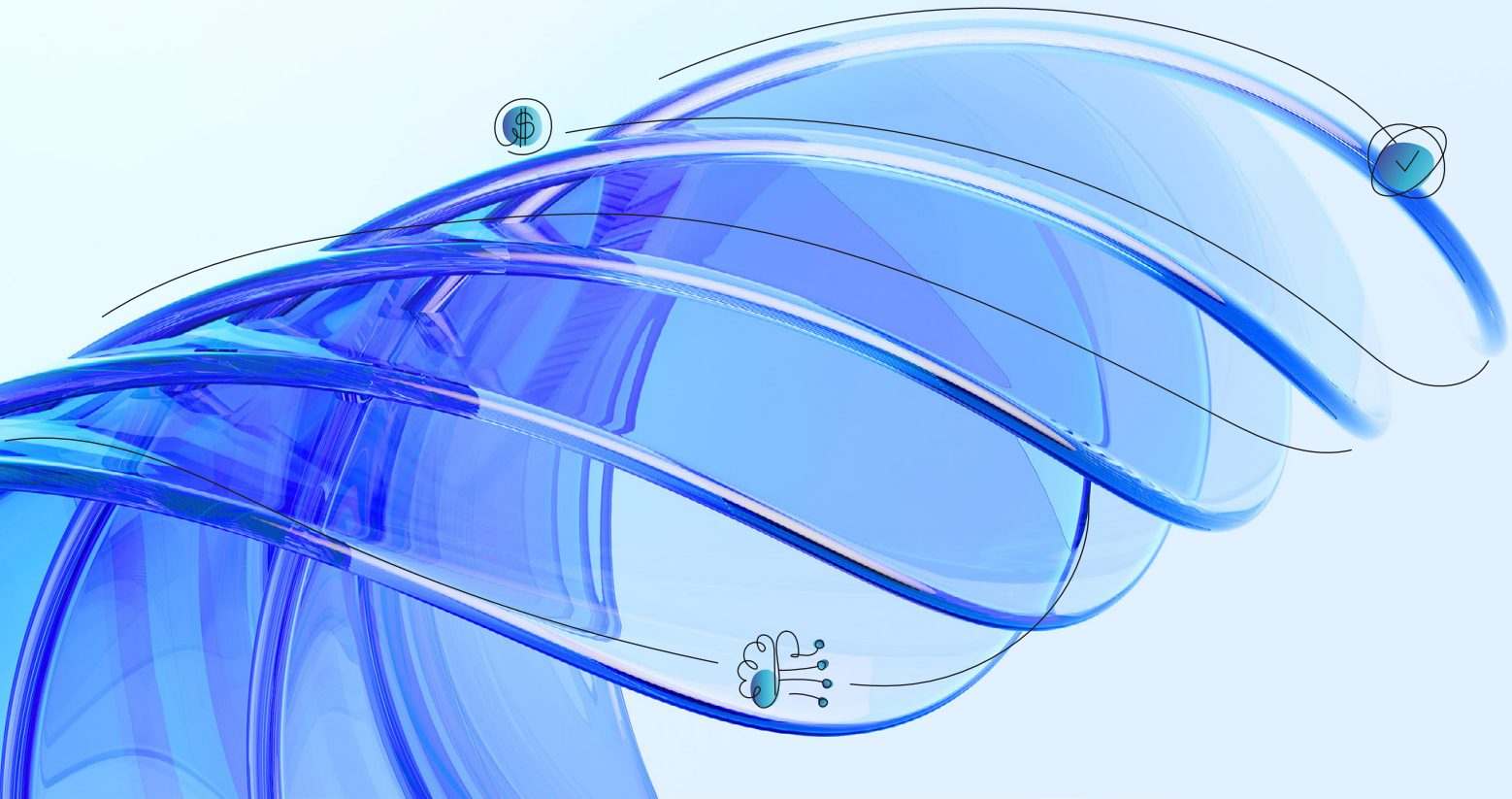
TEAM8™

THE PULSE ON DATA & AI

Key Findings from Team8's 2025 **Data Village Survey**

May 2025



WRITTEN BY

Aviad Harell
Managing Partner,
Team8

Kalman Heims
Strategy & Business
Research Manager,
Team8

Nathaniel Tavisal
Strategy & Business
Operations Manager,
Team8

Omri Sela
Business Development
Manager, Team8

CONTRIBUTORS

Nik Acheson
VP of Data Strategy,
Architecture & Engineering,
Highmark Health

Preeti Vaidya
VP, Data,
Morgan Health,
JPMorganChase

Diego de Aragão
SVP, Balance Sheet
Management & Analytics,
Citi

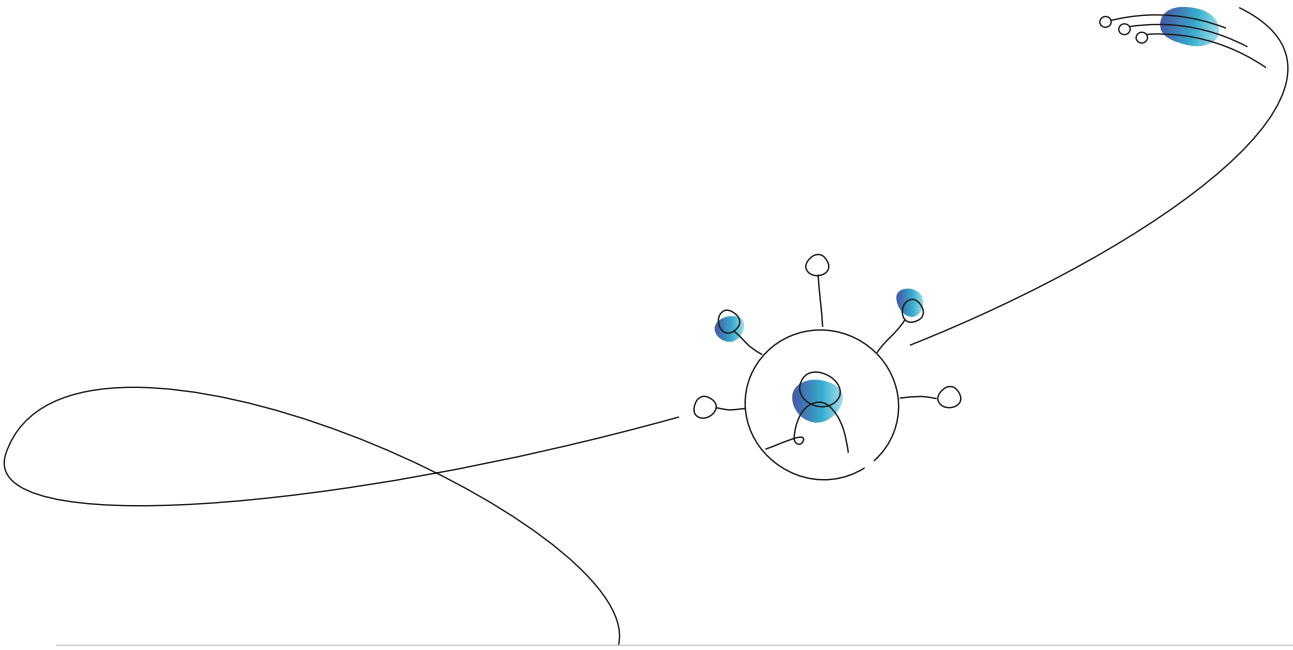
AJ Wilson
Chief Product Officer,
Data & Analytics,
Prudential

The Team8 Data Village is a community of data experts and thought leaders from the world's leading enterprises. The primary focus of the Village is to facilitate collaboration among the world's most prominent companies with the goal of sharing information and ideas, conducting intimate discussions on industry and technology trends and needs, and generating value for all parties.

By helping Team8 to identify real pain points and understand the requirements of large organizations,members of the Village are first in line to leverage solutions that are purpose-built by Team8's portfolio companies to support their needs.

To contact the Team8 Data Village, please email datavillage@team8.vc

Table of Contents



Executive Summary & Key Takeaways

This report is the second annual Data Survey from Team8 and presents insights learned from a survey of dozens of data and analytics executives and leaders. Respondents hailed from varied companies and industries, from scaling tech companies to Fortune 500 enterprises, and were asked to address key trends and challenges in their daily work.

Our analysis helps understand the nature of modern data and AI teams - including spending patterns, top priorities and AI adoption. These findings help contextualize the ways in which organizations are navigating today's data landscape and the strategies they are employing when adopting data best practices in their organizations.

Key Findings

Budget

Similar to last year's survey, the vast majority of organizations surveyed are planning on increasing their budgets for 2025. Notably, 29% of organizations plan to increase their spend by over 10%. We also have seen the rise of exploratory AI budgets that are beginning to turn into a permanent and significant spend category.

Tooling

The focus on AI has only grown stronger with a staggering 62% of data and AI leaders prioritizing investments in their AI/ML platforms for this year. Following this, data and AI leaders are prioritizing investments in data lakehouse architectures and data governance – two fundamental areas for successful AI adoption.

Pain Points

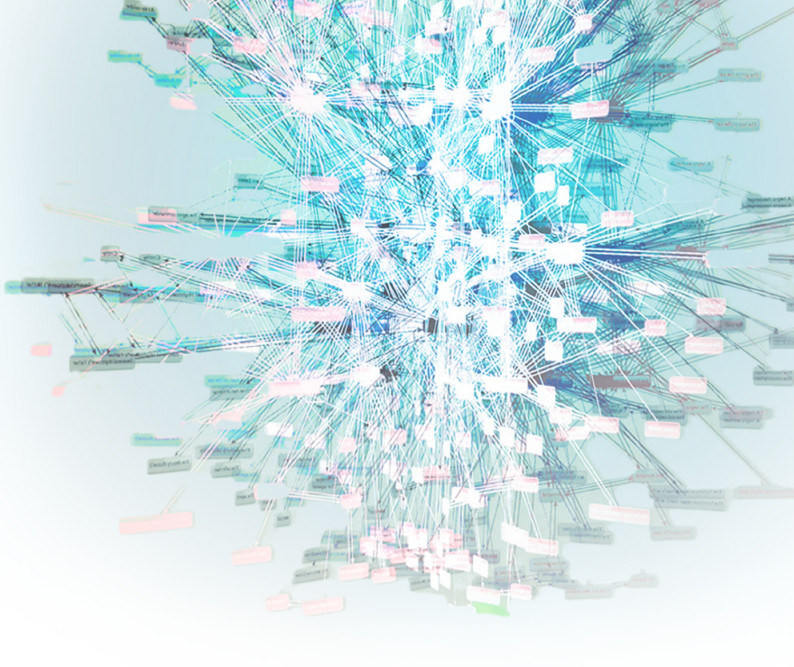
Similar to last year, data quality remained the top pain point for data and AI leaders. Data quality has again proven its importance in order to successfully implement AI in organizations. This was followed by increasing efficiency within data teams, which follows a larger mandate of efficiency for many companies at the moment. Tied with efficiency, production AI and deploying AI models at scale is a top pain for data and AI leaders this year.

AI Adoption

We were surprised at the scale of AI adoption among surveyed organizations. 42% of organizations are already deploying multiple AI applications in production. This is a major step up from a year ago when most companies were just beginning with early POCs and deployments. We are still in the early innings and see no signs of this trend slowing down.

01

Data Spend



How much do businesses plan to spend on data in 2025?

We categorized spending into two key areas: data infrastructure – covering expenses related to warehousing, compute, cloud services, and storage – and data analytics, AI, and governance tooling, encompassing licenses for business intelligence, catalogs, observability, ETL, MLOps and other tooling.

Similar to last year's results, in nearly every organization we surveyed, data infrastructure investment in 2025 exceeded spending on data analytics and governance tooling, sometimes by a significant ratio.

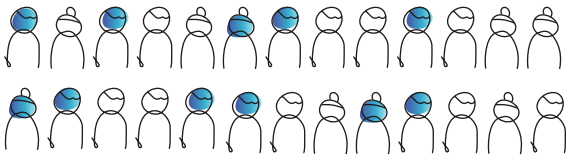
Let's look at how this spending differed among organizations of different sizes:

<250
Employees

For organizations under 250 employees, data infrastructure spend was much smaller, with most companies spending less than $1M in total spend. These are price sensitive organizations, and often don't have the need or ability to spend large amounts on building out a full data platform.

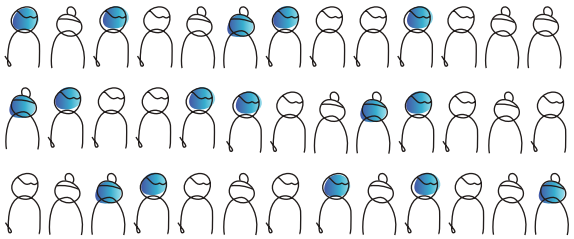
250-1,000
Employees

Organizations ranging from 250 to 1,000 employees demonstrated a noteworthy uptick in spending compared to organizations with under 250 employees, both in infrastructure and tooling. At this size, many organizations begin to deal with the complexities of managing large amounts of data across multiple business units, requiring the need for more specialized data tooling. This group of companies typically spent $1M - $3M on data infrastructure, and slightly less on data tooling.



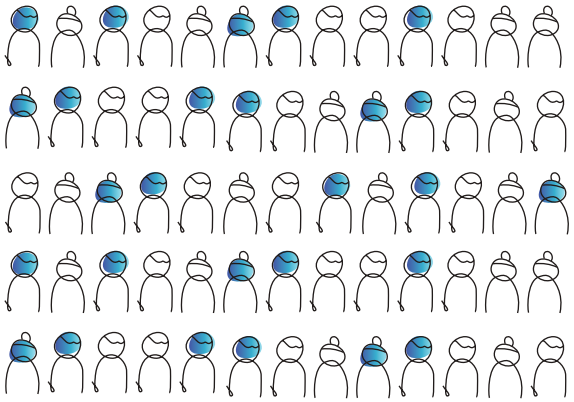
1,000-10,000
Employees

This group of companies accounted for nearly ⅓ of survey respondents, and demonstrated substantial spend increase over their sub-1000 employee peers. 38% of respondents in this category reported infrastructure spending over $5M. A similar 31% of respondents reported data and AI tooling spend over $3M. Still, there is a large variation between companies with multiple companies reporting under $500k in infrastructure spend and a couple companies topping $30M.



10,000-50,000
Employees

Organizations from 10,000-50,000 employees had a slight increase in spend compared to those topping out at 10,000 employees, but with a smaller jump than the previous cohorts. Nearly 30% of companies in this category topped $10M in infrastructure spending. Yet, variations between organizations remained large.



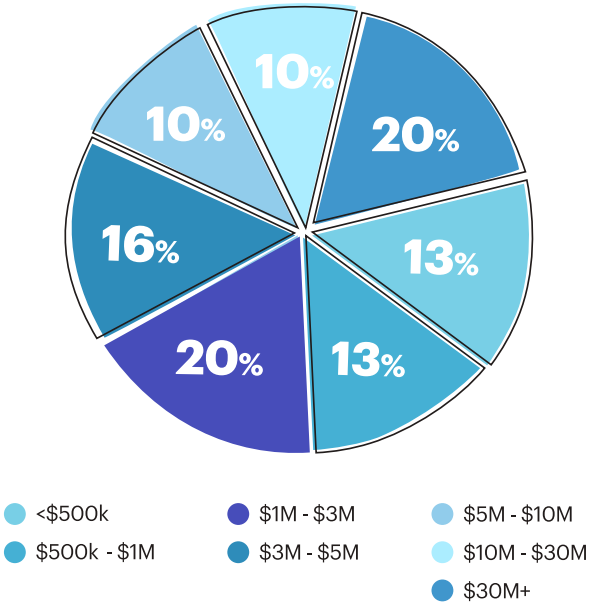
>50,000
Employees

Organizations with over 50,000 employees averaged over $30M on infrastructure spend, the highest option we provided on the survey. A number of companies also indicated they spend over $30M in data analytics, AI, and governance tooling. This was a surprising sum, representing the massive demand and increase in spend across all layers of the stack.

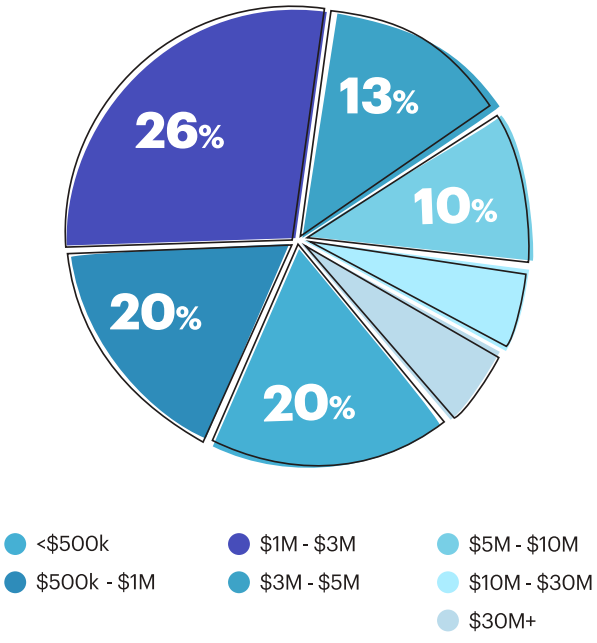
Industries

On an industry basis, financial services had far and away the largest amount of spend. This follows public reports that companies such as CapitalOne are some of Snowflake's largest customers. In fact, over 60% of the companies that plan to spend over $30M in the next year fall into the financial services category. Following financial services are a near even split between retail / e-commerce, industrial and manufacturing, and health and pharma.

What is your organiztion's annual budget for infrastructure spend (compute cost, cloud usage, storage, etc.) for 2025?



What is your organiztion's annual budget for data analytics, AI, governace tooling (various licenses such as BI, catalog, observability, ETL, MLOps) for 2025? This does not include infrastructure spend in the above question.



Data Stack

How Much Are Businesses Planning to Spend on Data in 2025?

Any talk of a slowdown in data and AI spend has long receded with organizations looking to ramp up spend and adopt AI across various use cases in the business. As many astute commenters in the data and AI industry have predicted, organizations have woken up to the fact that successful AI deployments require a solid data foundation. This, in turn, has accelerated data and AI infrastructure over the past year with 52% of companies planning to increase investments in the space over the next year. Notably, 29% of companies surveyed plan to increase their spend by over 10%.

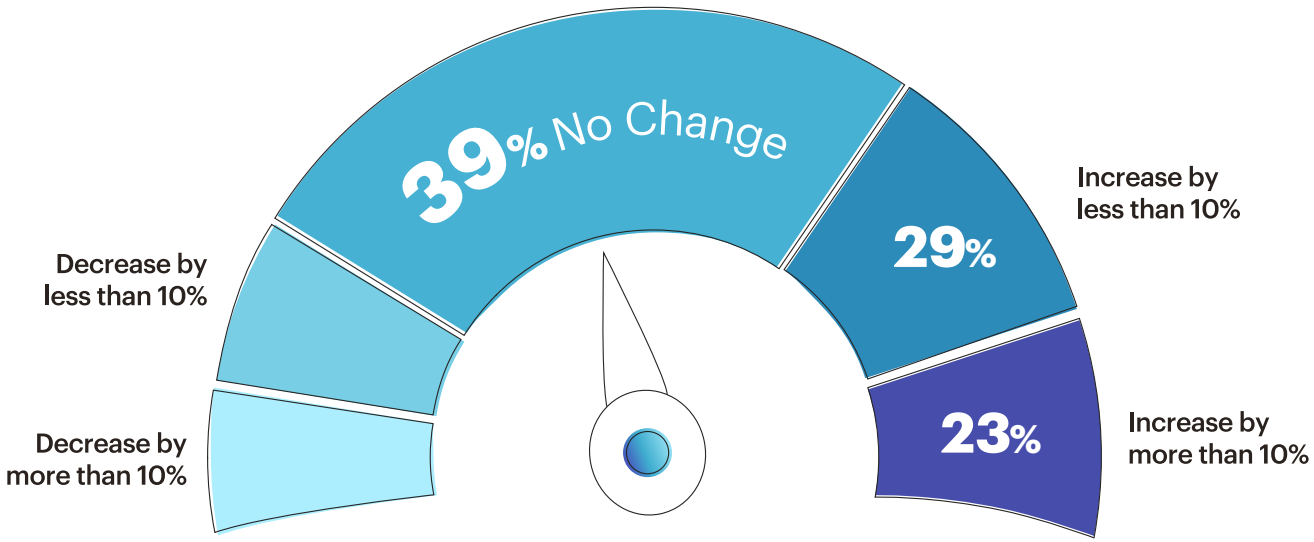
We have also seen the rise of exploratory AI budgets that are typically used to test and adopt new AI tools across business functions. Data leaders have told us that they have received mandates to help deploy AI, and are often being pushed to do so by business teams and executives. While these exploratory budgets are great for startups to get in the door, products will need to continue to provide enduring value once these exploratory budgets begin to get turned off. Last year we posed the following questions, which remain increasingly relevant when it comes to justifying long-term technology investments:

1. Does it contribute to revenue growth of existing offerings or a new innovative offering?

2. Does it help the company do more with less or become more efficient?

On the other hand, under 10% of surveyed companies plan to decrease their spend in data and AI over the next year. Meanwhile, 39% of organizations expect budgets to remain stable. Later on in the analysis, we will jump into the categories that data leaders are prioritizing for new spend.

How do you see your budget changing for 2025, as compared to this year?



Which Tools Do Businesses Expect to Purchase or Expand in 2025?

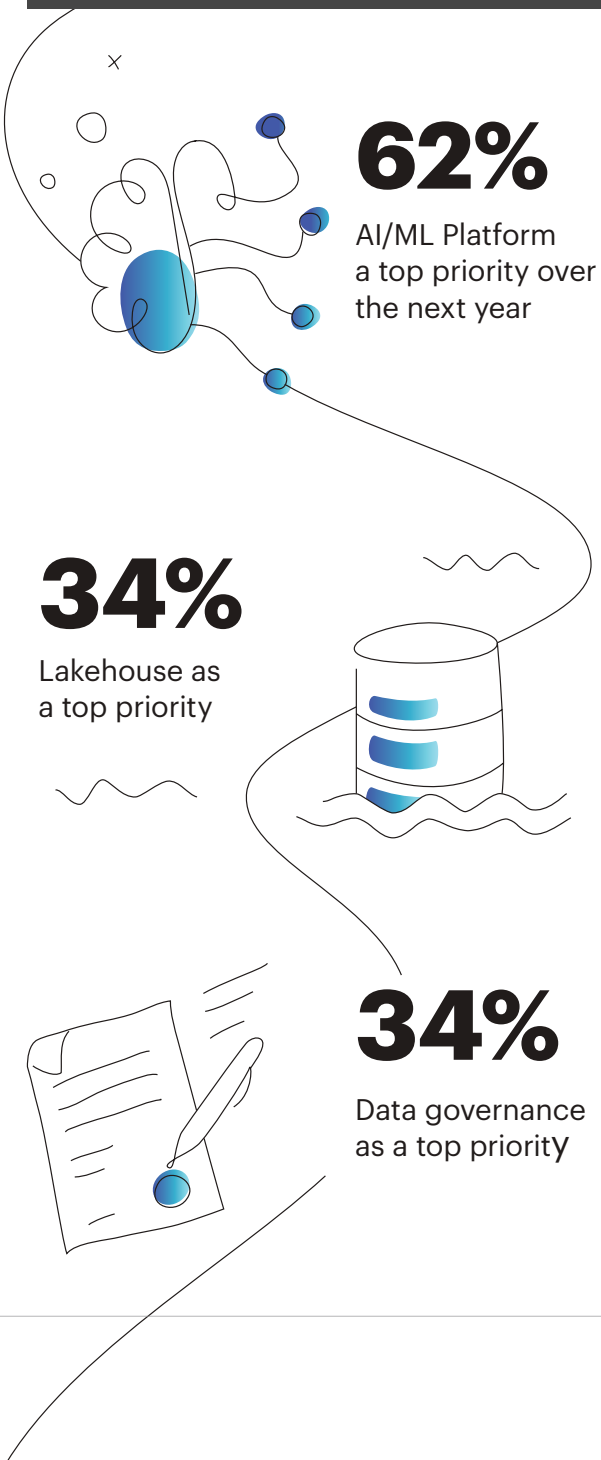
When asked about their priorities for expanding tool usage, deployment, or budget over the next year, the top priority for data and analytics leaders across the board is continuing to build out the AI and ML platform. A whopping 62% of all respondents indicated this was a top priority over the next year, an even larger percentage than the 48% from last year. We have continued to hear in numerous conversations that data and analytics leaders see themselves as responsible for setting organizations up for AI adoption.

Another leading priority for data and AI leaders over the next year is investing in their lakehouse platforms. **34% of leaders mentioned building a lakehouse as a top priority**, a notable growth from last year. A key driver of this trend is the widespread adoption of Apache Iceberg. This is a project Team8 has been following closely. We've begun to see multiple organizations re-architecting their infrastructure with Iceberg due to various benefits such as performance benefits and support for multiple query engines.

In a two way tie, **34% of data executives marked data governance at the top of their priorities**. This category includes both data classification and policy management. Historically, data teams have had lighter security standards, but this is beginning to change. GDPR and various data privacy frameworks have put a spotlight on data management practices. At the same time, a series of notable breaches over the past year involving data platforms has alerted security teams to the significant risks. This is an area we expect to see significant investment going forward.

Driving a lakehouse strategy and roll out plan is a top priority. We're delivering with more excellence, leveraging Iceberg to decrease tech and manual overhead, while giving flexibility towards the future. Consolidation with the lakehouse makes our ecosystem easier to run and scale, but we're also doing it in a way that reduces lock-in and makes the future easier to respond and adjust to for our customers and members.

Nik Acheson
VP of Data Strategy, Architecture & Engineering
HIGHMARK HEALTH



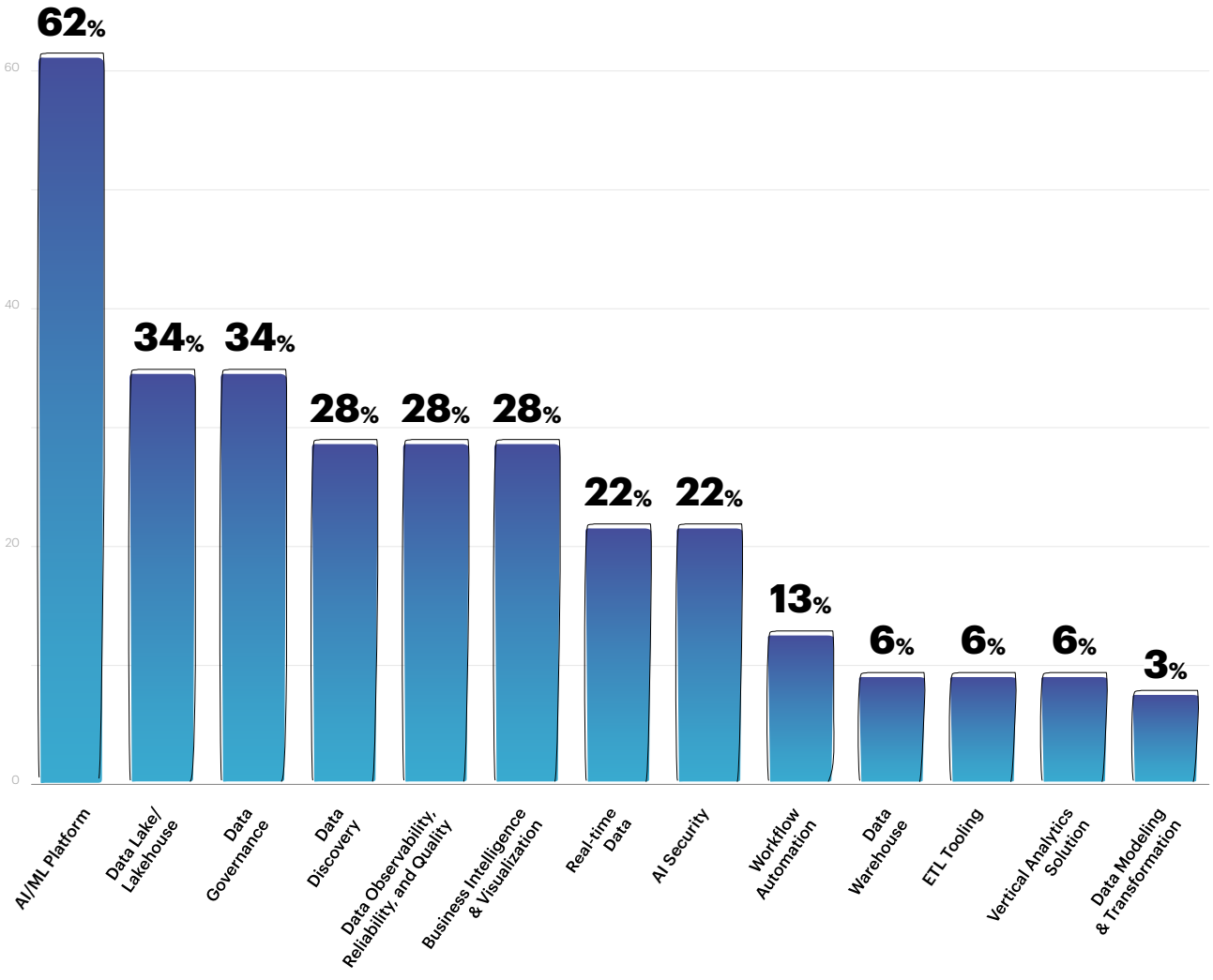
28% of data leaders placed data discovery at the top of their priorities. Within discovery, we included cataloging, lineage, and glossary tools, which have increasingly converged over the past couple years. This is still an unsolved problem for most leaders we speak with. However, AI is a major catalyst to help solve this problem. We see opportunity to enable far smarter analysis of unstructured metadata, giving data teams unprecedented visibility into their data environments.

An additional **28% of data and AI leaders placed data observability, reliability, and quality as a top priority**. Last year this was the second largest pain for data leaders with 45% of our sample placing this as a top priority. This may indicate that existing solutions are beginning to address the pain. This echoes the generally positive sentiment we hear in calls about existing tools in this domain.

A new entry to the list of top priorities, **AI observability was a top priority for 28% of surveyed data and AI leaders**. This makes sense given the rapid adoption of AI across organizations of all industries and sizes over the past year. We expect this space to become an even greater priority as organizations deploy more AI workloads into production over the coming years. An adjacent solution, **AI security was a top priority for 22%** of data and AI executives.

On the other end of the spectrum, data and AI leaders are no longer prioritizing investments in a couple areas. In particular, data modeling and transformation, ETL, and business intelligence tooling all received less than 10% as a top priority. We expect this is due to the strong maturity of tooling in these domains, and lower amounts of innovation in new solutions.

Of the solution categories below, which 2-3 are you expecting to expand your organization usage/deployment/budget next year?

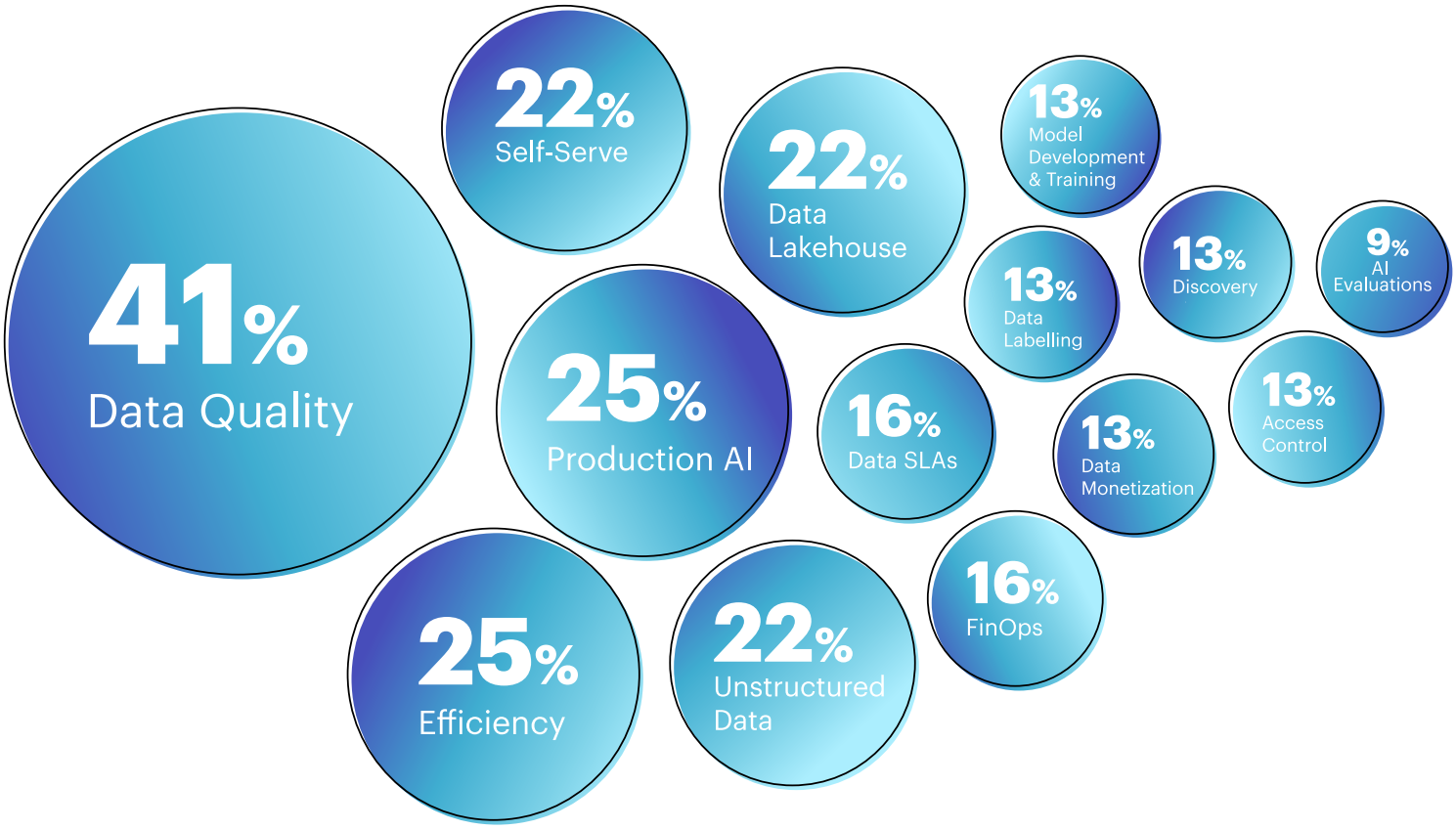


Pain Points

The AI revolution is just a couple years old, and companies are just beginning to scratch the surface of what it means to operate AI systems at scale. This year, data and AI leaders noted a significant shift in their top challenges, with a strong focus on both facilitating AI adoption and maximizing its benefits across organizations.

Some of the challenges that have been ranked by data and analytics leaders stem from technological gaps and tooling capabilities and limitations, while others fall into the realm of more "soft" issues. These "soft" pains are intricately linked to the pace of tool adoption and utilization within organizations, making them just as prominent as the more technical challenges.

What are the most acute pain points your organizations are not meeting needs? Mark the top three that apply to your organization.



Data Quality

Topping the list of pain points for data and AI leaders was data quality, which ranked similarly high in last year's survey. 41% of respondents placed the issue as a top priority.

After speaking with dozens of data and AI leaders, we were not surprised to see this as the greatest pain point. The pain remains on both the analytics and AI front, with high quality data sets being a crucial foundation for any downstream applications. Last year, we highlighted the growing importance of viewing data quality as an end-to-end problem from data production through usage. We expect continued focus in this space as companies continue to leverage their data.

Efficiency

Companies across every sector have prioritized efficiency across business units and teams over the past couple of years. While data teams are often tasked with supporting or implementing efficiency initiatives, these goals have finally begun to reach data teams themselves. 25% of respondents indicated that automating workflows across data and analytics teams was a top pain for their organization.

In conversations with data leaders, we often hear that teams operate with siloed information, inefficient processes, and often end up repeating work. Our portfolio company, Solid Data, was built to address this largely untapped need.

Unstructured Data

The vast majority of company data exists in semi-structured or unstructured formats. The often cited number is that over 90% of data is unstructured, from a 2023 IDC report.

Regardless of the exact total, there remains significant potential to utilize company data that lives in various formats. Examples that most companies have include invoices, contracts, meeting recordings, marketing content, messaging, documentation, and many more. Historically, companies have focused on structured data such as transactions or inventory data, but LLMs promise to unlock various new data stores.

At scale, data quality issues aren't just about bad inputs—they stem from broken contracts between systems, inconsistent semantics, and silent schema drift. Solving it requires more than validation checks; it takes shared ownership, metadata rigor, and automated observability baked into the pipeline.

Diego de Aragão
SVP, Balance Sheet Management & Analytics

citi

Self Serve

22% of data leaders mentioned self serve as a top pain point for their teams. In particular, this involves enabling business users to work with data.

Historically, there has been a large valley between data teams and their "customers" in various business teams. Various attempts have been made to solve this problem over the years such as business intelligence tools or embedding analysts in business teams. AI and conversational interfaces promise to aid in solving this pain. However, like most things this requires a strong foundation of modeled, quality, documented data.

> *Data teams are under pressure to drive efficiency across the business, but they also need to run lean themselves. That means fewer bespoke pipelines, more reusable assets, and tighter alignment with decision-makers—because you can't make the org efficient if your own house is inefficient.*

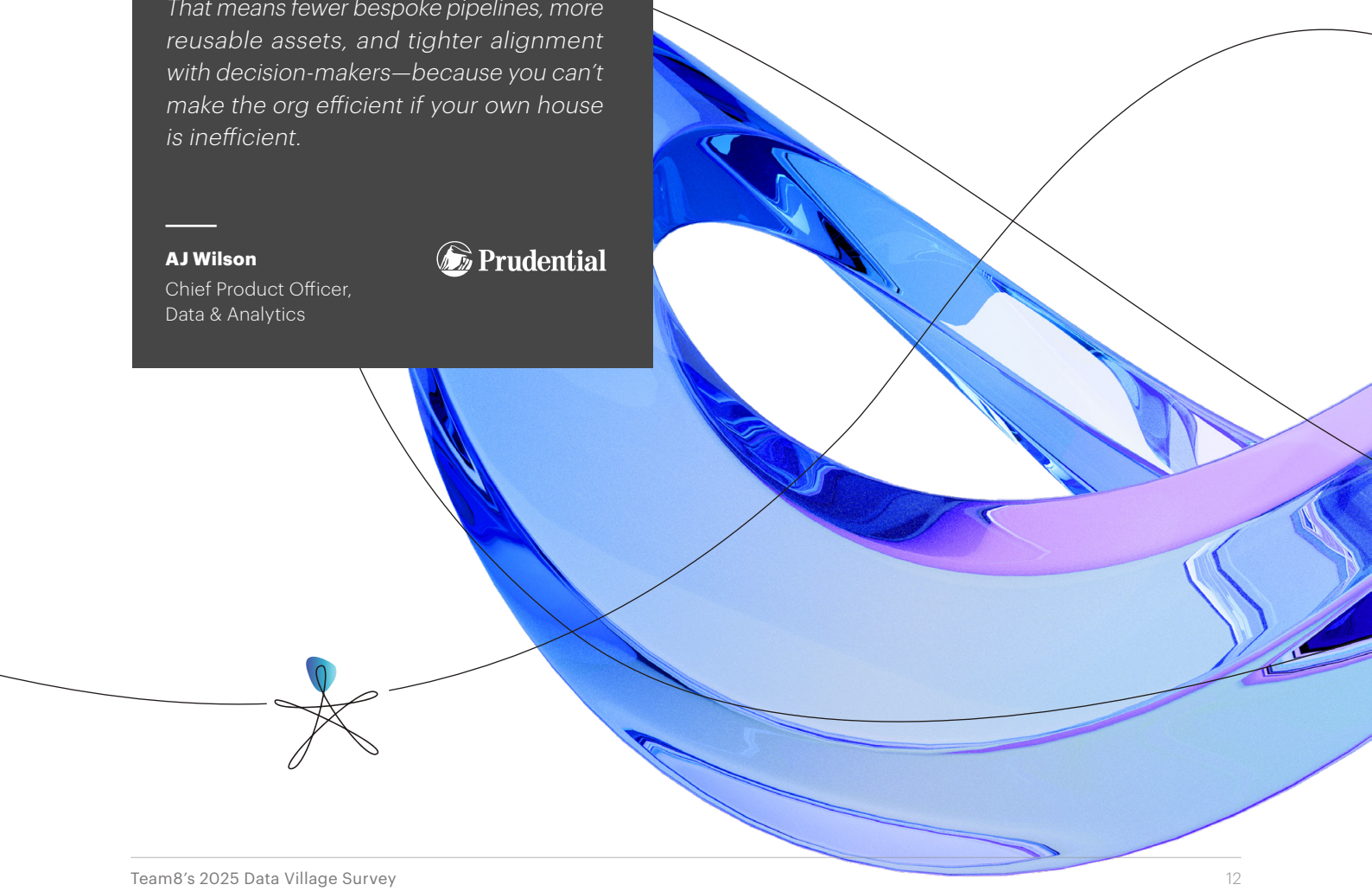
AJ Wilson
Chief Product Officer,
Data & Analytics

Prudential

Production AI

2025 is shaping up to be the year companies begin to roll out AI use cases at scale. However, there exist a number of growing pains that come with large scale AI implementations. Among these include large scale deployments and quota management. 25% of respondents mentioned production AI as a leading pain point.

Big tech companies have already released plans for over $300B in capital expenditures relating to AI data centers in the next year. This is in preparation for the massive expected demand resulting from inference time compute and better reasoning capabilities of leading AI models. We are just at the beginning of learning how to deploy AI systems at scale and see this as one of the most important areas for data leaders over the coming years.

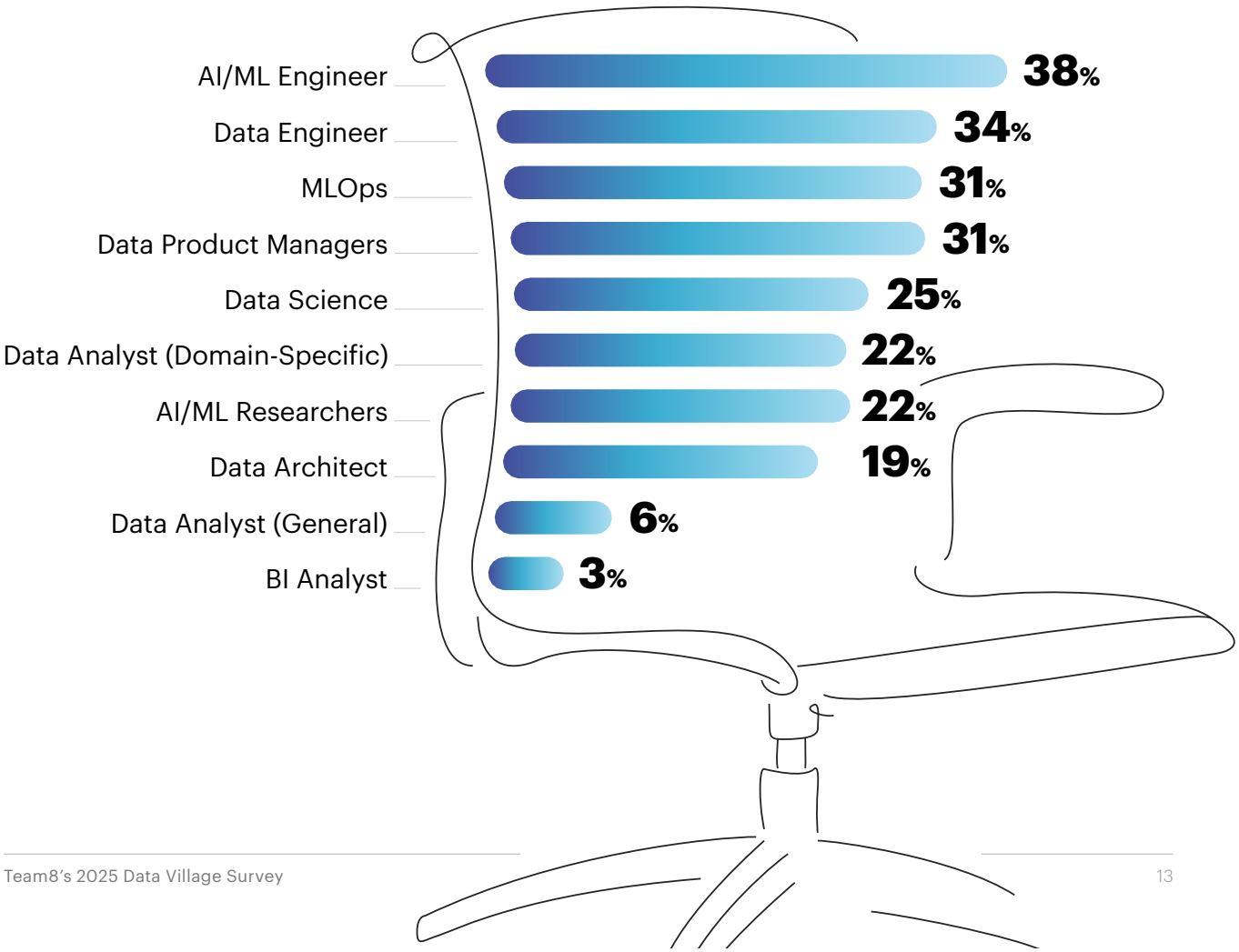


Hiring

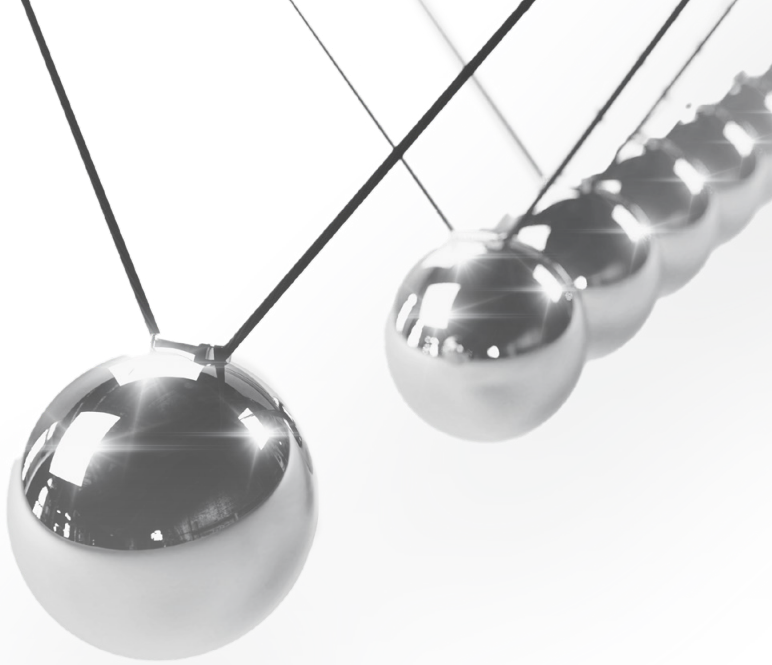
Of the key roles you are currently hiring for, which are the hardest to fill?

We asked survey respondents what key roles that they are actively hiring for are the hardest to fill. The purpose of this question was to understand where teams are making investments in headcount and where skills gap might remain. The results were not entirely surprising and echo the sentiments we have heard from our calls. In particular, data and AI leaders are focused on hiring engineering talent. 38% of leaders included data engineers and AI/ML engineers as the most in demand role. .

However, a somewhat surprising 31% of leaders put data product managers as a top priority role they were looking to hire. The data product manager is still an emerging role, and its rise mirrors the growth in treating data as a product to be consumed both internally and externally. In the future, it's likely data teams will start to more closely resemble software engineering teams with clearly defined products, customers, and SLAs. Other AI roles were also highly in demand by data and AI leaders including MLOps (31%), data scientists (25%), and AI/ML researchers (22%). Data architects also remain in high demand with 19% of data leaders placing this role at the top of their wishlist. On the other hand, companies are not struggling to hire generalist data analysts (6.3%) or BI analysts (3.2%).



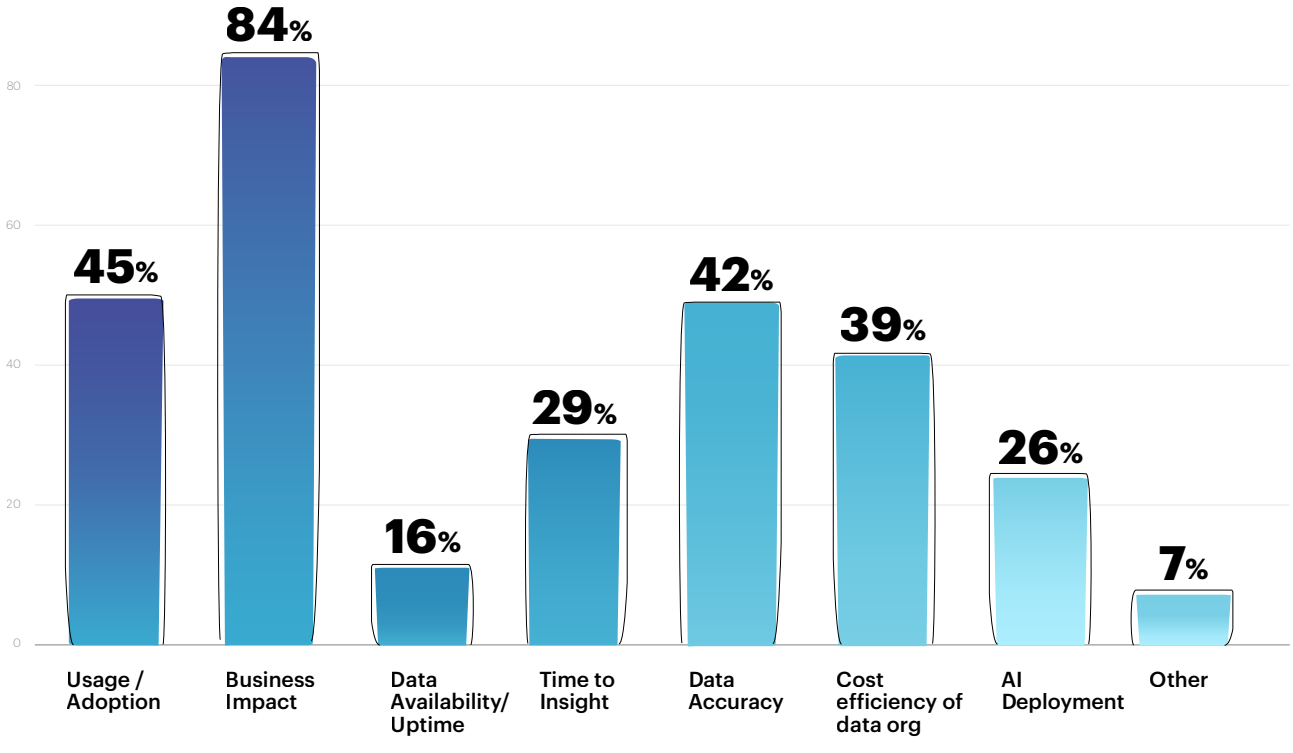
Data Team KPIs



This year, we asked data leaders how they are measured, including their core KPIs. We have often heard that data teams are increasingly being more directly measured on their business impact, and our survey results reinforce this fact. A massive 84% of data leaders are measured by their direct impact on overall business metrics such as increasing revenue or decreasing costs.

We have spoken with a number of leaders who have had multi-million dollar impacts on the business, but this required deep partnership and focus on direct business initiatives. In a distant second, 45% of data leaders are measured on the adoption of the products or models their teams build. Many studies have connected adoption of data-driven decision making to meaningful business outcomes, but this can often remain difficult to measure at a more granular level. Other important metrics that matter to data leaders typically revolve around the operations and performance of their teams. These include data accuracy (42%), cost efficiency of the data organization (39%), and time-to-insight (29%). A notable riser in the KPIs is AI adoption, with 26% of data leaders measured by this metric. We expect to see this number increase in the coming years.

How are you measured? What are your KPIs? (You may select more than one answer)

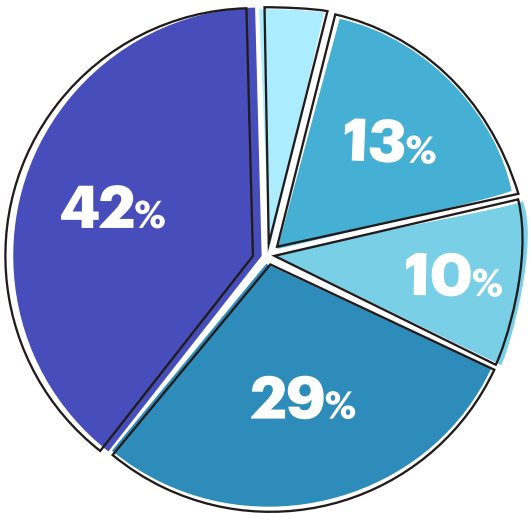


AI Adoption

AI Maturity

We asked respondents how far along in their journey they were in adopting AI. To our surprise, organizations have significantly picked up the pace of progress in their adoption. 42% of respondents indicated they have deployed multiple AI applications in production, the highest level of maturity offered on the survey. Another 29% of respondents are mostly in the POC phase but have deployed at least one application in production. Another 13% are in early POCs with no AI applications currently deployed. Following this, 10% of companies are attempting POCs but are still mostly focused on their data infrastructure investments. Just 6% of companies are focused entirely on their data infrastructure and not mature enough to transition to AI adoption.

What is your organization's maturity when it comes to AI adoption?



- Deploying multiple AI applications in production
- Mostly POCs, first AI application in production
- Early POCs
- Attempting POCs, focused on data infrastructure
- Not mature enough, focused on data infrastructure

AI Use Cases

We asked survey respondents which business units have been leading AI adoption in their companies. Engineering has proven to be the earliest adopters of AI across most organizations. Quickly, engineering co-pilots were released by leading engineering tools, and independent vendors have also seen rapid adoption. Among survey respondents, engineering was a leading use case for 65% of organizations. Behind engineering, the go-to-market and customer service functions have also seen major adoption with marketing (33%), customer success (29%), and sales (23%) being early adopters. Other business functions that have seen early adoption include operations (19%) and finance (10%). We expect these numbers to continue to increase dramatically over the next couple years.

Which business units are leading in AI adoption in your company? (You may select more than one answer)

65% Engineering

33% Marketing

29% Customer Success

23% Sales

19% Operations

10% Finance

7% Human Resources

7% Supply Chain

07

Data Team Structure

Data Team Size

Data team size largely correlates with the size of the overall organization, according to the survey results. Naturally, larger organizations deal with greater amounts of data and more complex requirements.

- For organizations under 1,000 employees, size of the data team varied significantly from under 5 members to nearly 50 employees.
- For organizations of 1,000 to 10,000 employees, the most common size of data teams grew to 51-100 working in data engineering/analytics/AI across the organization.
- Among organizations of over 10,000 employees, data teams most commonly fell in the range of 250-500 employees, with multiple larger organizations having teams of over 500 data professionals.

What is the breakdown of roles in data teams?

We found that on average data teams had a pretty even split of employees across functions. We asked survey respondents to break down the size of their teams based on the percentage of data engineering vs. analytics vs. data science/ML roles. Data analysts are the largest function today, and currently represent 35% of roles on the average data team. Closely following this is data engineers, which currently account for 34% of the average team. 25% of data and AI teams today consist of data science and ML professionals, and another 6% fall into different categories. Based on the results from section 3, we may begin to see a shift in these proportions towards more engineering and AI/ML focused roles over the coming years.

As AI adoption becomes vital for future-ready enterprises, building agile, cross-functional teams is not just a strategic advantage but essential for achieving business objectives in a rapidly evolving AI landscape. Data Product Owners, who bridge the gap between technological innovation and business needs, are key to this transformation and can help ensure that Data and AI initiatives align with business goals.

Preeti Vaidya
VP, Data

morganhealth

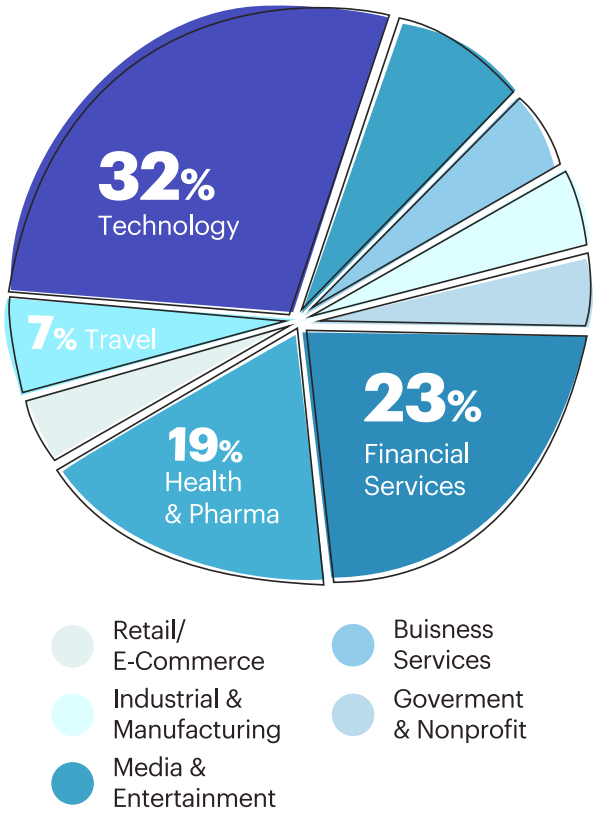
Survey Respondent Information

In addition to the data-related insights presented in this report, we also gathered general information about the participating data and analytics leaders' companies and teams. Analyzing the aggregated data revealed the following findings:

Industry

The industry with the most of the surveyed data and analytics leaders (32%) work in technology companies, followed by a smaller portion of respondents from financial services (23%), and health and pharma (19%). About 7% state that they work in media and entertainment, and a similar portion work in the industrial and manufacturing industries. A smaller portion of respondents are from other sectors, such as transportation, marketing, gaming, retail, and government.

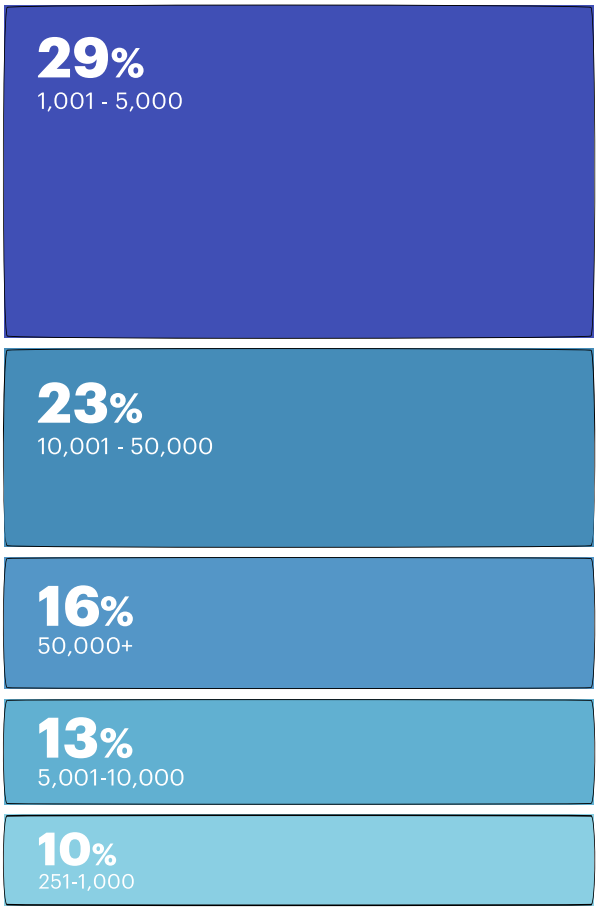
What industry does your organization primarily operate in?



Company Size

A plurality of respondents (29%) work in organizations with 1,001-5,000 employees. We also had a significant number of respondents from larger organizations of 5,001-50,000 employees (36%) and major enterprises (16%). The remaining 19% of respondents represent companies with under 1,000 employees.

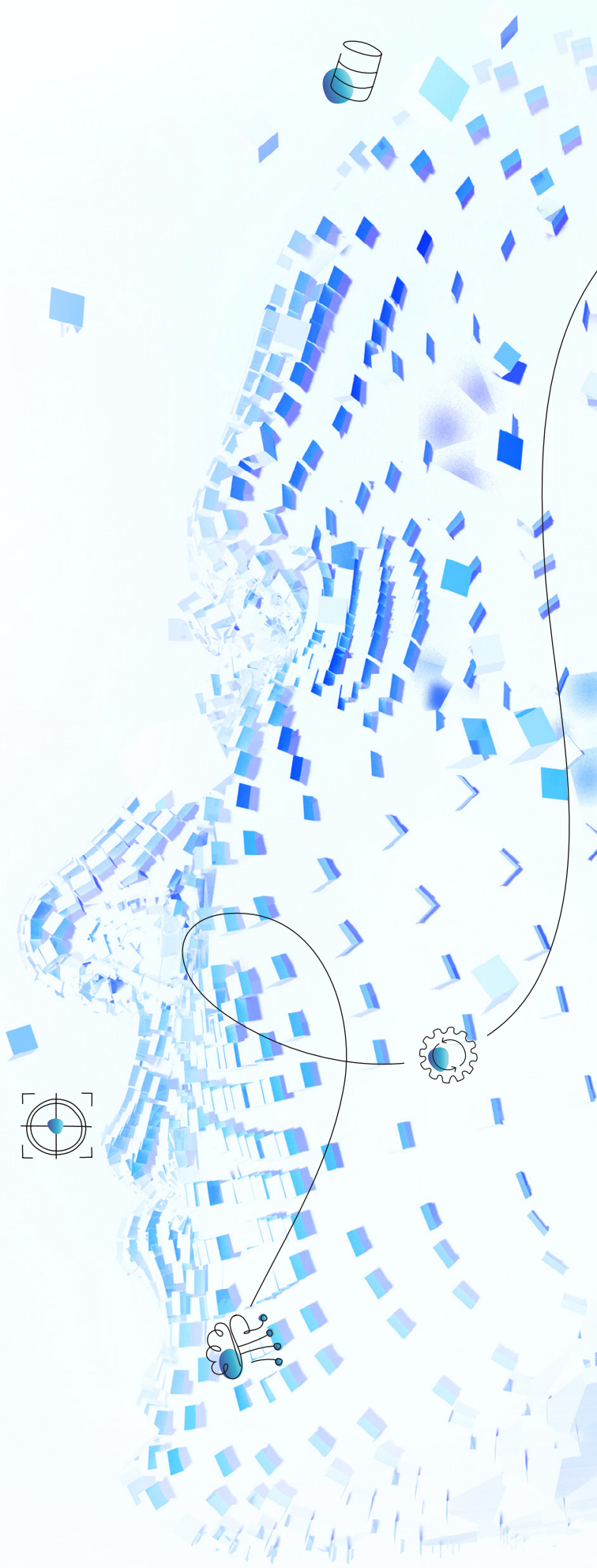
How many employees do you have in your company?



Looking Forward

Every year, we attempt to gather a broad view on the direction of the data and AI landscape via our survey of Team8 Data Village members. This year's survey had additional new questions on AI, including adoption, use cases, and tooling. It's clear that we are in the early innings of an industry supertrend, and in many organizations data and AI leaders will be tasked with building and deploying AI in production. We just scratched the surface with these insights but invite the industry to collaborate with us further as we learn and develop best practices at scale.

Last year, we posed the question of ROI of data teams, and this continues to remain a hot topic in the industry. AI provides an opportunity for data teams to move beyond analytics functions into strategic revenue-generation engines. Particularly, we see data teams at the core of building the infrastructure for both externally and internally facing AI products. This enables data teams to more closely align themselves with company KPIs, improving the operational and strategic value of data and AI investments.



Glossary

Term	Definition	Section
Data Discovery	Catalog, lineage, Glossary	2.2
Data Governance	Classification, Policy Management	2.2
Real-Time Data	Event Streaming & Processing, Real-time Analytics Database	2.2
Data Lakehouse	Unifying data infrastructure	3
Self Serve	Enabling business users to work with data	3
Data Quality	Quality of data sets & data cleaning	3
Data SLAs	Uptime, monitoring	3
Discovery	Too much data & don't know what data to use	3
Efficiency	Automating workflows across data and analytics teams	3
Access Control	Controlling and enabling access to data and AI	3
FinOps	Controlling costs of data and AI	3
Data Monetization	Enabling business to sell data and insights	3
Unstructured Data	Integrating and modeling unstructured data	3
Data Labeling	Custom data annotation	3
Model Development & Training	Building and fine-tuning AI models	3
AI Evaluations	Assessing performance and accuracy of AI models	3
Production AI	Deploying AI systems at scale, quota management	3

Endnotes

1. Snowflake Inc. (SNOW) Deutsche Bank Technology Conference Transcript, SA Transcripts, 2022

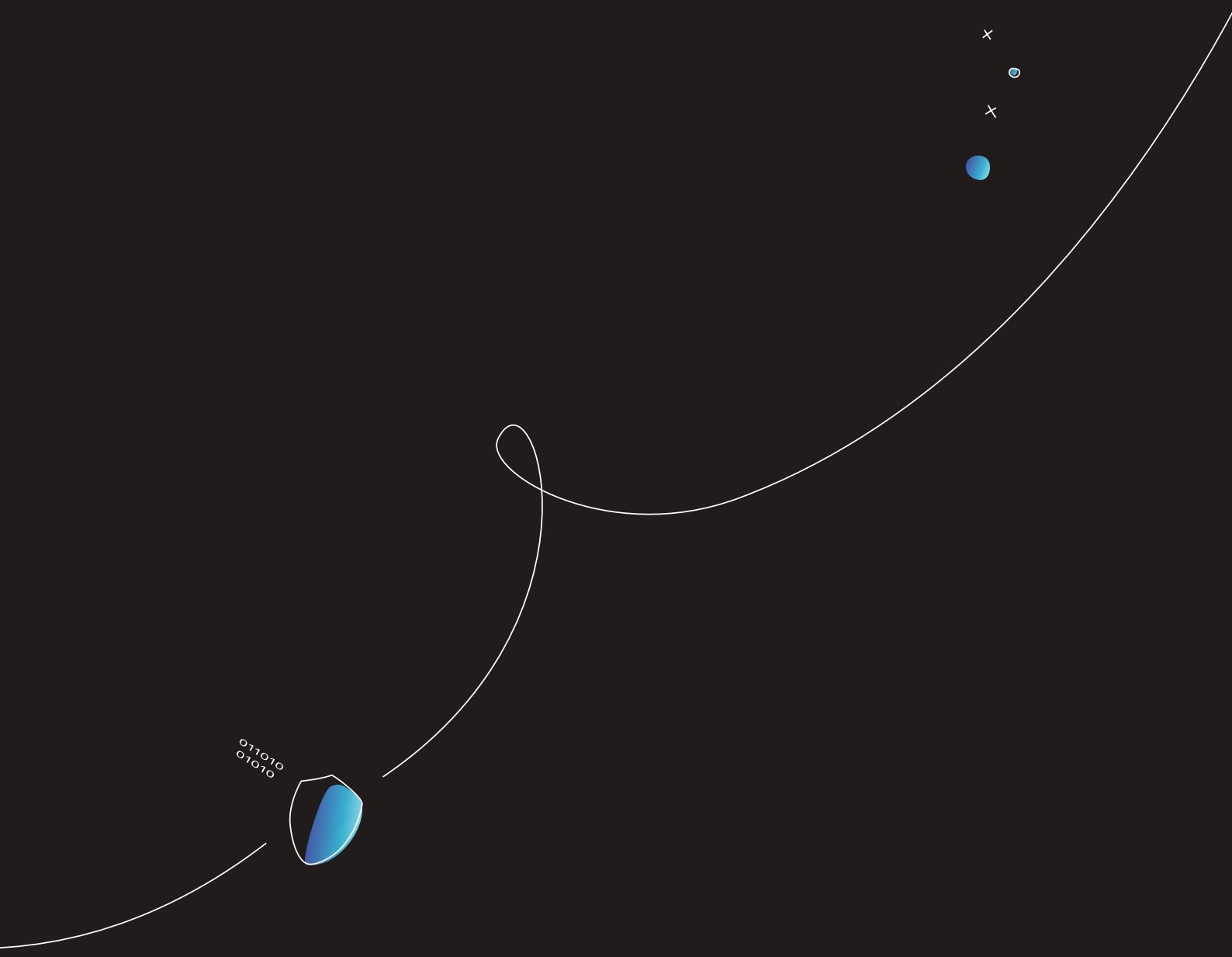
Building the Next Generation of Data & AI Companies

At Team8, we believe this is a pivotal time to be building data companies that define how enterprises successfully utilize their data. We're excited to be backing leading teams in data looking to solve challenges that data and analytics leaders are facing today. We'd like to thank our Data Village for sharing their insights with us, and hope this analysis can provide some clarity into the current state of the market.

If you have any questions or would like to get in touch, please reach out to us at datavillage@team8.vc - we're happy to chat!

Calling all Data Executives to join our community of leading data expert

Apply Here



For more information

Contact us at: datavillage@team8.vc | www.team8.vc

TEAM8™