

 CBINSIGHTS

# The future of large language models

October 2023

## WEBINAR

# About the analyst

Deepashri Varadharajan is passionate about studying new business models and opportunities that AI is helping shape, and how industries are adapting to a rapidly changing landscape.

Deepashri received her undergraduate degree in electronics and communications engineering from VIT University in India and an MS from the Columbia University Graduate School of Journalism.



**Deepashri Varadharajan**  
Director, AI & New Tech  
research



CB Insights allowed us to build a **clear view of the comprehensive AI landscape** and draw out trends through a company by company analysis.



**David Titus**

Former Principal and Team Lead, IBM

**We collect an immense amount of quantitative and qualitative information...**

**...to create technology company and market intelligence...**

**...that we can deliver to you in a number of ways.**



### Company Briefings

Company briefings that illuminate:

- Pricing
- Competitive differentiation
- Revenue
- Headcount
- Product demos



### Software Buyer Interviews

Analyst-led interviews with software buyers that cover:

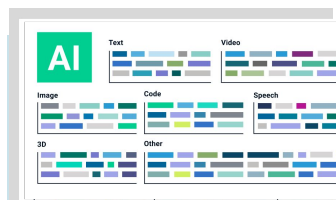
- What drove evaluation
- Who was evaluated
- Who you selected & why
- Pricing
- Post-deployment experience
- CSAT (customer satisfaction)



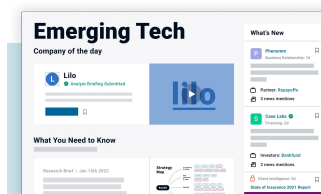
### Machine Intelligence

Our software crawls millions of sources daily to extract & build:

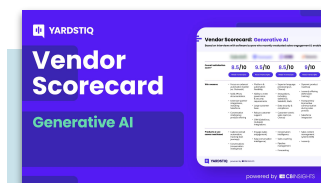
- Company profiles
- Business relationships (vendors, partners, customer)
- Customer references
- Market sizings
- Transaction data (financing, M&A, IPOs, valuations)



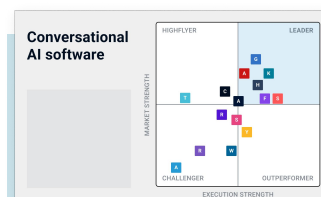
### Tech market maps



### Tech market profiles



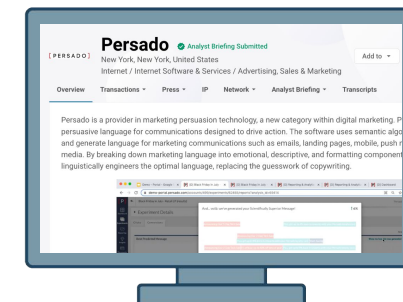
### Vendor comparison scorecards



### Vendor rankings

### CBI Application

Web & mobile interface



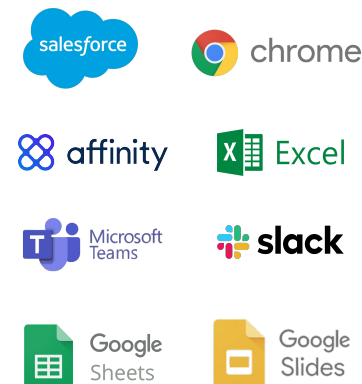
### API & Data Feeds

Integrate our data into internal and commercial applications



### Integrations

Bring our data to the places you work every day





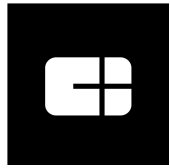
# Webinar agenda

---

3 Key trends to watch  
in large language models

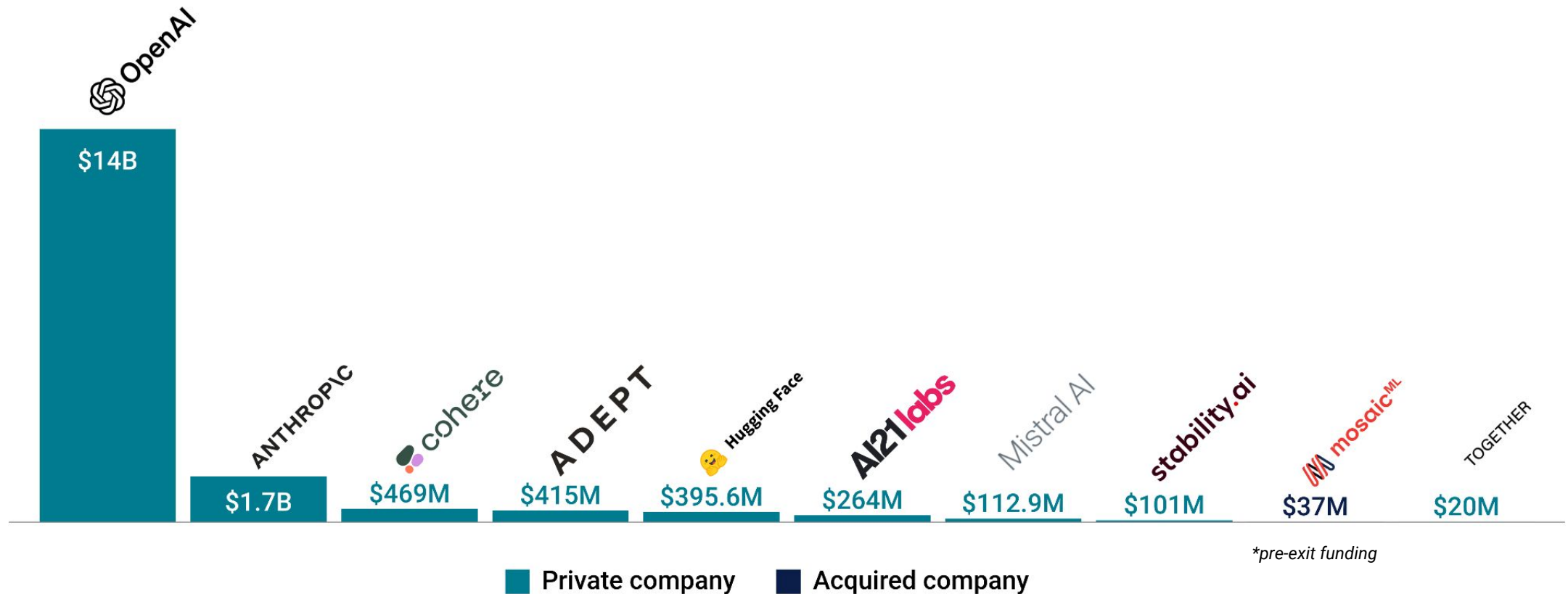
## Key trends

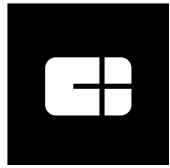
- 1. No clear winner yet in foundation models**
2. LLM infrastructure market is rapidly growing
3. Rise of open-source LLMs



# LLM startups emerging to compete with OpenAI

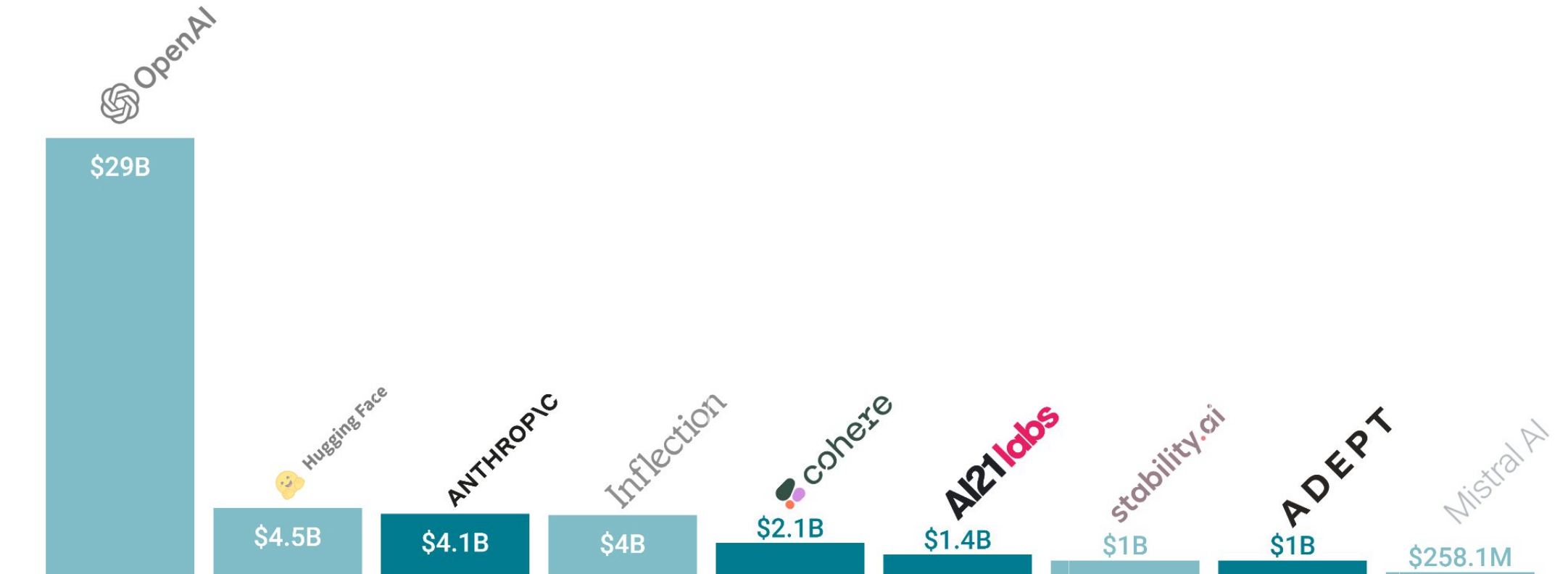
Disclosed equity funding to LLM developers





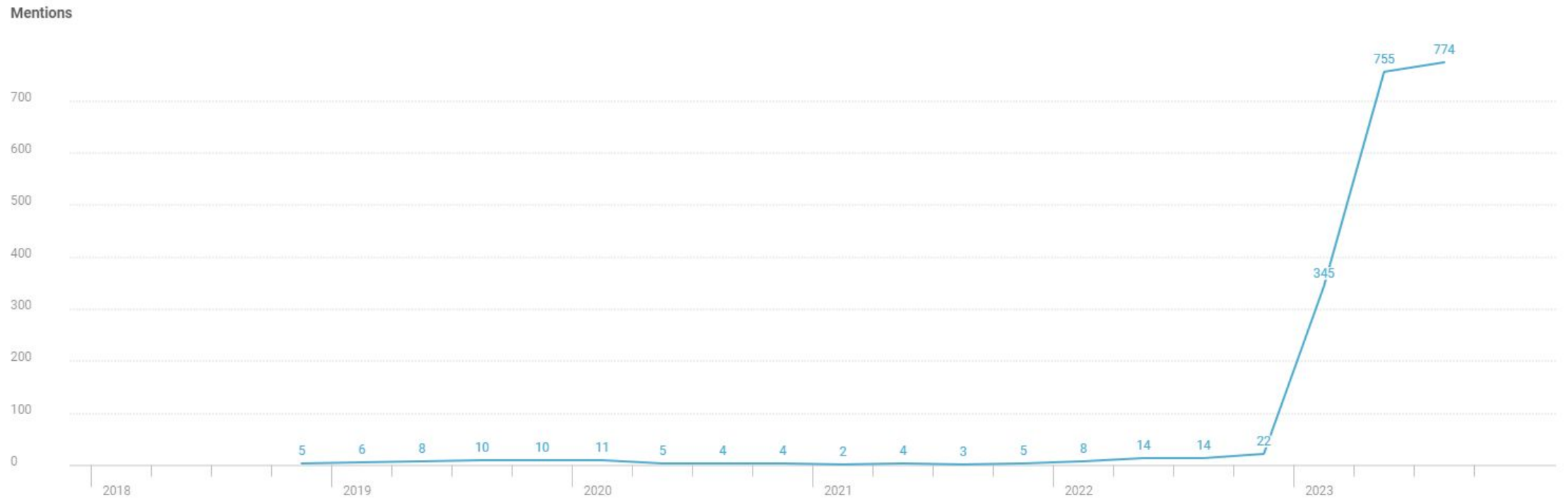
# Four developers joined the unicorn club in 2023

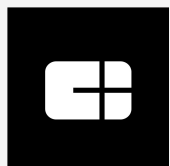
Disclosed valuations of LLM developers



# Executive interest in LLMs is at an all-time high...

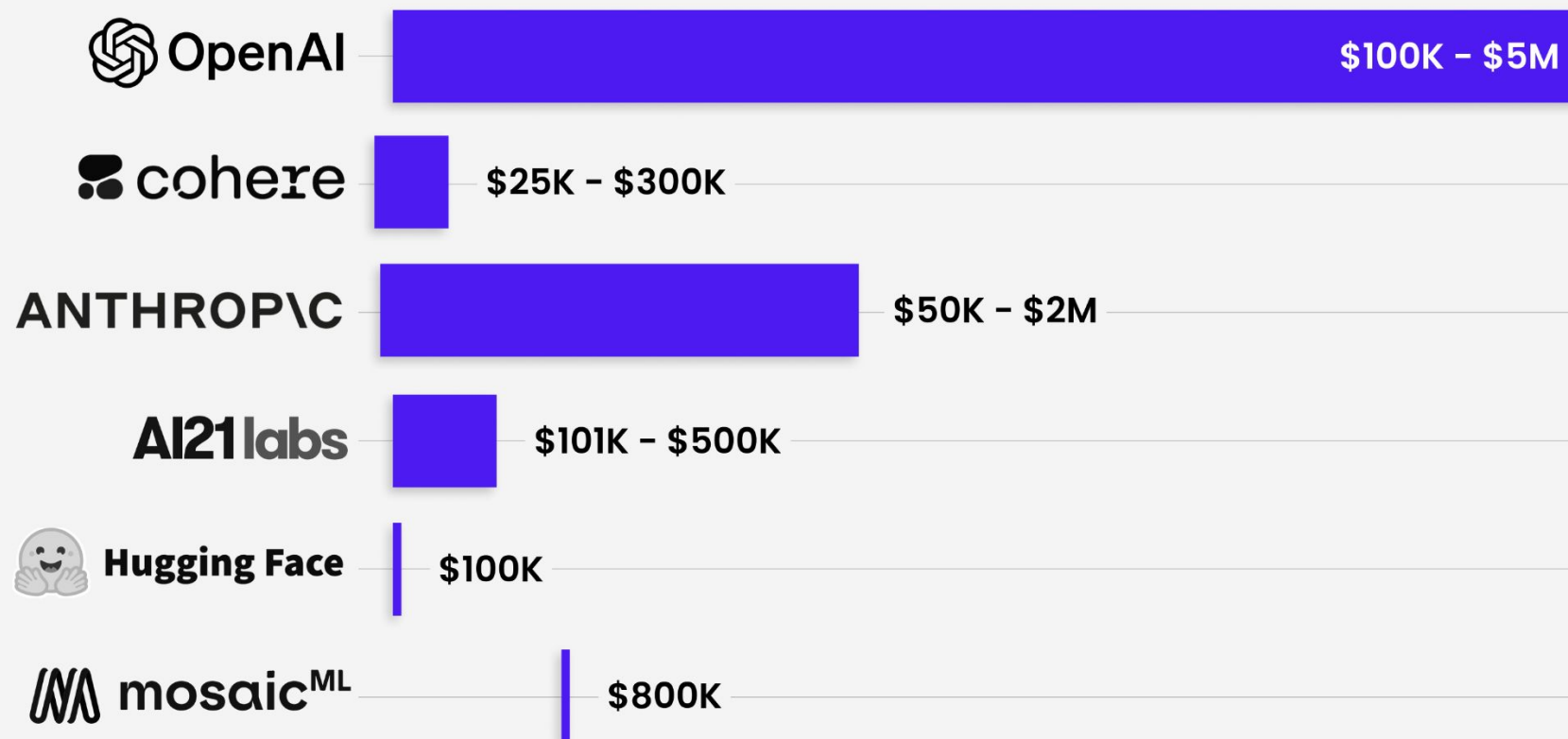
Earnings call mentions of LLM-related terms



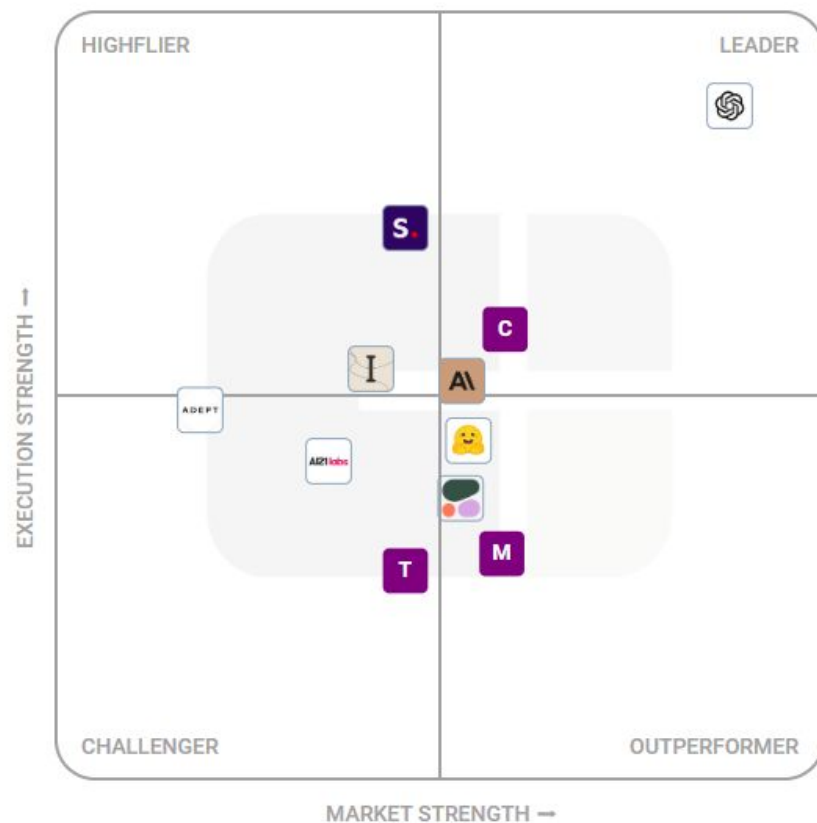


# ...with budget allocations entering the millions

Annual spend ranges displayed



# Vendors are competing on multiple fronts to become the go-to model developer



## Key KPIs for evaluation

- Pricing and deployment
- Privacy & ethics
- Core NLP tech
- Speed & accuracy
- Token limits

## ESP ranked companies

### Leader



OpenAI



Contextual AI



Anthropic

### Highflier



Stability AI



Inflection AI

### Outperformer



Hugging Face



Cohere



Mistral AI

### Challenger



AI21 Labs



Together



Adept

# Azure integration gave OpenAI a leg up with this healthcare buyer

Buyer evaluated healthcare-specific LLMs

“Advantage is with, for example, John Snow Labs, it's a very sort of clinically trained model... It's not just trained on wiki pages or like general text. In that sense, I think it's much better...in terms of entity recognition and things like that.”

VP, Machine Learning at Fortune 500 company



YARDSTIQ

powered by CBINSIGHTS

Advantage of OpenAI

“But the limitations, I would think, are these models are not getting trained in on the volume of data anywhere as close to what ChatGPT is trained on... It [OpenAI deployment] was pretty minimal overhead... the goal... is to essentially enable the use of ML tools that are available from the Azure subscription at the Enterprise level...”



YARDSTIQ

powered by CBINSIGHTS

VC-backed entertainment company using AI21 Labs

# Pricing will influence this gaming startup's intent to renew

## AI21 labs

CSAT score: 7

"For the vendor to get to a 9.5 or 10 satisfaction score, they would need to continue to compete on cost level.

OpenAI just lowered their costs by 10x this past week, so that's a big change...

The vendor would need reinforcement learning as a must; that's what ChatGPT is doing so well."

C-level executive at VC-backed entertainment company



YARDSTIQ

powered by



CBINSIGHTS

# This buyer chose Anthropic for higher safety standards

Purchase amount: \$50K/year

## Key requirements

“The specific requirement was safety requirements around ensuring that no private data or personal data is exposed. Anthropic model performed higher than other models in this regard. There were other requirements around groundedness, specificity, sensibleness that we evaluated across these models....Then, there is the nonfunctional metrics like speed and efficiency, throughput, latency that we also evaluated Anthropic on...”

 **YARDSTIQ** powered by  **CBINSIGHTS**

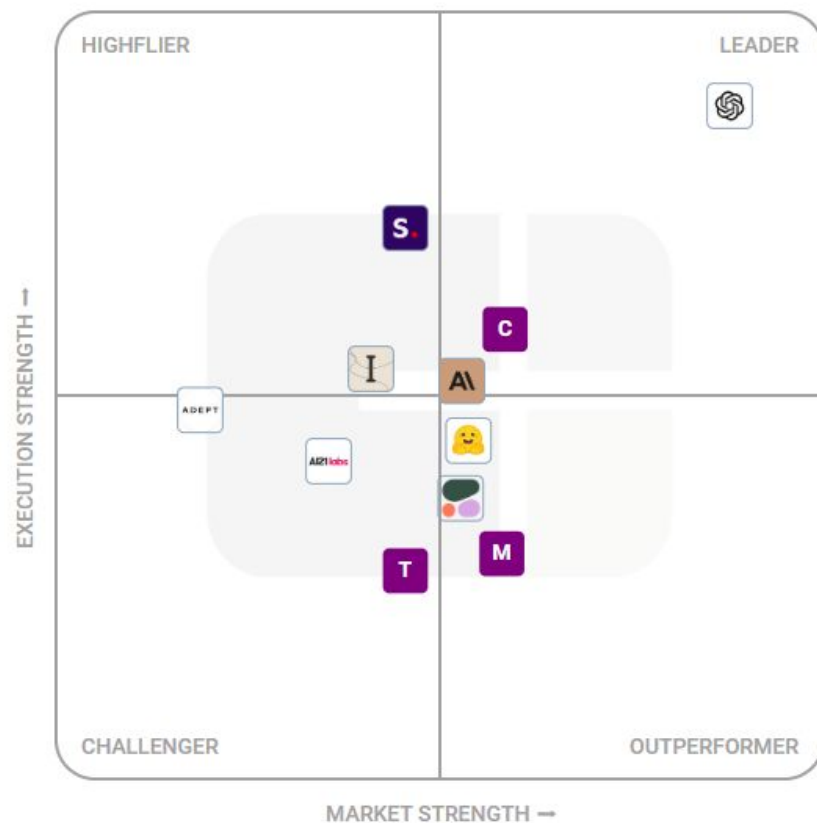
## Strengths & weaknesses

“Strengths, I would say, the ethical considerations of privacy and bias, fairness... their model outperformed the other models, including GPT-3 and ChatGPT... In terms of weaknesses, the specificity of the model output and the interestingness of the model output... I think that other weakness also was in terms of speed and efficiency, like latency, and once you ask a question, how long does it take to fully respond.”

Senior manager of data science at a model-as-a-service platform

 **YARDSTIQ** powered by  **CBINSIGHTS**

# Vendors are competing on multiple fronts to become the go-to model developer



## Key KPIs for evaluation

- Pricing and deployment
- Privacy & ethics
- Core NLP tech
- Speed & accuracy
- Token limits

## ESP ranked companies

### Leader



OpenAI



Contextual AI



Anthropic

### Highflier



Stability AI



Inflection AI

### Outperformer



Hugging Face



Cohere



Mistral AI

### Challenger



AI21 Labs



Together



Adept

## Key trends

1. No clear winner yet in foundation models
- 2. LLM infrastructure market is rapidly growing**
3. Rise of open-source LLMs

Why executives are buying  
genAI infrastructure tools

# Better training data for generative AI applications



“...we had, basically, messy data and we needed a better way of providing, of doing better training data for like generative content ... We have a lot of machine learning models that are fueling us here. In both cases, it was the desire to have, basically, a higher level of data hygiene in our training data.”

Chief Product Officer at IT company

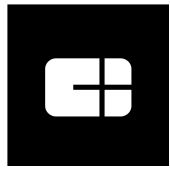
Why executives are buying  
genAI infrastructure tools

# Observe performance of models in production



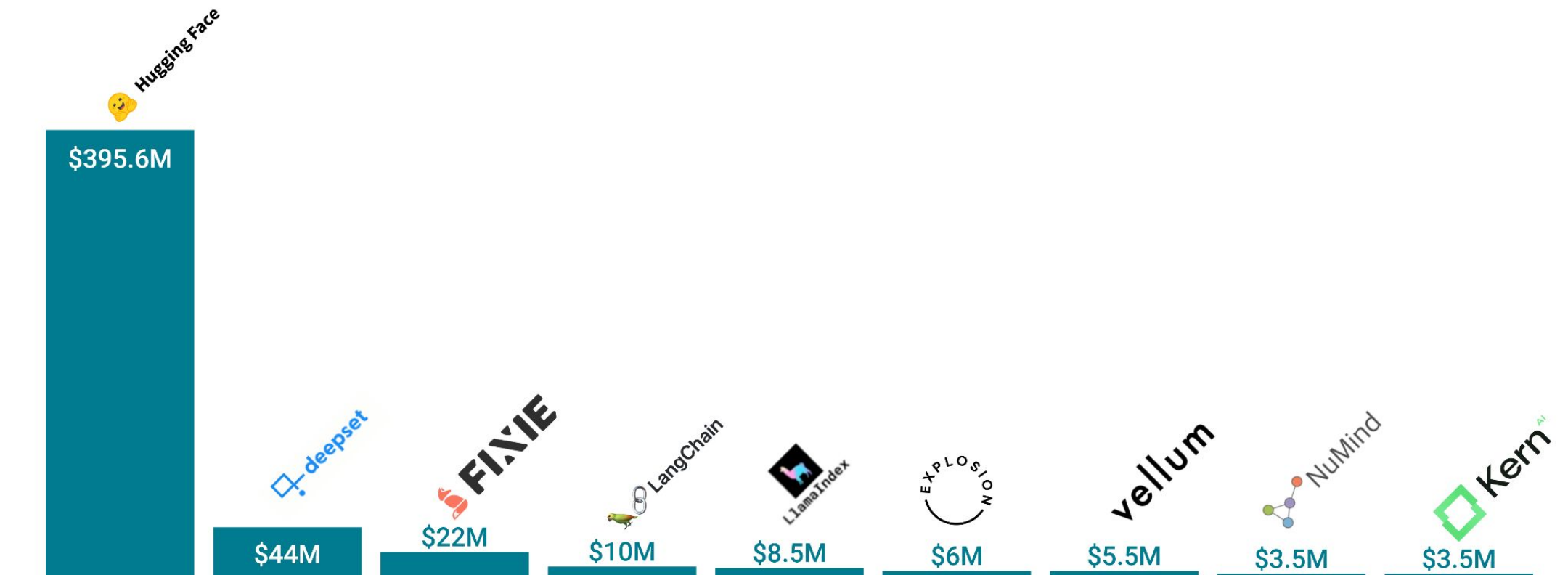
I led a small data science team that created a lot of models in production. As we scaled, our observability of our machine learning models in production was limited, and we felt blind to issues or ways to improve the model once in production. We tried to build a solution in-house, which showed the difficulty of the challenge.

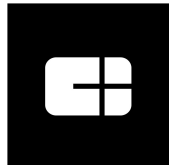
Senior Manager at \$10M+ funded data analytics platform



# New vendors are emerging for LLM fine-tuning & customization

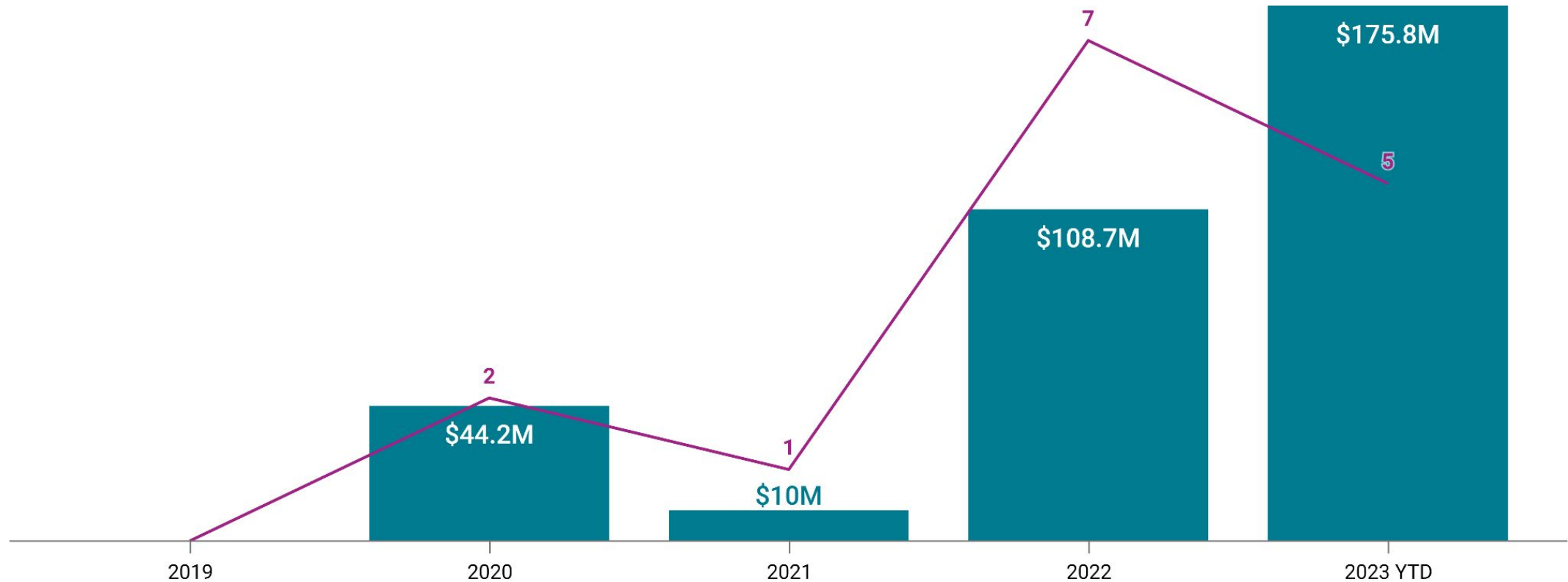
Disclosed equity funding

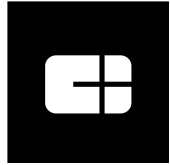




# Vector database startups raised record funding amid LLM boom

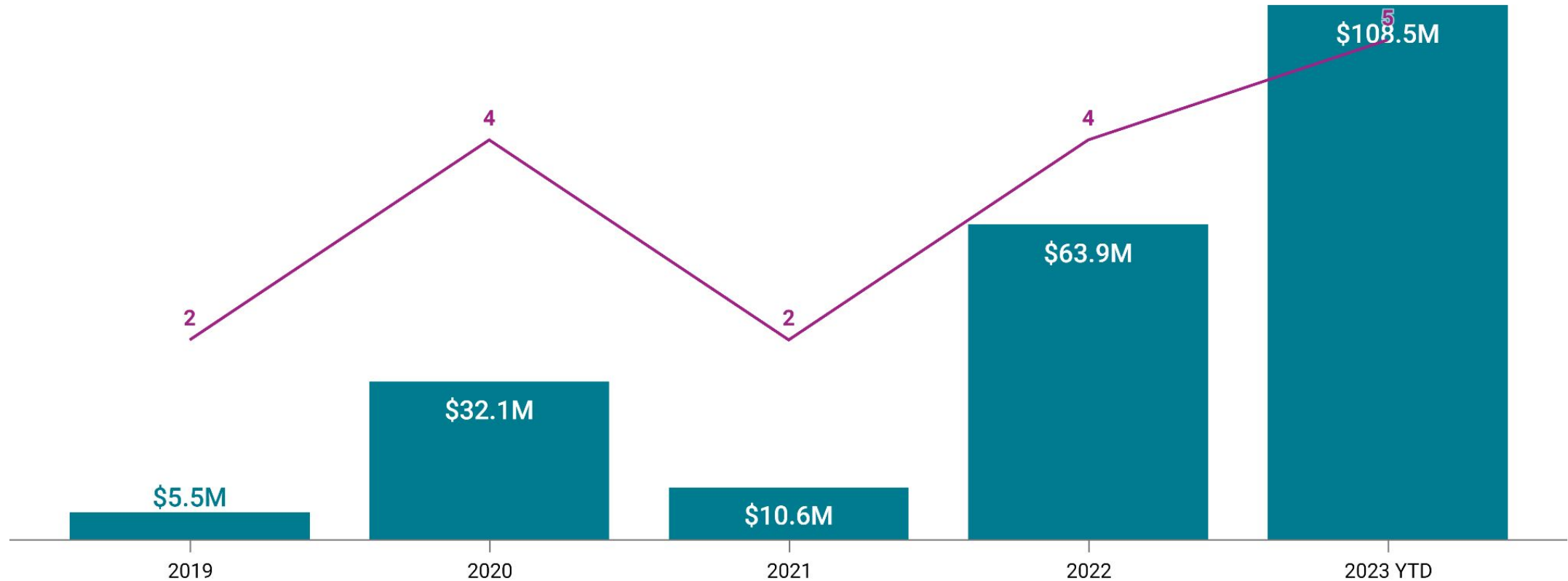
Disclosed equity funding





# Machine learning security is gaining traction

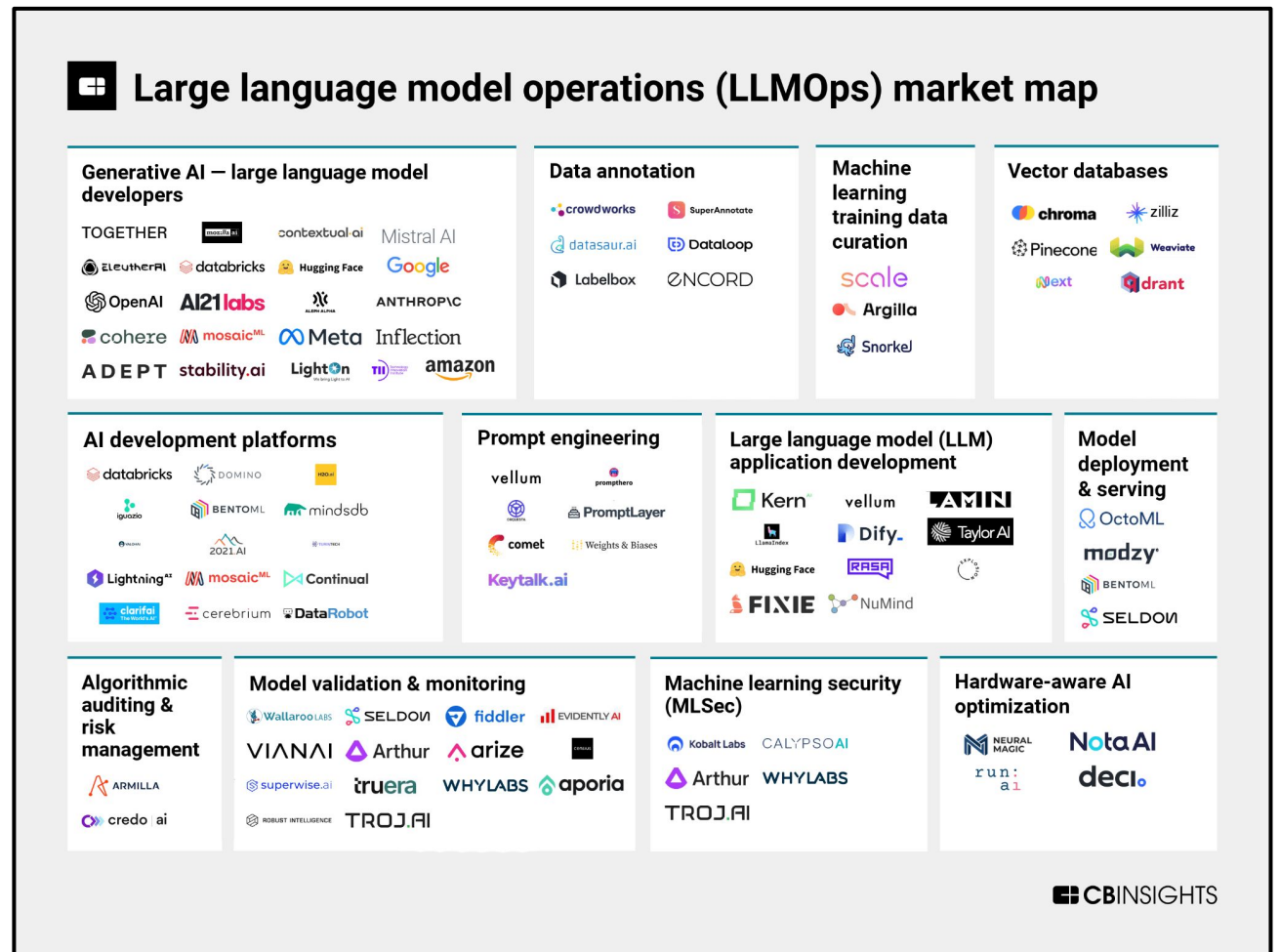
Disclosed equity funding to MLSec startups



# We mapped out end-to-end LLM development tools

## Market map

90+ large language model operations (LLMOps) vendors

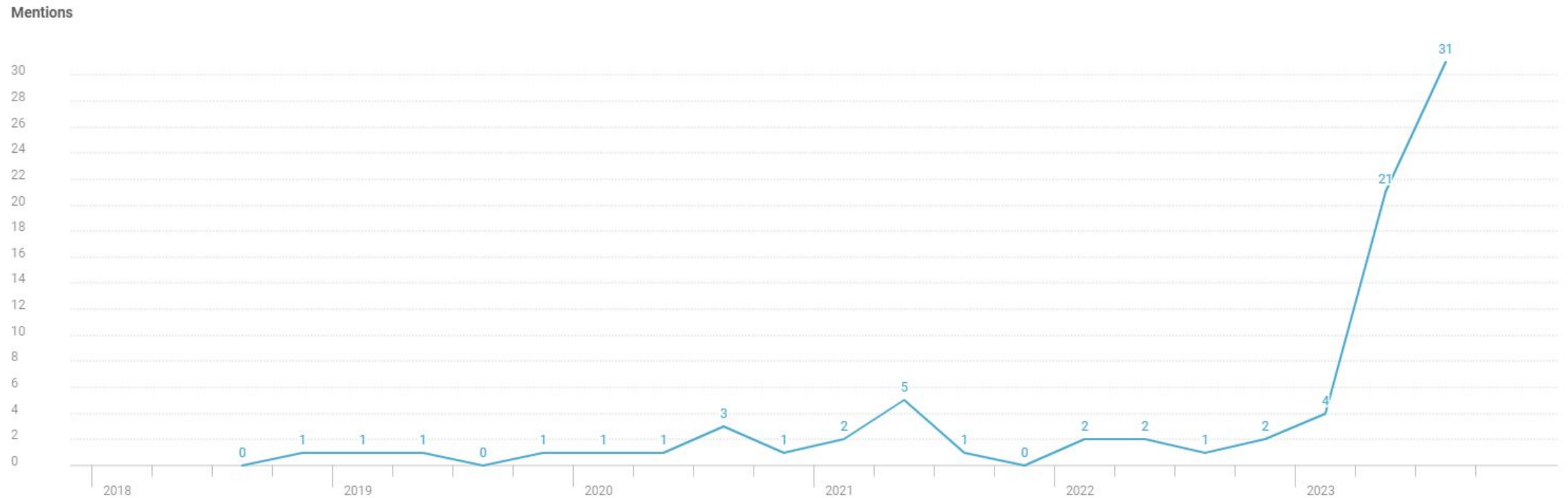


## Key trends

1. No clear winner yet in foundation models
2. LLM infrastructure market is rapidly growing
- 3. Rise of open-source LLMs**

# Need for transparency and rapid innovation is fueling the open-source AI movement

Earnings calls mentions of open-source AI



# Meta is leveling the playing field with Llama-2

## Qualcomm

**“Notably, we recently announced a collaboration with Meta on Llama 2-based AI implementations on flagship smartphones and PCs that will enable developers to create new and exciting Gen AI applications using the AI capabilities of Snapdragon platforms beginning in 2024.”**

Qualcomm CEO Cristiano Amon  
Q3'23 earnings calls

## Microsoft

**“And Azure AI is ushering in new born-in-the-cloud AI-first workloads with the best selection of frontier and open models, including Meta's recent announcements supporting Llama on Azure and Windows, as well as OpenAI. ”**

Microsoft CEO Satya Nadella  
Q3'23 earnings calls



**“AI is the flavor of the day. And thanks to ChatGPT great launch, everyone has discovered this potential. We can expect new changes by the day in the tech world. Just take yesterday, Meta launch of Llama 2, which will be available on Azure free of charge, including for commercial purposes.”**

Publicis CEO Arthur Sadoun  
Q3'23 earnings calls

## Llama-2 vs ChatGPT

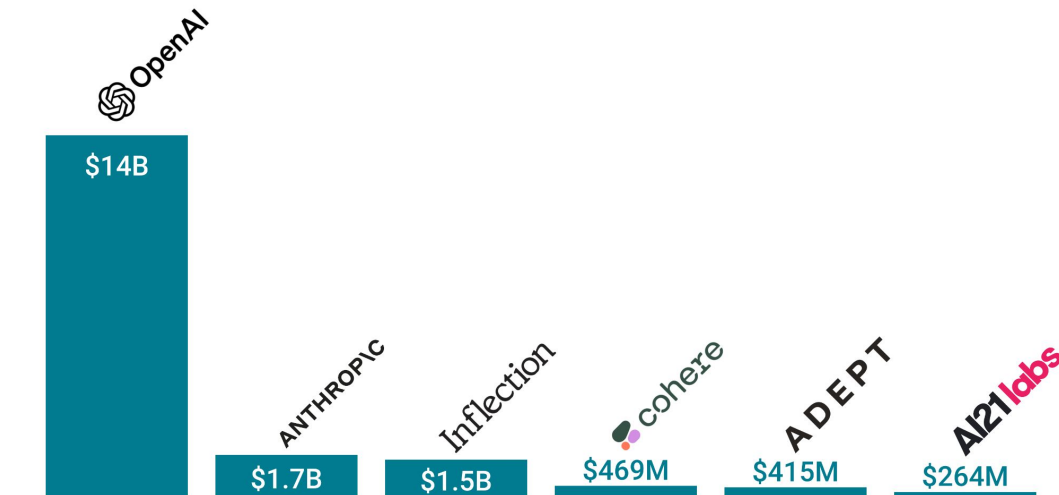
**“...Meta's invention... it may be cheaper than the OpenAI [model] because it's open source. Then we believe the performance of the Hugging Face and also the Llama-2 is also comparable to the OpenAI [model]. Maybe just a little bit weaker than that, but maybe the overall...ROI is quite a good deal.”**

— Cloud. Data & AI lead at Fortune 500 company

# The private market is split into open vs closed

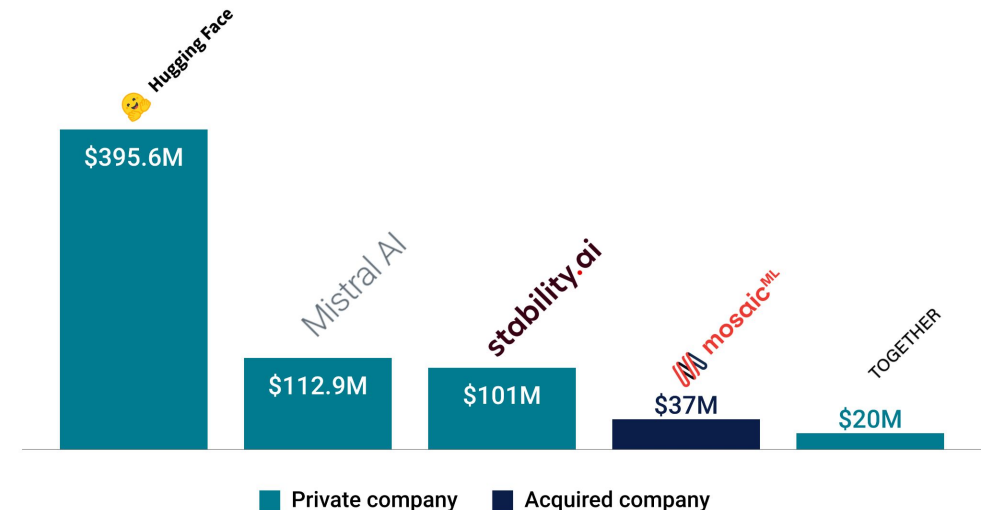
Disclosed equity funding to LLM developers

## Closed-source LLMs



*\*some developers may offer open-source versions of their models but keep their core models proprietary*

## Open-source LLMs

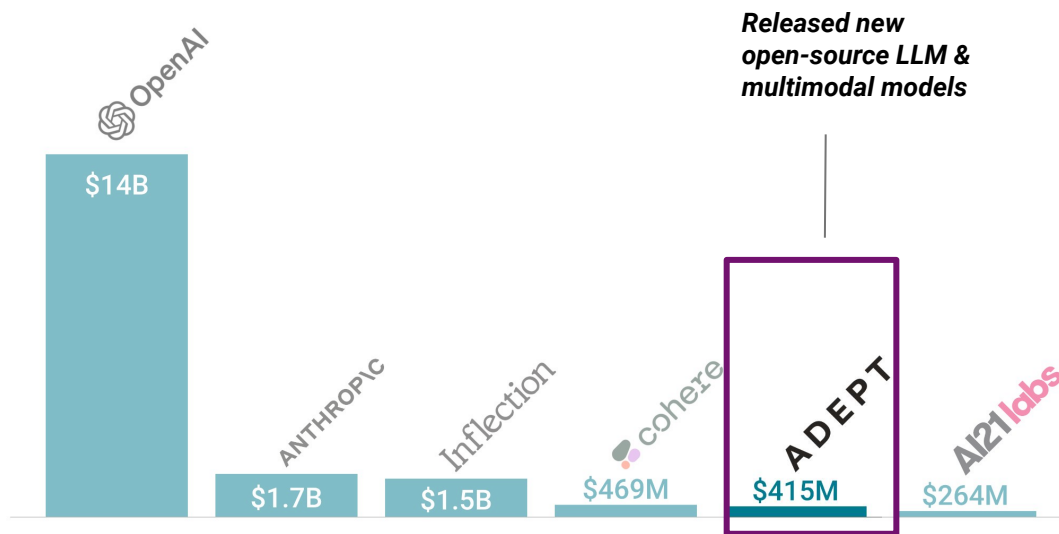


*\*excludes open-source developers that have not raised equity funding*

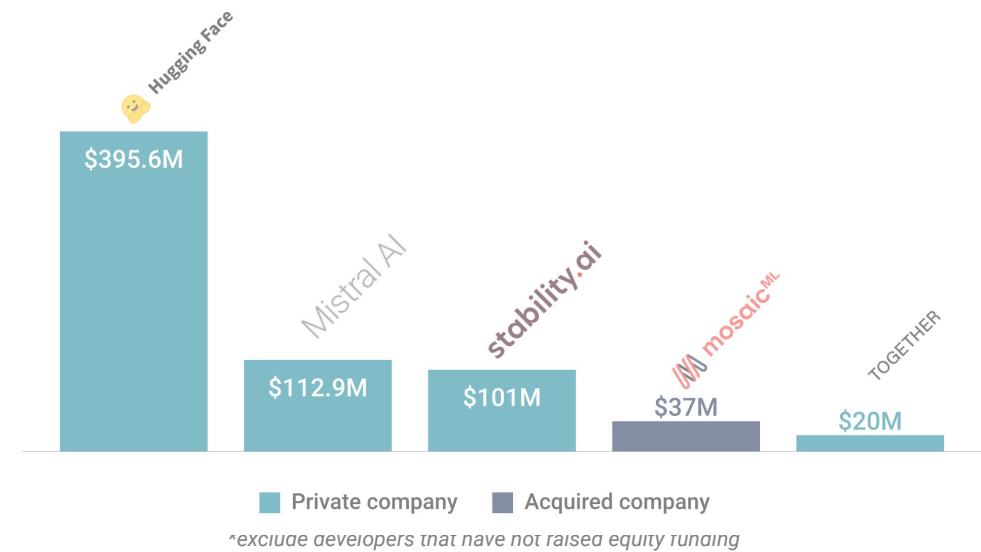
# The private market is split into open vs closed

Disclosed equity funding to LLM developers

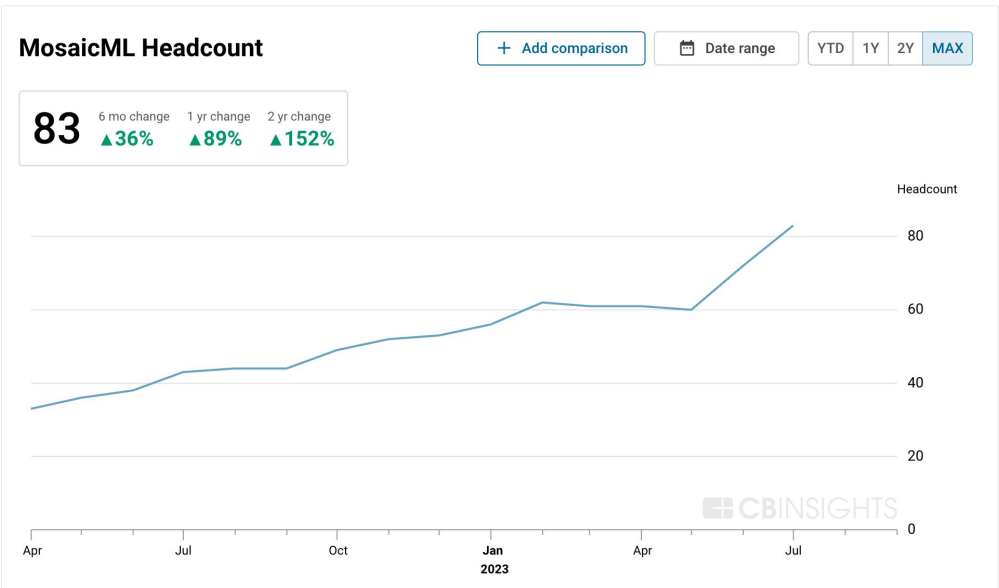
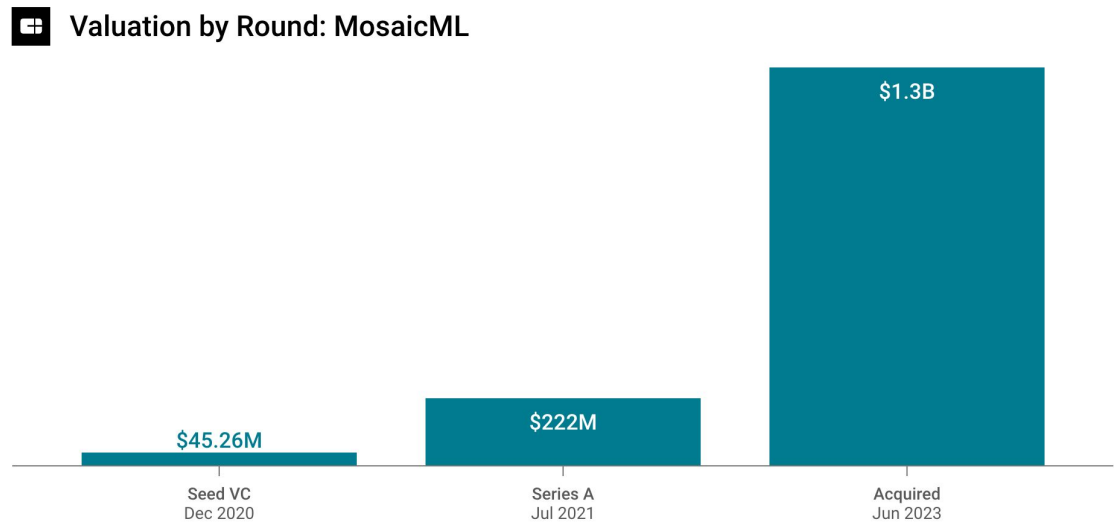
## Closed-source LLMs



## Open-source LLMs



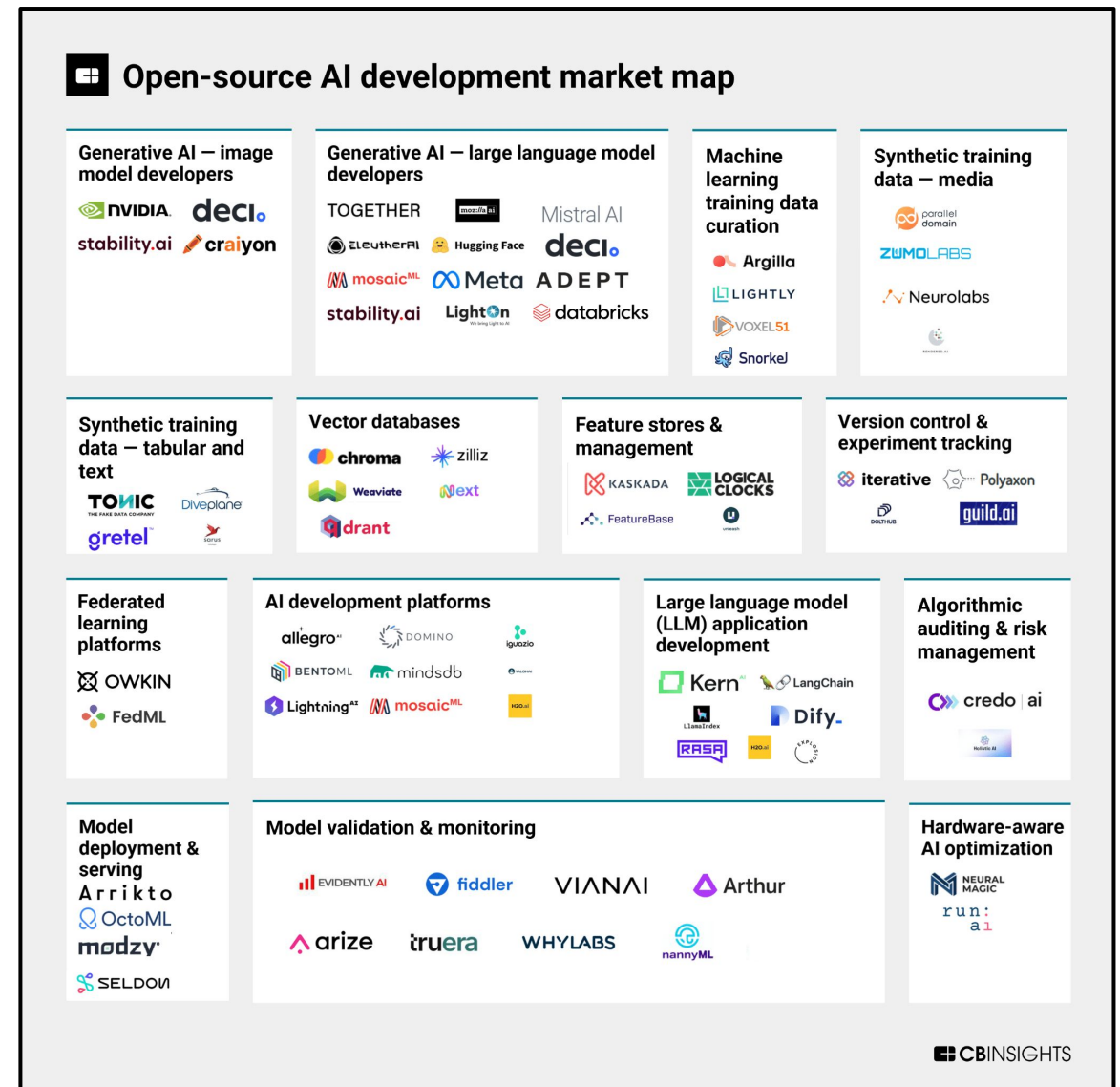
# Databricks acquired open-source LLM developer MosaicML for \$1.3B in Q2'23



# We mapped out open-source AI development tools

## Market map

70+ open-source machine learning operations (MLOps) vendors





[@cbinsights](#)

[@deeplearning](#)

[dvaradharajan@cbinsights.com](mailto:dvaradharajan@cbinsights.com)