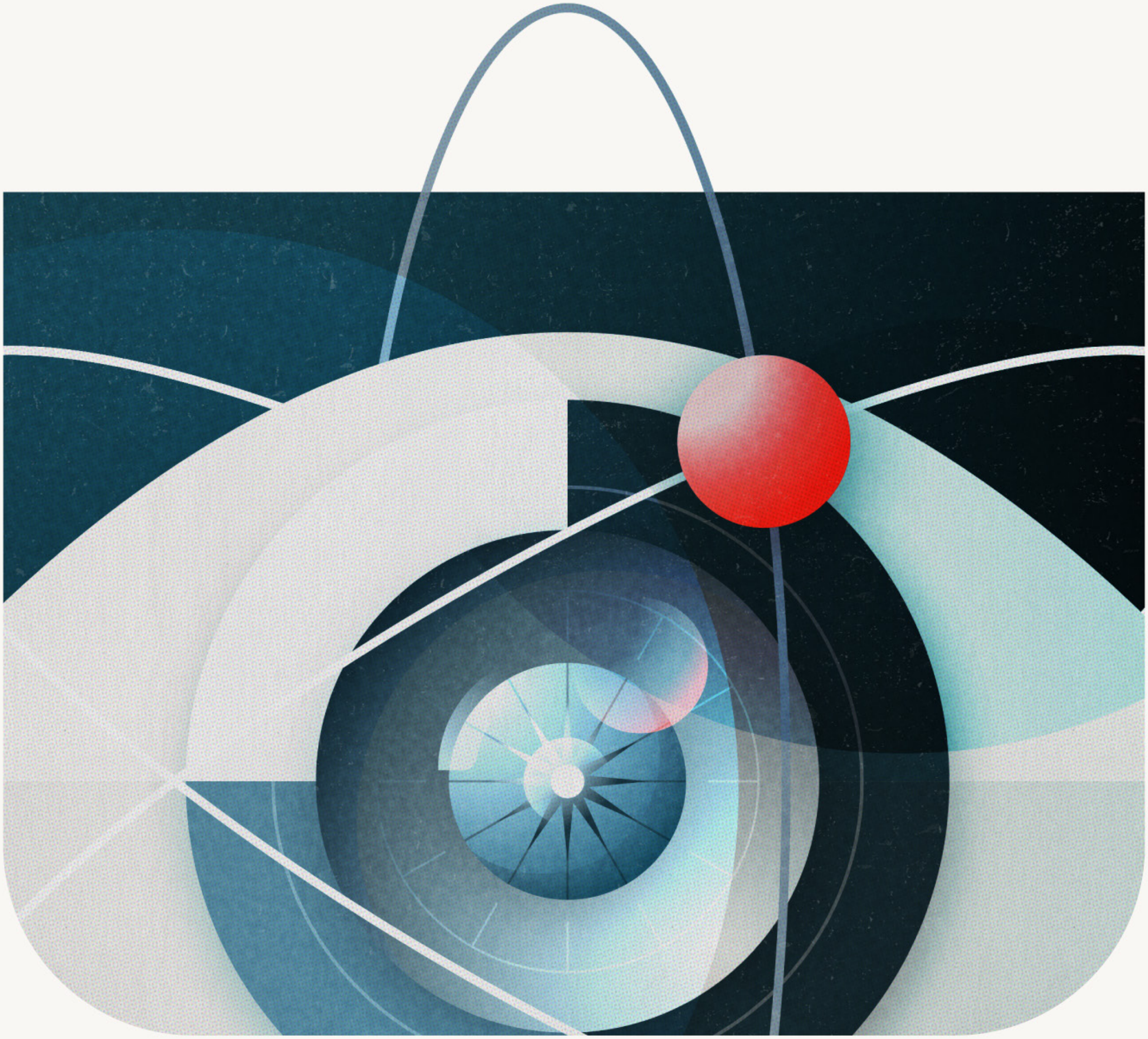
databricks

State of Data+AI

Data intelligence and the race to customize LLMs



Organizations rush to democratize data and AI

Introduction

Generative AI is ushering in a new era of innovation, creativity and productivity. Just 18 months after it entered mainstream conversations, companies everywhere are investing in GenAI to transform their organizations. Enterprises are realizing that their data is central to delivering a high-quality GenAI experience for their users. The urgent question among business leaders now is: *What's the best and fastest way to do that?*

With siloed data and AI platforms, it's difficult for teams to accelerate their GenAI projects — whether they are using natural language to ask questions of their data or are building intelligent apps with their data. We believe that data intelligence platforms will result in radical democratization across organizations. This new category of data platforms uses GenAI to more easily secure and leverage data, and lower the technical bar to create value from it. Among our own customers, there's already a clear acceleration of AI adoption.

The *State of Data + AI* report provides a snapshot of how organizations are prioritizing data and AI initiatives. The insights shared come from more than 10,000 global customers — now including over 300 of the Fortune 500 — using the Databricks Data Intelligence Platform. Discover how the most innovative organizations are succeeding with machine learning, adopting GenAI and responding to evolving governance needs.

This report is designed to help companies develop effective data strategies in the evolving era of enterprise AI.

Major Findings

11x more AI models were put into production this year

After years of being stuck experimenting with AI, companies are now deploying substantially more models into the real world than a year ago.

On average, organizations became over 3 times more efficient at putting models into production.

Natural language processing is the most-used and fastest-growing machine learning application.

70% of companies leveraging GenAI use tools and vector databases to augment base models

In less than one year of integration, LangChain became one of the most widely used data and AI products.

Companies are hyperfocused on customizing LLMs with their private data using retrieval augmented generation (RAG).

RAG requires vector databases, which grew 377% YoY. (Usage inclusive of both open source and closed LLMs.)

76% of companies using LLMs choose open source, often alongside proprietary models

Many companies select smaller open source models when considering trade-offs between cost, performance and latency.

Only 4 weeks after launch, Meta Llama 3 accounts for 39% of all open source model usage.

Highly regulated industries are the surprise GenAI early adopters. Financial Services, the leader in GPU usage, is moving the fastest, with 88% growth over 6 months.

Methodology:
How did Databricks create this report?

The 2024 *State of Data + AI* is built from fully aggregated, anonymized data collected from our customers based on how they are using the Databricks Data Intelligence Platform and its broad ecosystem of integrations.

This report focuses on trends in machine learning, the adoption of GenAI, integrations and use cases. The customers in this report represent every major industry and range in size from startups to many of the world's largest enterprises. Unless otherwise noted, this report presents and analyzes data from February 1, 2023, to March 31, 2024, and usage is measured by number of customers. When possible, we provide year-over-year (YoY) comparisons to showcase growth trends over time.

Machine Learning

AI is in production

ORGANIZATIONS RACE TO PUT ML MODELS INTO PRODUCTION

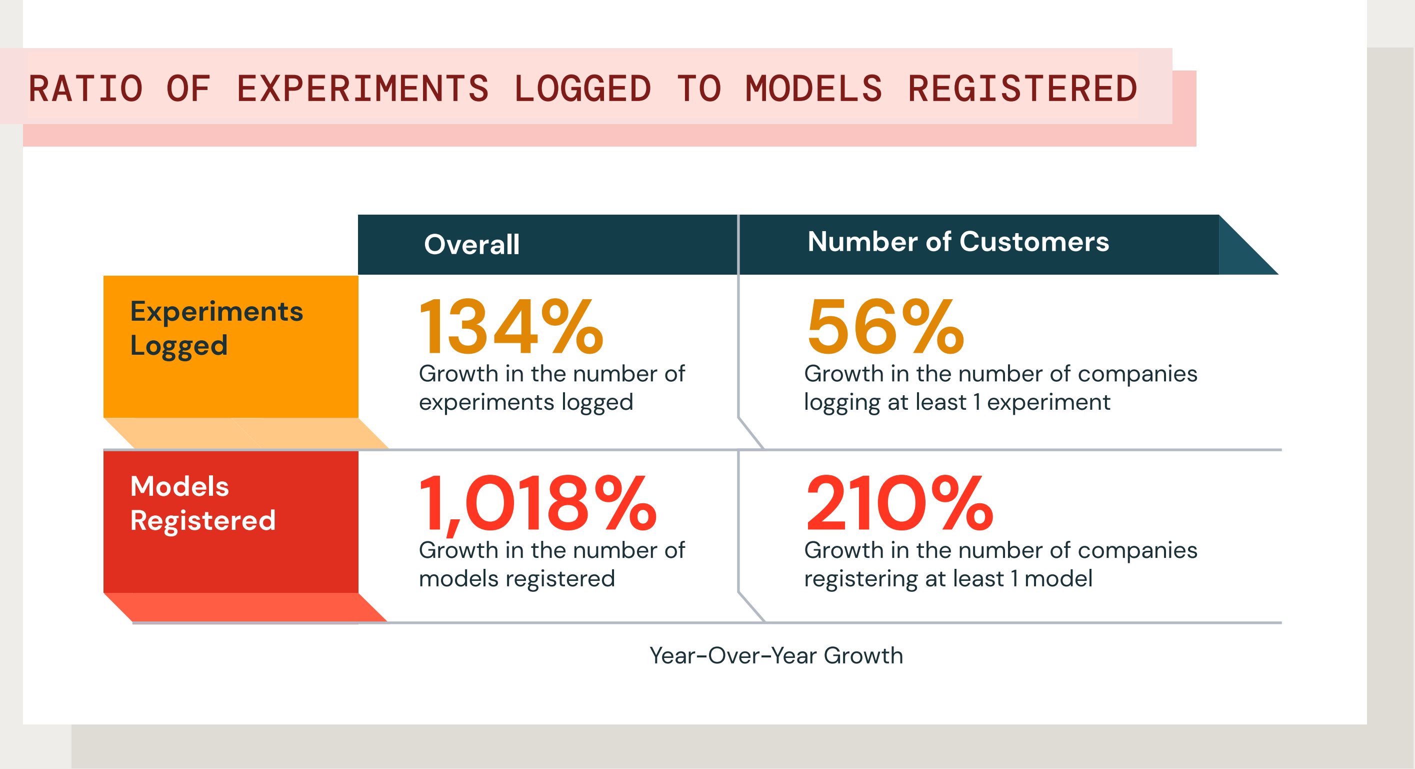
This year, we've seen a shift from experimentation to production applications of AI. As machine learning (ML) takes off, companies are learning to navigate two distinct halves of the ML model lifecycle. Organizations first create their ML models through the process of *experimental testing*, trying out different algorithms and hyperparameters to get to the best models, before putting these models into *production*. In this process, teams have two competing goals: ensuring the experimentation phase is as time-efficient as possible, while only putting rigidly tested models into production.

Deploying models in production has historically had many challenges: disparate data and AI platforms, complex deployment workflows, lack of access controls for governance, inability to monitor and more. Our data reveals how companies are overcoming these challenges with the introduction of data intelligence platforms.

Companies accelerate production ML

Data from MLflow (an open source MLOps platform developed by Databricks) shows how frequently our customers are *logging* models (representing experimentation) and *registering* models (representing production).

The results? Not only do we see more experimentation, but companies are also becoming substantially more efficient at getting into production.



RATIO OF EXPERIMENTS LOGGED TO MODELS REGISTERED

	Overall	Number of Customers
Experiments Logged	134% Growth in the number of experiments logged	56% Growth in the number of companies logging at least 1 experiment
Models Registered	1,018% Growth in the number of models registered	210% Growth in the number of companies registering at least 1 model

Year-Over-Year Growth

Figure 1:
The YoY growth of models registered has far outpaced the growth of experiments logged, indicating companies are moving from experimentation to production.

A giant leap: 11x more models went into production

The volume of models has grown substantially in measurable ways.

THE NUMBER OF COMPANIES INVESTING IN ML HAS SKYROCKETED

Our data shows that 56% more companies are logging experimental models compared to a year ago, but 210% more are registering models. This indicates many companies that were focused on experimenting last year have now moved into production.

THE NUMBER OF ML MODELS IS UP ACROSS COMPANIES

After years of intense focus on experimentation, organizations are now charging into production. 1,018% more models were registered this year, far outpacing experiments logged, which grew 134%. We see this trend at the company level as well. The average organization registered 261% more models and logged 50% more experiments this year.

THE TAKEAWAY

ML is core to how companies innovate and differentiate. And as companies continue to build their confidence, we expect to see this trend continue in the coming years. The newer field of GenAI is still in the testing phase, but companies are starting to make traction.

Companies become 3x more efficient at putting models into production

ML efficiency has real value that can be measured in time, money and resources. While model development and experimentation are crucial, ultimately these models need to be deployed to real-world use cases to drive business value.

We looked at the ratio of logged-to-registered models across all customers to assess progress. In February 2023, the ratio of logged-to-registered models was 16-to-1. This means that for every 16 experimental models, one model gets registered for production. By the end of the data range, the ratio of logged-to-registered models dropped sharply to 5-to-1, an improvement of 3x.

The takeaway? Companies are becoming significantly more efficient at getting models into production, spending fewer resources on experimental models that never provide real-world value.

OVERALL RATIO OF LOGGED-TO-REGISTERED MODELS

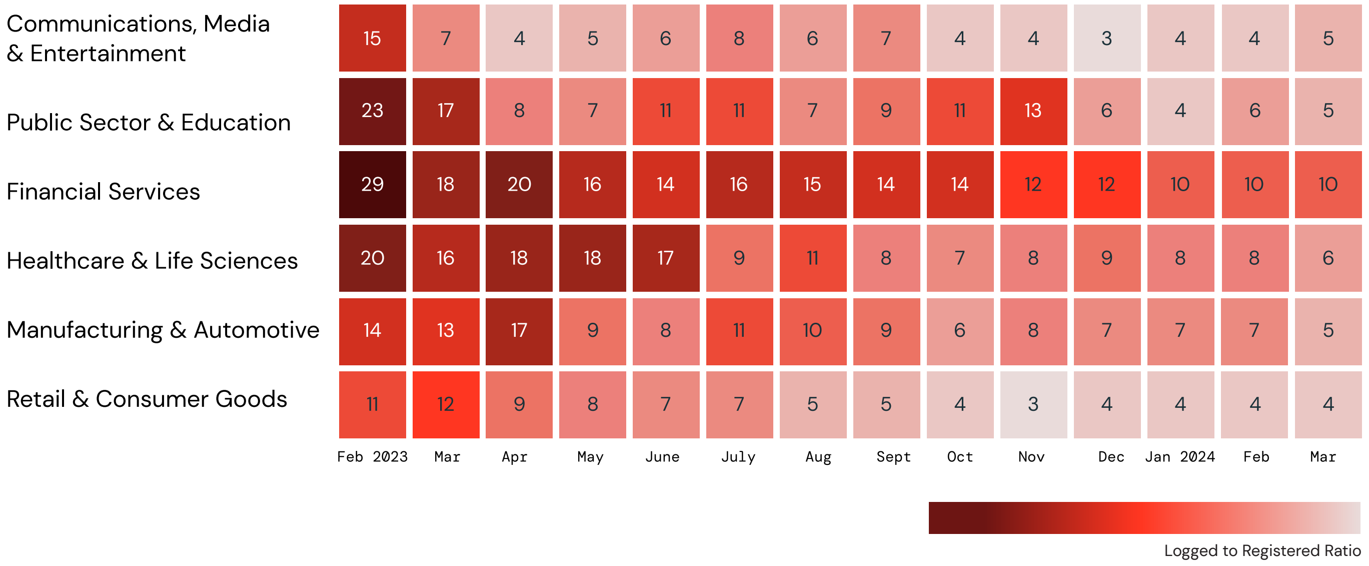
February 2023	March 2024
16:1	5:1

Efficiency at an industry level

Industries have different datasets, strategic goals and risk profiles. Therefore, we expect to see variations in their ML approach, including their mix of ML experimentation and production.

We analyzed six key industries to better understand these trends.

RATIO OF EXPERIMENTS LOGGED TO REGISTERED RATIO, BY INDUSTRY



	Feb 2023	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	Jan 2024	Feb	Mar
Communications, Media & Entertainment	15	7	4	5	6	8	6	7	4	4	3	4	4	5
Public Sector & Education	23	17	8	7	11	11	7	9	11	13	6	4	6	5
Financial Services	29	18	20	16	14	16	15	14	14	12	12	10	10	10
Healthcare & Life Sciences	20	16	18	18	17	9	11	8	7	8	9	8	8	6
Manufacturing & Automotive	14	13	17	9	8	11	10	9	6	8	7	7	7	5
Retail & Consumer Goods	11	12	9	8	7	7	5	5	4	3	4	4	4	4

Figure 2:
The ratio of logged-to-registered models steadily decreased between February 1, 2023–March 31, 2024, indicating that companies deployed more experimental models in production.

NOTE: Due to changes in the Model Registry API and tracking, this year's data does not directly correlate with last year's logged and registered model chart.

THE MOST EFFICIENT INDUSTRY, RETAIL, PUTS 25% OF THEIR MODELS INTO PRODUCTION

Retail & Consumer Goods reached a ratio of one model in production for every four experimental models, the most efficient of our featured industries. As outlined in the MIT Technical Review Insights report, Retail & Consumer Goods has long been an early-AI driver due to competitive pressure and consumer expectations.

Efficiency gain:
Financial Services became nearly 3x more efficient at getting models into production

FINANCIAL SERVICES SEES THE SHARPEST EFFICIENCY GAIN

Financial Services is the most testing-heavy industry. At the beginning of 2023, on average they logged 29 experiments for every one model registered. They became nearly 3x more efficient, ending March 2024 at a ratio of 10-to-1. The stakes for production ML are higher in regulated industries, which makes lengthy testing cycles critical.

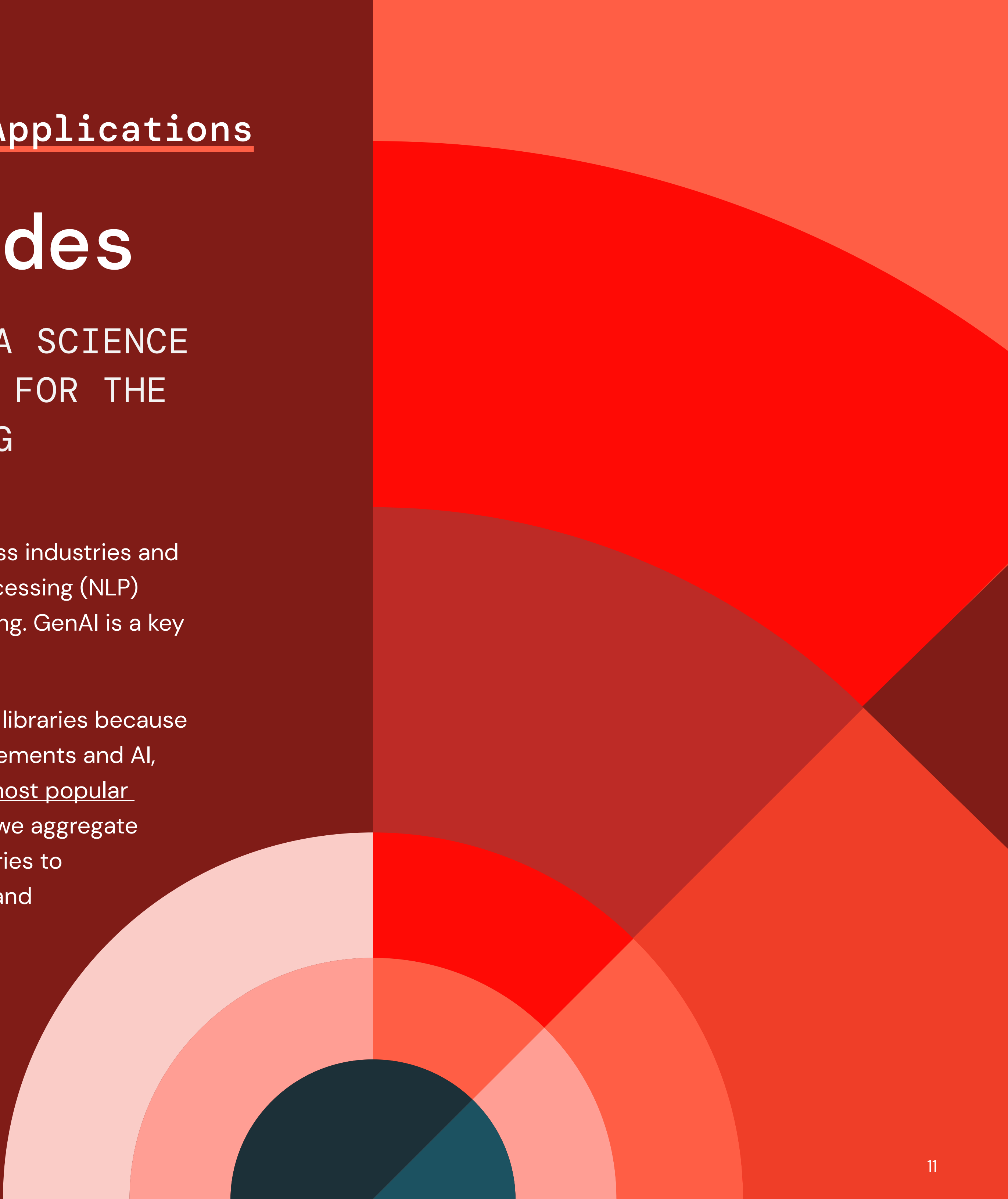
Why were more companies able to get more models into production this year? One factor is likely the availability of data intelligence platforms, which provide a standardized, open environment for practitioners across the ML lifecycle. Companies are able to execute each stage — from data preparation and model training to real-time serving and monitoring — on one platform while ensuring data governance, privacy and security. This increases the quality of output and supports production readiness.

NLP explodes

NLP IS THE TOP DATA SCIENCE AND ML APPLICATION FOR THE SECOND YEAR RUNNING

Unstructured data is ubiquitous across industries and regions, making natural language processing (NLP) techniques essential to derive meaning. GenAI is a key use case of NLP.

The following charts focus on Python libraries because they're at the forefront of ML advancements and AI, and consistently rank as one of the most popular programming languages. In our data, we aggregate the usage of specialized Python libraries to determine the top five data science and ML (DS/ML) applications used within organizations.



TOP DS/ML APPLICATIONS, BY INDUSTRY

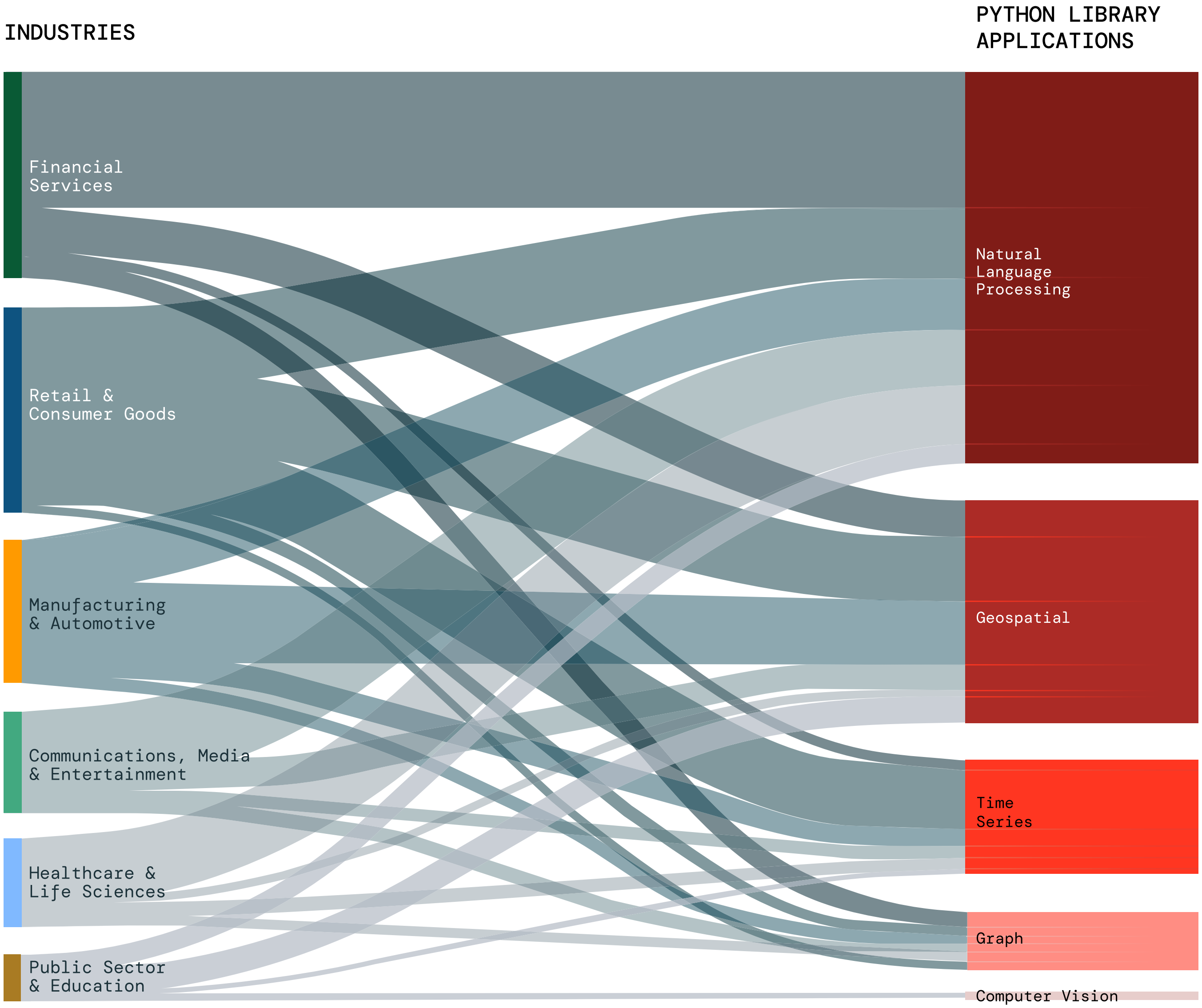


Figure 3:
NLP is the most commonly used Python library application, leveraged heavily by all our featured industries.

NOTE: This chart reflects the unique number of notebooks using ML libraries in each category. It does not include libraries used in tools for data preparation and modeling.

For the second year in a row, our data shows NLP is the top DS/ML application; 50% of specialized Python libraries used are associated with NLP.

Data teams are also eager to leverage Geospatial and Time Series applications. Geospatial libraries, which are often used for location-based analysis to customize user experiences, are the second most popular use case, accounting for 30% of Python library usage.

HEALTHCARE & LIFE SCIENCES HAS THE HIGHEST ADOPTION OF NLP

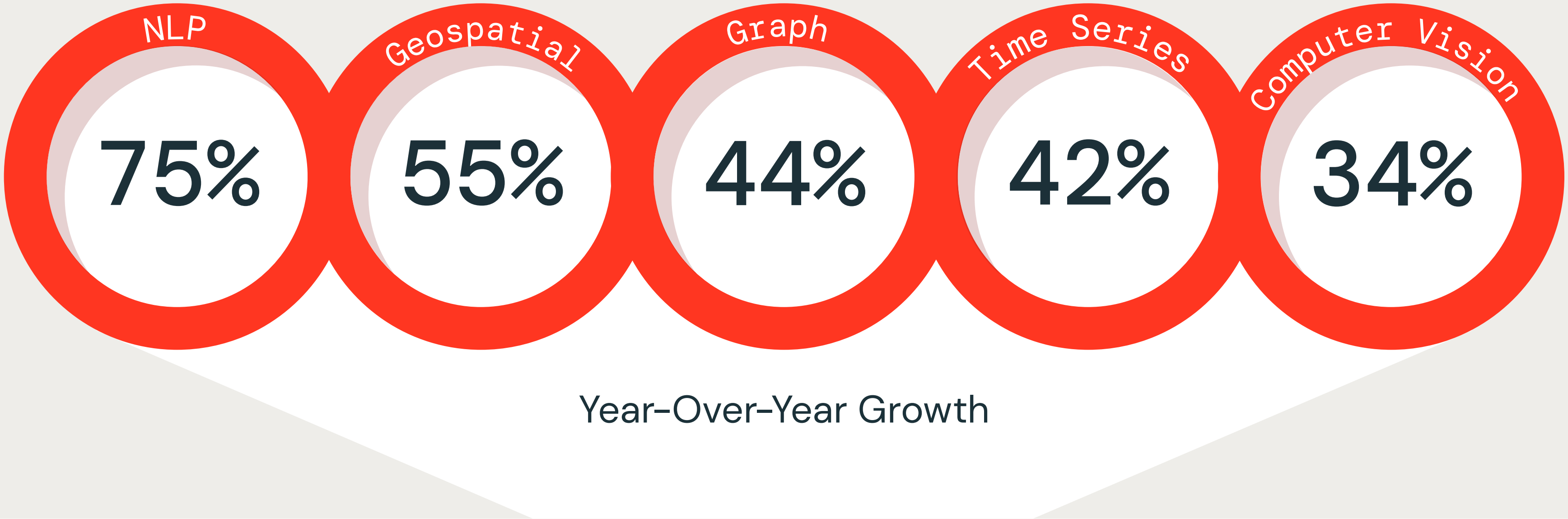
Among our featured industries, Healthcare & Life Sciences has the highest proportion of Python library usage devoted to NLP, at 69%. According to a survey done by Arcadia with the Healthcare Information and Management Systems Society, the healthcare industry generates 30% of the world's data volume and is growing faster than any other industry. NLP can support the analysis of clinical research, accelerate the process of bringing novel drugs to market and increase sales and marketing commercial effectiveness.

50% of specialized Python libraries used are associated with NLP

NLP, the most widely used DS/ML application, isn't slowing down

With the rise of AI-driven applications, there's a growing demand for NLP solutions across industries. While NLP dominates the use of Python libraries, it also has the highest growth of all applications at 75% YoY.

Fastest-Growing DS/ML Applications



FASTEST-GROWING DS/ML APPLICATIONS, BY INDUSTRY

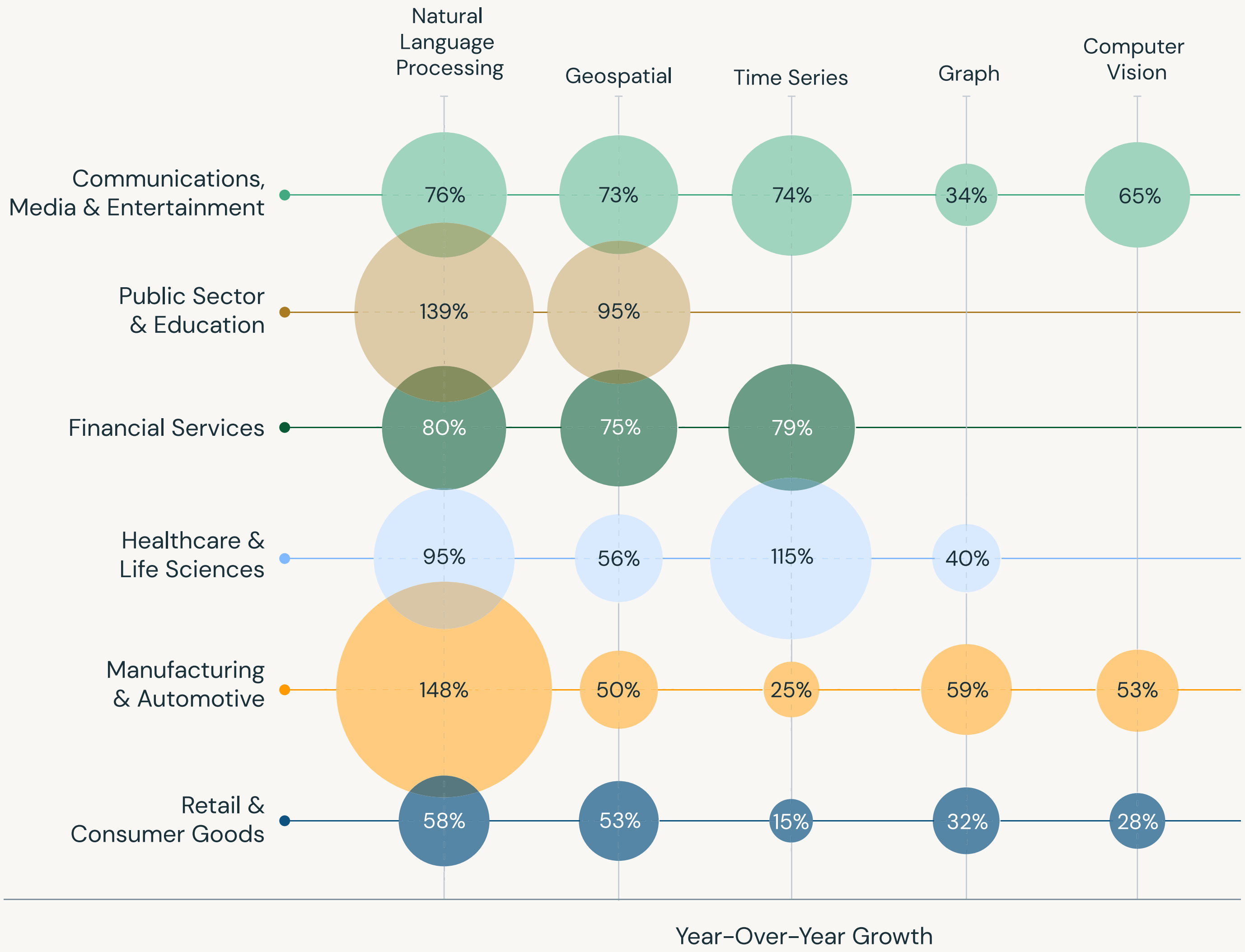


Figure 4:
NLP is experiencing the highest growth among applications. The largest YoY growth is the increase in Manufacturing & Automotive's use of NLP, at 148%.

ALL INDUSTRIES INVEST HEAVILY IN NLP

Among our featured industries, Manufacturing & Automotive had the largest gains in use of NLP, with a 148% YoY increase. NLP — which helps the industry do everything from analyzing feedback from customers to monitoring quality control to powering chatbots — enables companies to improve operational efficiency. Public Sector & Education's growth of NLP over the past year follows close behind, at 139% YoY.

FROM WILDFIRES TO BIRD FLU, CURRENT EVENTS CORRESPOND WITH ML GROWTH

Geospatial is the other application that grew significantly across all six industries. Companies are increasingly searching for patterns, trends and correlations in location-based data. The high rate of Geospatial growth from Public Sector & Education may relate to disaster management and emergency response planning.

The third highest rate of growth across all applications and industries is the adoption of Time Series libraries among Healthcare & Life Sciences, at 115% YoY. Time Series supports patient risk predictions, supply forecasting and drug discovery. In a 2023 review done by the NIH, they determined "time-series analysis allows us to do easily and, in less time, precise short-term forecasting in novel pandemics by estimating directly from data."[1]

[1] Applications of Time Series analysis in epidemiology: Literature review and our experience during COVID-19 pandemic, October 16, 2023.

THE MODERN DATA AND AI STACK

Evolution to GenAI

TOP DATA AND AI PRODUCTS SHOW THE NEXT PHASE OF GEN AI

Data leaders are always searching for the best tools to deliver their AI strategies. Our Top 10 Data and AI Products showcase the most widely adopted integrations on the Databricks Data Intelligence Platform. Our categories include DS/ML, data governance and security, orchestration, data integration and data source products.

Among our top products, 9 out of 10 are open source. Organizations are choosing more flexibility while avoiding proprietary walls and restrictions. As we'll discuss later in the report, we also see a growing popularity of open LLMs.

TOP 10 DATA AND AI PRODUCTS

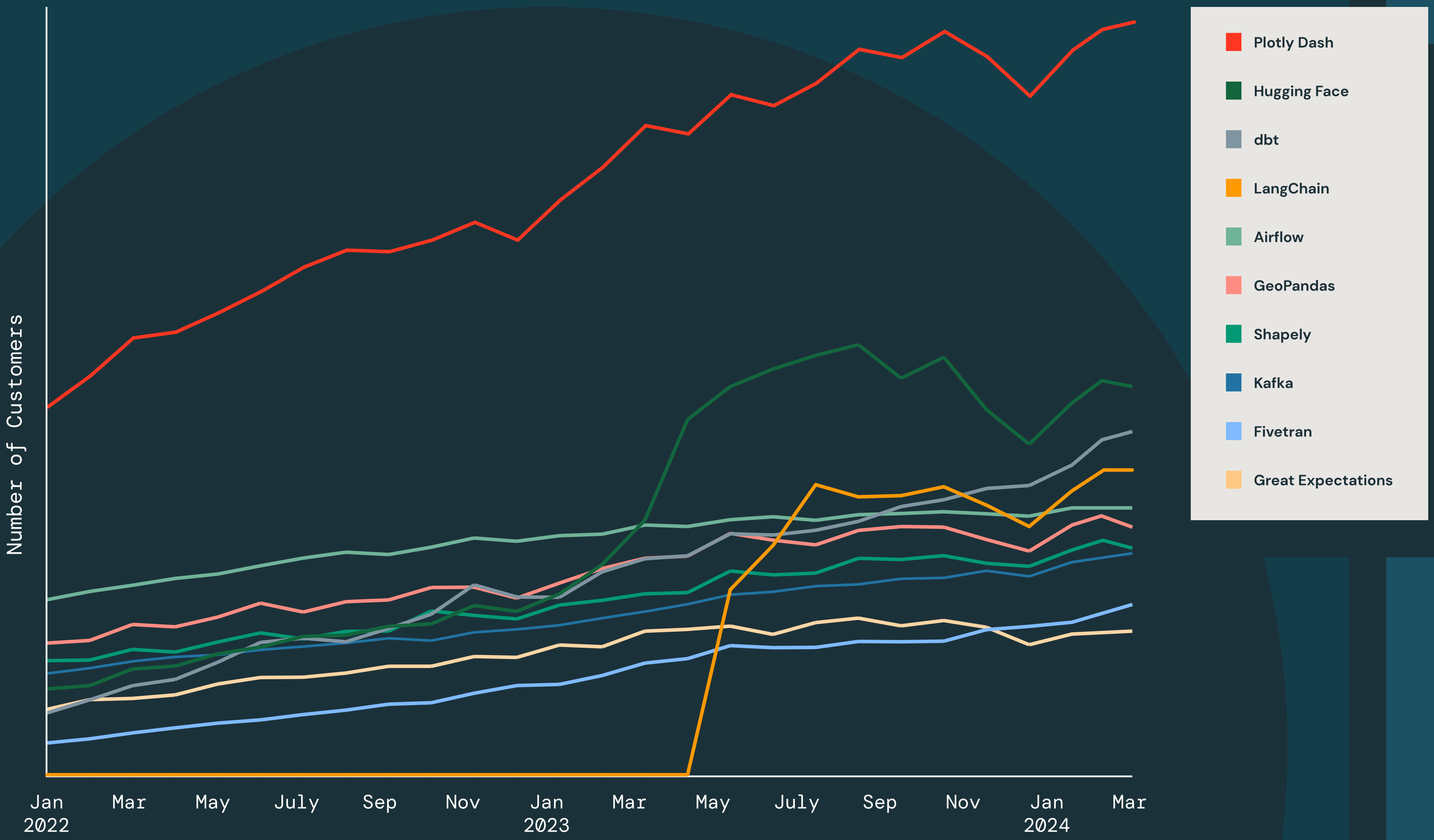


Figure 5: Our top 10 Data and AI Products span the categories of DS/ML, data governance and security, orchestration, data integration and data source products.

PLOTLY DASH MAINTAINS TOP POSITION

Plotly Dash is a low-code platform that enables data scientists to easily build, scale and deploy data applications. Products like Dash help companies deliver applications faster and more easily to keep up with dynamic business needs. For more than 2 years, Dash has held its position as No. 1, which speaks to the growing pressure on data scientists to develop production-grade data and AI applications.

HUGGING FACE TRANSFORMERS JUMPS TO NO. 2

Hugging Face Transformers ranks as the second most popular product used among our customers, up from No. 4 a year ago. Many companies use the open source platform's pretrained transformer models together with their enterprise data to build and fine-tune foundation models. This supports a growing trend we're seeing with RAG applications.

LANGCHAIN BECOMES A TOP PRODUCT ONLY MONTHS AFTER INTEGRATION

LangChain — an open source toolchain for working with and building proprietary LLMs — jumped into the top ranks last spring and rose to No. 4 in less than one year of integration. When companies build their own modern LLM applications and work with specialized transformer-related Python libraries to train the models, LangChain enables them to develop prompt interfaces or integrations to other systems.

COMPANIES INVEST IN PRODUCTS TO BUILD HIGH-QUALITY DATASETS

The prominence of three data integration products in our top 10 indicates companies are focused on building trusted datasets. dbt (data transformation), Fivetran (automation of data pipelines) and Great Expectations (data quality) all have steady growth. Most notably, dbt jumped two spots in the past year.

Rising Star

John Snow LABS

John Snow Labs is an AI and NLP company that helps Healthcare & Life Science organizations build, deploy and operate AI projects. Using advanced NLP, ML models and GenAI, John Snow Labs is instrumental in making disease diagnosis, drug discovery and patient care more accurate.

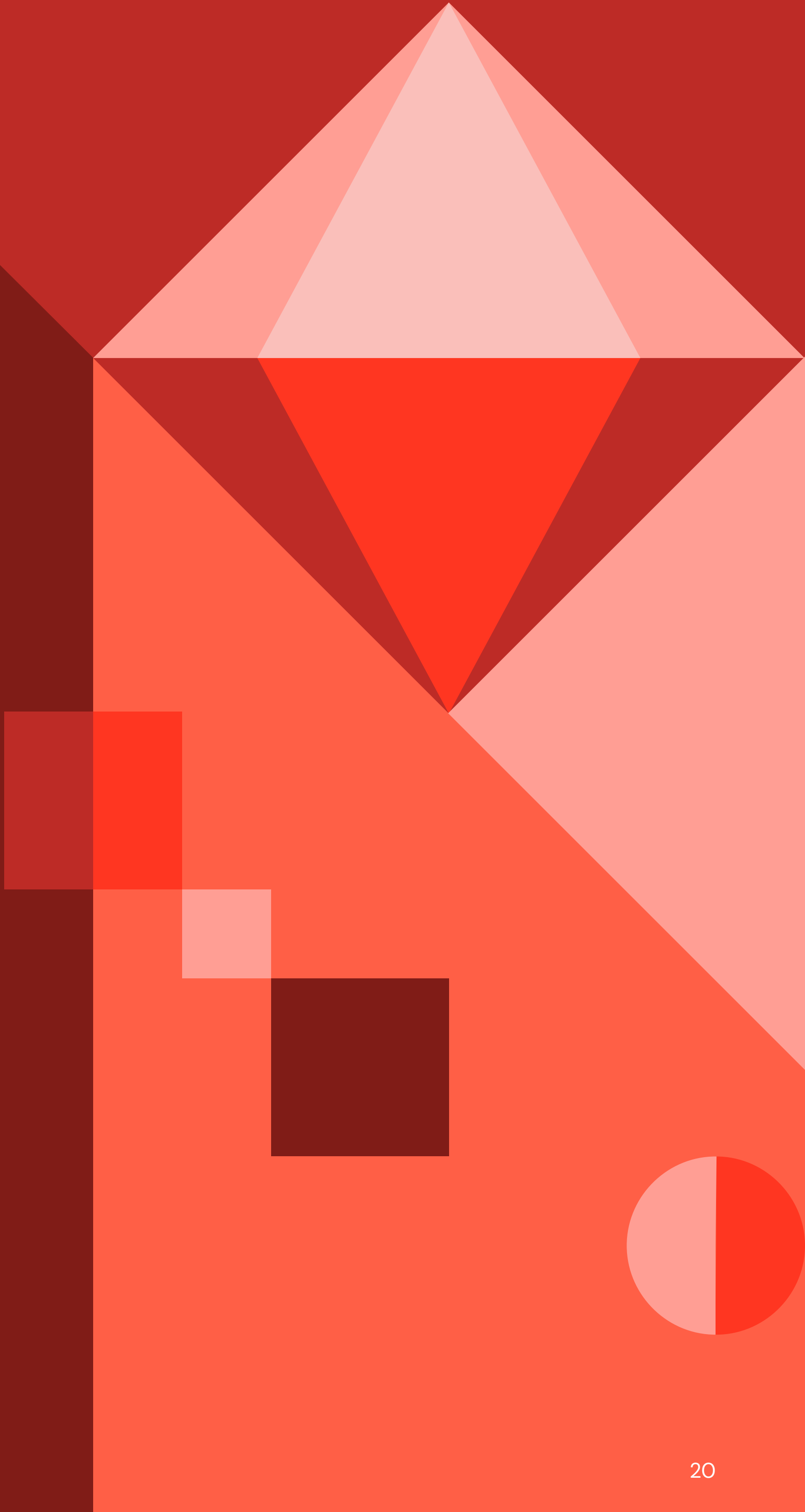
John Snow Labs deserves recognition because while it's predominantly used by Healthcare & Life Sciences, it ranks at No. 15 among our data and AI products. The company's widely used Spark NLP library supports a variety of NLP tasks, such as text classification, entity recognition and sentiment analysis, making it useful in other verticals, such as Financial Services.

Vector Databases

Enterprises rush to customize LLMs

LLMs support a variety of business use cases with their language understanding and generation capabilities. However, especially in enterprise settings, LLMs alone have limitations. They can be unreliable information sources and are prone to providing erroneous information, called hallucinations. At the root, stand-alone LLMs aren't tailored to the domain knowledge and needs of a specific organization.

Our data confirms that more companies are turning to RAG instead of relying on stand-alone LLMs. RAG enables organizations to use their own proprietary data to better customize LLMs and deliver high-quality GenAI apps. By providing LLMs with additional relevant information, the models can give more accurate answers and are less likely to hallucinate.



RAG leads the way for GenAI in the enterprise

Last year, our LLM Python Libraries chart revealed the hot trajectory of SaaS LLMs, which grew 1,310% in just over 5 months. SaaS LLMs like GPT-4 are trained on massive text datasets and went mainstream less than 2 years ago.

This year, vector database adoption is tearing up our chart. The entire vector database category has grown 377% YoY, and 186% just since the Public Preview of Databricks Vector Search.[2]

WHAT IS RAG?

Retrieval augmented generation (RAG) is a GenAI application pattern that finds data and documents relevant to a question or task and provides them as context for the LLM to give more accurate responses.

HOW DO VECTOR DATABASES AND RAG WORK TOGETHER?

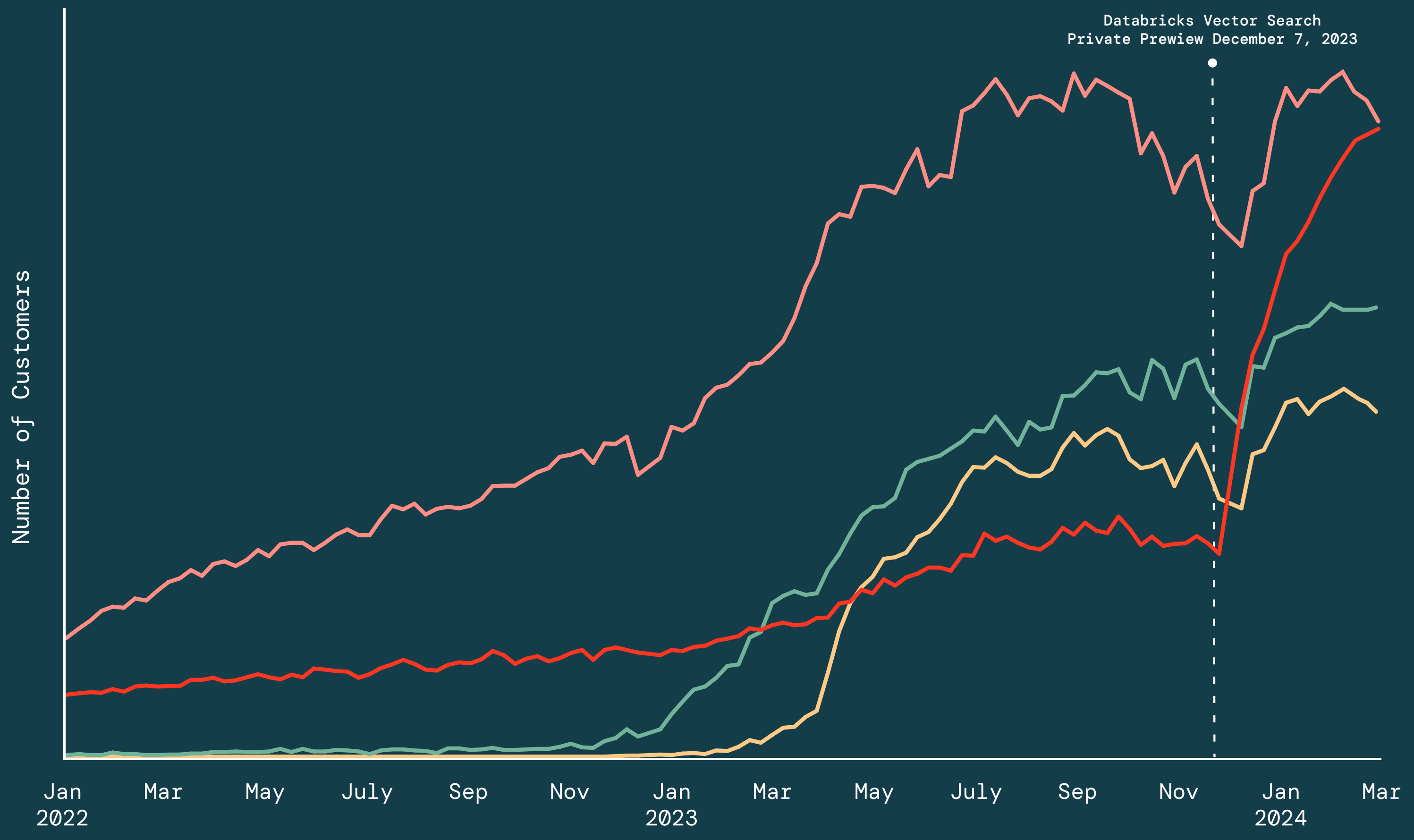
Vector databases generate representations of predominantly unstructured data. This is useful for information retrieval in RAG applications to find documents or records based on their similarity to keywords in a query.

RAG applications have a lot of advantages over off the shelf. RAG has quickly emerged as a popular way to incorporate proprietary, real-time data into LLMs without the costs and time requirements of fine-tuning or pretraining a model.

The exponential growth of vector databases suggests that companies are building more RAG applications in order to integrate their enterprise data with their LLMs.

[2] Databricks Vector Search went in Public Preview on December 7, 2023.

USE OF LLM PYTHON LIBRARIES



LLM DEFINITIONS

Transformer training: Libraries for training transformer models (e.g., Hugging Face Transformers)

SaaS LLMs: Libraries for accessing API-based LLMs (e.g., OpenAI)

LLM tools: Toolchains for working with and building proprietary LLMs (e.g., LangChain)

Vector databases: Vector/KNN indexes (e.g., Pinecone and Databricks Vector Search)

Figure 6: Since Public Preview of Databricks Vector Search, the entire vector database category has grown 186%, far more than any other LLM Python library.

NOTES: Customers may be using more than one tool in a category and, thus, may be counted more than once.
Usage is measured by package use connecting to external vector database services and by API calls on our platform.
Trend lines have been averaged between December 18 and January 1 to address seasonal dips.

COMPANIES ARE BECOMING MORE SOPHISTICATED IN BUILDING LLMs

Last year, customers were jumping into LLMs with off-the-shelf models. We still see 178% YoY growth in the number of customers using SaaS LLMs. But companies are beginning to take more control over their LLMs and build tools specific to their needs.

The continuing growth of vector databases, LLM tools and transformer-related libraries shows that many data teams are choosing to build vs. buy. Companies increasingly invest in LLM tools, such as LangChain, to work with and build proprietary LLMs. Transformer-related libraries like Hugging Face are used to train LLMs, and still claim the highest adoption by number of customers. Use of these libraries grew 36% YoY. Together, these trend lines indicate a more sophisticated adoption of open source LLMs.

377% YoY growth
in the number of customers
using vector databases

Companies prefer smaller open source models

USE OF OPEN SOURCE LLMs

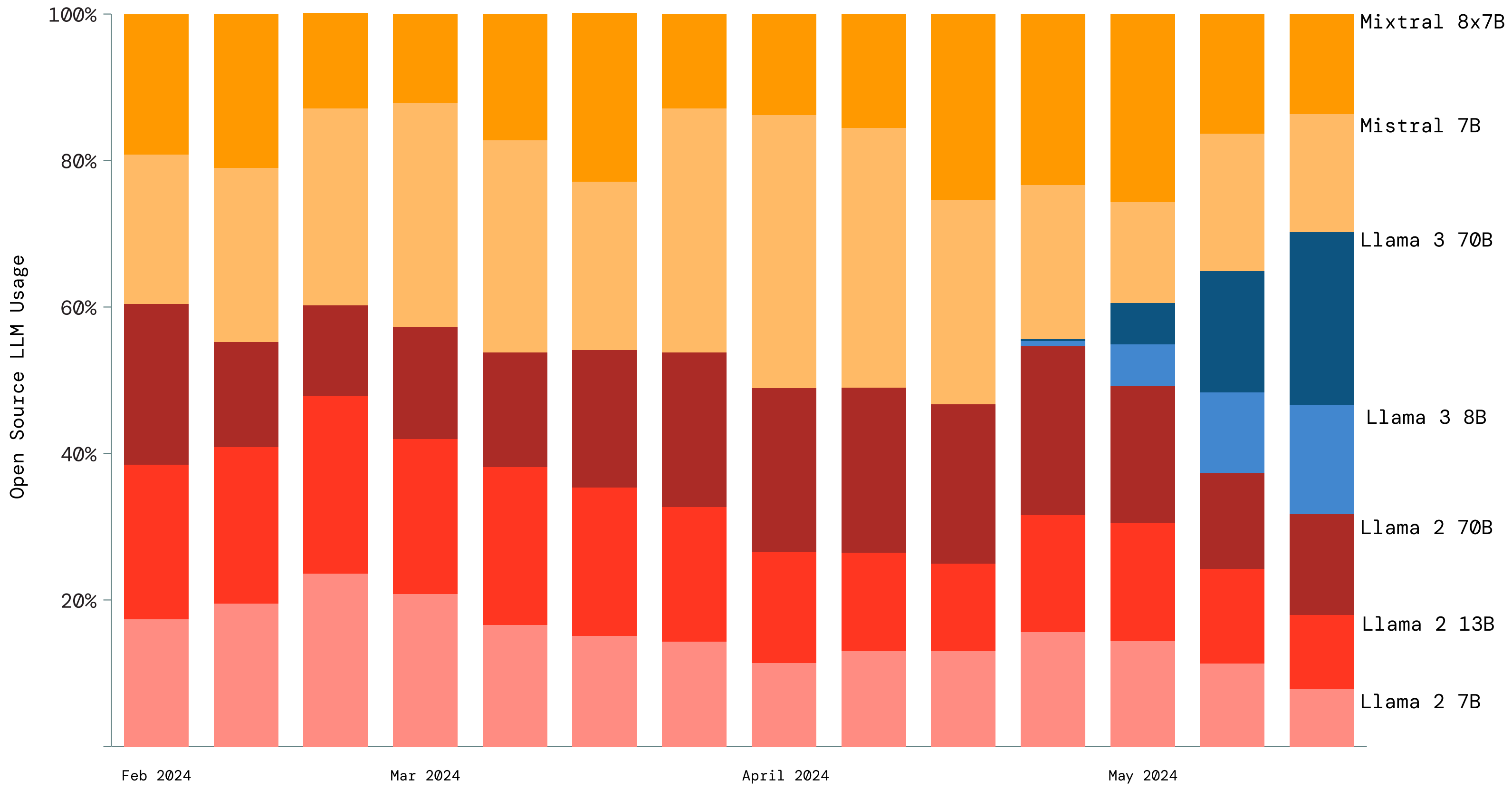


Figure 7: Relative adoption of Mistral and Meta Llama open source models in Databricks' foundation model APIs.

NOTE: Chart extended to May 19, 2024, to accommodate the Meta Llama 3 launch.

One of the biggest benefits of open source LLMs is the ability to customize them for specific use cases — especially in enterprise settings. We often hear the question: *What's the most popular open source model?* In practice, customers often try many models and model families. We analyzed the open source model usage of Meta Llama and Mistral, the two biggest players. Our data shows that the open LLM space is fluid, with new state-of-the-art models getting rapid adoption.

With each model, there is a trade-off between cost, latency and performance. Together, usage of the two smallest Meta Llama 2 models (7B and 13B) is significantly higher than the largest, Meta Llama 2 70B. Across Meta Llama 2, Llama 3 and Mistral users, 77% choose models with 13B parameters or fewer. This suggests that companies care significantly about cost and latency.

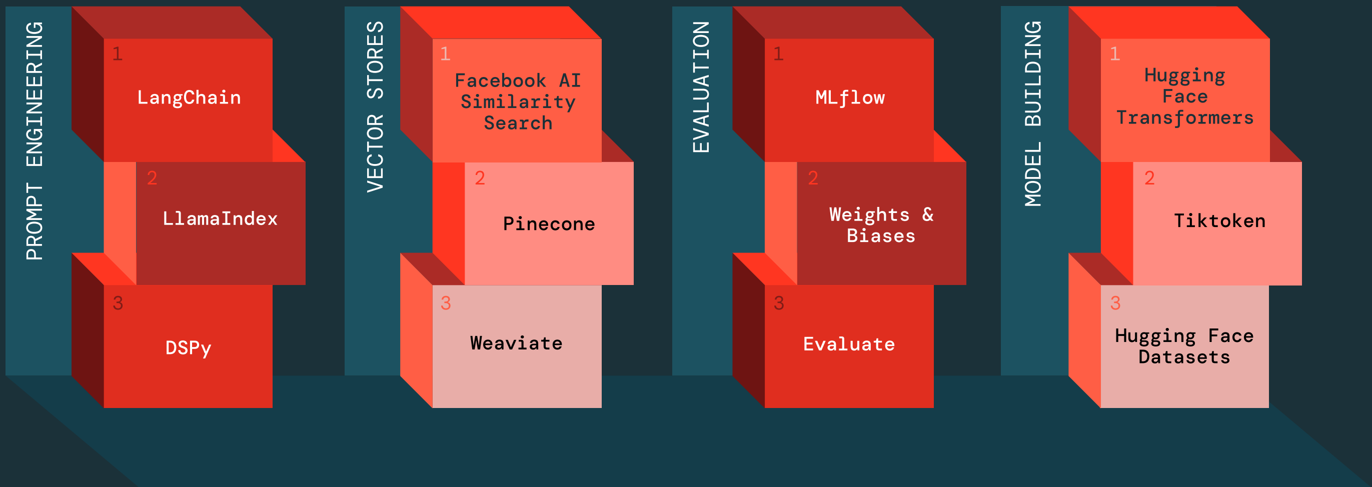
76%
of companies that use LLMs are choosing open source models, often alongside proprietary models.

70%
of companies that leverage GenAI are using tools, retrieval and vector databases to customize models.

COMPANIES ARE QUICK TO TRY NEW MODELS

Meta Llama 3 launched on April 18, 2024. Within its first week, organizations already started leveraging it over other models and providers. Just 4 weeks after its launch, Llama 3 accounted for 39% of all open source LLM usage.

Top GenAI Python Packages



PROMPT ENGINEERING
1 LangChain
2 LlamaIndex
3 DSPy

VECTOR STORES
1 Facebook AI Similarity Search
2 Pinecone
3 Weaviate

EVALUATION
1 MLflow
2 Weights & Biases
3 Evaluate

MODEL BUILDING
1 Hugging Face Transformers
2 Tiktoken
3 Hugging Face Datasets

Generative AI

Highly regulated industries are early adopters

Highly regulated industries have the reputation of being risk averse and hesitant to adopt new technologies. There are multiple reasons, including strict compliance requirements, ingrained legacy systems that are costly to replace and the need for regulatory approval before implementation.

While all industries are embracing new AI innovations, two highly regulated industries — Financial Services and Healthcare & Life Sciences — are keeping pace with, and often surpassing, their less-regulated counterparts.

In December 2023, Databricks released foundation model APIs, providing instant access to popular open source LLMs, such as Meta Llama and MPT models. We expect the interest in open source to grow significantly as models continue to rapidly improve, as shown by the recent launches of Llama 3.

HARNESSING OPEN LLMs FOR INDUSTRY-SPECIFIC NEEDS

Manufacturing & Automotive and Healthcare & Life Sciences take the lead in adopting foundation model APIs with the highest average usage per customer. In manufacturing, supply chain optimization, quality control and efficiency are deemed the most promising use cases.

A recent report from MIT Tech Review Insights shares that, among those surveyed, CIOs in Healthcare & Life Sciences believe GenAI will bring value to their organizations. Open source LLMs enable highly regulated industries like Healthcare & Life Sciences to integrate GenAI while maintaining the utmost control of their data.

Foundation Model APIs Usage, by Industry

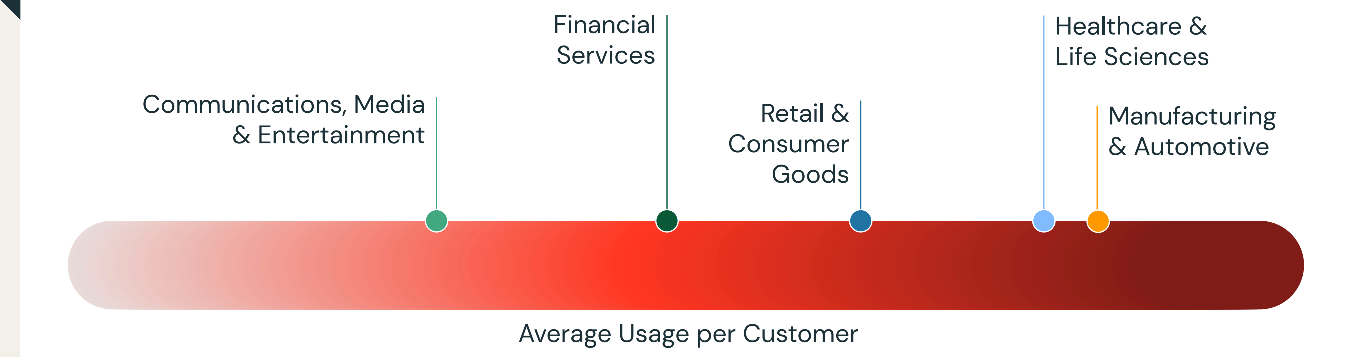


Figure 8: Manufacturing & Automotive and Healthcare & Life Sciences lead the adoption of foundation model APIs with the highest average usage per customer.

NOTE: Date Range: January 2024 to March 2024.

CPUs vs. GPUs: Financial Services' commitment to LLMs grows 88% in 6 months

CPUs are general-purpose processors designed to handle a wide range of tasks quickly, but they are limited in how many tasks they can handle in parallel. CPUs are used for classic ML. GPUs are specialized processors that can parallel-process thousands or millions of separate tasks at once. GPUs are necessary to train and serve LLMs.

We looked at CPU and GPU usage and growth among our Model Serving customers to understand how they're approaching AI. The GPUs in our data are predominantly associated with LLMs.

FINANCIAL SERVICES DOMINATES GPU USAGE

Financial Services, one of the most regulated industries, isn't shying away from GenAI. It has by far the highest average usage of GPUs per company, as well as the highest GPU growth, at 88% over the past 6 months. LLMs support business-critical use cases, including fraud detection, wealth management, and investor and analyst applications.

MODEL SERVING, CLASSIC ML VS. LLM

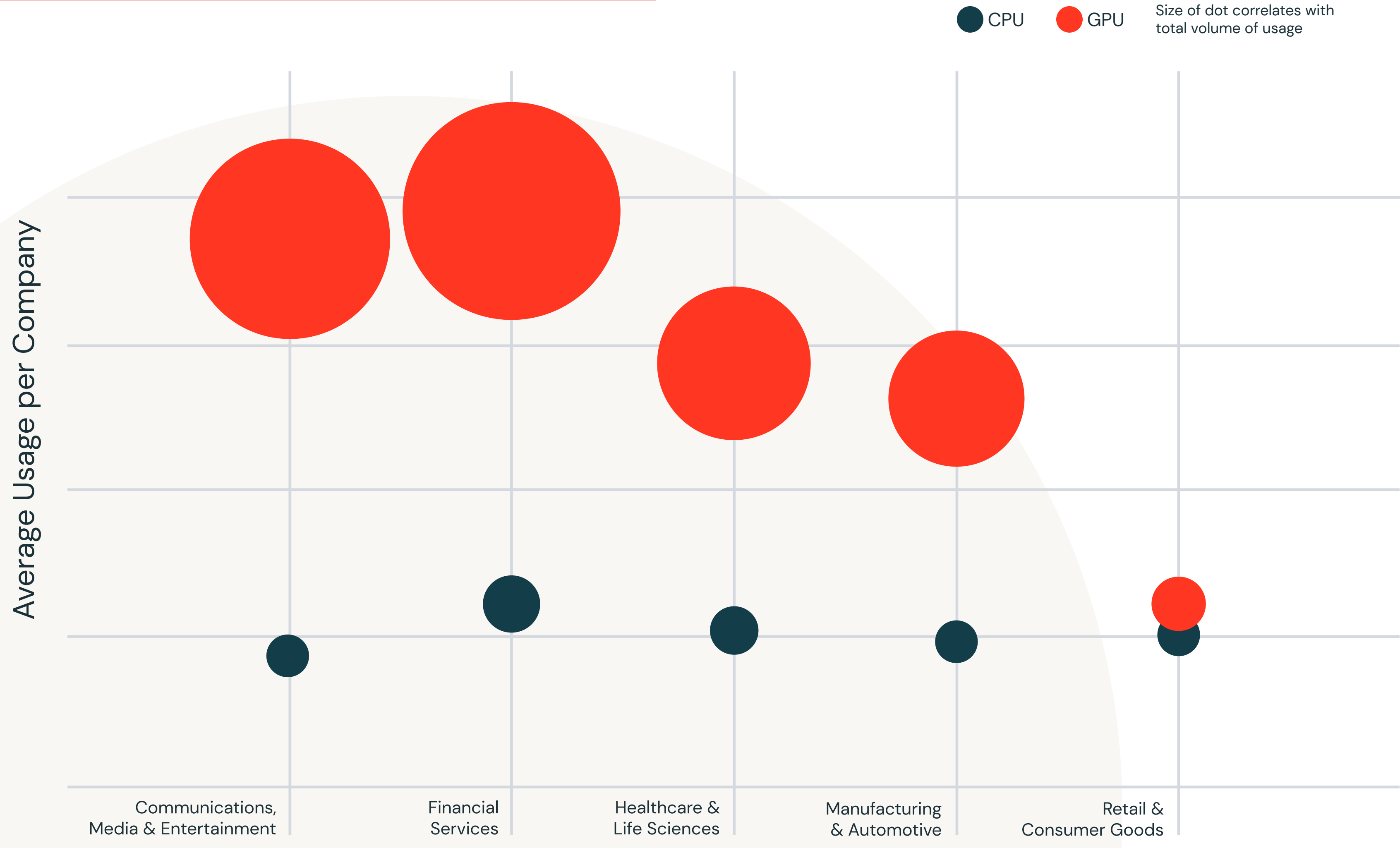


Figure 9: Financial Services has the highest average usage of both CPU and GPUs.

NOTE: Data range: January 2024 to March 2024.

Highly regulated industries lead the adoption of unified governance

AI security and governance are critical to establishing trust in an organization's AI initiatives. They help data practitioners develop and maintain products while adhering to precise guidelines and standards. Unified governance solutions, like Databricks Unity Catalog, span all data and AI assets, and make it easier for organizations to train and deploy GenAI models on their private data.

According to Gartner, AI trust, risk and security management are the top trends in 2024 that will factor into business and technology decisions. Now more than ever, leaders want to leverage data and AI to transform their organizations. We see this reflected in the adoption of unified governance among our customers.

FINANCIAL SERVICES IS AT THE FOREFRONT OF DATA AND AI GOVERNANCE

Regulatory and security compliance is engrained in the culture of Financial Services organizations. According to survey data from the CIO vision 2025 report by MIT Technology Review Insights, financial institutions are expected to see the highest investment growth in data management and infrastructure, estimated at “74% between now and 2025, according to financial industry respondents, compared with 52% for the sample as a whole.”

ADOPTION OF UNITY CATALOG, BY INDUSTRY

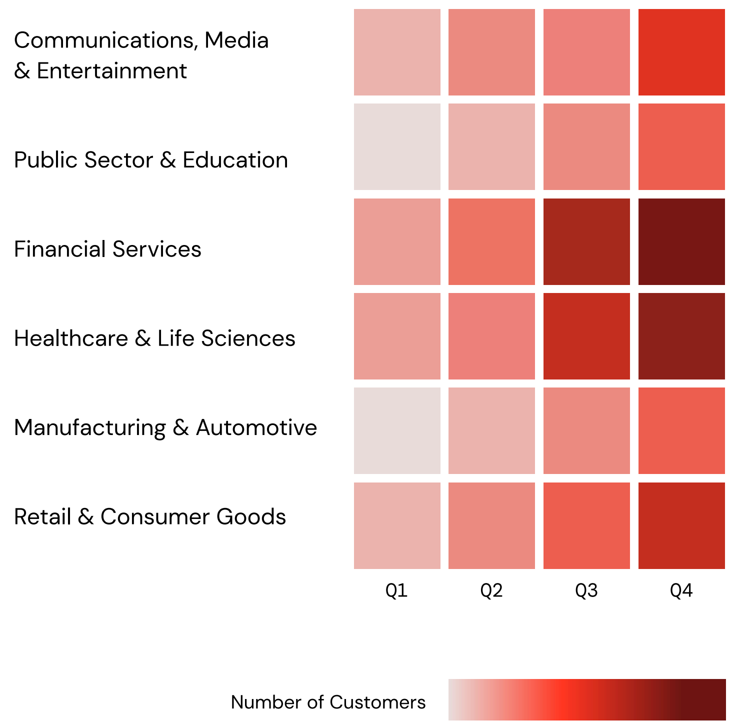


Figure 10: Financial Services leads the adoption of Unity Catalog for unified data and AI governance.

NOTE: Data range: February 1, 2023, to January 31, 2024.

ADOPTION OF SERVERLESS MODEL SERVING, BY INDUSTRY

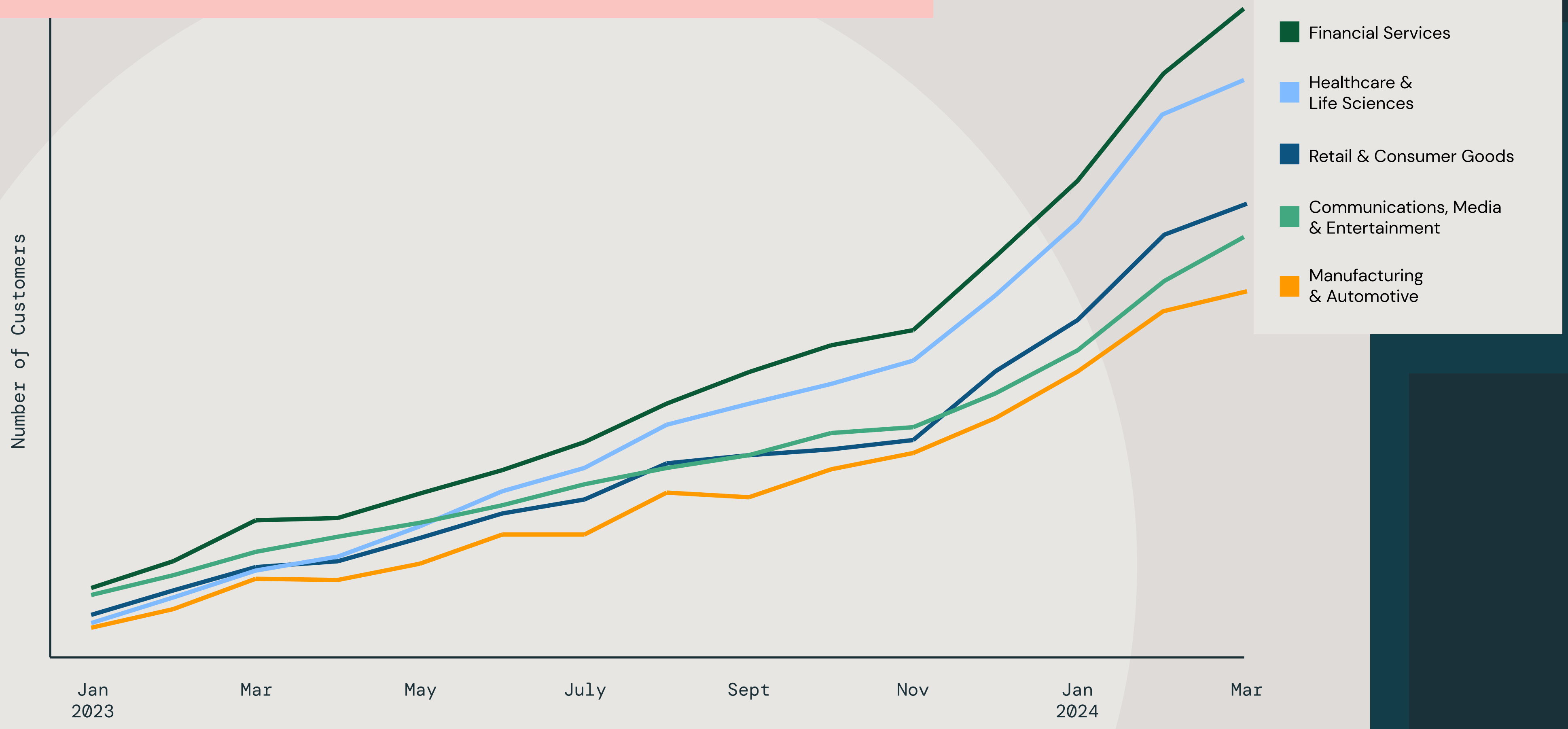


Figure 11: Financial Services leads the adoption of serverless products, followed by Healthcare & Life Sciences.

NOTES: This includes Model Serving on Serverless Endpoints, Databricks SQL, Lakehouse Monitoring and Serverless jobs. In November 2023, Serverless became available to additional regional cloud platforms.

Companies shift to serverless to build real-time ML applications

Real-time ML systems are revolutionizing how businesses operate by providing the ability to make immediate predictions or actions based on incoming data. But they need a fast and scalable serving infrastructure that requires expert knowledge to build and maintain.

Serverless model serving automatically scales up or down to meet demand changes, reducing cost as companies only pay for their consumption. Companies can build real-time ML applications ranging from personalized recommendations to fraud detection. Model serving also helps support LLM applications for user interactions.

We have seen steady growth in the adoption of serverless data warehousing and monitoring, which also scales with demand.

Financial Services, the largest adopter of serverless products, grew usage 131% over 6 months. This industry strives to predict the markets, and real-time prediction provides stronger market analysis.

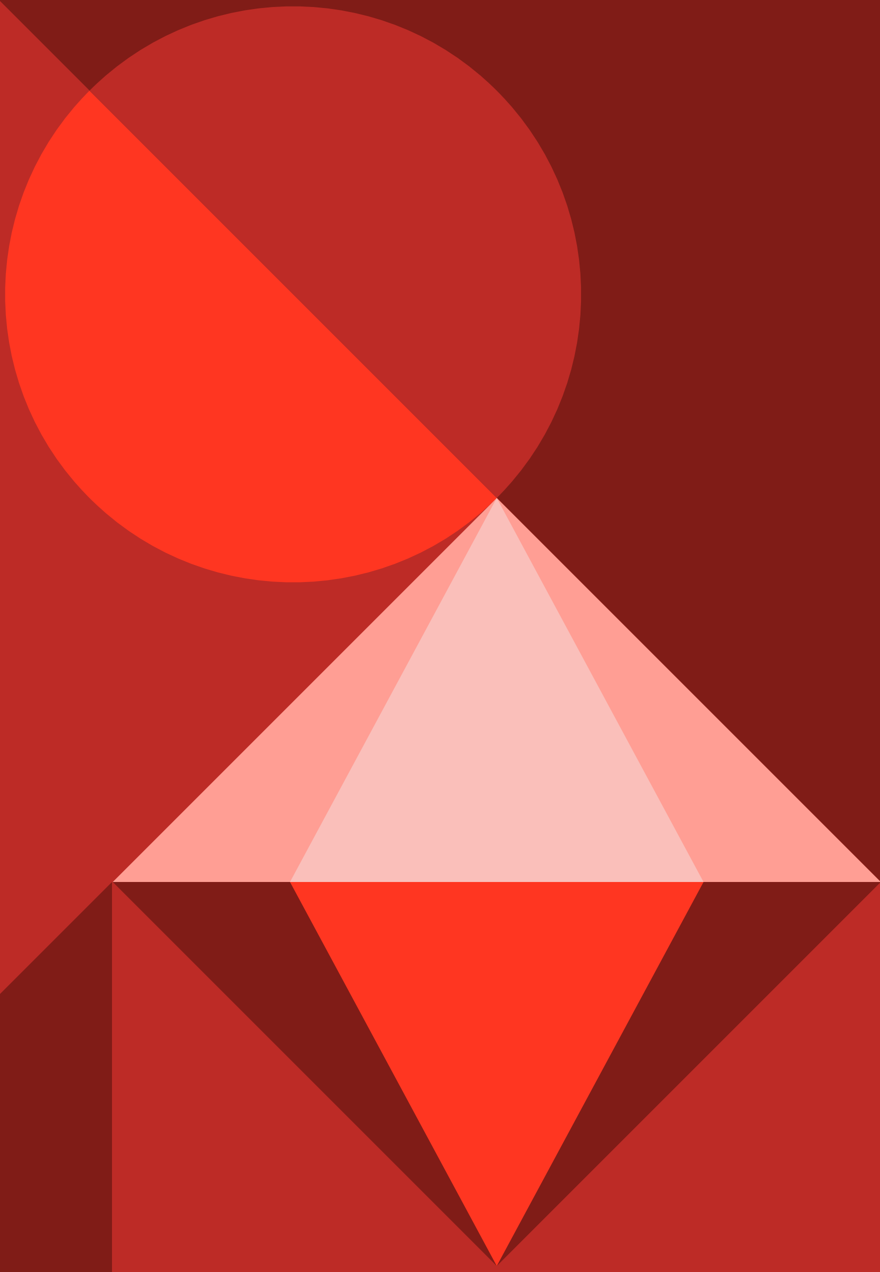
Healthcare & Life Sciences grew usage of serverless products 132% over 6 months. The industry has moved from No. 4 to No. 2 over the past year. Healthcare & Life Sciences experiences significant fluctuations in data processing requirements, especially during peak times or when dealing with large datasets such as genomic data or medical imaging.

Conclusion

Data science and AI are propelling companies toward greater efficiency, and GenAI is opening up a new landscape of possibilities. With data intelligence platforms, there is one cohesive, governed place for the entire organization to use data and AI. Our data shows companies across all industries are embracing these tools, and early adopters may come from industries you may not expect.

Organizations have realized measurable gains in putting ML models into production. Companies are increasingly adopting and using NLP to unlock insights from data. They are using vector databases and RAG applications to integrate their own enterprise data into their LLMs. Open source tools are the future, as they continue to rank high among our most popular products. Companies are strategizing with unified data and AI governance.

The takeaway: The winners in every industry will be those who most effectively use data and AI.



About Databricks

Databricks is the data and AI company. More than 10,000 organizations worldwide — including Block, Comcast, Condé Nast, Rivian, Shell and over 60% of the Fortune 500 — rely on the Databricks Data Intelligence Platform to take control of their data and put it to work with AI. Databricks is headquartered in San Francisco, with offices around the globe, and was founded by the original creators of Lakehouse, Apache Spark™, Delta Lake and MLflow.

To learn more, follow Databricks on LinkedIn, X and Facebook.

databricks