PitchBook

EMERGING TECH RESEARCH

Artificial Intelligence & Machine Learning Report

VC trends and emerging opportunities

Q1 2024

Contents

Institutional Research Group

Analysis

Brendan Burke Senior Analyst, Emerging Technology
brendan.burke@pitchbook.com
pbinstitutionalresearch@pitchbook.com

Data

Matthew Nacionales Senior Data Analyst

Publishing

Report designed by **Jenna O'Malley** and **Julia Midkiff**

Published on May 23, 2024

For previous updates as well as our complete AI & ML research, please see the designated analyst workspace on the PitchBook Platform.

Vertical update

AI has resisted simple narratives about market dominance as use cases and deployment options expand in the field. At surface level, scale effects continue to show their primacy: Meta's leadership in computational resources has yielded best-in-class open-source models, and Microsoft reduced its dependence on OpenAI through its talent acquisition of the Inflection AI team while investing in Mistral AI. Microsoft and Amazon dominate generative AI (GenAI) spending-intent surveys, and Alphabet held up its AI growth story in its Q1 2024 earnings report. IT giants can efficiently convert GenAI innovations into new business, as seen with Accenture and ServiceNow. Behind these results, NVIDIA's GPU Technology Conference (GTC) in March demonstrated the kingpin position of the company in this mafia with new product announcements expanding the company's dominance of the hardware market.

On the other hand, the need for efficient deployment offers market opportunities for customized data layers, the creation of new workflows with GenAI applications, and semiconductors that are complementary to NVIDIA's. Open-source vector databases are among the most popular projects in all of data analytics, and rapid revenue growth is emerging for customized GenAI applications, such as Glean and Perplexity, that combine large language models (LLMs) and deterministic models for leading use cases including search, customer support, and knowledge work assistance. Robotics presents a complementary opportunity to knowledge work and can be addressed by both NVIDIA's new research lab and startup models from Covariant and Figure. As part of the NVIDIA ecosystem, semiconductor companies such as Astera Labs, Celestial AI, and Groq have seen accelerating revenue growth. These companies address the networking and power constraints in cutting-edge datacenters, with chip fabrication and power generation looming as outstanding opportunities for VC investment going forward.

Foundation model innovations at challenger startups bear watching after the humbling results of Meta and OpenAI's new LLM variants. The opportunity to compete in frontier transformer models is dwindling for all but the most cutting-edge and well-resourced companies, such as Mistral AI and xAI. Even so, foundation model labs can look past the current paradigm to improve on LLMs. AI21 Labs recently released an advanced structured state space model based on academic research to enhance computational efficiency. Cohere released a dedicated model for retrieval-augmented generation tasks that leads the market in latency and has become popular on Hugging Face. Reka AI launched a natively multimodal model called Core across image, video, and audio before OpenAI did, leading the company to receive acquisition interest from its partner Snowflake. These approaches show more than a Big Tech horse race for benchmark performance on conversational tasks.

Our view that code automation guides the future of GenAI is reinforced by the wave of innovation in coding agents. Leading technical founders and aggressive VCs can bypass the traditional startup journey with well-funded stealth operations and massive early-stage megadeals, as seen with Augment and Cognition in the coding agent space. These approaches can benefit from an increasingly diverse range of computational methods; there is speculation that Augment's technology is built on open-source models, while Cognition uses OpenAI. In either case, agentic architectures take LLMs down the last mile to developer workflows through LLM-specific scripts and retrieval of existing customer-specific code samples. These novel approaches build on the commercial momentum from not only GitHub Copilot but its competitors including Codeium and Anysphere. Some of the most widespread rollouts of GenAI solutions are in the coding space for software developers, lending hope that these tools can scale before more generic knowledge work tasks.

Q1 2024 timeline

VC deal

February 16

Graphics processing unit (GPU) cloud startup Lambda raises a $320.0 million Series C led by US Innovative Technology Fund, increasing its valuation by 7.1x for a $1.5 billion post-money valuation.

News

March 4

Anthropic announces the Claude 3 model family, which rises to the top of LLM performance leaderboards with three levels of cost and performance. The announcement is closely followed by the release of the second tranche of Amazon's committed investment in the company.

VC deal

March 26

Silicon photonics innovator Celestial AI reaches a $1.2 billion post-money valuation in a round led by US Innovative Technology Fund.

VC deal

February 26

Foundation model innovator Mistral AI raises its valuation by 6.1x, resulting in a $2.0 billion post-money valuation, in a round led by Andreessen Horowitz, General Catalyst, and Lightspeed Venture Partners.

VC exit

March 20

Datacenter connectivity company Astera Labs achieves a $5.5 billion exit in an IPO after raising only $232.3 million in VC.

News

March 27

Databricks launches DBRX, a high-performance open-source LLM, to compete with low-cost large models like GPT-3.5 and large open-source models. The leadership of this model has since been challenged by Llama 3.

Q1 VC deal activity

1,779
total deals

$21.6B
total VC raised

-7.8%
deal value growth QoQ

2024 YTD summary

1,779
total deals

$21.6B
total VC raised

-31.2%
deal value growth QoQ

AI & ML landscape

1. Horizontal platforms
2. Vertical applications
3. Semiconductors
4. Autonomous machines

AI & ML VC ecosystem market map

This market map is an overview of venture-backed or growth-stage companies that have received venture capital or other notable private investments. Click to view the full map on the PitchBook Platform.

1 Horizontal platforms

AI automation platforms

Automation Anywhere · ThoughtSpot · DRIVENETS · SingleStore · DEVO · Icertis · BigPanda · harness · workato

AI core

DataRobot · dataiku · scale · neo4j · SANDBOX AQ · DATASTAX · sparkcognition · Weights & Biases · H2O.ai · DOMINO

Computer vision

trax · MEGVII 旷视 · BLACK SESAME TECHNOLOGIES · flock safety · 依图 | YITU · oosto · meero · XREAL · SCANDIT · EAGLE EYE NETWORKS

Foundation models

ANTHROP\C · databricks · MISTRAL AI_ · ALEPH ALPHA · cohere · ADEPT · AI21labs · MINIMAX · imbue

Natural language technology

Dataminr · attentive · gusto · AlphaSense · uniphore · Rokid · verbit · dialpad · grammarly · glean

2 Vertical applications

Consumer

ByteDance 字节跳动 · 便利蜂 · FAIRE · Contentsquare · 头条 · JELLYSMACK · animoca BRANDS · FLYR · NIANTIC · turnitin

Financial services

stripe · JDT 京东科技 · NAVAN · ramp · Mission Lane · lentra · Tradeshift · 量化派 QUANT GROUP · LIQUIDITY CERTAINTY IS A SCIENCE · Konfio

Healthcare

commure · TEMPUS · freenome · 平安医保科技 · HUMAN LONGEVITY · XtalPi · Generate:Biomedicines · NOOM · innovaccer

Industrial

eagleview · flexport. · Relativity · indigo · Plenty · FBN · BOWERY · IMPROBABLE · Plume · KoBold Metals

IT

FIGURE · Airtable · securonix · onetrust · ARCTIC WOLF · knock · checkr · BlueVoyant · HIGHSPOT · GONG

Transportation

metropolis · T3出行 · VIA · 斑马智行 Powered by AIOS · G7易流 · PATEO · kyte. · motive · CAMBRIDGE MOBILE TELEMATICS · FOR-U SMART FREIGHT

3 Semiconductors

GPU cloud

CoreWeave · Crusoe · Lambda · KAO DATA · together.ai · FOUNDRY · Genesis Cloud · salad · Cloudalize · colovore

Edge AI software

摩尔线程 MOORE THREADS · SiMa.ai · auradine · celona · Clobotics · black ore · ARMADA · EDGE IMPULSE · TAALAS

Intelligent sensors & devices

TERMINUS 特斯联 · DREAME · Verkada · ambiq · Aqara · humane · Density · ECHODYNE · Inflection

Processor design

SambaNova · 壁仞科技 BIREN TECHNOLOGY · cerebras · GRAPHCORE · blaize · groq · HAILO · tenstorrent

4 Autonomous machines

Autonomous vehicles

WAYMO · 华人运通 HUMAN HORIZONS · AIWAYS · WeRide 文远知行 · WAYVE · momenta · pony.ai · 阿维塔 AVATR · Applied Intuition

Intelligent robotics

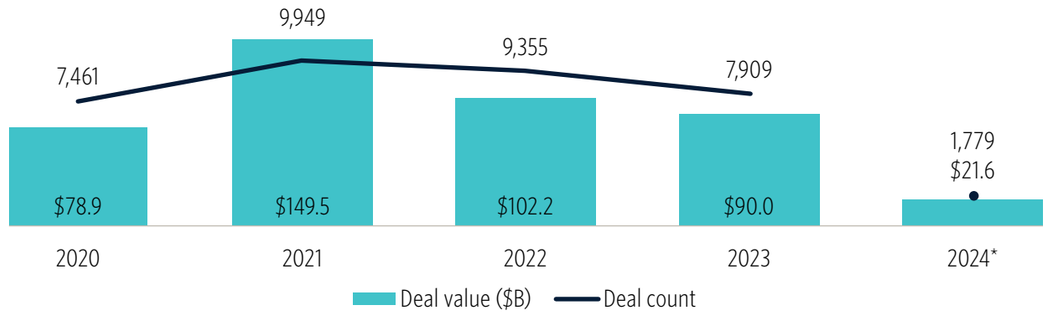
ANDURIL · nuro · Shield AI · FIGURE · Skydio · 达闼 · Geek+ · AGILE ROBOTS · EPIRUS · covariant

VC activity

Q1 2024 extended AI's run of outlier-fueled deal totals. Out of the $21.6 billion in AI & ML VC funding in the quarter, we attributed $5.3 billion to foundation model megadeals for Anthropic, Mistral AI, and xAI. Thus far in Q2 2024, xAI and Mistral AI are already continuing the trend of megadeals, suggesting that financial VCs can continue the momentum set by corporate VC investors last year. Though it invested $650 million while hiring the team, Microsoft's takeover of Inflection AI did not add to this total, but the deal demonstrated the strong position of corporates with the companies they have invested in. Leading LLM performance and access to unique datasets can justify large deals, with pure talent concentration becoming less important given the underperformance of some promising teams. The focus on early-stage leaders shows up in the median early-stage pre-money valuation exceeding the late-stage median, reaching $55.0 million across 78 recorded valuations.

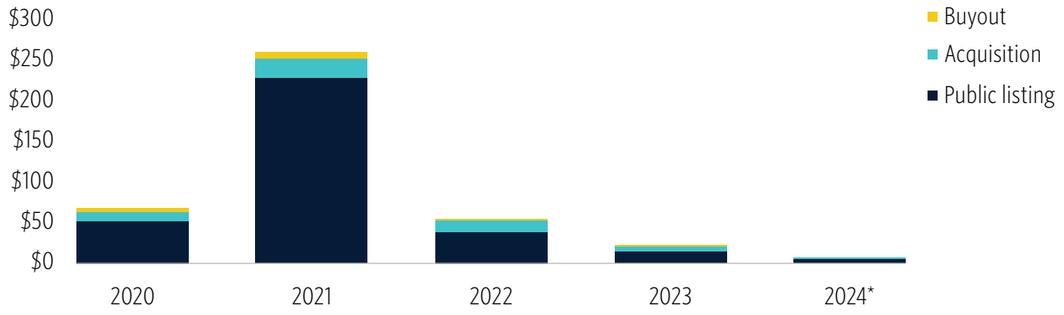
Q1 extended the theme of hardware acceleration being the primary outlet for exits in the current market environment. After MosaicML's outlier acquisition last year and Arm's IPO, Astera Labs reached public markets with an outstanding revenue multiple that set a high-water mark for an AI semiconductor valuation premium. Momentum from datacenter financing and the company's partnerships with Intel and NVIDIA suggest that revenue growth acceleration is likely in the short run. Not all AI companies will see this demand based on revenue forecasts and strategic alignment with market leaders. Thus far in Q2, NVIDIA announced an acquisition of Run:ai in the hardware acceleration as well, demonstrating the urgency of hardware challenges that we have documented. Some GenAI companies faced small acquisitions due to limited momentum, including BirchAI, Claypot AI, and Inflection AI. We expect the theme of semiconductors producing larger outcomes than software to continue with a potential listing for Cerebras. Foundation model startup acquisitions seem likely given the dominance of leading research labs yet can still achieve high deal values, as evidenced by MosaicML's $1.3 billion valuation last year.

AI & ML VC deal activity



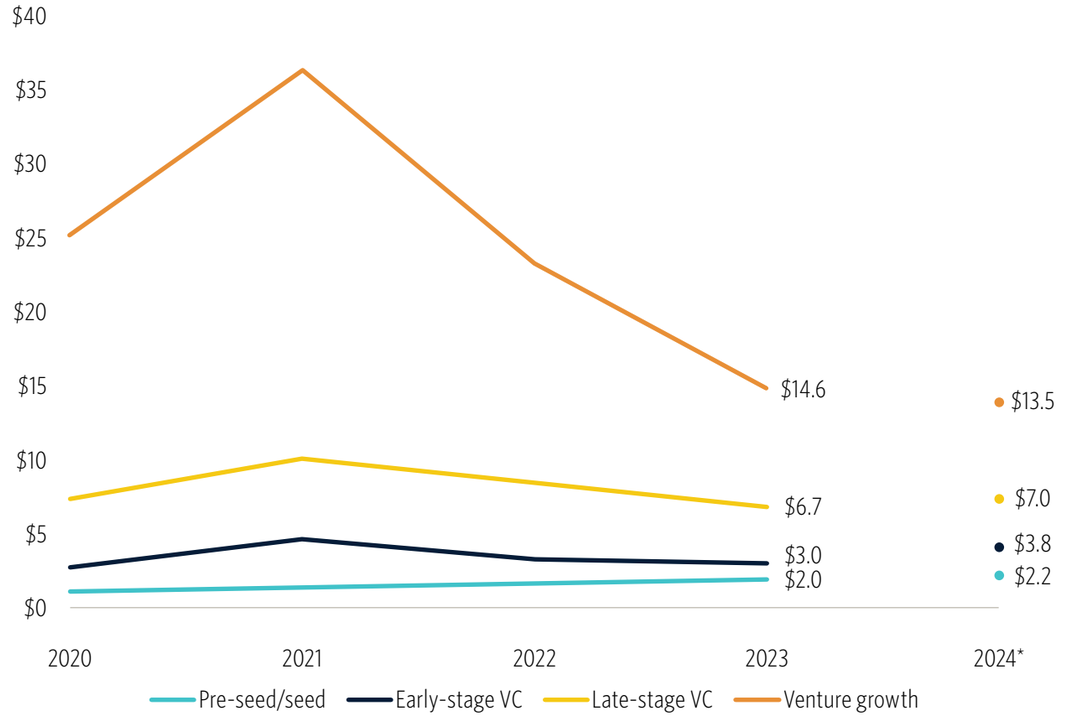
Source: PitchBook • Geography: Global • *As of March 31, 2024

AI & ML VC exit value ($B) by type



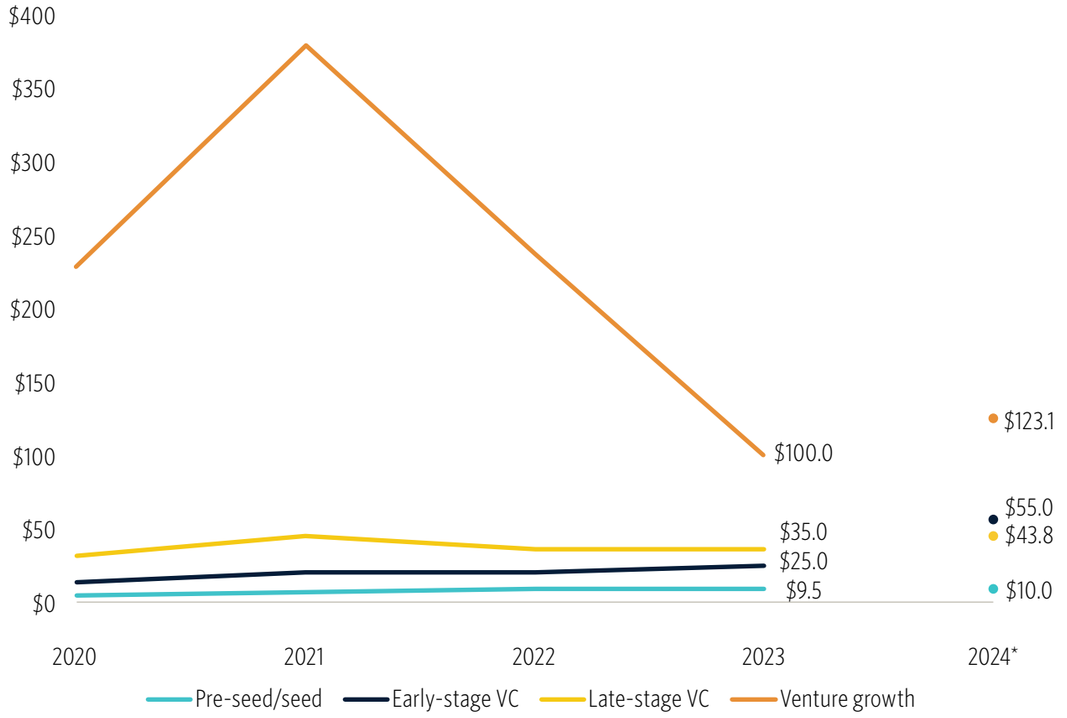
Source: PitchBook • Geography: Global • *As of March 31, 2024

Median AI & ML VC deal value ($M) by stage



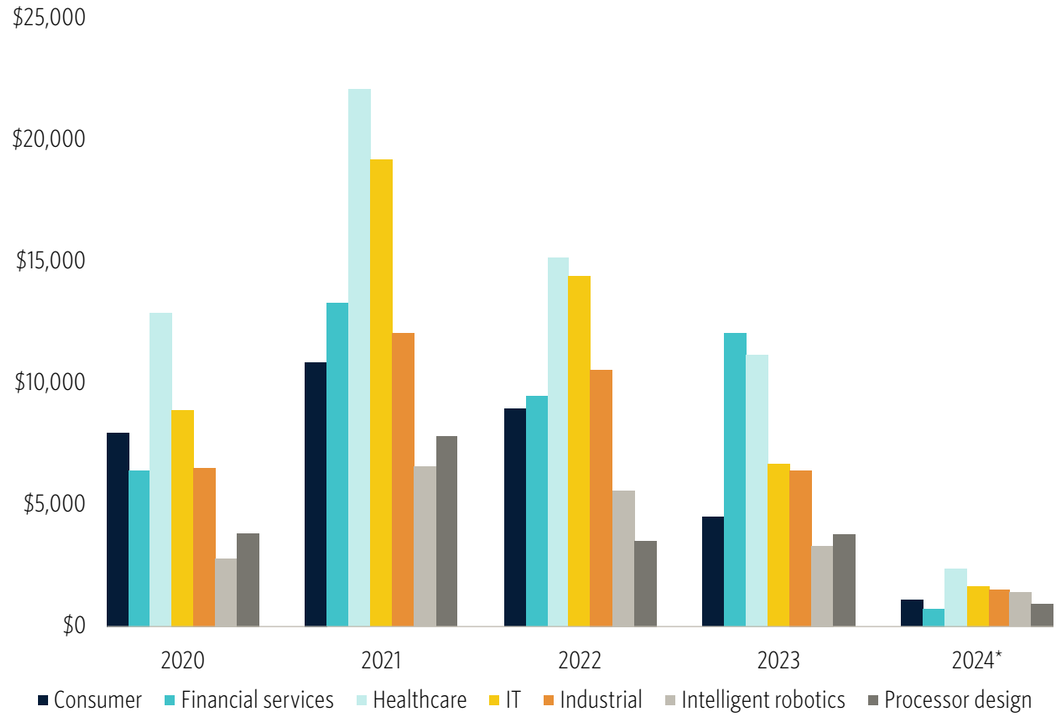
Source: PitchBook • Geography: Global • *As of March 31, 2024

Median AI & ML VC pre-money valuation ($M) by stage



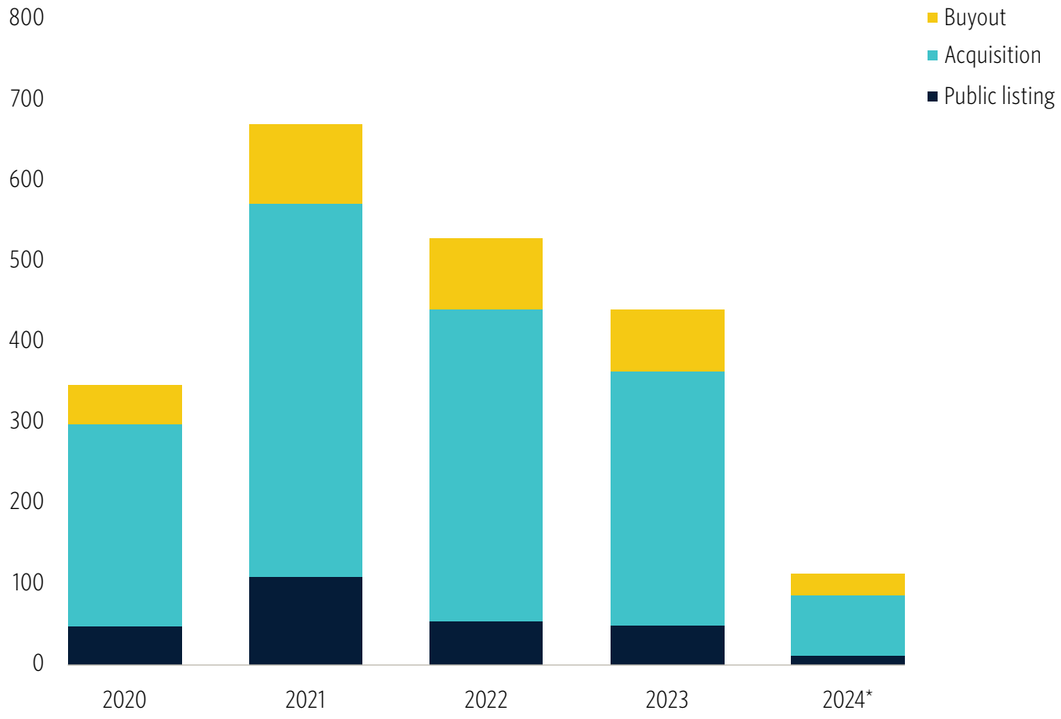
Source: PitchBook • Geography: Global • *As of March 31, 2024

AI & ML VC deal value ($M) by vertical application



Source: PitchBook • Geography: Global • *As of March 31, 2024

AI & ML VC exit count by type



Source: PitchBook • Geography: Global • *As of March 31, 2024

Key AI & ML early-stage VC deals by deal value in Q1 2024*

Company	Close date	Segment	Category	Deal value ($M)	Post-money valuation ($M)	Lead investor(s)	Valuation step-up (post to pre)
MiniMax AI	March 4	Horizontal platforms	AI core	$600.0	$2,500.0	Alibaba Group	N/A
Mistral AI	February 26	Horizontal platforms	AI core	$431.3	$2,000.0	Andreessen Horowitz, General Catalyst, Lightspeed Venture Partners	6.1x
xAI	January 11	Horizontal platforms	AI core	$134.7	N/A	IPO Club, Bossanova Investimentos	N/A
Rebellions	January 29	Semiconductors	Processor design	$124.7	$664.9	KT Cloud	N/A
Magic	February 12	Vertical applications	IT	$117.3	$500.0	Nat Friedman	3.2x
Together	March 13	Horizontal platforms	AI core	$106.0	$1,250.0	Salesforce Ventures	2.0x
Sierra	January 26	Horizontal platforms	Natural language technology	$85.0	N/A	Sequoia Capital	N/A
ElevenLabs	January 22	Vertical applications	Consumer	$80.0	$1,000.0	Andreessen Horowitz, Daniel Gross, Nat Friedman	9.2x
Perplexity	March 6	Vertical applications	Consumer	$73.6	$1,000.0	Daniel Gross, NVIDIA	1.8x
Hippocratic AI	March 18	Vertical applications	Healthcare	$55.0	$500.0	General Catalyst, Premji Invest	2.0x

Source: PitchBook • Geography: Global • *As of March 31, 2024

Key AI & ML late-stage VC deals by deal value in Q1 2024*

Company	Close date	Segment	Category	Deal value ($M)	Post-money valuation ($M)	Lead investor(s)	Valuation step-up (post to pre)
Anthropic	March 27	Horizontal platforms	AI core	$4,000.0	$18,400.0	Amazon.com, 7GC & Co and SquareOne Capital	N/A
Anthropic	January 30	Horizontal platforms	AI core	$750.0	N/A	Menlo Ventures	N/A
Figure AI	February 27	Autonomous machines	Intelligent robotics	$675.0	$2,675.0	Bezos Expeditions, Microsoft, NVIDIA	N/A
Lambda	February 16	Semiconductors	Processor design	$320.0	$1,520.0	US Innovative Technology Fund	7.1x
Glean	February 27	Horizontal platforms	Natural language technology	$200.0	$2,200.0	Kleiner Perkins, Lightspeed Venture Partners	2.0x
AI2 Incubator	March 8	Horizontal platforms	AI core	$200.0	N/A	N/A	N/A
BioAge	February 1	Horizontal platforms	AI core	$194.3	$394.3	Sofinnova Investments	0.8x
Celestial AI	March 26	Semiconductors	Processor design	$175.0	$1,175.0	US Innovative Technology Fund	2.9x
Abridge	March 29	Vertical applications	Healthcare	$150.0	$850.0	Lightspeed Venture Partners, Redpoint Ventures	3.4x
Observe	March 27	Horizontal platforms	AI core	$125.0	N/A	Sutter Hill Ventures	N/A

Source: PitchBook • Geography: Global • *As of March 31, 2024

Key AI & ML VC exits by exit value in Q1 2024*

Company	Close date	Segment	Category	Exit value ($M)	Exit type	Acquirer(s)
Astera Labs	March 20	Semiconductors	Processor design	$4,885.7	Public listing	N/A
Good Chemistry	January 5	Horizontal platforms	AI core	$75.0	Acquisition	SandboxAQ
GCT Semiconductor	March 26	Semiconductors	Processor design	$461.0	Public listing	Concord Acquisition III
Zapata Computing Holdings	March 28	Horizontal platforms	AI core	N/A	Public listing	Andretti Acquisition
BirchAI	March 26	Horizontal platforms	Natural language technology	N/A	Buyout	BPEA EQT, Sagility
Corvus	January 2	Vertical applications	Financial services	$435.0	Acquisition	Travelers Companies
Inspekto	February 15	Horizontal platforms	Computer vision	N/A	Acquisition	Siemens
Einblick	January 29	Horizontal platforms	AI automation platforms	N/A	Acquisition	Databricks
Claypot AI	January 22	Horizontal platforms	AI core	N/A	Acquisition	Voltron Data
DarwinAI	January 1	Horizontal platforms	AI core	N/A	Acquisition	Apple

Source: PitchBook • Geography: Global • *As of March 31, 2024

Top strategic acquirers of AI & ML companies since 2019*

Investor	Deal count	Investor type
Accenture	21	Corporation
Apple	17	Corporation
Cisco Systems	11	Corporation
Meta Platforms	11	Corporation
International Business Machines	9	Corporation
ServiceNow	8	Corporation
Snowflake	8	Corporation
Microsoft	8	Corporation
DataRobot	7	VC-backed company
Intel	7	Corporation

Source: PitchBook • Geography: Global • *As of March 31, 2024

Top post-seed VC investors in AI & ML companies since 2022*

Investor	Deal count	Early-stage VC	Late-stage VC	Venture growth	Investor type
Alumni Ventures	101	53	44	4	VC
Sequoia Capital	80	46	28	6	VC
Andreessen Horowitz	71	34	27	10	VC
Tiger Global Management	67	22	36	9	VC
Bossanova Investimentos	66	23	35	8	VC
FJ Labs	57	29	26	2	VC
Gaingels	55	24	27	4	VC
10X Capital	50	20	25	5	VC
HongShan	50	34	15	1	VC
Khosla Ventures	50	24	21	5	VC

Source: PitchBook • Geography: Global • *As of March 31, 2024

Key VC horizontal platform companies by total VC raised to date*

Company	VC ($M) raised to date	Category	IPO probability	M&A probability	No exit probability
OpenAI	$10,310.0	AI core	90%	6%	4%
Databricks	$4,181.9	AI core	95%	3%	2%
DataRobot	$1,048.1	AI core	57%	41%	2%
Automation Anywhere	$973.0	AI automation platforms	93%	5%	2%
Dataiku	$851.9	AI core	9%	89%	2%
ThoughtSpot	$801.4	AI automation platforms	97%	1%	2%
Scale AI	$602.7	AI core	74%	24%	2%
Aleph Alpha	$519.6	AI core	13%	84%	3%
Cohere	$440.0	AI core	35%	62%	3%
Hugging Face	$394.7	AI core	54%	43%	3%

Source: PitchBook • Geography: Global • *As of March 31, 2024

Note: Probability data is based on PitchBook VC Exit Predictor methodology.

Emerging opportunities

Inference engines

Inference will generate greater returns than training for compute providers and software vendors.

Information security LLMs

Information security leaders are not yet taken with GenAI features but can benefit from more advanced data science infrastructure enabled by GenAI.

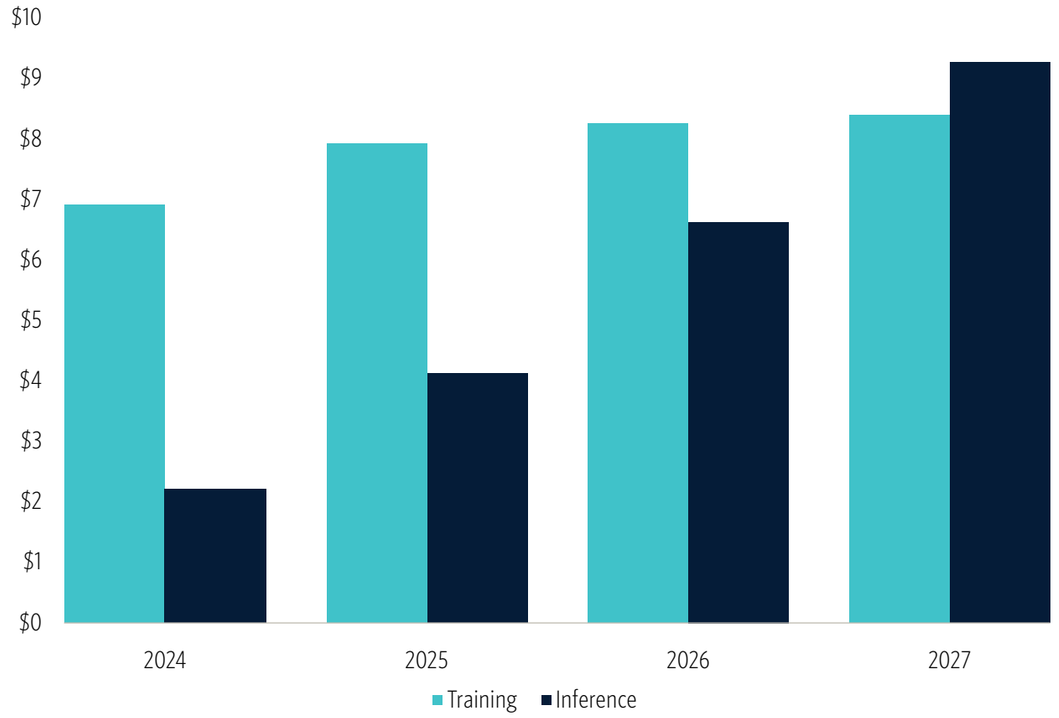
Inference engines

Overview

Inference engines refer to the software libraries and semiconductor accelerators that reduce the time to run AI models that have already been trained. In comparison, training involves teaching a generative model about a large dataset, where it learns to understand and replicate complex patterns and data structures. This phase requires substantial computational power and is typically conducted on high-performance hardware equipped with advanced GPUs or specialized AI servers. Inference, on the other hand, occurs when a trained GenAI model is used to generate new content, such as text, images, or music, based on input it receives. Inference engines include specialized core processors and number formats designed to efficiently process the complex mathematical operations required by AI models. These engines are optimized for speed and power efficiency, enabling real-time response capabilities. Semiconductors in inference engines help facilitate the deployment of GenAI across a wide array of platforms—from cloud-based servers that handle massive parallel requests to edge devices such as smartphones and Internet of Things devices, which require low latency and power efficiency.

Revenue growth in servers for inference is on pace to exceed that of training servers by 2027, encouraging investments in alternative inference architectures. Due to GenAI model deployment, GenAI inference servers are on track to grow 13.1x this year to $2.2 billion, according to IDC research on the server and storage market.[1] Inference servers can utilize advanced CPUs and unique bills of materials. Vendors of inference servers include AIME, Cisco, Dell, Gigabyte, SuperMicro, and Thinkmate. Inference servers can also be repurposed from legacy GPU racks.

GenAI server infrastructure spending estimate ($B) by stage*



Source: IDC • Geography: Global • *As of April 21, 2024

1: "The Infrastructure Market for Generative AI — 2024 Update," IDC, Peter Rutten and Madhumitha Sathish, April 21, 2024.

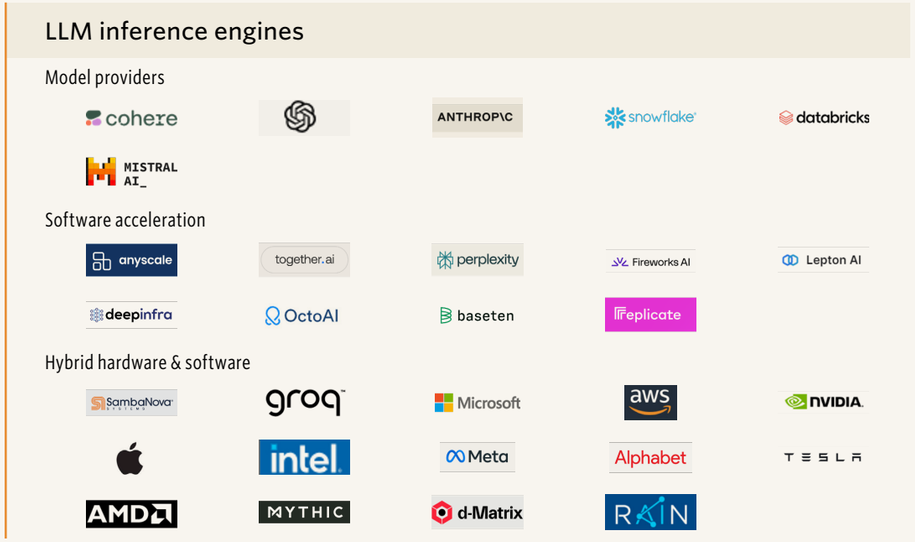
Market direction

Inference will generate greater returns than training for compute providers and software vendors. In the past fiscal year, 40% of NVIDIA's datacenter came from inference-optimized semiconductors and supporting software. CEO Jensen Huang focused on inference capabilities at NVIDIA's GTC in March, disclosing the new Blackwell computing platform with claims of 25x reductions in cost and energy. He also announced that the business is now generating 70% of revenue from inference instead of training. Even this percentage suggests an outsized concentration of revenue in training, given that an individual model should see 10x to 20x the amount of compute to run inference at scale than needed to train it, according to industry estimates. Much of this inference can occur on alternative hardware, meaning that NVIDIA is still heavily focused on training instead of on the inference that occurs with the models it creates. Even so, we expect the company to continue shifting its revenue mix toward inference.

Power constraints make inference speedup a critical global problem. The International Energy Agency forecasts that datacenter electricity demand will double from 2022 to 2026.[2] In 2026, AI-specific datacenters may consume 10x the power they did in 2023, stretching utility capacity and political will to approve new sites. Power constraints have weighed on datacenter stocks, with a short seller's report on Equinix power constraints contributing to a 20% share price decline after publication. The short seller, Hindenburg Research, cited former executives who believed the datacenter provider lacked the power capacity to deliver on AI commitments.[3] In this context, power savings from more advanced semiconductors like Blackwell must be

2: "Electricity 2024: Analysis and Forecast to 2026," International Energy Agency, January 24, 2024.
3: "Equinix Exposed: Major Accounting Manipulation, Core Business Decay and Selling an AI Pipe Dream as Insiders Cashed Out Hundreds of Millions," Hindenburg Research, March 20, 2024.

LLM inference engines market map



weighed against skyrocketing costs. With leading-edge datacenter providers, this will encourage innovation in engineering and retrofitting. The US Department of Energy is already focusing on this problem via the Energy Efficiency Scaling for 2 Decades program, and the newly funded National Semiconductor Technology Center can support innovation in AI-specific approaches.

Trending startups

Startups can capture market share and developer mind share from NVIDIA for efficient inference. Chip startups have already competed favorably with NVIDIA inference for bidirectional encoder representations from transformers (BERT) LLMs. SambaNova Systems disclosed significant speedups over NVIDIA hardware for BERT training and inference, including a 5.8x latency reduction.[4] In Q3 2023, the company launched a new generation of chips configured for midsize LLMs, predicting a world where customers will be running a variety of fine-tuned models in parallel. This focus on individual enterprise customers can create an adjacent market to NVIDIA's dominance of the hyperscaler market.

In the inference semiconductor niche, Groq is achieving scale with a customized language processing unit (LPU) inference accelerator that was developed when the company was founded in 2016. The LPU is fabricated at legacy 14-nanometer nodes that, when combined with custom accelerator software, can achieve 10x improvements in both LLM output speed and power savings over NVIDIA hardware. Since its recent announcements about record speeds on LLM tokens per second and a partnership with Aramco Digital, the company has achieved letters of interest and memorandums of understanding for 240,000 of its LPUs in the past two months, per our briefing with management.[5] The company forecasts 1.5 million shipments for 2025, factoring in a recent partnership with Norway-based edge datacenter host Earth Wind & Power. This is a similar volume to NVIDIA's shipments, though with more competitive pricing per unit for alternative cloud providers. The company also counts 150,000 developers among its community, having built 21,000 applications on its architecture. We have long believed that startups can stand out in AI inference rather than training, and this rapid scaling demonstrates that complementary hardware and compiler software to NVIDIA GPUs will be needed for the fast-growing inference market.

As hardware options multiply, agnostic compiler platforms will be needed to route workloads to underutilized assets. In Q1, OctoAI launched OctoStack, a GenAI software platform. The platform offers high efficiency with 4x better GPU utilization and up to 50% lower operating costs compared with leading alternatives, as well as flexibility across models and hardware form factors. OctoAI has benefited from the popularity of open-source models and can become a version of NVIDIA's CUDA for a range of chip architectures. The platform builds on the open-source Apache TVM project. GPU utilization has been a differentiator for acquisitions of MosaicML and Run:ai. Given that these platforms now reside within the Databricks and NVIDIA ecosystems, respectively, there is an opportunity to attract developers to an agnostic platform.

4: "Breakthrough Efficiency in NLP Model Deployment," SambaNova Systems, September 20, 2020.

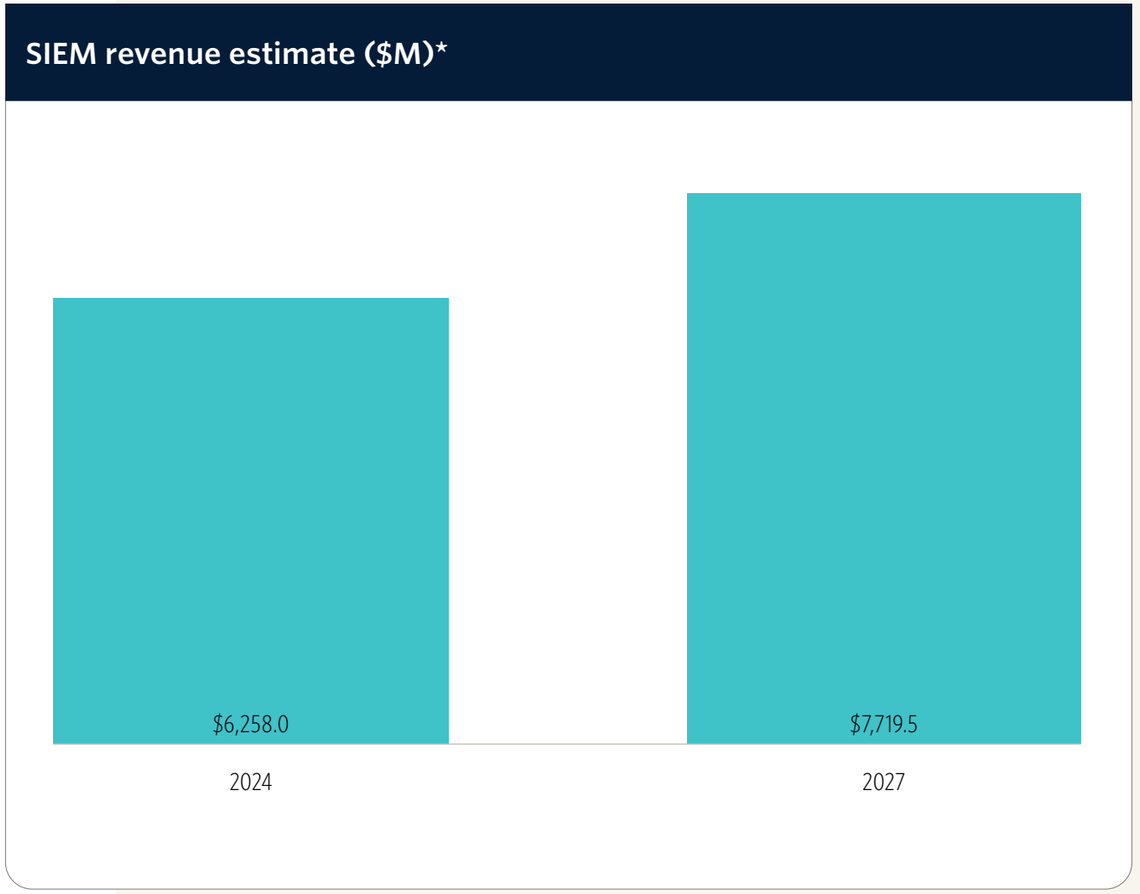
5: Jonathan Ross, CEO of Groq, briefing with Brendan Burke, April 15, 2024.

Information security LLMs

Overview

AI is a controversial topic in the information security (infosec) community given the history of deep neural networks creating more work for security teams and failing to provide automation. However, our research suggests that some of the best applications of advanced ML reside in the industry due to the volume of machine data and common attack patterns available in threat intelligence. These custom models have produced some of the largest and fastest-growing companies in all of enterprise software, including CrowdStrike and SentinelOne. The new wave of GenAI promises to create similarly epochal companies, especially given the need for large models to keep pace with machine-generated attacks from similarly sized models, as seen in LLM evaluations by OpenAI and Anthropic.

LLMs can best address the market opportunity for security information and event management (SIEM) due to the legacy data analysis capabilities in the market. The market will surpass $6.0 billion in 2024, yet SIEM platforms require significant augmentation. Exabeam has carved out a niche as a user analytics solution on top of SIEM platforms and recently launched a new product for Microsoft SIEM augmentation, which looks at only a sliver of Azure-relevant security data without significant price increases. Splunk and Amazon Security Lake have a diverse ecosystem of analytics augmentation solutions, including Tego Cyber's threat hunting engine. Given this complex ecosystem, startups can create a better security data analytics experience, and some have already begun to, with disruptive architectures from startups such as Anvilogic, Hunters, and



Source: IDC • Geography: Global • *As of August 23, 2023

Panther. The market is undergoing consolidation with a merger between LogRhythm and Exabeam and the acquisition of IBM's SIEM assets by Palo Alto Networks, opening opportunity for seamless connections between databases and security analytics. GenAI is already playing this role in normalizing data sources for analytics and can do the same for infosec.

Market direction

Data engineering in the infosec community can enable more advanced data science applications. Our briefing with Dave Palmer, partner of specialist VC firm Ten Eleven Ventures, indicated that big new waves in infosec tend to mirror shifts in computing architecture.[6] The shift to LLM computing will create the opportunity to both leverage the latest in computing hardware for infosec analytics, such as GPU clusters, while defending the latest types of data and infrastructure configurations. Security practitioners are experimenting with emerging database architectures, including ClickHouse and Snowflake. This can enable security teams to leverage proprietary database instances rather than shipping sensitive data to a cloud vendor, which can also save on costs. Additionally, traditional malware detection can benefit from integration of diverse data sources, including identity policies and qualitative threat intelligence feeds, which can be enabled by GPU computing and LLM analytics.

Security leaders are not yet taken with GenAI capabilities but can benefit from the more advanced data science infrastructure that GenAI enables. In an IDC survey of security leaders, only 11% of respondents viewed GenAI as a top-three feature for ubiquitous SIEM platforms.[7] Other preferred capabilities can benefit from GenAI, however, including real-time detection engines, connectors to all security data sources, and automation capability within the SIEM. Additionally, some of the fastest-growing SIEM vendors are rapidly integrating GenAI search, including Elastic and Securonix. Splunk's AI assistant remains in preview, suggesting that competitors integrating advanced data analytics can build on top of them and create valid alternatives.

Trending startups

The recent RSA Conference, a flagship show for the infosec industry, gave credit to startups moving beyond securing AI instances to building custom models to improve security analyst workflows. The conference's Innovation Sandbox was won by Reality Defender, a startup built on advanced GenAI research into deepfake creation. The company's models know how GenAI content differs from the real world, down to fine-grained patterns, based on proprietary data collection and model fine-tuning. The company stands out for the volume of Ph.D. researchers it employs who specialize in audio and video models but work in other fields as well. Fellow Sandbox participant Harmonic Security developed a cluster of small fine-tuned LLMs to label data targeted for GenAI applications, enabling scaled review of sensitive data that will be missed by existing data loss prevention systems. Judges asked fair questions about the ability of these products to keep up with the latest in AI, and contestants conceded that new models may be disruptive and will require future-proofing.

The conference highlighted three startups building custom LLMs for security operations center automation: Culminate, Dropzone AI, and Radiant Security. Culminate and Dropzone focus on incident report creation, while Radiant focuses on alert triage and automated remediation. These

6: Dave Palmer, General Partner of Ten Eleven Ventures, briefing with Brendan Burke, May 7, 2024.
7: "SIEM Users Rank Important Features," IDC, Michelle Abraham, May 2, 2024.

companies are aligned in seeking to automate Tier 1 analysts, which refers to junior security staff who can be hard to hire. We have featured Dropzone AI in past reports, and in April, the company raised a $16.9 million Series A led by AI specialist investor Theory Ventures. Culminate received favorable feedback from VC panelists, including infosec specialists from Ballistic Ventures, Bain Capital Ventures, and AI specialist Conviction VC, eliciting suggestions to focus on middle-market capabilities. Radiant Security raised a $15.0 million Series A in Q4 2023 and has begun announcing commercial partnerships. We believe that these companies can generate powerful data learning effects through LLMs' intelligence about diverse coding formats and can become security platforms of the future. AI will never be wholly accepted in the infosec community given the primacy of people and processes, yet we believe that developer adoption of AI coding copilots can spread to infosec practitioners for workflow augmentation over the next two years.

Select company highlights

ANTHROP\C

Overview

Anthropic was founded in 2021 by a group of seven former senior members of OpenAI. Several of the founders worked at Google Research before OpenAI. The founders split off from OpenAI with a desire to train safe AI systems that could mitigate the worst risks of superintelligence, including human extinction. To serve this mission, the company was established as a public-benefit corporation. Since its founding, Anthropic has developed proprietary AI systems and commercialized a series of LLMs called Claude.

Claude leverages transformer architecture and a reinforcement learning process referred to as Constitutional AI. Constitutional AI tests the alignment of model outputs against a list of ethical rules and principles. In this way, the model self-improves by selecting outputs that conform to the company's ethics. The Constitutional AI process improves the safety of responses without the extensive human intervention seen with other approaches, as in OpenAI's use of human feedback for outsourced reinforcement learning. We have not tracked any patent for this technique. Because of this process, Anthropic does not shift its response types as often as ChatGPT, which faces complaints of censorship and model drift as the model changes over time.

Key company information

Founded
2021

Employees
375

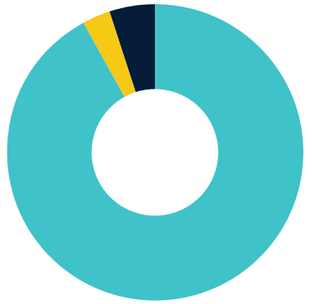
Total raised
$7.9B

Last financing valuation
$18.4B

Last financing
Raised $4.0B in a Series D

Lead investor(s)
Alphabet, Amazon, LG CNS, Calm Ventures, SK Telecom, 7GC & Co., SquareOne Capital, Salesforce Ventures, Zoom Ventures, Spark Capital, Sam Bankman-Fried, Jaan Tallinn

Exit Predictor



IPO: 92% probability **M&A:** 3% probability **No exit:** 5% probability

Success: 95% probability

Note: Probability data is based on PitchBook VC Exit Predictor methodology.

Leadership

The company is co-led by CEO and co-founder Dario Amodei and President and co-founder Daniela Amodei. The two are siblings, and neither reports to the other. At OpenAI, Dario focused on research, and Daniela on safety and policy. Dario earned a Ph.D. in biophysics from Princeton University and worked as a research scientist at Baidu and Google. Daniela comes from a nontechnical background; she was a risk manager at Stripe and led human resources, business operations, and technical safety teams at OpenAI.

Other key co-founder executives include Head of Compute Engineering Tom Brown and Head of Policy Jack Clark. Brown led the engineering of OpenAI's GPT-3 and was formerly a chief technology officer at a startup and a Google researcher. Clark was a tech journalist and led communications at OpenAI.

Competitors

Anthropic competes in the foundation model category of AI against OpenAI, Microsoft, Google, Meta, and the open-source community. OpenAI, Meta, and Google are developing comparable LLMs. Recently, open-source foundation model startup Mistral AI matched Anthropic's performance despite starting two years later with fewer resources. Google has also made rapid progress with its Gemini Ultra model. OpenAI's GPT-5 is tracking to set a new bar for performance this year, which will require Claude to keep pace via successive model releases.

Anthropic partners with Amazon Web Services (AWS) and Google via cloud investments. In October 2023, the company raised a $4.0 billion investment led by Amazon that involved both cloud credits and equity investment. Later that month, the company raised a $2.0 billion round led by Google. Amazon then released the second tranche of its commitment in March 2024. Anthropic has established an enterprise customer base through collaboration with AWS and channel partners, including Accenture, and forecasts 325% revenue growth in 2024, standing out from other OpenAI competitors. Concerns about Anthropic's model performance have been alleviated by the launch of Claude 3; on research team LMSYS' chatbot leaderboard, which aggregates crowdsourced opinions of model quality, the model currently ranks second, tied with Gemini and behind the latest versions of GPT-4. We believe the company can stand out in enterprise deployments because of its integration with the leading public cloud, AWS, and its focus on baked-in safety without the need for volatile reinforcement learning from human feedback.

Anthropic will be a primary beneficiary of existing AWS customers leveraging Bedrock. Additionally, significant AWS consulting relationships from Accenture, Deloitte, and Slalom will encourage the use of Anthropic models. Anthropic, AWS, and Accenture now have an official collaboration to deploy Claude across industries. Accenture is the leading AI services firm by market share. Given that consultants present the first port of call for enterprise deployments, Anthropic can benefit from close relationships via AWS.

Outlook

Anthropic presents a legitimate alternative to OpenAI given the company's research talent, partnership with AWS, and focus on safety issues that will be a critical concern for enterprise customers. The company disclosed that it had reached $200.0 million in annual recurring revenue (ARR) in 2023. Sacra estimates the company reached $316.0 million in March 2024.[8]

Enterprise accounts drive this financial total, with anchor customers including Notion, Quora, and DuckDuckGo. The company forecasts $850.0 million in ARR in 2024, demonstrating the traction the company's models have in enterprises and the potential for Anthropic to become a larger enterprise vendor than OpenAI. Current investment opportunities include secondary shares from prior transactions along with a special-purpose vehicle from Menlo Ventures.

8: "Anthropic," Sacra, n.d., accessed May 10, 2024.

Financing history

Series A	Series B	Series C	Series D1	Series D
May 28, 2021	**April 29, 2022**	**May 15, 2023**	**October 27, 2023**	**March 27, 2024**
Total raised $124.0M	**Total raised** $580.0M	**Total raised** $450.0M	**Total raised** $2.0B	**Total raised** $4.0B
Pre-money valuation $426.0M	**Pre-money valuation** $2.4B	**Pre-money valuation** $4.1B	**Pre-money valuation** $12.8B	**Pre-money valuation** $14.9B
Investor Jaan Tallinn	**Investor** Sam Bankman-Fried	**Investor(s)** Salesforce Ventures, Zoom Ventures, Spark Capital	**Investor(s)** Wisdom Ventures, Factorial Funds, Alphabet	**Investor(s)** Amazon, 7GC & Co., and SquareOne Capital

Overview

Founded in 2013, Databricks offers a data science platform that includes AI-as-a-service functionality with a suite of data science tools for data engineering, data warehousing, and ML algorithms. The company grew out of the open-source Apache Spark data science community and created an extensible product, the Unified Data Analytics Platform, that can ingest data from enterprise silos and prepare it for cluster-based computing. Once an effective open-source product was in place, the company moved to a closed-source model and rapidly increased revenue beginning in 2016.

Databricks now offers a product suite on top of Apache Spark that includes an AI platform called Lakehouse AI, which is optimized for GenAI with data storage, a model training pipeline, and model-serving and monitoring; a data lake, which integrates unstructured data in a central database for analytics; and data security. The company has focused on the GenAI opportunity, training a custom LLM called DBRX that can run more efficiently than GPT-3.5 and comparable open-source models. Databricks recently acquired Einblick for data analysis using GenAI and Lilac for GenAI data preparation. In Q1, the company invested in Mistral AI and Adaptive AI, contributing to the company's vision of fine-tuned open-source models for its customers.

Key company information

Founded
2013

Employees
8,101

Total raised
$4.2B

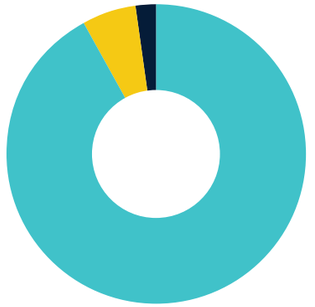
Last financing valuation
$43.2B

Last financing
Raised $684.6M in a Series I

December 2023
Leader in Gartner Magic Quadrant for Cloud Database Management Systems

Lead investor(s)
NVIDIA, T. Rowe Price, Counterpoint Global, Franklin Templeton Investments, Andreessen Horowitz, NEA

Exit Predictor



IPO: 92% probability **M&A:** 6% probability **No exit:** 2% probability

Success: 98% probability

Note: Probability data is based on PitchBook VC Exit Predictor methodology.

Leadership

Databricks' management has high technical expertise capable of generating product-led growth. The company was founded by seven early contributors to Apache Spark, and all the co-founders remain at the company. CEO and co-founder Ali Ghodsi earned a Ph.D. in distributed computing and was vice president of engineering before being promoted to CEO at the start of the company's growth phase in 2016. The acquisition of MosaicML brought in the Vice President of Generative AI, Naveen Rao, who has credibility in the open-source AI community. The board features co-founders and VC investors, including Andreessen Horowitz co-founder Ben Horowitz, yet still lacks the public-company leadership we believe is necessary to go public.

Competitors

Databricks stands out in the AI & ML market by supporting customized AI model training using open-source models atop its data lake, functionality that distinguishes it from legacy vendors such as SAP, Microsoft, and AWS and allows it to be more AI oriented than next-generation data-warehouse vendor Snowflake. Because of its ability to support batch processing for ML analytics and other high-value workloads, Databricks is capturing market share from incumbents such as Alteryx, IBM, Oracle, and SAS. Databricks runs on Microsoft Azure and AWS, although those hyperscalers would likely prefer to natively support their GenAI workflows. The company's Unity Catalog product gives it market credit for data governance—a critical problem for data scientists new to AI.

Outlook

Databricks has achieved the necessary scale and growth to realize high performance in public markets. The company neared $2.0 billion in 2023 with 50% revenue growth.[9] The company will need to defend itself against ML-optimized challengers that can carry out more efficient streaming data analysis and in-memory analytics within its database. Furthermore, it will benefit from analytics built on its database becoming essential to high-value industries, including IT and industrial. The company remains innovative and active in future-proofing its business, leading us to believe that, in the long run, Databricks will be valued more highly than relational database incumbents based on the growth opportunities of data lake architecture for streaming data.

9: "Databricks," Sacra, n.d., accessed May 10, 2024.

Financing history

Series A

September 24, 2013

Total raised
$14.0M

Pre-money valuation
$30.0M

Lead investor
Andreessen Horowitz

Series B

June 30, 2014

Total raised
$33.4M

Pre-money valuation
$220.0M

Lead investor
NEA

Series C

December 15, 2016

Total raised
$60.0M

Pre-money valuation
$460.0M

Lead investor
NEA

Series D

September 25, 2018

Total raised
$140.0M

Pre-money valuation
$800.0M

Lead investor
Andreessen Horowitz

Series E

January 11, 2019

Total raised
$250.0M

Pre-money valuation
$2.5B

Lead investor
Andreessen Horowitz

Series F

October 22, 2019

Total raised
$400.0M

Pre-money valuation
$5.8B

Lead investor
Andreessen Horowitz

Series G

February 9, 2021

Total raised
$1.0B

Pre-money valuation
$27.0B

Lead investor
Franklin Templeton Investments

Series H

August 31, 2021

Total raised
$1.6B

Pre-money valuation
$36.4B

Lead investor
Counterpoint Global

Series I

November 13, 2023

Total raised
$684.6M

Pre-money valuation
$42.5B

Lead investor(s)
NVIDIA, T. Rowe Price

About PitchBook Industry and Technology Research

Independent, objective, and timely market intel

As the private markets continue to grow in complexity and competition, it's essential for investors to understand the industries, sectors, and companies driving the asset class.

Our Industry and Technology Research provides detailed analysis of nascent tech sectors so you can better navigate the changing markets you operate in—and pursue new opportunities with confidence.

PitchBook Data, Inc.

John Gabbert Founder, CEO

Nizar Tarhuni Vice President, Institutional Research and Editorial

Paul Condra Head of Emerging Technology Research

Additional research

Eric Bellomo
eric.bellomo@pitchbook.com
Gaming
E-Commerce

Brendan Burke
brendan.burke@pitchbook.com
Data Analytics
Information Security
Artificial Intelligence & Machine Learning

Aaron DeGagne
aaron.degagne@pitchbook.com
Medtech
Digital Health

Alex Frederick
alex.frederick@pitchbook.com
Agtech
Foodtech

Jonathan Geurkink
jonathan.geurkink@pitchbook.com
Supply Chain Tech
Mobility Tech

Kazi Helal
kazi.helal@pitchbook.com
Biopharma
Pharmatech

Derek Hernandez
derek.hernandez@pitchbook.com
Enterprise SaaS
Infrastructure SaaS

Ali Javaheri
ali.javaheri@pitchbook.com
Emerging Spaces

Robert Le
robert.le@pitchbook.com
Insurtech
Crypto

John MacDonagh
john.macdonagh@pitchbook.com
Carbon & Emissions Tech
Clean Energy Tech

Rebecca Springer
rebecca.springer@pitchbook.com
Healthcare Services
Healthcare IT

Rudy Yang
rudy.yang@pitchbook.com
Enterprise Fintech
Retail Fintech