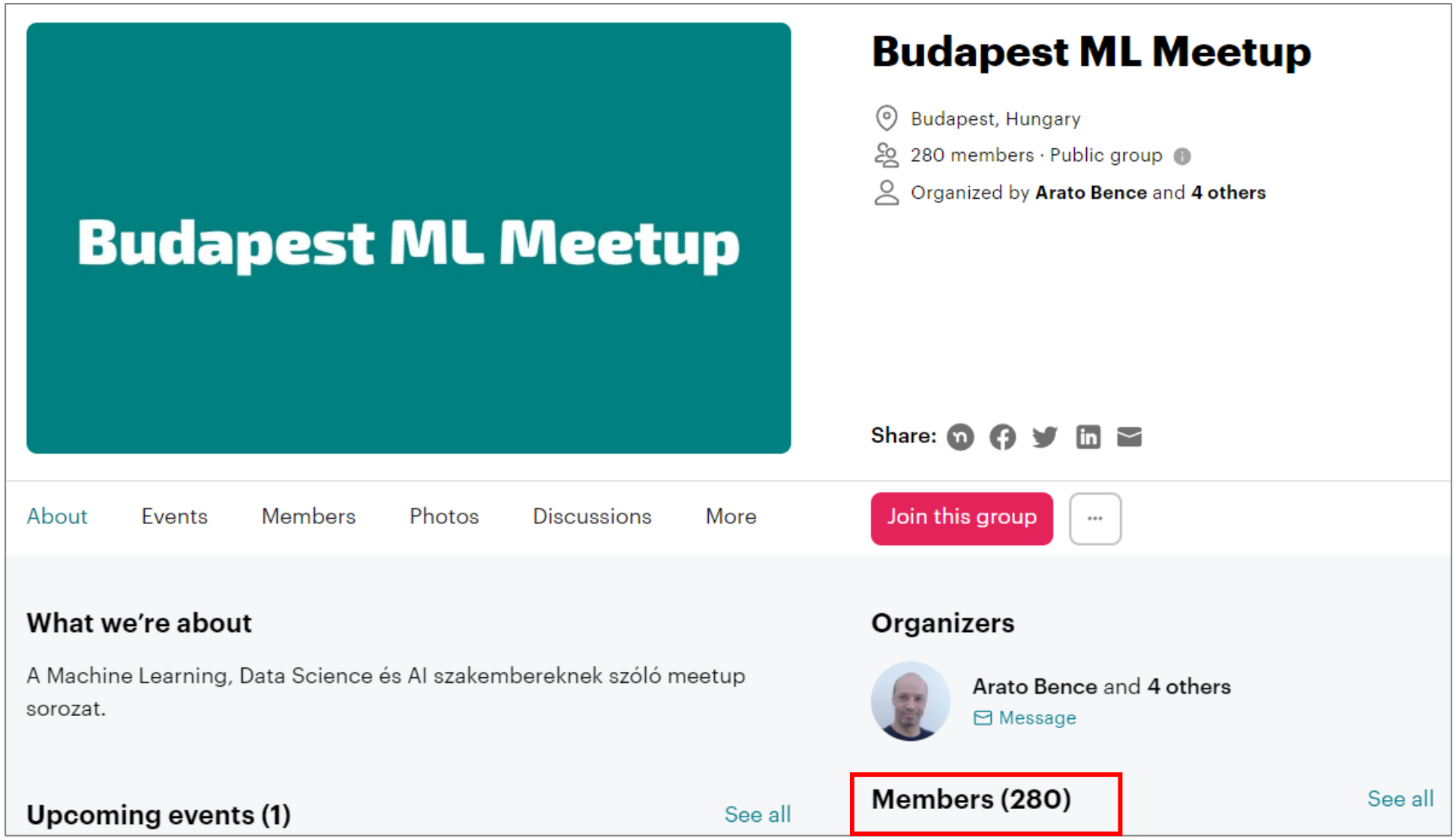
Oltyán Gábor
PwC Hungary

2024. január 17. · szerda · 18:00 - 20:00

PwC Hungary · 1055 · Budapest · Bajcsy-Zsilinszky út 78.

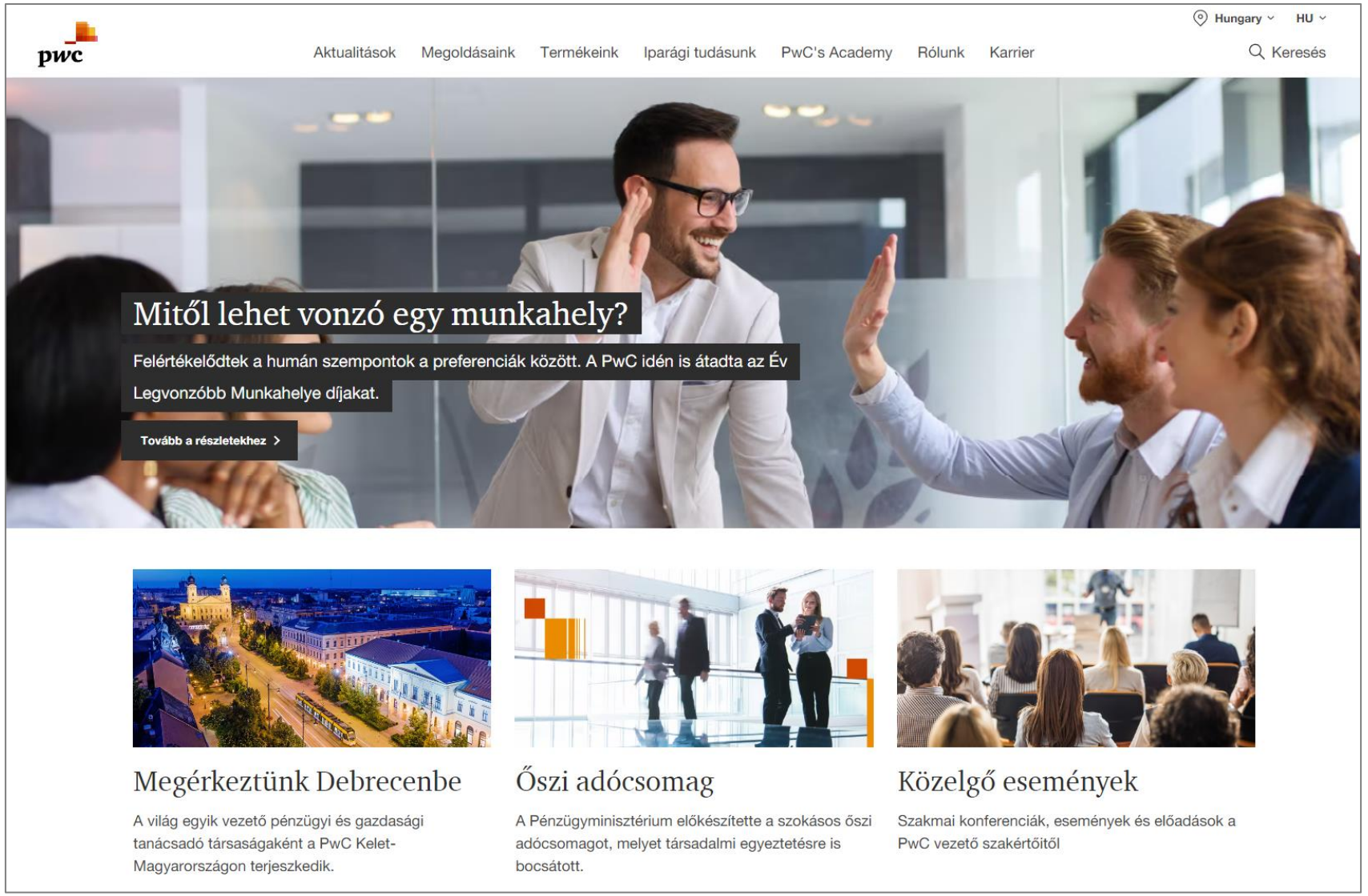
Bevezető

Welcome!



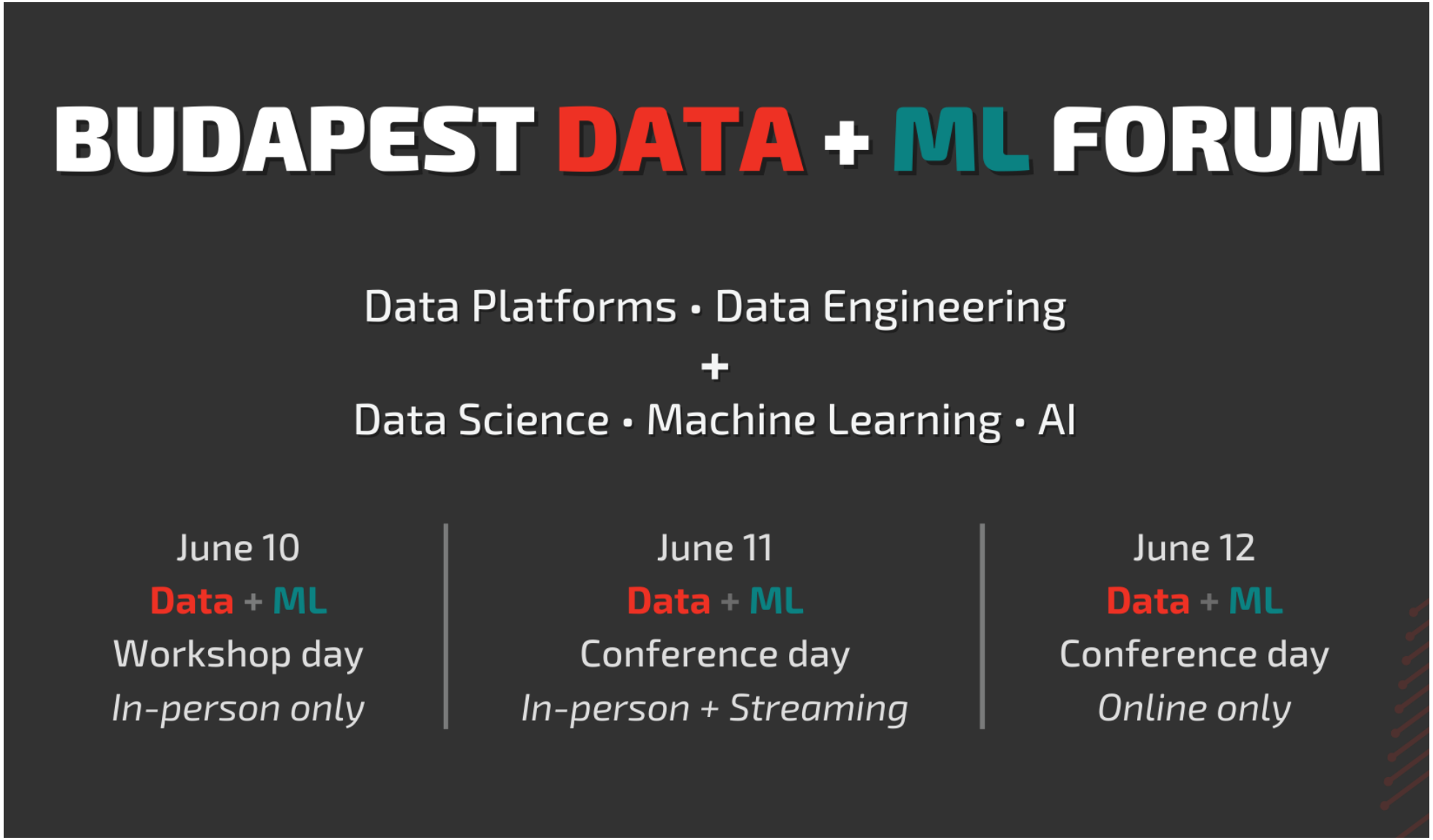
mlmeetup.hu

A mai meetup támogatója



pwc.hu

Data+ML Forum 2024

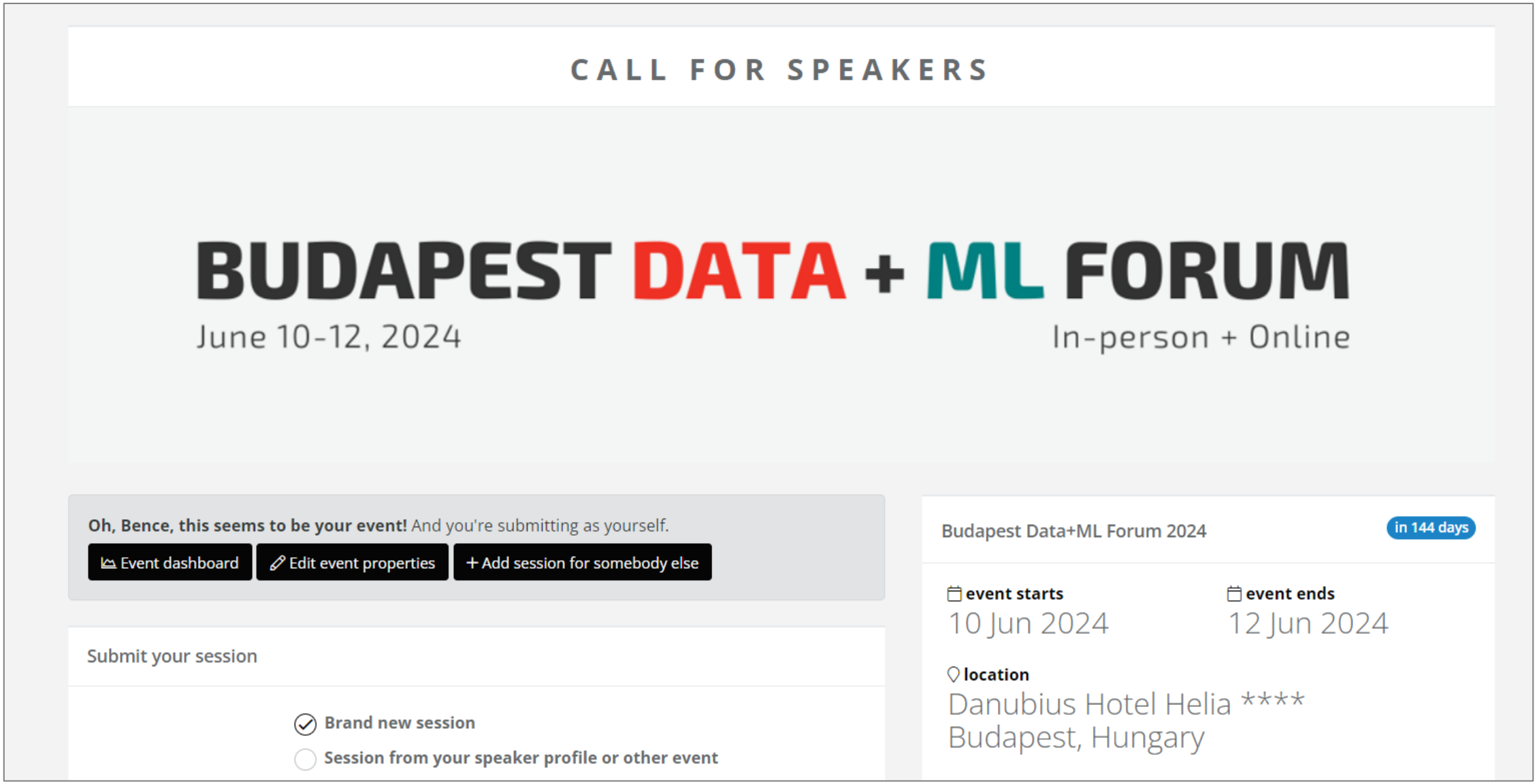


https://budapestdata.hu/2024/

Data+ML Forum 2024

- 2024. június 10-12 (hétfő-szerda) a Hotel Héliában
  - Hétfő: workshopok
  - Kedd: személyes konferencianap (streaming lesz)
  - Szerda: online konferencianap
- Előadói pályázat
  - Data Science, ML, AI témájú előadások
  - Technológiai előadások, esettanulmányok, legjobb gyakorlatok
  - Angol vagy magyar nyelv
- Jegyértékesítés január végétől
- Szponzorok és kiállítók felkérés alatt

Előadói pályázat



https://sessionize.com/budapestdataml24/

Ismét lesz PyData track

EVENTS RESOURCES PyData ABOUT SPONSOR PYDATA

About PyData

An educational program of NumFOCUS, a 501(c)(3) nonprofit charity

PyData Mission

PyData provides a forum for the international community of users and developers of data analysis tools to share ideas and learn from each other. The global PyData network promotes discussion of best practices, new approaches, and emerging technologies for data management, processing, analytics, and visualization. PyData communities approach data science using many languages, including (but not limited to) Python, Julia, and R.

We aim to be an accessible, community-driven conference, with novice to advanced level presentations. PyData tutorials and talks bring attendees the latest project features along with cutting-edge use cases.

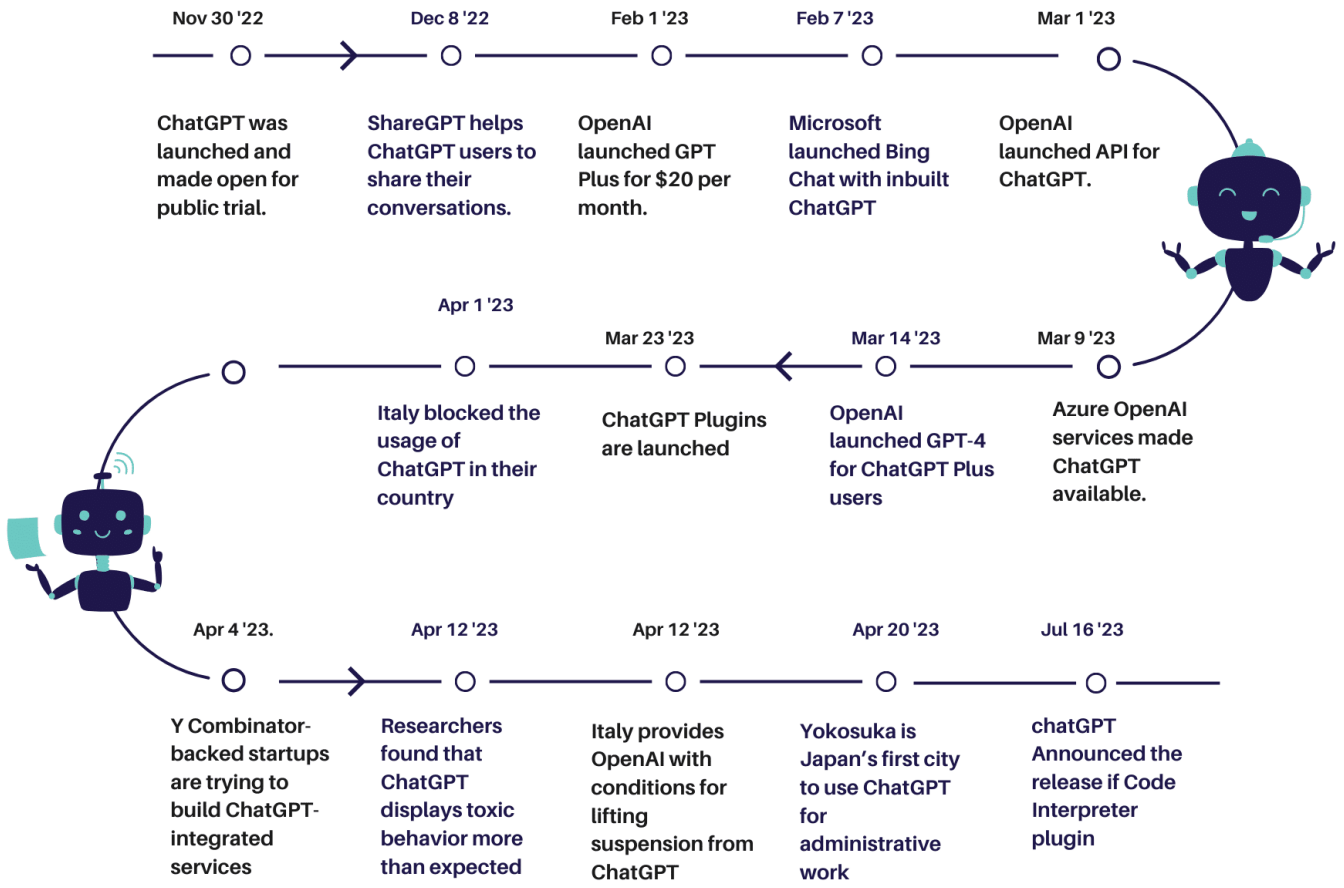
https://pydata.org/about/

Az AI és ML őrült éve

Ahogy indult

ChatGPT

ChatGPT Timeline



ChatGPT Details Statistics | © Copyright demandsage

https://www.demandsage.com/chatgpt-statistics/

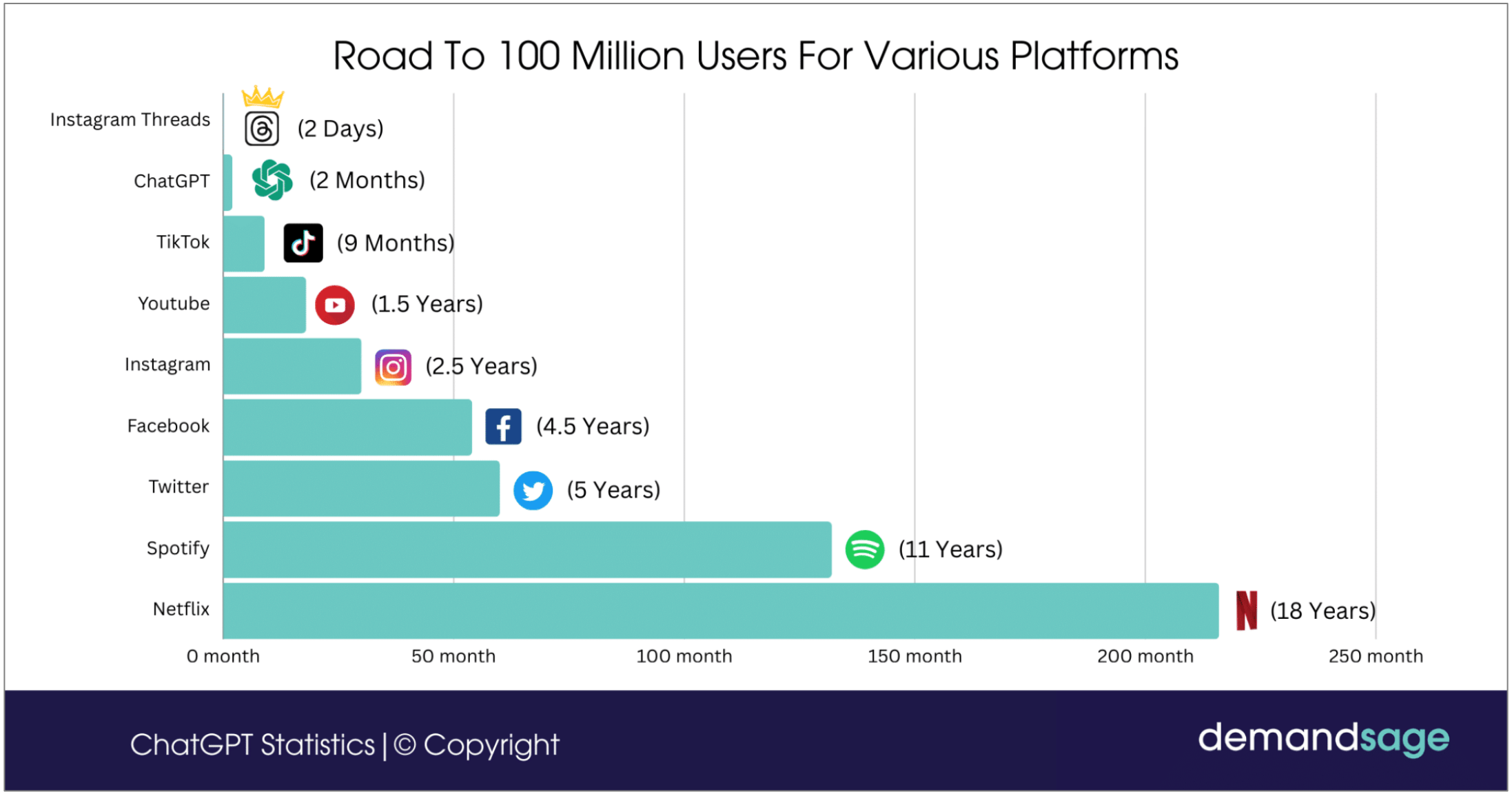
ChatGPT

ChatGPT Statistics 2024: Our Top Picks

- ChatGPT has over 180 million monthly users and 100 million weekly active users.
- ChatGPT reached 1 million users in just five days after its launch.
- Nearly 15% of the ChatGPT users are from the United States.
- More than 92% of Fortune 500 companies are using ChatGPT.
- ChatGPT is forecasted to generate a revenue of $1 billion in 2024.
- OpenAI spends approximately $700,000 every day to operate ChatGPT.
- ChatGPT gets over 1.7 billion page visits every month on average.
- 7 nations, including China and Russia, cannot access ChatGPT.
- 53% of the readers failed to identify ChatGPT written content as AI-written.

https://www.demandsage.com/chatgpt-statistics/

ChatGPT



https://www.demandsage.com/chatgpt-statistics/

Tőkeemelések

Tőkeemelések

Év, hónap	Cég	Befektető	Összeg (USD)
2023.01	OpenAI	Microsoft	Up to 10B
2023.02	Anthropic	Google	300M
2023.05	Anthropic	Series C (több befektető)	450M
2023.06	Inflection.ai	Több befektető	1.3 B
2023.06	Mistral AI	Több befektető	113M
2023.08	Anthropic	SK Telecom	100M
2023.09	Databricks	Series I (több befektető)	500M +
2023.09	Anthropic	Amazon	1.25B + 2.75B *
2023.10	Anthropic	Google	500M + 1.5B*
2023.11	Aleph Alpha	Series B (több befektető)	500M
2023.12	Mistral AI	Több befektető	414M

OpenAI

Microsoft invests billions more dollars in OpenAI, extends partnership

Kyle Wiggers @kyle_l_wiggers / 3:15 PM GMT+1 • January 23, 2023

Comment

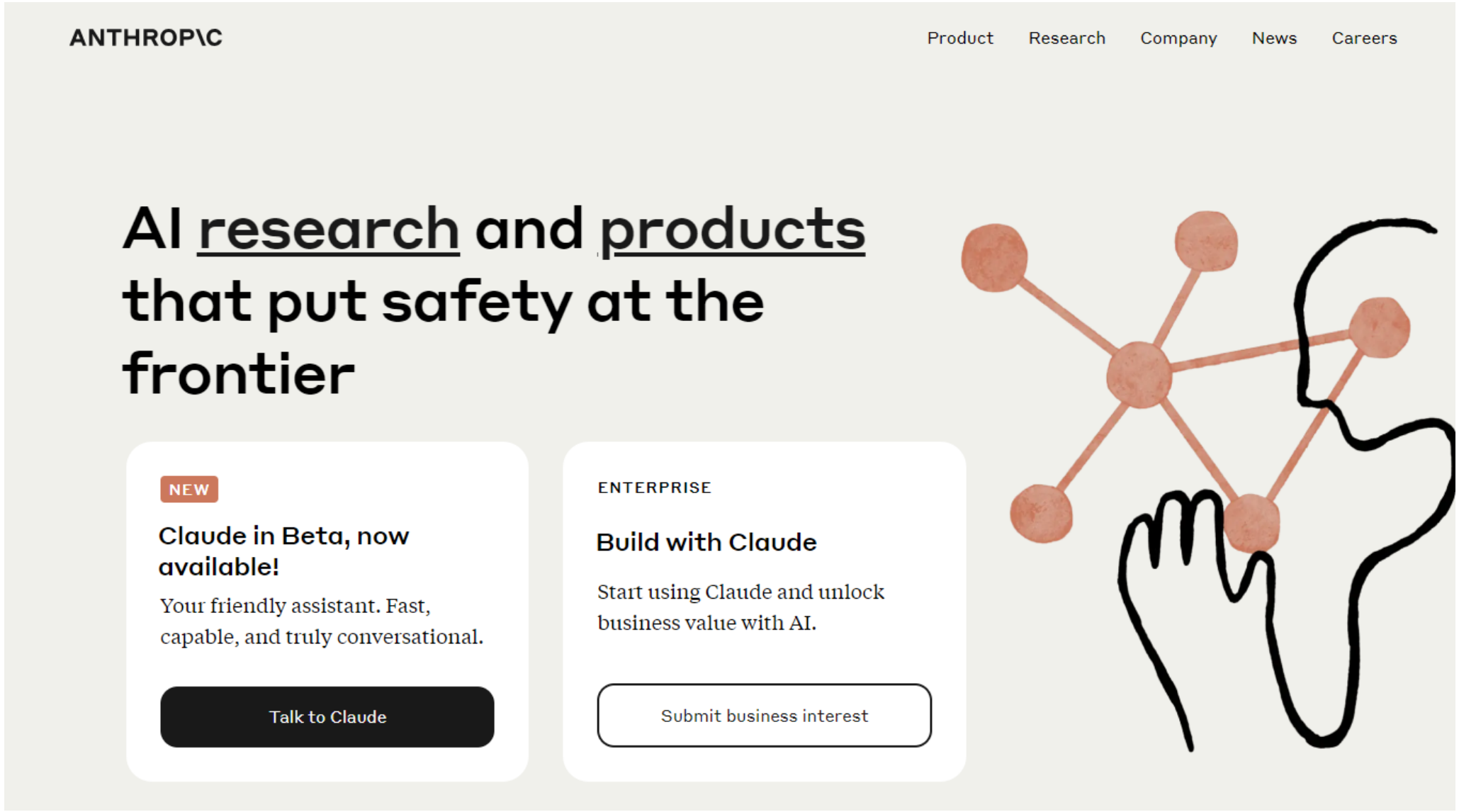
Image Credits: OpenAI

Microsoft today said that it's extending its partnership with OpenAI, the startup behind art- and text-generating AI systems like ChatGPT, DALL-E 2 and GPT-3, with a "multi-year, multi-billion-dollar" investment. OpenAI says that the infusion of new capital — the exact amount of which wasn't disclosed — will be used to continue its independent research and develop AI that's "safe, useful and powerful."

The optics aren't the best for Microsoft, which just last week announced plans to lay off 10,000 employees as a part of broader cost-cutting measures. But they'd been telegraphed by the company earlier this month — in an interview with The Wall Street Journal, Microsoft CEO Satya Nadella said that Microsoft planned to make OpenAI's foundational systems available as commercials platforms so that any entity in any industry can build on them.

https://techcrunch.com/2023/01/23/microsoft-invests-billions-more-dollars-in-openai-extends-partnership/

Anthropic



https://www.anthropic.com/

Mistral AI

France's Mistral AI blows in with a $113M seed round at a $260M valuation to take on OpenAI

Ingrid Lunden @ingridlunden / 6:02 PM GMT+2 • June 13, 2023

Comment

Image Credits: John Seaton Callahan / Getty Images

AI is well and truly off to the races: a startup that is only four weeks old has picked up a $113 million round of seed funding to compete against OpenAI in the building, training and application of large language models and generative AI.

Mistral AI, based out of Paris, is co-founded by alums from Google's DeepMind and Meta and will be focusing on open source solutions and targeting enterprises to create what CEO Arthur Mensch believes is currently the biggest challenge in the field: "To make AI useful." It plans to release its first models for text-based generative AI in 2024.

Lightspeed Venture Partners is leading this round, with Redpoint, Index Ventures, Xavier Niel, JCDecaux Holding, Rodolphe Saadé and Motier Ventures in France, La Famiglia and Headline in Germany, Exor Ventures in Italy, Sofina in Belgium, and First Minute Capital and LocalGlobe in the UK all also participating.

https://techcrunch.com/2023/06/13/frances-mistral-ai-blows-in-with-a-113m-seed-round-at-a-260m-valuation-to-take-on-openai

Mistral AI

Mistral AI raises 385 mln euros in second round in seven months

By **Supantha Mukherjee**

December 11, 2023 9:27 AM GMT+1 · Updated a month ago

Dec 11 (Reuters) - French artificial intelligence company Mistral AI said on Monday it had raised 385 million euros ($414.41 million) in its second funding round in seven months, led by investors such as Andreessen-Horowitz and LightSpeed Ventures.

While the company did not disclose its valuation, a source familiar with the matter put it at 2 billion euros - an amount that would put the French startup among the top AI companies in the world.

Paris-based Mistral AI, founded by former Meta (META.O) and Google (GOOGL.O) AI researchers, had in June raised 105 million euros just a month after being set up.

The company also launched Mixtral 8x7B, an open model that seeks to be Europe's answer to AI platforms such as OpenAI's ChatGPT and Google's Bard. The Mistral AI platform is available in beta access and will be operational by early 2024.

https://www.reuters.com/technology/mistral-ai-raises-385-mln-euros-second-round-seven-months-2023-12-11/

Aleph Alpha

Lidl owner and Bosch Ventures co-lead $500M Series B into German AI startup Aleph Alpha

Mike Butcher @mikebutcher / 3:08 PM GMT+1 • November 6, 2023

Comment

Image Credits: Getty Images

German AI startup Aleph Alpha has raised a Series B funding round of $500 million from a consortium of seven new investors, as well as existing investors from previous rounds.

Founded in 2019, Aleph Alpha's funding may be dwarfed by that of Microsoft-backed OpenAI ($11.3 billion), but the startup makes great play of the fact that its clients have "full sovereignty" over the implementation of AI into their businesses. So although it's tempting to compare Aleph Alpha with other foundational models like OpenAI, it's much closer to startups like France's Mistral ($112 million in funding), which works with large corporates to deploy LLMs internally.

The consortium is led by the Innovation Park Artificial Intelligence (Ipai). The round was co-led by Schwarz Group (the owners of the Lidl supermarket chain) and Bosch Ventures. Other new investors include Berlin-based Christ&Company Consulting, Hewlett Packard Enterprise and SAP, as well as Burda Principal Investments. Existing institutional investors also participated.

https://techcrunch.com/2023/11/06/lidl-owner-and-bosch-ventures-co-lead-500m-series-b-into-german-ai-start-up-aleph-alpha/

OpenAI Saga

OpenAI Saga

Blog

OpenAI announces leadership transition

November 17, 2023

Authors
OpenAI ↓

Announcements

Chief technology officer Mira Murati appointed interim CEO to lead OpenAI; Sam Altman departs the company.

Search process underway to identify permanent successor.

The board of directors of OpenAI, Inc., the 501(c)(3) that acts as the overall governing body for all OpenAI activities, today announced that Sam Altman will depart as CEO and leave the board of directors. Mira Murati, the company's chief technology officer, will serve as interim CEO, effective immediately.

A member of OpenAI's leadership team for five years, Mira has played a critical role in OpenAI's evolution into a global AI leader. She brings a unique skill set, understanding of the company's values, operations, and business, and already leads the company's research, product, and safety functions. Given her long tenure and close engagement with all aspects of the company, including her experience in AI governance and policy, the board believes she is uniquely qualified for the role and anticipates a seamless transition while it conducts a formal search for a permanent CEO.

https://openai.com/blog/openai-announces-leadership-transition

OpenAI Saga

Sam Altman
@sama · Follow
i loved my time at openai. it was transformative for me personally, and hopefully the world a little bit. most of all i loved working with such talented people.
will have more to say about what's next later.
10:46 PM · Nov 17, 2023
97.4K
Reply
Share
Read 6.8K replies

OpenAI Saga

Sam Altman
@sama

first and last time i ever wear one of these

10:03 PM · Nov 19, 2023 · **20.7M** Views

3.8K 9K 100K 2.6K

Satya Nadella
@satyanadella

We remain committed to our partnership with OpenAI and have confidence in our product roadmap, our ability to continue to innovate with everything we announced at Microsoft Ignite, and in continuing to support our customers and partners. We look forward to getting to know Emmett Shear and OAI's new leadership team and working with them.

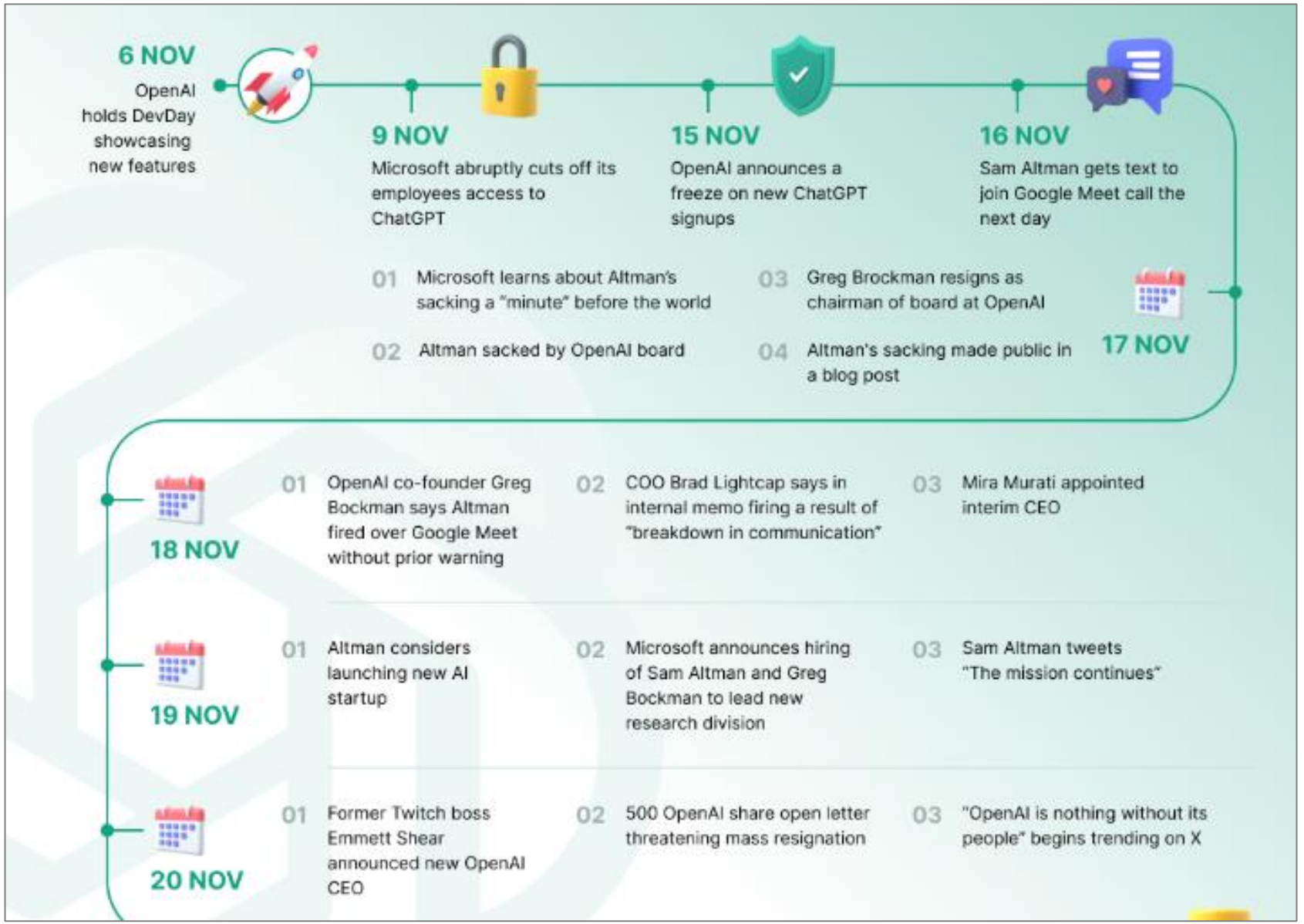
And we're extremely excited to share the news that Sam Altman and Greg Brockman, together with colleagues, will be joining Microsoft to lead a new advanced AI research team. We look forward to moving quickly to provide them with the resources needed for their success.

8:53 AM · Nov 20, 2023 · **41.3M** Views

4.7K 27K 94K 5.5K

OpenAI Saga

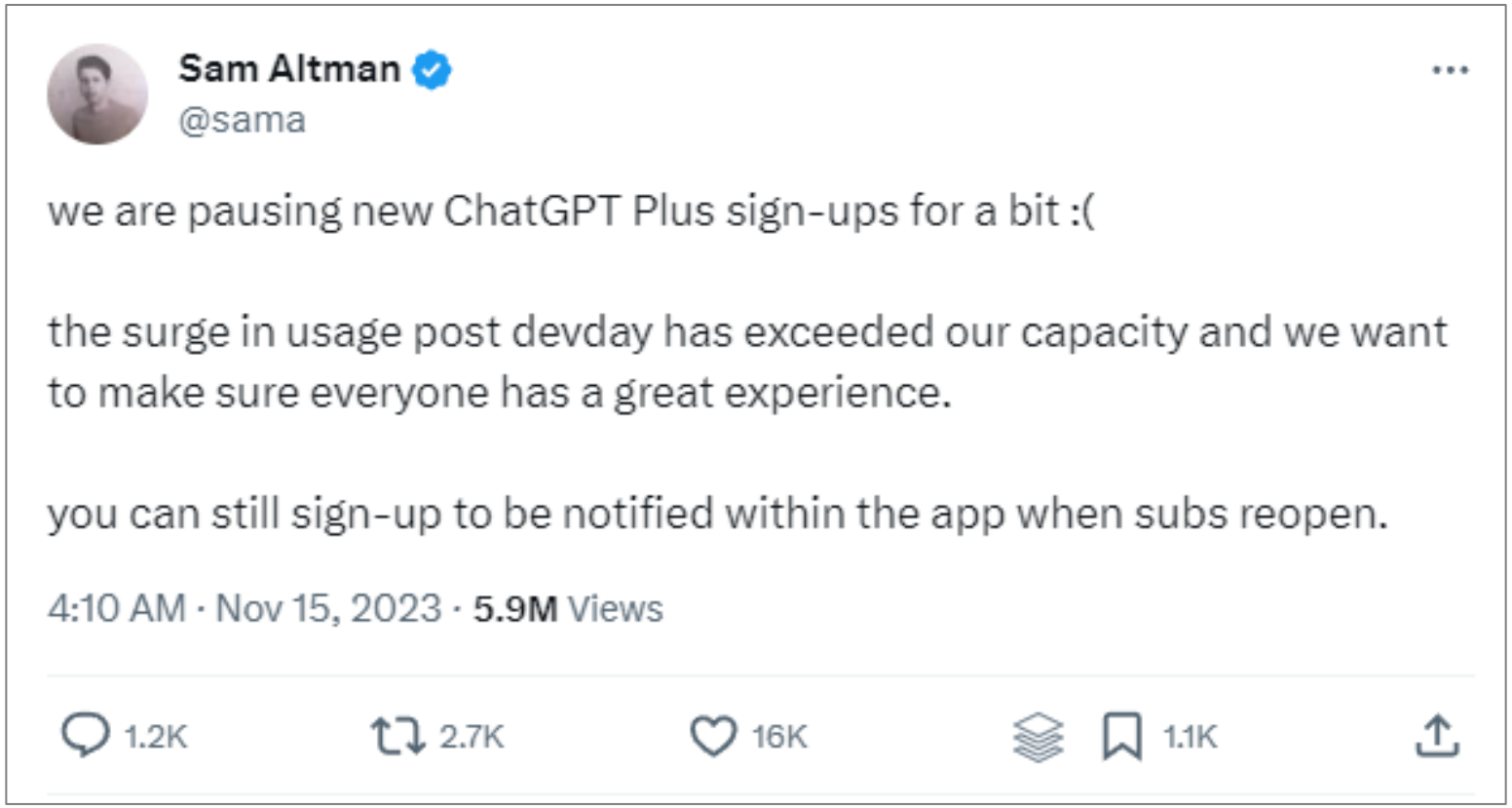
https://twitter.com/OpenAI/status/1727206187077370115



https://venngage.com/blog/sam-altman-timeline/

GPU hiánygazdaság

GPU hiánygazdaság



Sam Altman ✓
@sama

we are pausing new ChatGPT Plus sign-ups for a bit :(

the surge in usage post devday has exceeded our capacity and we want to make sure everyone has a great experience.

you can still sign-up to be notified within the app when subs reopen.

4:10 AM · Nov 15, 2023 · **5.9M** Views

1.2K 2.7K 16K 1.1K

https://twitter.com/sama/status/1724626002595471740

GPU hiánygazdaság

OpenAI blocks new ChatGPT Plus subscriptions — and creates black market for selling existing accounts

Here's why you shouldn't buy ChatGPT access from an eBay reseller.

By Matt Binder on November 16, 2023

ChatGPT Plus resellers have surged since OpenAI paused new sign-ups for the $20 per month service. Credit: Omar Marques/SOPA Images/LightRocket via Getty Images

OpenAI, the AI giant behind the popular ChatGPT chatbot, announced some big updates earlier this month, ranging from a more powerful GPT-4 Turbo language model to a custom GPT marketplace.

These upgrades resulted in a "surge of usage," according to OpenAI CEO Sam Altman. And because of this, the company decided to temporarily pause sign-ups for its $20-per-month ChatGPT Plus subscription service.

https://mashable.com/article/chatgpt-plus-black-market-ebay-resellers

GPU hiánygazdaság

2023 Year in Review: The Great GPU Shortage and the GPU Rich/Poor

NVIDIA becomes the 5th hyperscaler to join the $T+ market capitalization club driven by surging demand from clouds, GPU clouds, big tech and AI startups

CHRIS ZEOLI
JAN 4, 2024

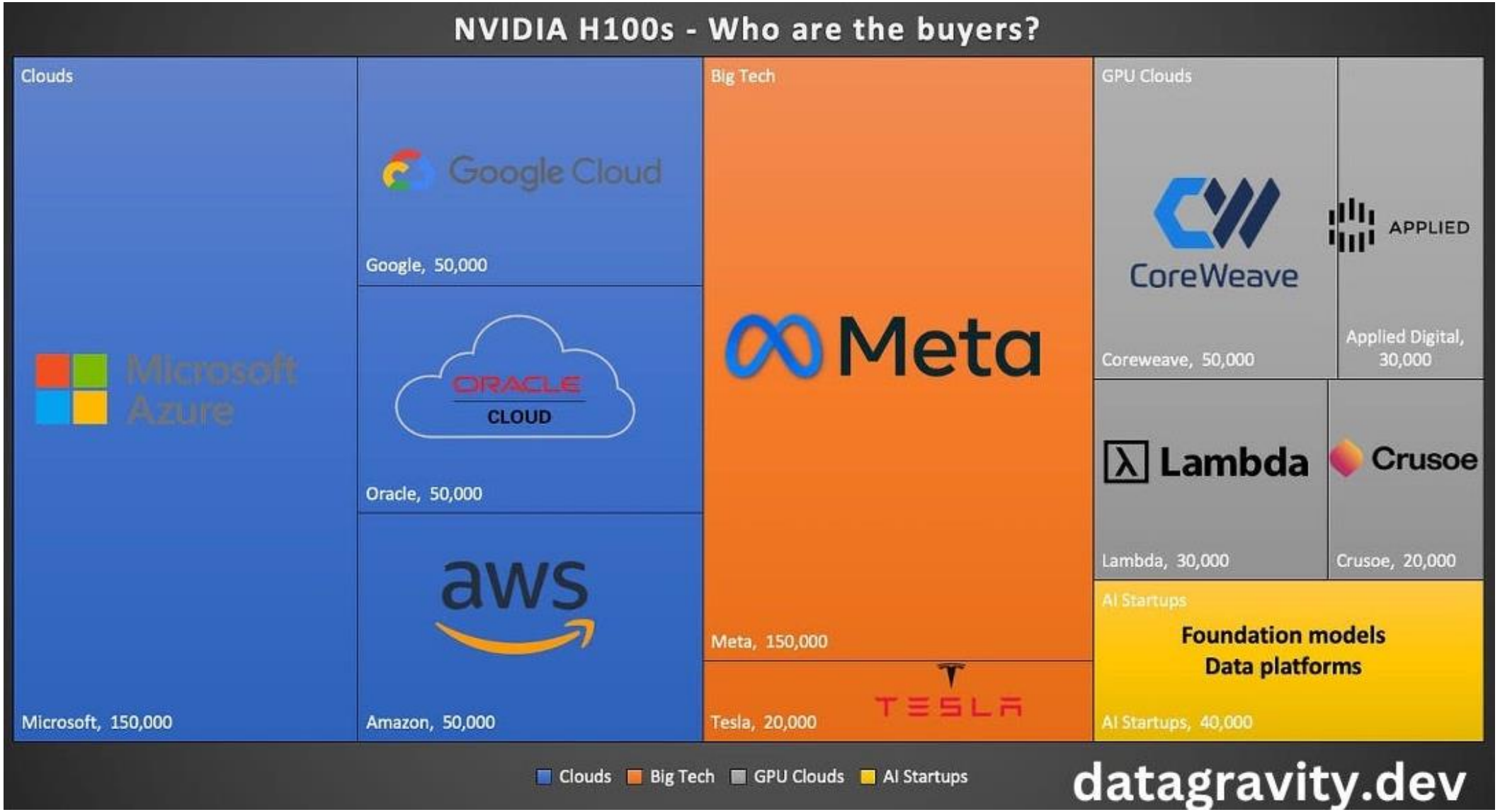
6

Share

In the first half of 2023, OpenAI had all the headlines with the rise of ChatGPT. By May however, the spotlight changed towards the lower-level infrastructure powering LLMs: the GPUs. NVIDIA reported its Q1 earnings, highlighting the company's role as the true backbone of AI in the data center and taking its market capitalization to $1T. It joined elite company of only 4 other tech companies: Apple, Microsoft, Google and Amazon. Today, NVIDIA's market capitalization is >10X that of all foundation model providers, highlighting the key role its chips play in the data center.

https://www.datagravity.dev/p/2023-year-in-review-the-great-gpu

GPU hiánygazdaság



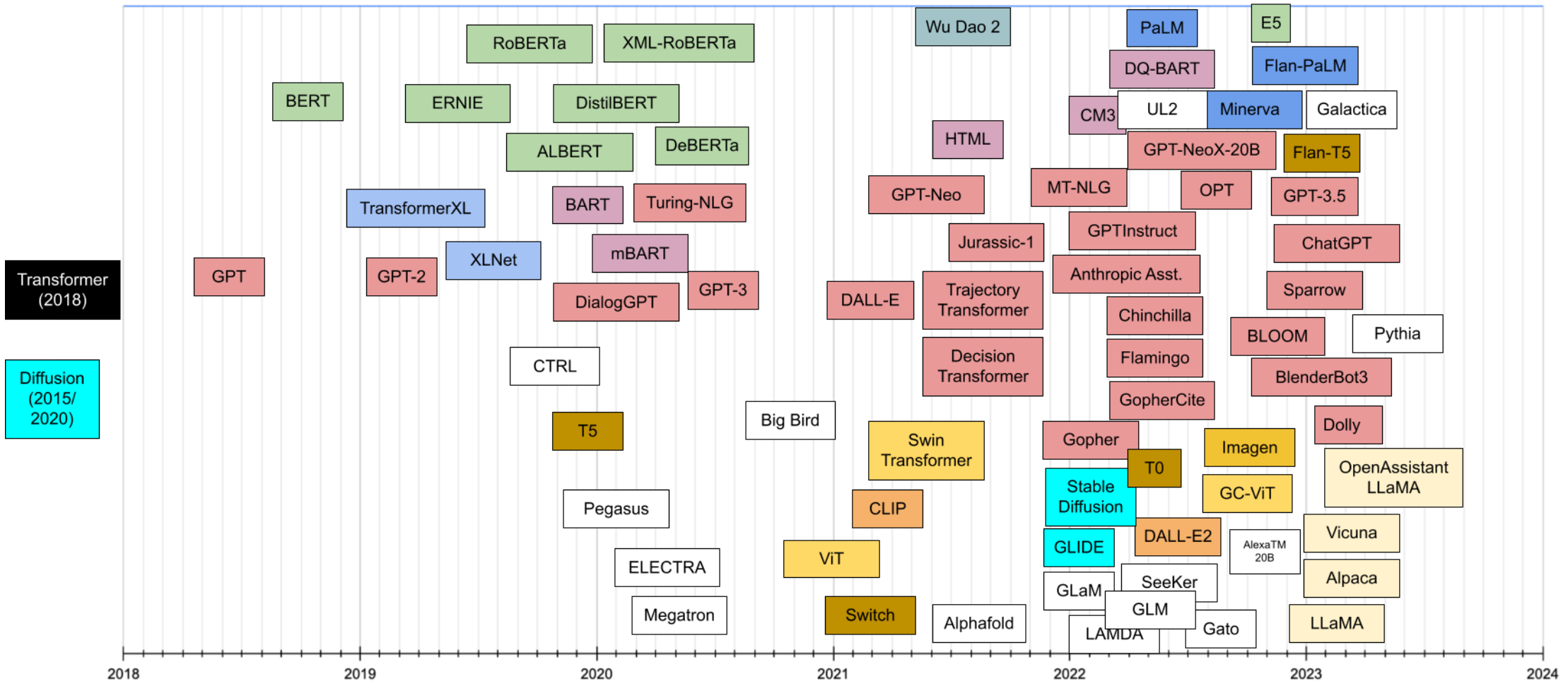
https://www.datagravity.dev/p/2023-year-in-review-the-great-gpu

GPU hiánygazdaság

https://twitter.com/sama/status/1734984269586457078

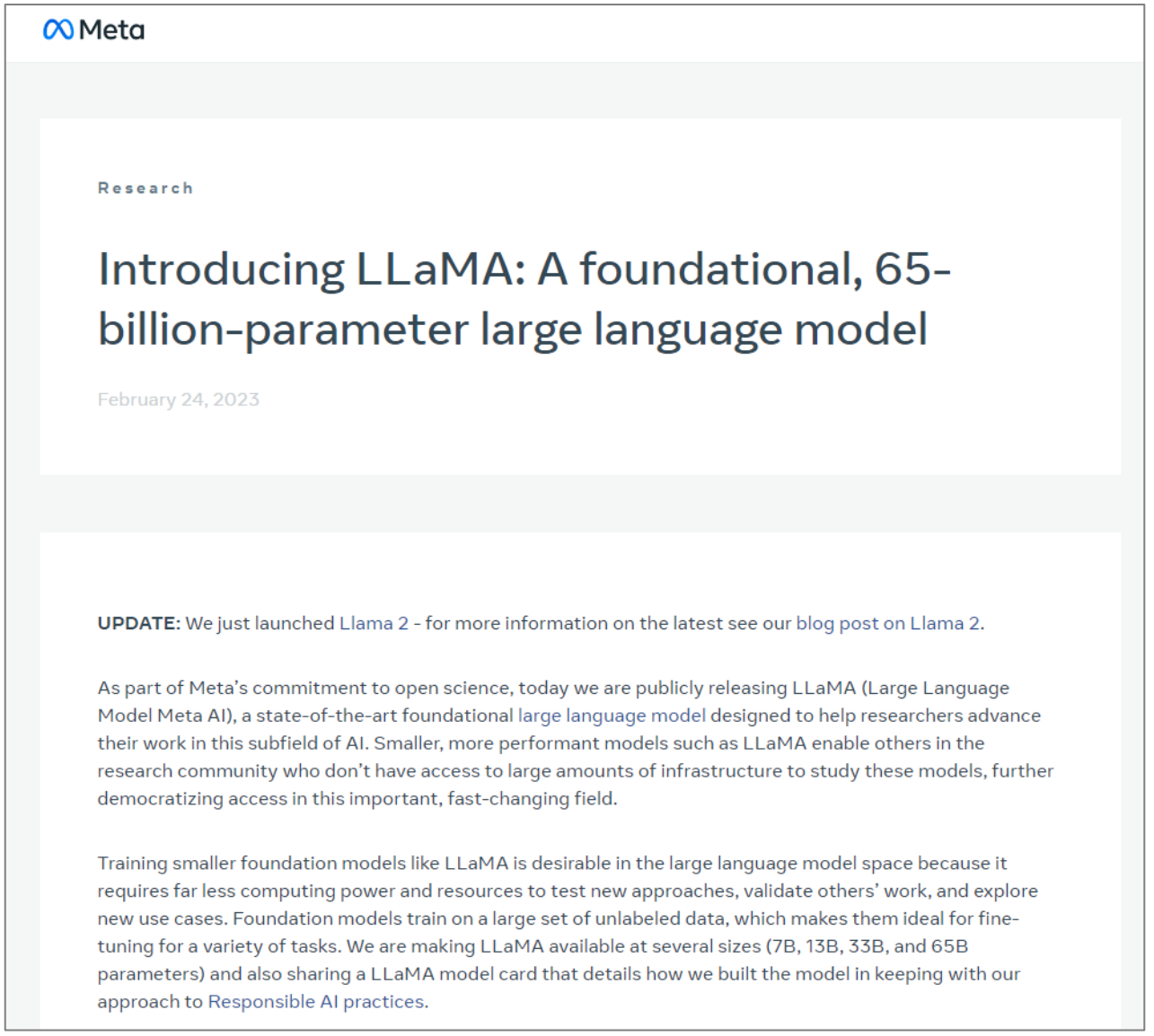
Open Source LLMs

Open Source LLMs



https://blog.genzeon.com/evaluating-open-source-large-language-models-a-practical-guide

Open Source LLMs



Meta

Research

Introducing LLaMA: A foundational, 65-billion-parameter large language model

February 24, 2023

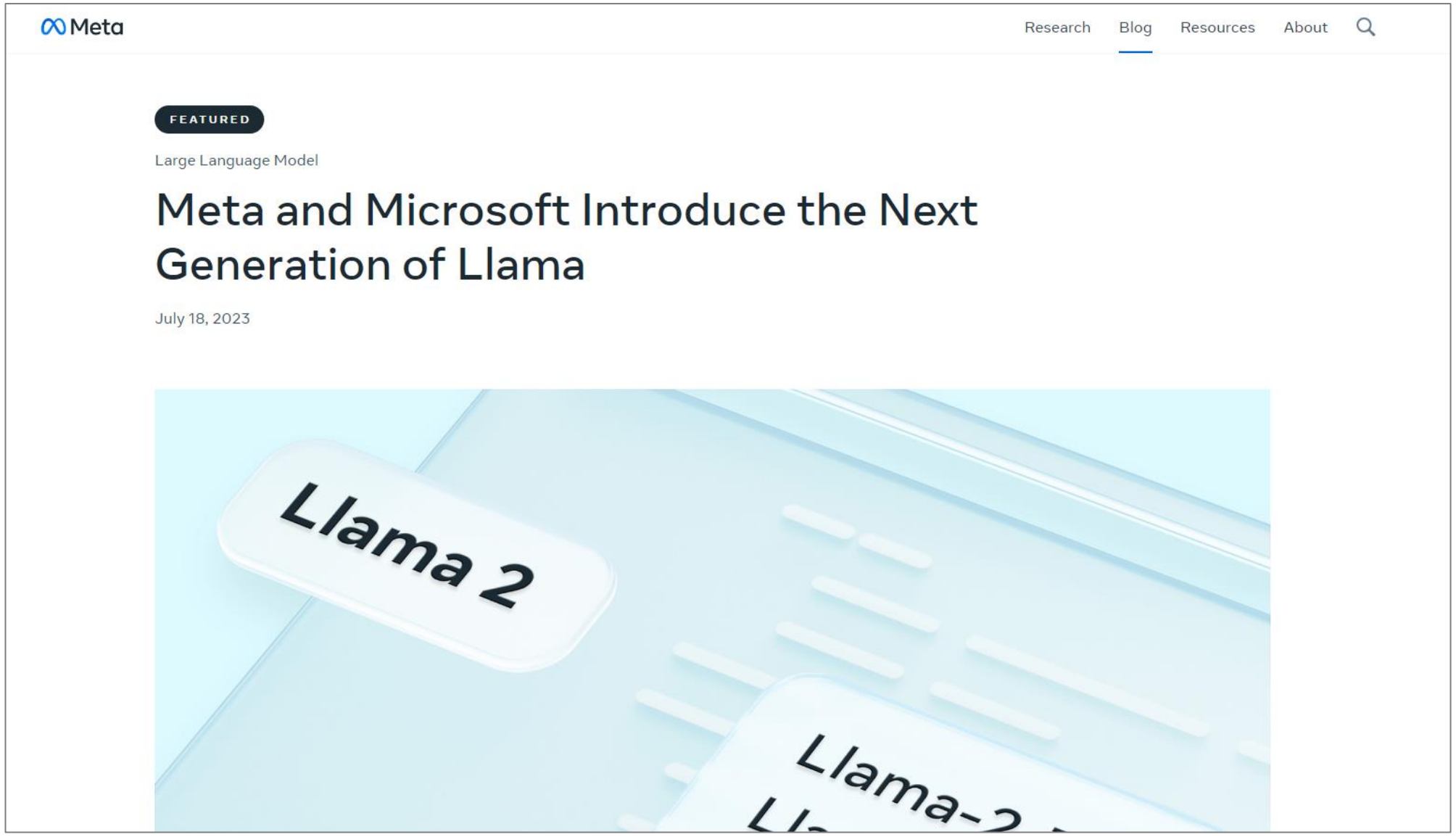
UPDATE: We just launched Llama 2 - for more information on the latest see our blog post on Llama 2.

As part of Meta's commitment to open science, today we are publicly releasing LLaMA (Large Language Model Meta AI), a state-of-the-art foundational large language model designed to help researchers advance their work in this subfield of AI. Smaller, more performant models such as LLaMA enable others in the research community who don't have access to large amounts of infrastructure to study these models, further democratizing access in this important, fast-changing field.

Training smaller foundation models like LLaMA is desirable in the large language model space because it requires far less computing power and resources to test new approaches, validate others' work, and explore new use cases. Foundation models train on a large set of unlabeled data, which makes them ideal for fine-tuning for a variety of tasks. We are making LLaMA available at several sizes (7B, 13B, 33B, and 65B parameters) and also sharing a LLaMA model card that details how we built the model in keeping with our approach to Responsible AI practices.

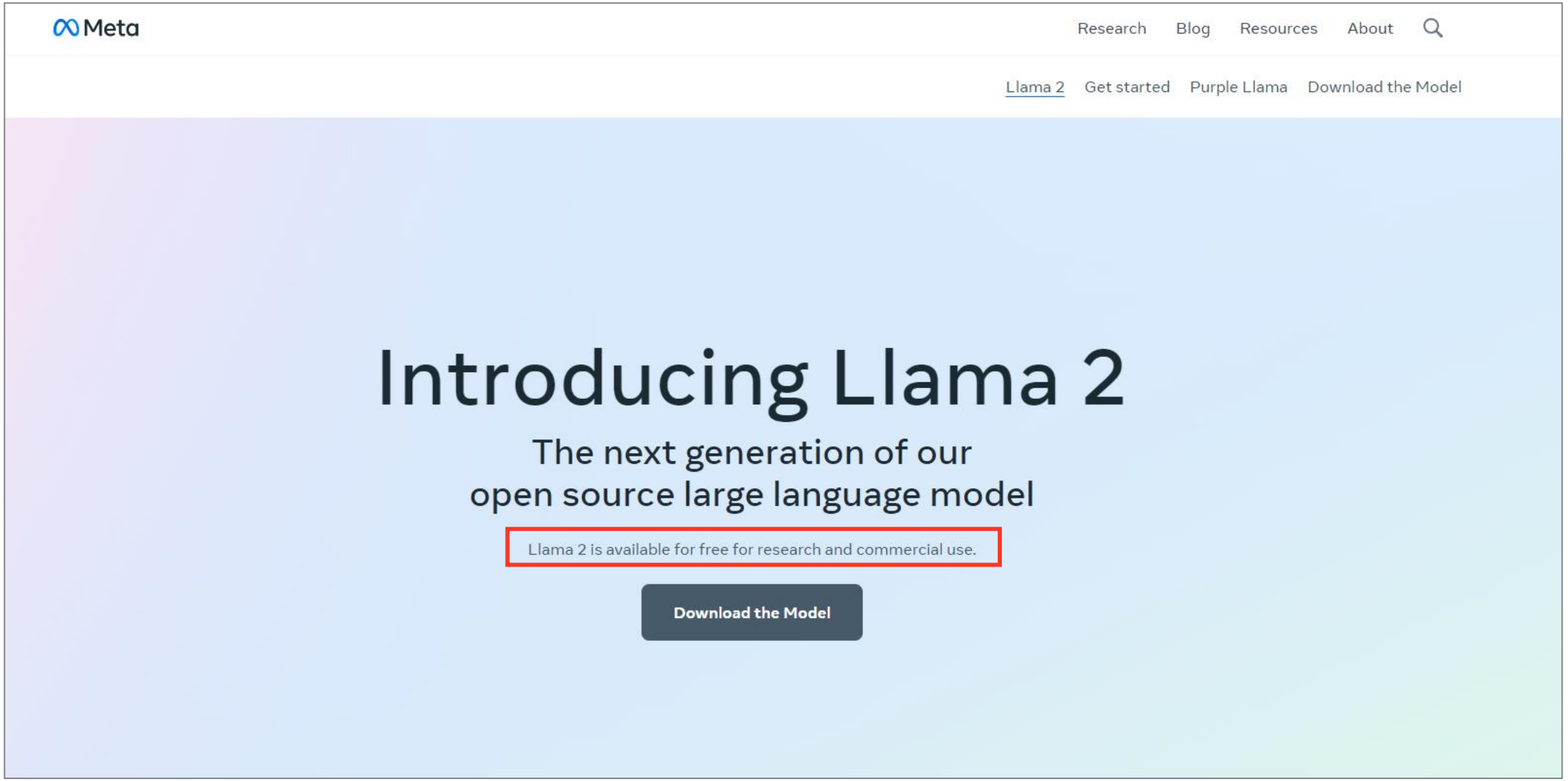
https://ai.meta.com/blog/large-language-model-llama-meta-ai/

Open Source LLMs



https://ai.meta.com/blog/llama-2/

Open Source LLMs



https://ai.meta.com/llama/

Open Source LLMs

Mistral AI's Latest Mixture of Experts (MoE) 8x7B Model

Published 3 weeks ago on December 15, 2023
By **Aayush Mittal**

SEARC

Mistral AI which is a Paris-based open-source model startup has challenged norms by releasing its latest large language model (LLM), MoE 8x7B, through a simple torrent link. This contrasts Google's traditional approach with their Gemini release, sparking conversations and excitement within the AI community.

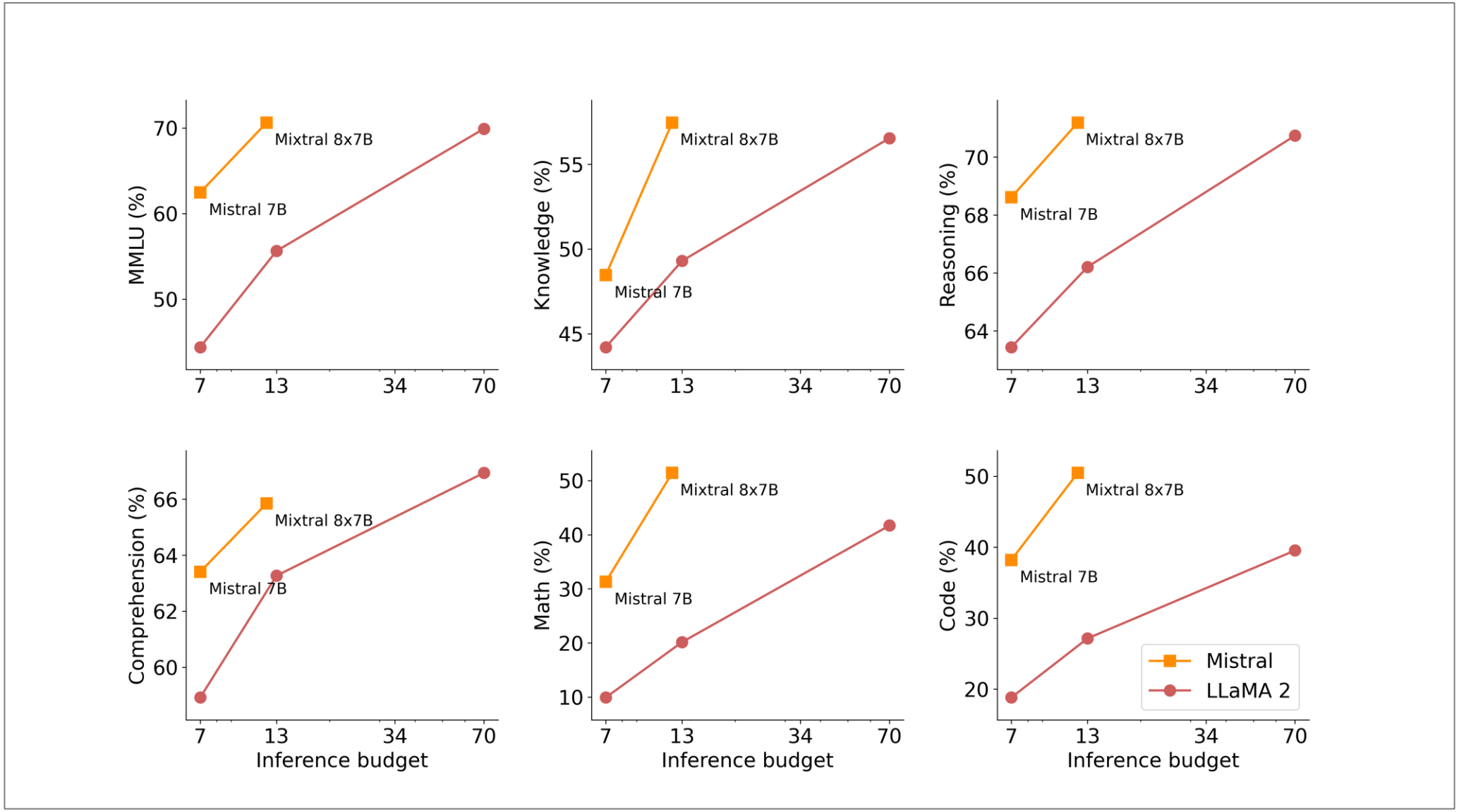
https://www.unite.ai/mistral-ais-latest-mixture-of-experts-moe-8x7b-model/

Open Source LLMs

	LLaMA 2 70B	GPT - 3.5	Mixtral 8x7B
MMLU (MCQ in 57 subjects)	69.9%	70.0%	**70.6%**
HellaSwag (10-shot)	87.1%	85.5%	86.7%
ARC Challenge (25-shot)	85.1%	85.2%	**85.8%**
WinoGrande (5-shot)	**83.2%**	81.6%	81.2%
MBPP (pass@1)	49.8%	52.2%	**60.7%**
GSM-8K (5-shot)	53.6%	57.1%	**58.4%**
MT Bench (for Instruct Models)	6.86	**8.32**	8.30

https://mistral.ai/news/mixtral-of-experts/

Open Source LLMs



https://mistral.ai/news/mixtral-of-experts/

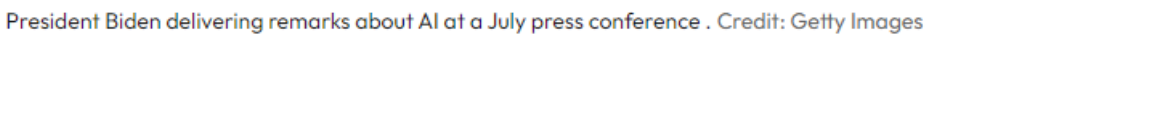
Jog

USA

White House drops an AI regulation bombshell: 10 new mandates that'll shake up the industry

Open AI, Google, Microsoft, and other prominent AI players must answer to the new AI legislation.

By Cecily Mauran and Kimberly Gedeon on October 30, 2023



President Biden delivering remarks about AI at a July press conference . Credit: Getty Images

The White House just announced a thunderous executive order tackling AI regulation. These directives are the "strongest set of actions any government in the world has ever taken" to protect how AI affects American citizens, according to White House Deputy Chief of Staff Bruce Reed.

The Biden administration has been working on plans to regulate the untethered AI industry. The order builds on the Biden-Harris blueprint for an AI Bill of Rights as well as voluntary commitments from 15 leading tech companies to work with the government for safe and responsible AI development.

Instead of waiting for Congress to pass its own legislation, the White House is storming ahead with an executive order to mitigate AI risks while capitalizing on its potential.

https://mashable.com/article/white-house-drops-ai-regulation-bombshell

EU AI Act

Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI

Press Releases IMCO LIBE 09-12-2023 - 00:04

- Safeguards agreed on general purpose artificial intelligence
- Limitation for the of use biometric identification systems by law enforcement
- Bans on social scoring and AI used to manipulate or exploit user vulnerabilities
- Right of consumers to launch complaints and receive meaningful explanations
- Fines ranging from 35 million euro or 7% of global turnover to 7.5 million or 1.5% of turnover

https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai

EU AI Act

Banned applications

Recognising the potential threat to citizens' rights and democracy posed by certain applications of AI, the co-legislators agreed to prohibit:

- biometric categorisation systems that use sensitive characteristics (e.g. political, religious, philosophical beliefs, sexual orientation, race);
- untargeted scraping of facial images from the internet or CCTV footage to create facial recognition databases;
- emotion recognition in the workplace and educational institutions;
- social scoring based on social behaviour or personal characteristics;
- AI systems that manipulate human behaviour to circumvent their free will;
- AI used to exploit the vulnerabilities of people (due to their age, disability, social or economic situation).

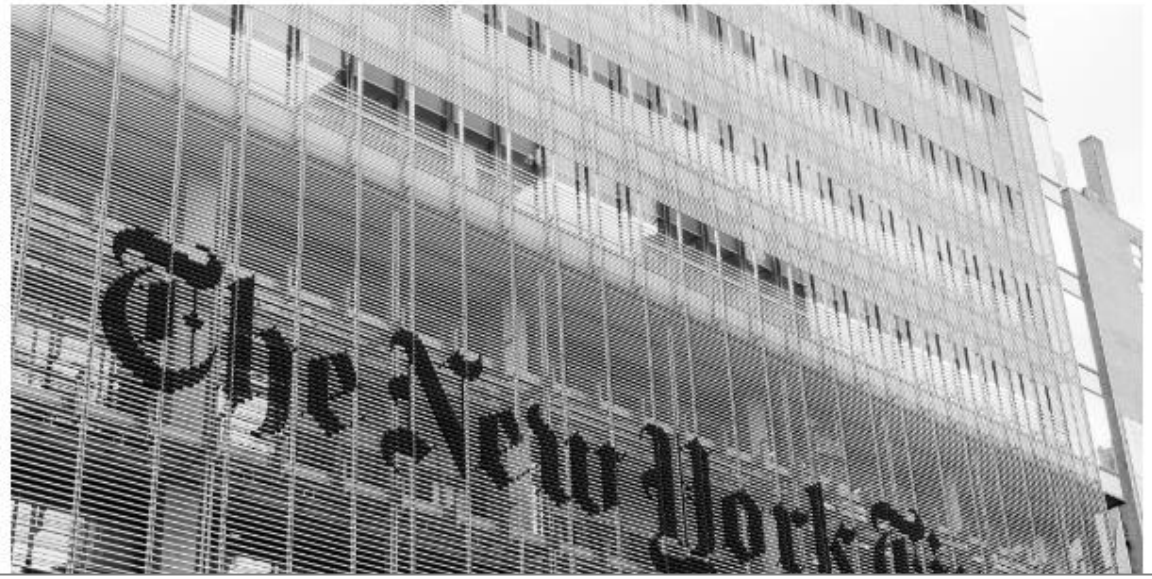
https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai

New York Times vs. OpenAI

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.

Share full article 1.3K



New York Times vs. OpenAI

The New York Times wants OpenAI and Microsoft to pay for training data

Kyle Wiggers @kyle_l_wiggers / 10:24 PM GMT+1 • December 27, 2023

Comment

Image Credits: mbbirdy / Getty Images

The New York Times is suing OpenAI and its close collaborator (and investor), Microsoft, for allegedly violating copyright law by training generative AI models on Times' content.

In the lawsuit, filed in the Federal District Court in Manhattan, The Times contends that millions of its articles were used to train AI models, including those underpinning OpenAI's ultra-popular ChatGPT and Microsoft's Copilot, without its consent. The Times is calling for OpenAI and Microsoft to "destroy" models and training data containing the offending material and to be held responsible for "billions of dollars in statutory and actual damages" related to the "unlawful copying and use of The Times's uniquely valuable works."

"If The Times and other news organizations cannot produce and protect their independent journalism, there will be a vacuum that no computer or artificial intelligence can fill," reads The Times' complaint. "Less journalism will be produced, and the cost to society will be enormous."

https://techcrunch.com/2023/12/27/the-new-york-times-wants-openai-and-microsoft-to-pay-for-training-data/

New York Times vs. OpenAI

The Times is the first major American media organization to sue the companies, the creators of ChatGPT and other popular A.I. platforms, over copyright issues associated with its written works. The lawsuit, filed in Federal District Court in Manhattan, contends that millions of articles published by The Times were used to train automated chatbots that now compete with the news outlet as a source of reliable information.

The suit does not include an exact monetary demand. But it says the defendants should be held responsible for "billions of dollars in statutory and actual damages" related to the "unlawful copying and use of The Times's uniquely valuable works." It also calls for the companies to destroy any chatbot models and training data that use copyrighted material from The Times.

A Python világa

Scikit-learn

Gaël Varoquaux

Mon 27 November 2023

←Home

People underestimate how impactful Scikit-learn continues to be

under scikit-learn open-source machine learning Read time: 3 min. Tweet

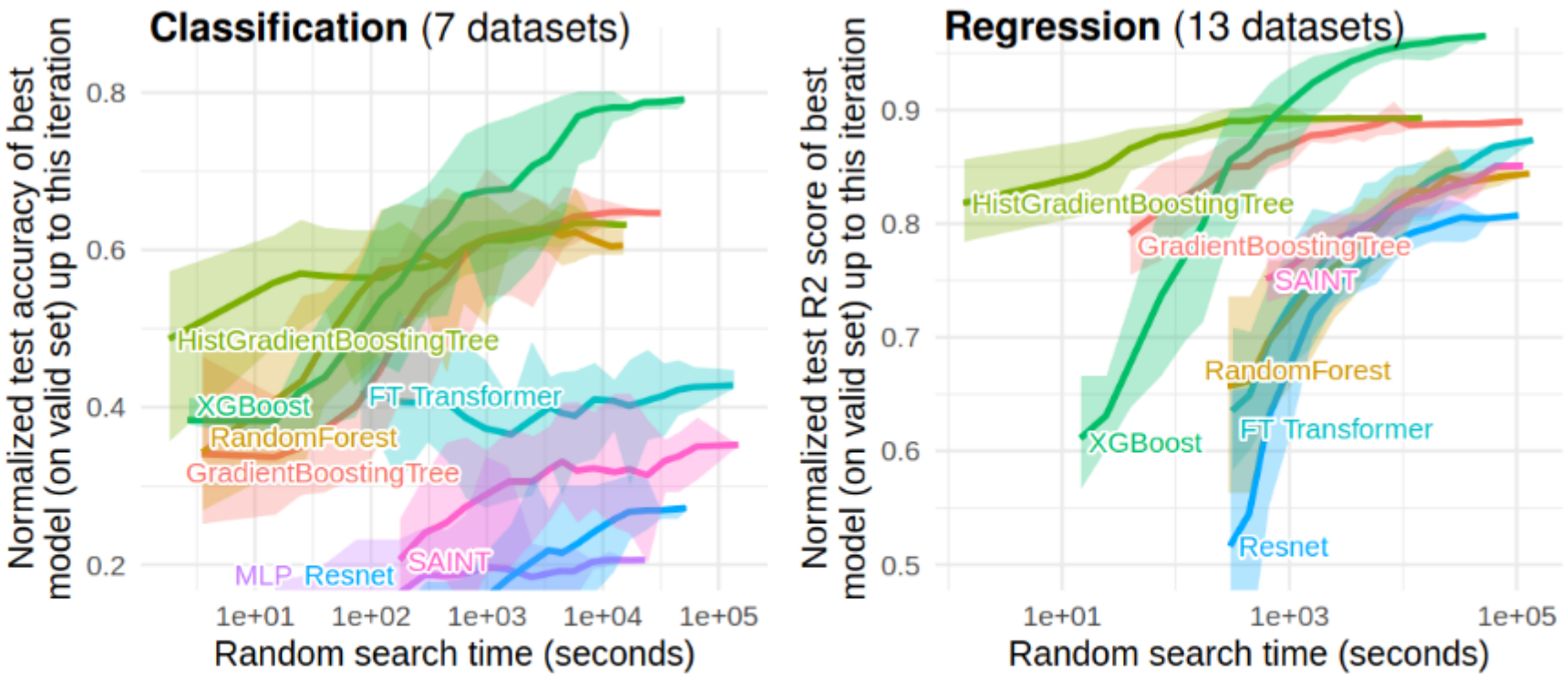
François Chollet rightfully said that people often underestimate the impact of scikit-learn. I give here a few illustrations to back his claim.

A few days ago, François Chollet (the creator of Keras, the library that that democratized deep learning) posted:

> **François Chollet** @fchollet · Nov 22
> People underestimate how impactful Scikit-Learn continues to be.

https://gael-varoquaux.info/programming/people-underestimate-how-impactful-scikit-learn-continues-to-be.html

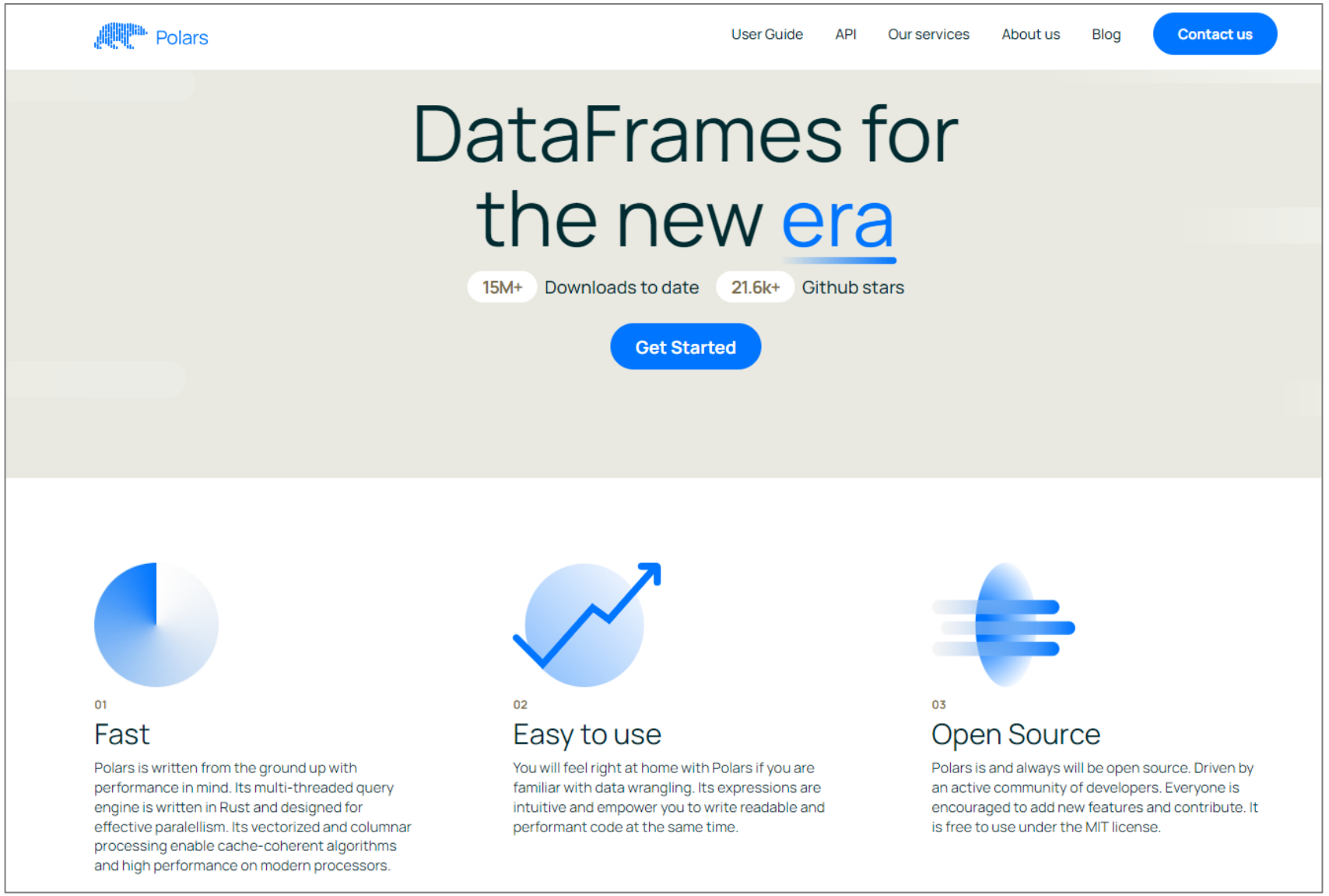
Scikit-learn



Benchmark comparing models on tabular data while tuning hyper-parameters *(from Grinsztajn et al 2023) Each value corresponds to the test score of the best model (on the validation set) after a specific time spent doing random search. The ribbon corresponds to the minimum and maximum scores on these 15 shuffles. Models HistGradientBoostingTree, GradientBoostingTree, and RandomForest come from scikit-learn. FTtransformer, Saint, ResNet and MLP are all deep learning architecture, with FT transformer and Saint models specifically developed for tabular data.*

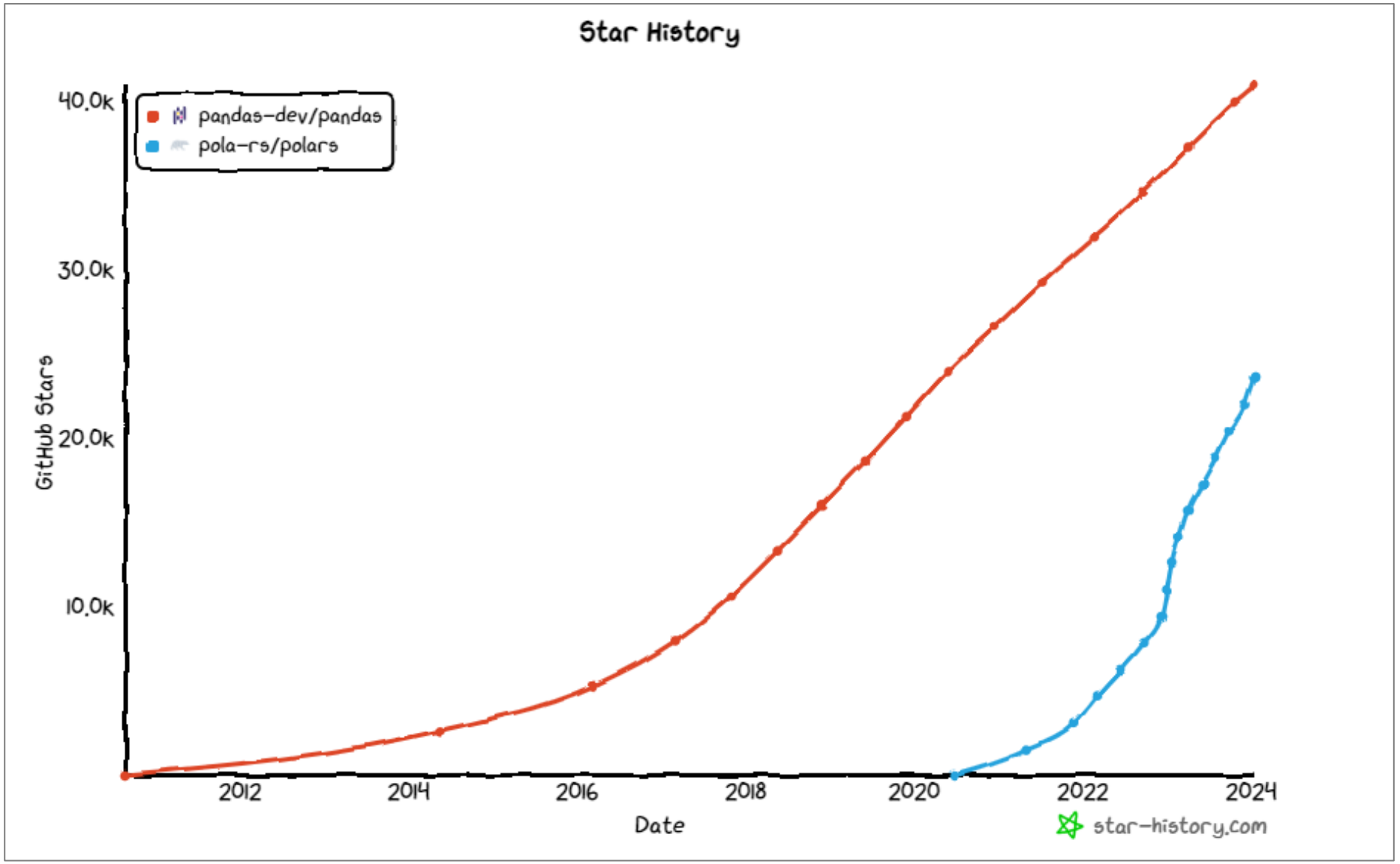
https://gael-varoquaux.info/programming/people-underestimate-how-impactful-scikit-learn-continues-to-be.html

Polars



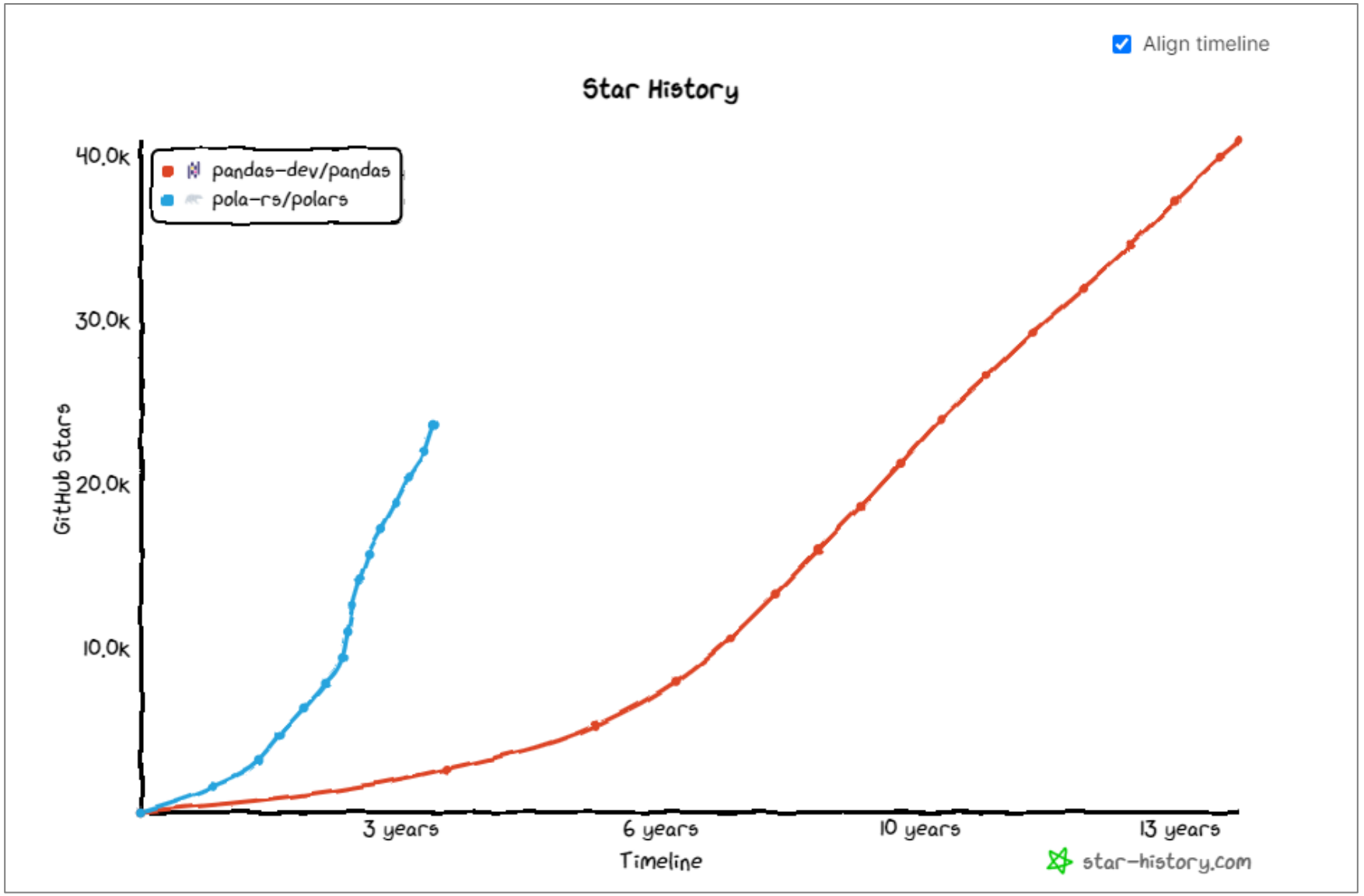
https://pola.rs/

Polars



https://star-history.com/

Polars



https://star-history.com/

Thank You