The State of Applied Machine Learning 2023

AN apply() REPORT FROM tecton

Table of contents

Welcome to the first-ever report on the state of applied machine learning.

This report is based on answers from over 1,700 respondents to a survey that Tecton sent to the global machine learning (ML) community and provides insights into both the challenges and opportunities presented by applied ML. It also takes a look at where things are headed in the next few years.

How companies use data has come a long way in the past decade. And there's a good reason for that: Data, a key cornerstone of ML, is being generated at ever-increasing volume and velocity. Traditional business intelligence and analytics cannot keep up with this accelerating pace of data. Instead, organizations have to turn to ML to automate the process of transforming data into valuable insights, and the companies that have been able to do so have been wildly successful.

Our survey found that companies in many industries are increasingly adopting applied ML for a wide range of use cases, including customer analytics, personalized recommendations, and fraud detection. At the same time, many are also facing multiple challenges on their journey to implementing applied ML, such as generating accurate training data, building production data pipelines, and demonstrating business ROI.

Because of the difficulty inherent in data, ML—and specifically applied ML—is still a challenge to get right for most organizations. However, despite the challenges, survey respondents indicate that their companies are committed to improving their applied ML capabilities, with a growing focus on improving model deployment time, adopting real-time analytics, and implementing central ML platforms to improve cross-team collaboration and organizational scalability.

The goal of this report is to identify the challenges and opportunities in the space, and pinpoint common trends across a diverse set of machine learning initiatives. Read on for key findings, recommendations, and a deeper dive into the trends that will shape the future of applied ML.

What is applied ML?

In short, all the ML a company uses is applied ML. It's real-world ML that has clear business value, whether it's boosting analytics capabilities to help companies make better decisions or ML models embedded in production applications to power predictive systems like fraud detection and personalized recommendations.

Even though most organizations are just getting started on their journey to applied ML, we already have an idea of its transformative potential through existing use cases like getting ETA and pricing predictions when calling a rideshare, and obtaining quotes for car and home insurance.

What is real-time ML?

Real-time ML is a subset of applied ML use cases. Real-time ML consists of running ML models in production to make real-time predictions and using those predictions to power production applications, with no human in the loop. These use cases are typically more advanced and complicated than batch ML use cases.

Real-time ML models can be powered by batch data only, but can also use streaming or real-time data to increase the accuracy of predictions by incorporating the freshest information available. *(Read more about real-time ML.)*

Key Findings

1. Applied AI/ML is a top priority for organizations

2. Companies are investing in their MLOps stack to address challenges

3. Real-time machine learning is picking up real traction

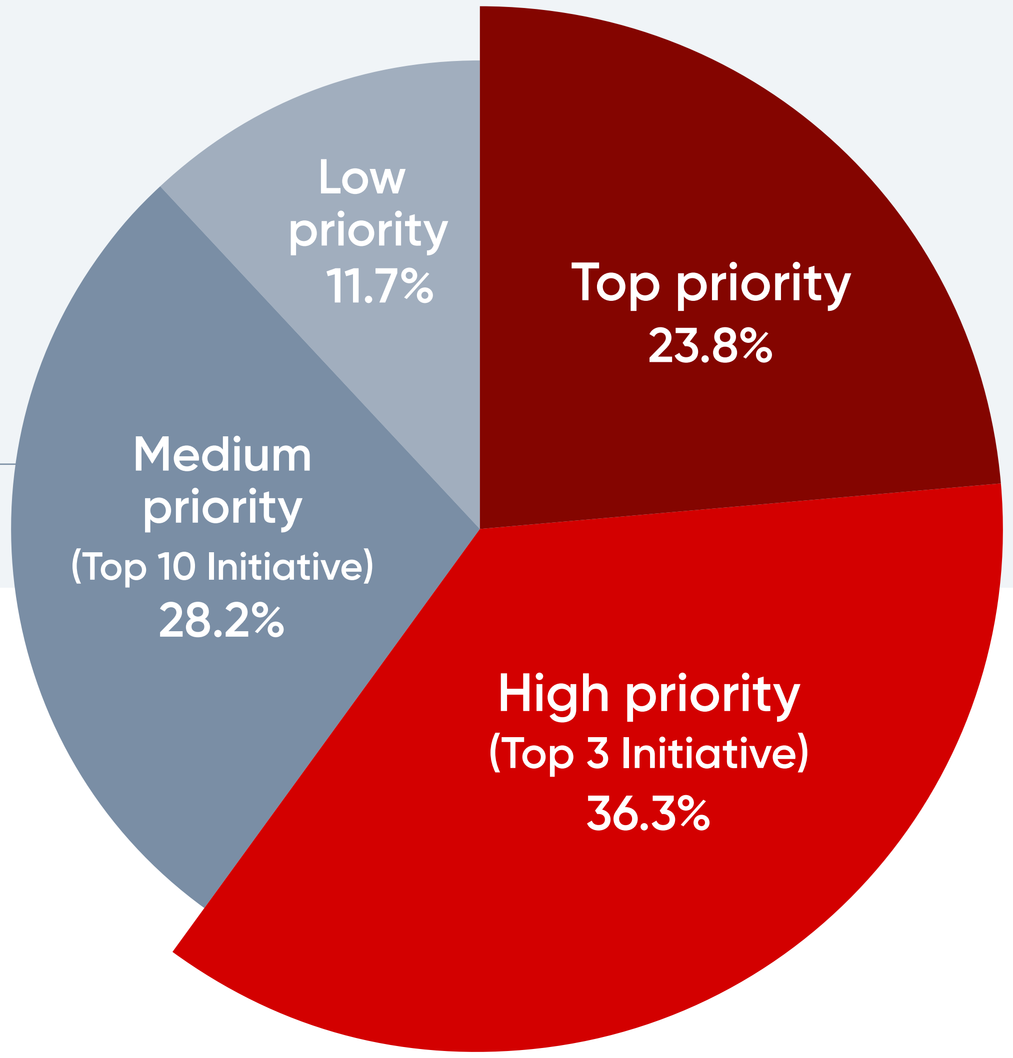
KEY FINDINGS 1

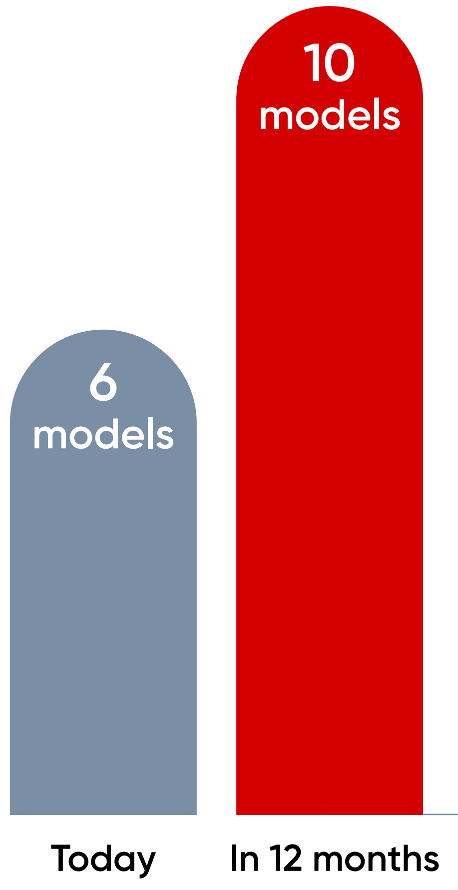
Applied AI/ML is a top priority for organizations

The Importance of Applied ML

Applied ML is declared as the **No. 1 company initiative** for 23.8% of respondents and a **top 3 initiative** for 60.1%.

How important is applied ML to your organization?





Number of Models in Production (Median)

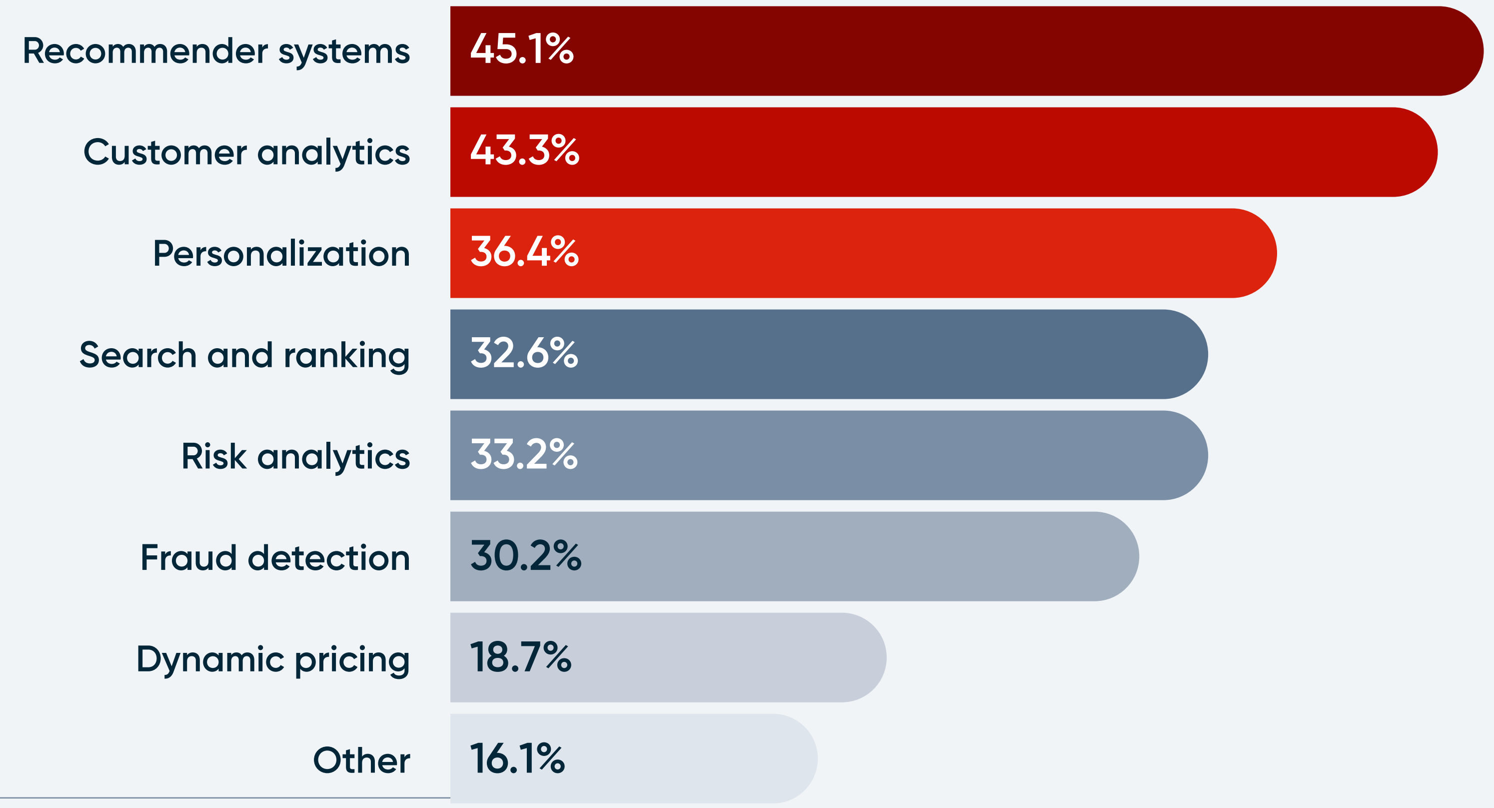
Roughly 50% of survey participants indicated their organizations have **6 or more models in production**, and companies are planning on deploying more models quickly. Based on survey responses, the projected median for the **number of models in production in the next 12 months will increase from 6 to 10.**

How many models in production do you have today? How many do you expect to have in 12 months?

Applied ML Use Cases

Companies are leveraging applied ML for **use cases that directly impact revenue**: The top 3 use cases are **recommender systems** (45.1% of respondents), **customer analytics** (43.3%), and **personalization** (36.4%).

Which types of use cases are these ML models used for?

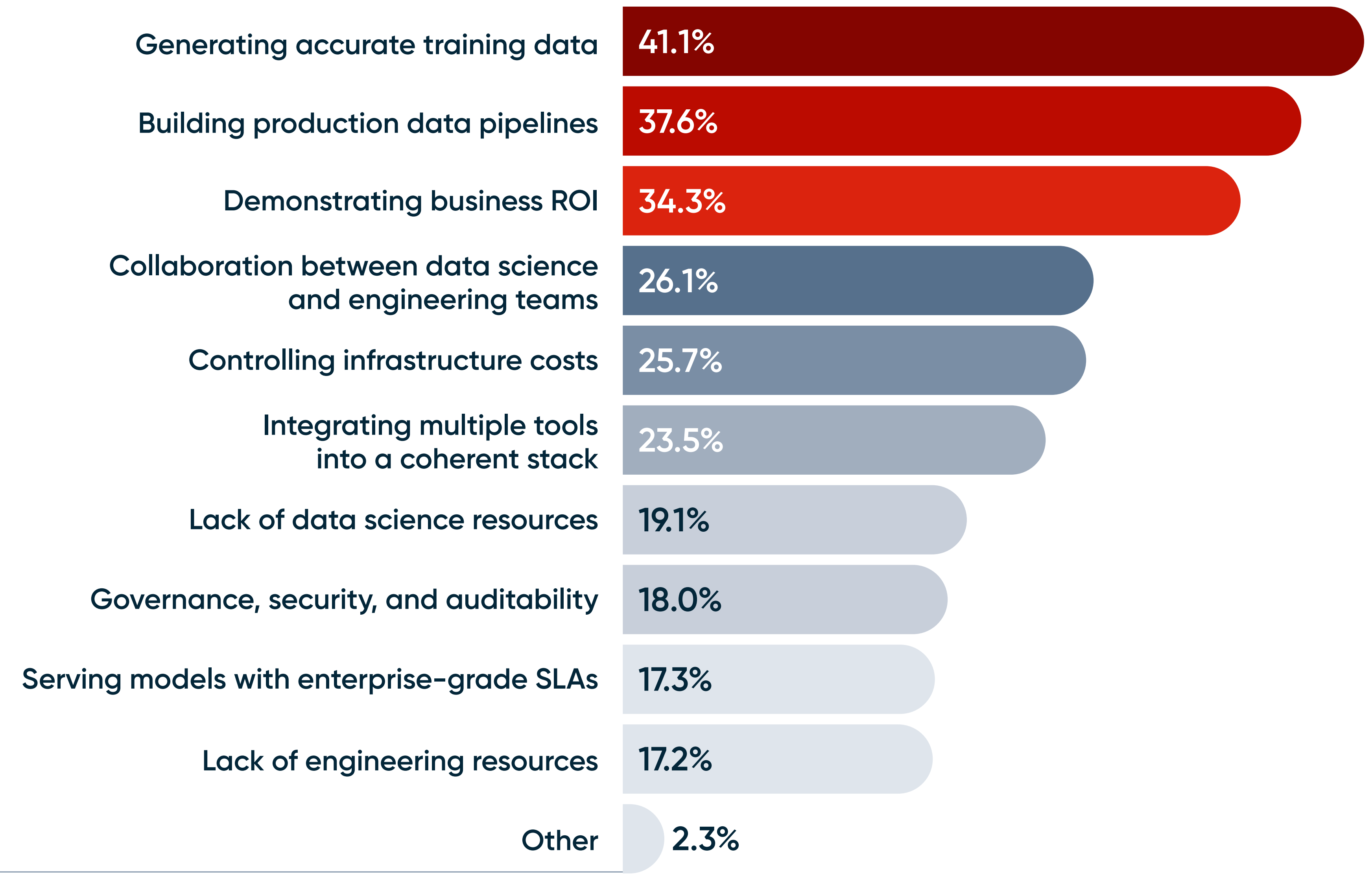


Companies are investing in their MLOps stack to address challenges

Deployment Challenges

The **top 3 challenges** encountered when deploying new models to production are:

- **Generating accurate training data** (41.1%)
- **Building production data pipelines** (37.6%)
- **Demonstrating business ROI** (34.3%)



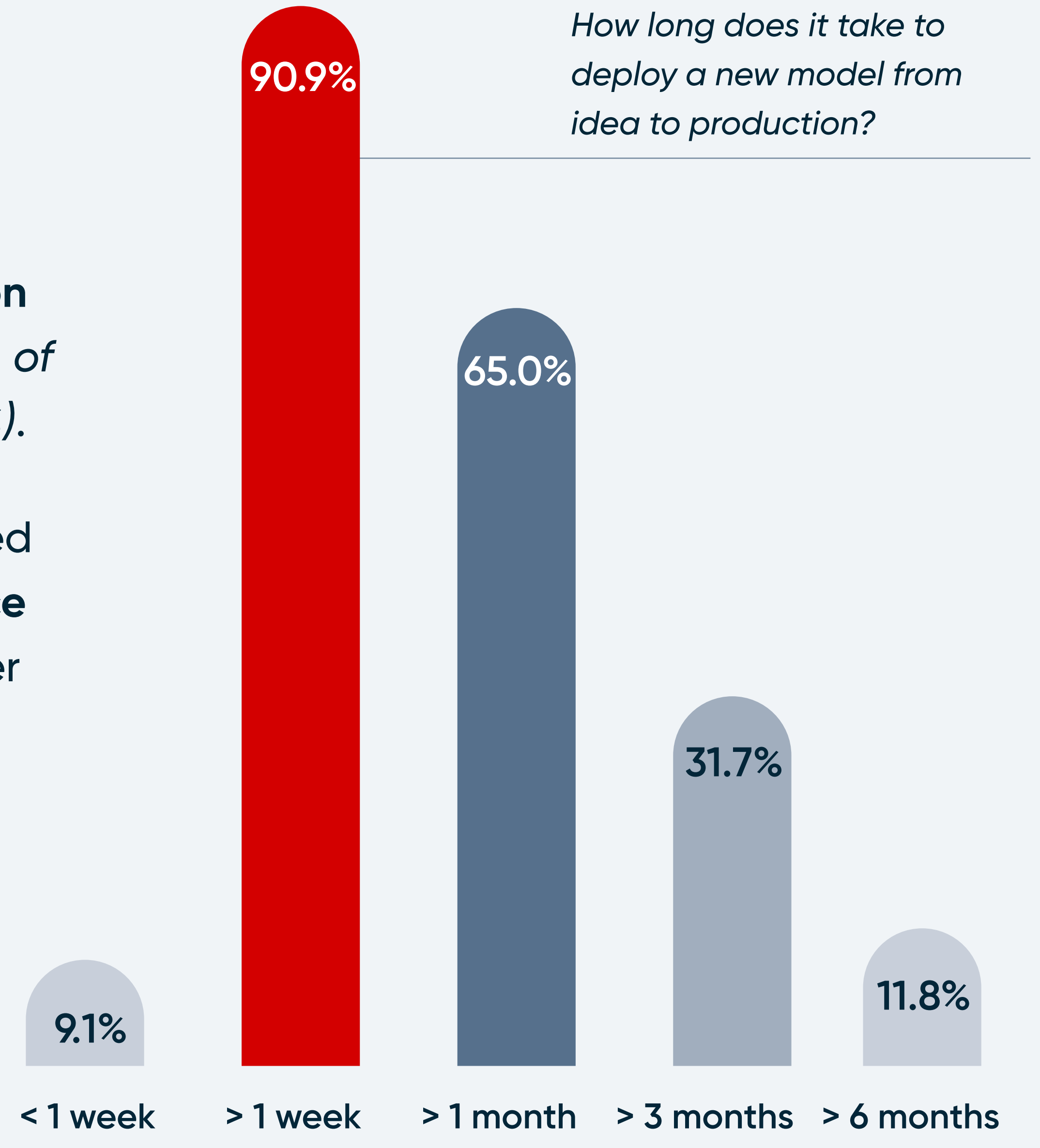
What are the 3 biggest challenges encountered when deploying new models to production?

Time to Deployment

Deploying a new model to production is a long process *(>1 month for 65.0% of respondents and >3 months for 31.7%).*

However, 54.5% of respondents shared that their organizations **aim to reduce deployment time** by at least 25% over the next 12 months.

How long does it take to deploy a new model from idea to production?



Does your company aim to reduce model deployment time within the next 12 months? If so, by how much?

54.5% of respondents aim to **reduce deployment time** by at least 25% over the next 12 months

Companies are investing in their MLOps stack to address challenges

The MLOps Tech Stack

Only 9.5% of respondents say their organization has a full MLOps stack today (5 out of 5 components).

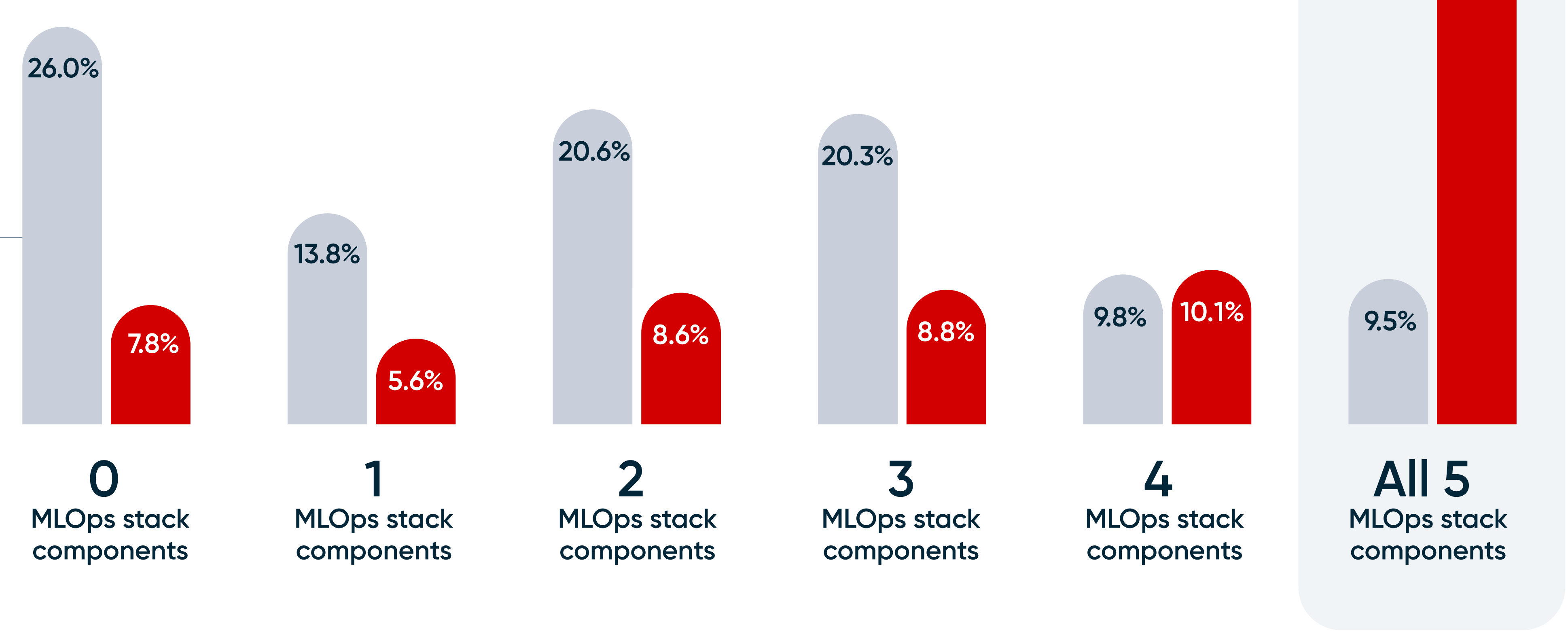
However, 59.1% of respondents indicate that their companies are **planning to have all 5 components within 12 months.**

The Feature Store / Feature Platform and Monitoring & Observability components will see the largest increases (~43 percentage points increase for both) in adoption in the next 12 months.

How many of the 5 components do you currently have in your MLOps tech stack? How many do you plan to have in the next 12 months?

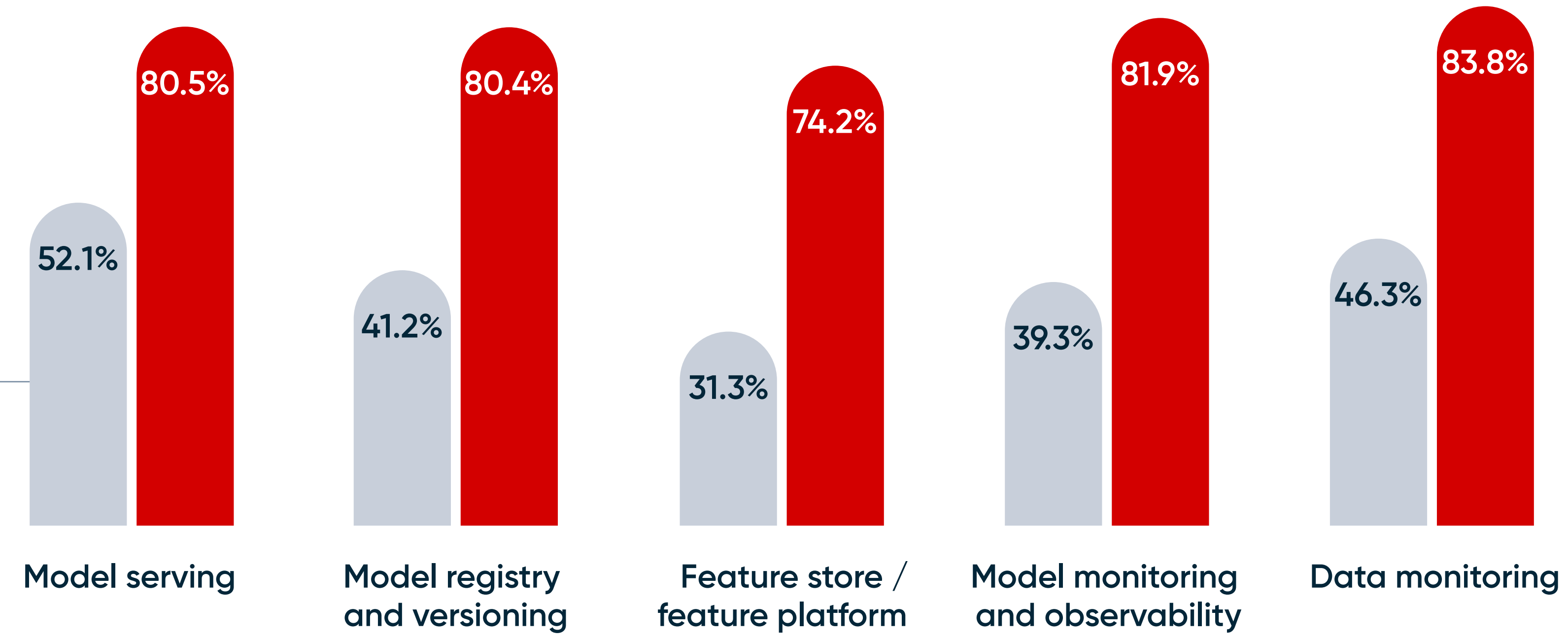
A. Model serving
B. Model registry and versioning
C. Feature store / feature platform
D. Model monitoring and observability
E. Data monitoring

Today In 12 Months



Which components of the MLOps stack does your company use today or plan to adopt within 12 months?

Today In 12 Months

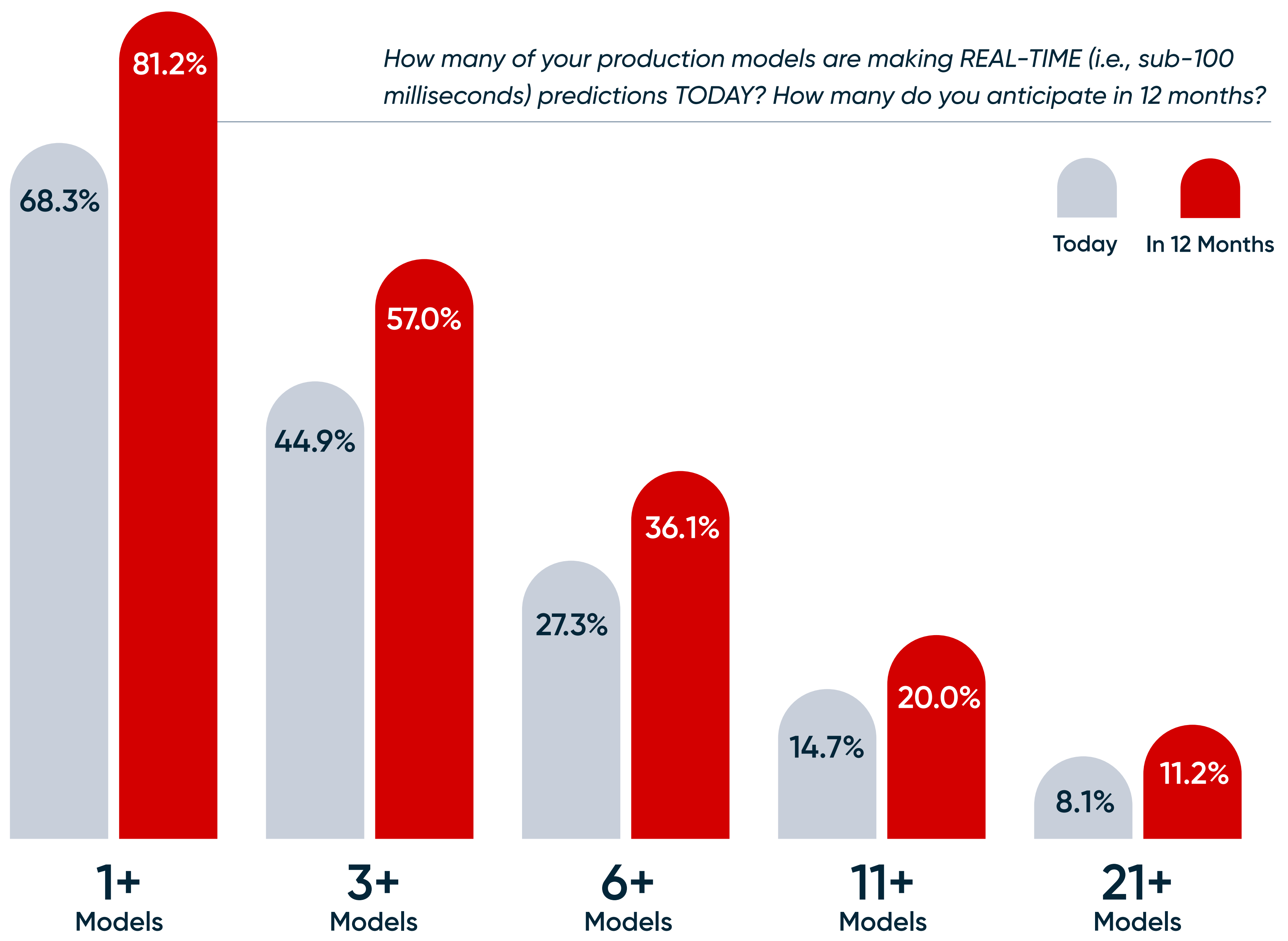


KEY FINDINGS 3

Real-time machine learning is gaining real traction

Number of models making real-time predictions

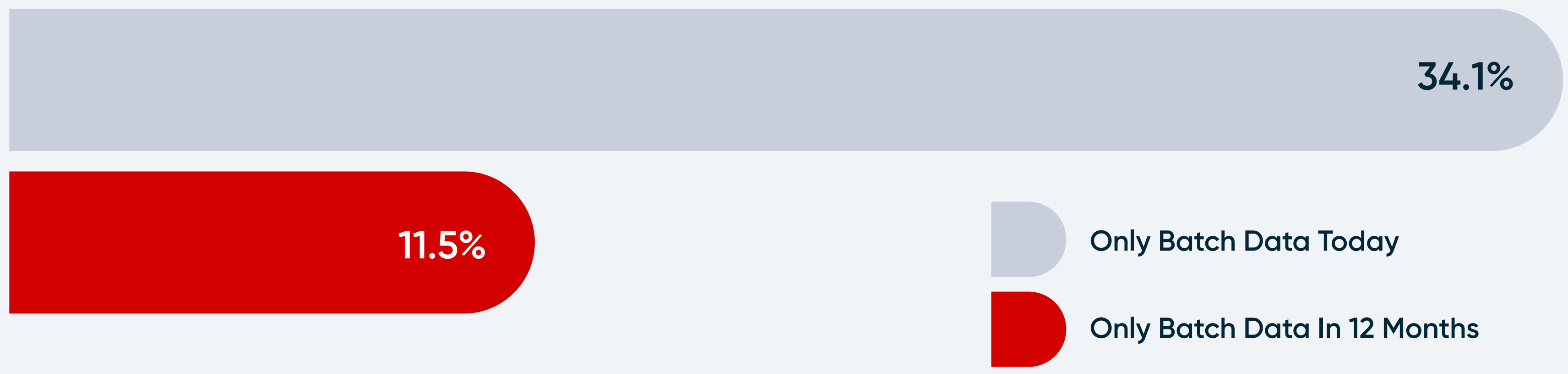
68.3% of ML respondents surveyed indicated their **teams already have at least one real-time (i.e., sub-100 milliseconds) ML model in production**, while 14.7% have more than 10.



How many of your production models are making REAL-TIME (i.e., sub-100 milliseconds) predictions TODAY? How many do you anticipate in 12 months?

Teams who use batch data only

34.1% of respondents say their ML teams use exclusively batch data to power their real-time ML models today, **but this number is expected to drop to 11.5% within 12 months** as they plan to adopt streaming or real-time data.



Which types of data sources does your company use to generate ML features today or plan to adopt within 12 months? ("Only batch data" results shown)

Recommendations

1 If you're not using real-time ML in your organization, consider starting now.

The majority of survey respondents (68.3%) are **already using real-time ML**, and that number will likely increase to 81.2% in 12 months. Evaluate your current use cases and the potential ROI of moving to real-time ML—competitors with similar use cases are probably already doing so.

2 Invest in your MLOps stack to streamline ML model deployment.

Survey participants shared that the biggest roadblocks are **generating training data and building production data pipelines**. MLOps solutions have matured a lot over the past few years and the right ones will help overcome these operational challenges.

3 Start by investing in ML use cases with easily demonstrable ROI.

Business ROI is named as a top 3 roadblock by many survey participants who are early in their applied ML journey. The survey found that **companies are prioritizing recommender systems, customer analytics, and personalization use cases**.

Over 1,700 people in various industries completed the survey.

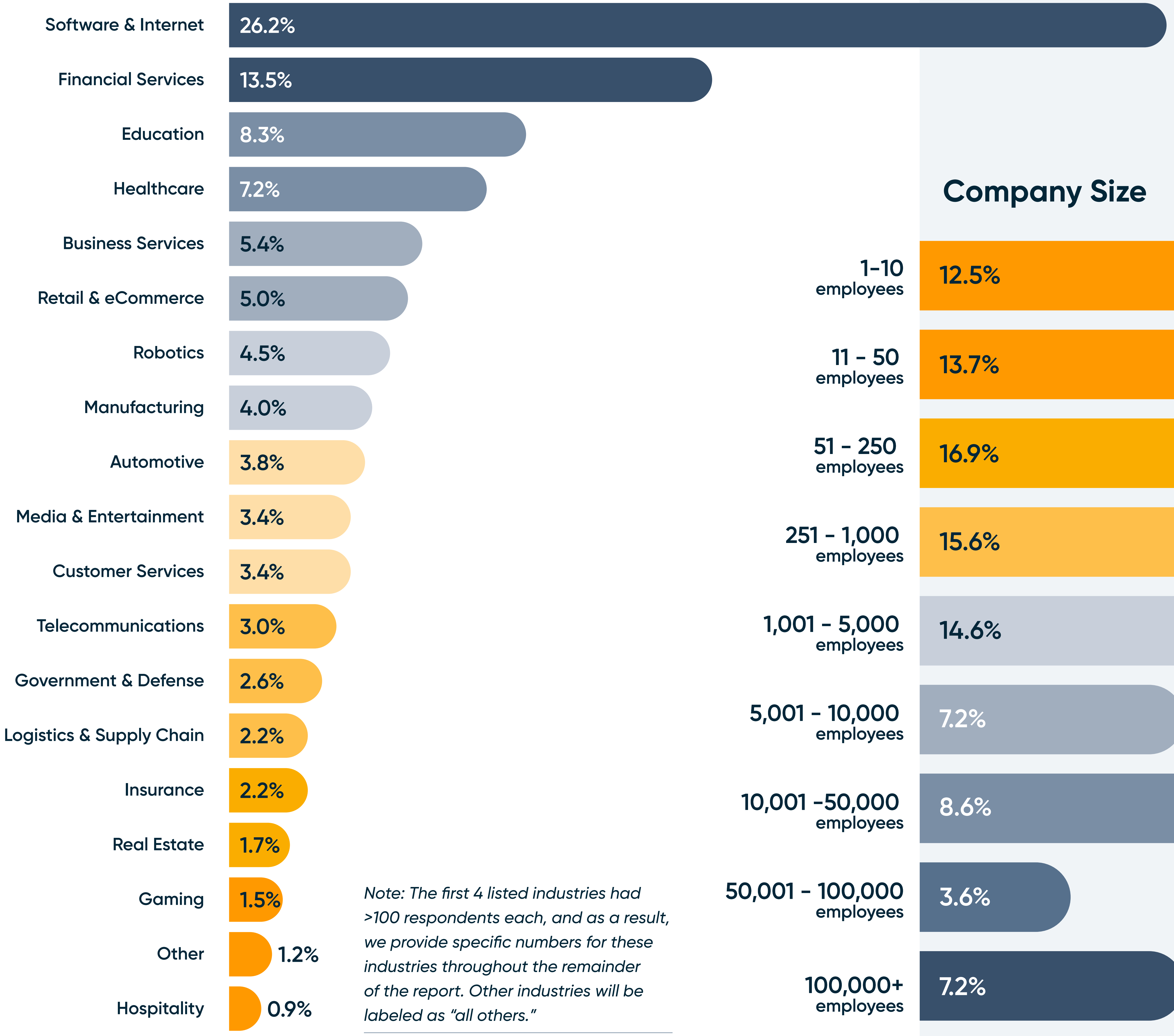
The survey was distributed on February 2, 2023, and remained available to participants until March 1, 2023. Respondents' job roles spanned a number of titles, including those in Data Science, ML Engineering, Software Engineering, Data Engineering, Product Management, and Architecture. The survey was promoted in relevant industry newsletters, emails, community channels, and social media. Gift cards were offered as incentives to encourage participation.

Over 1,700 people in various industries completed the survey. Respondents are mainly from Europe and North America, representing companies of all sizes, from small startups to enterprises with more than 100,000 employees. Note that due to the characteristics and structure of this survey, the results may not accurately represent ML teams across all industries. Both selection bias and observational bias could be present in the data. Despite the limitations of the survey, this report is unique in its focus on teams involved in building, validating, deploying, and monitoring ML models, and aims to serve as a guide on the adoption of ML across industries.

A Deeper Dive Into Survey Results

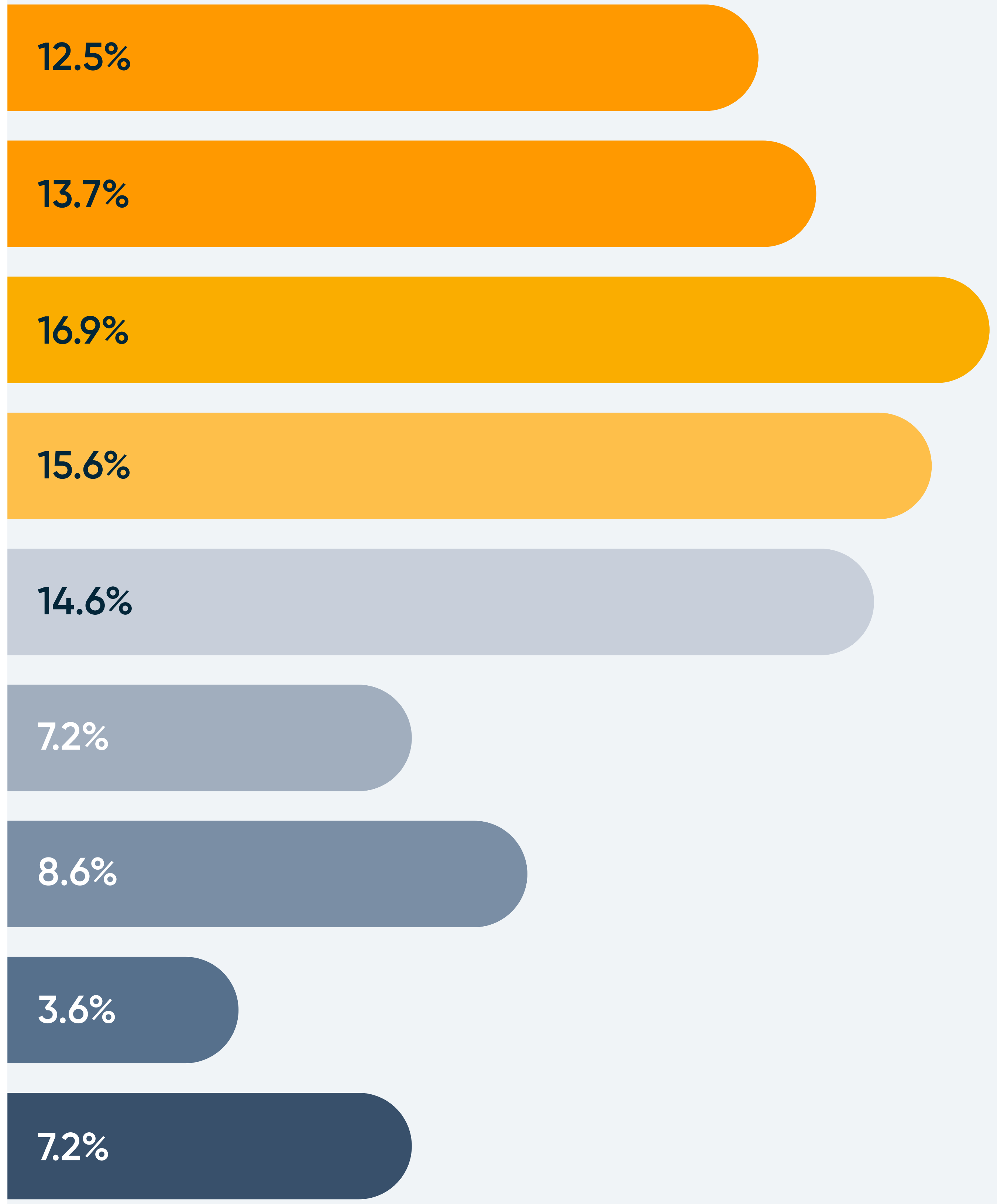
RESPONDENT DEMOGRAPHICS

Industry

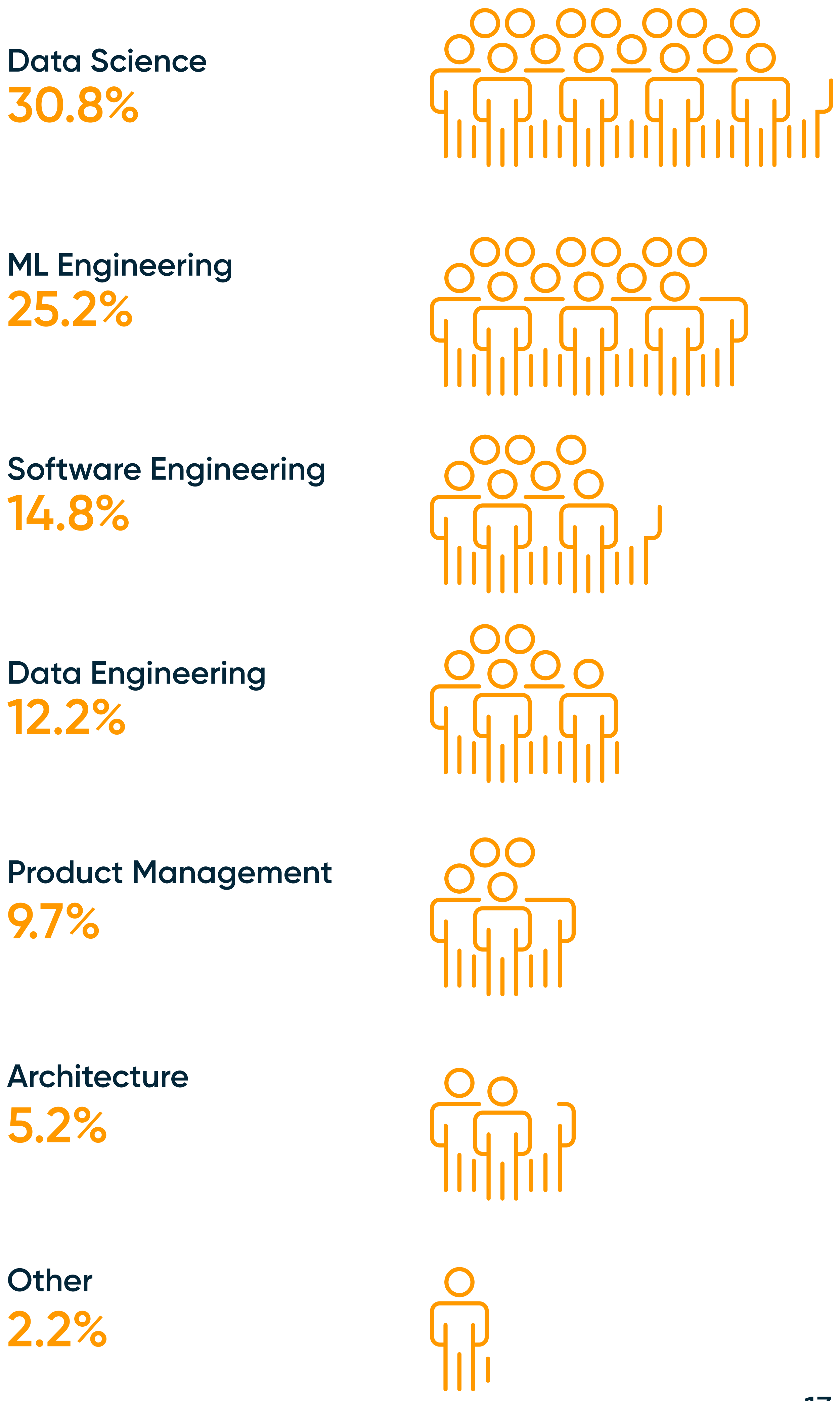


Note: The first 4 listed industries had >100 respondents each, and as a result, we provide specific numbers for these industries throughout the remainder of the report. Other industries will be labeled as "all others."

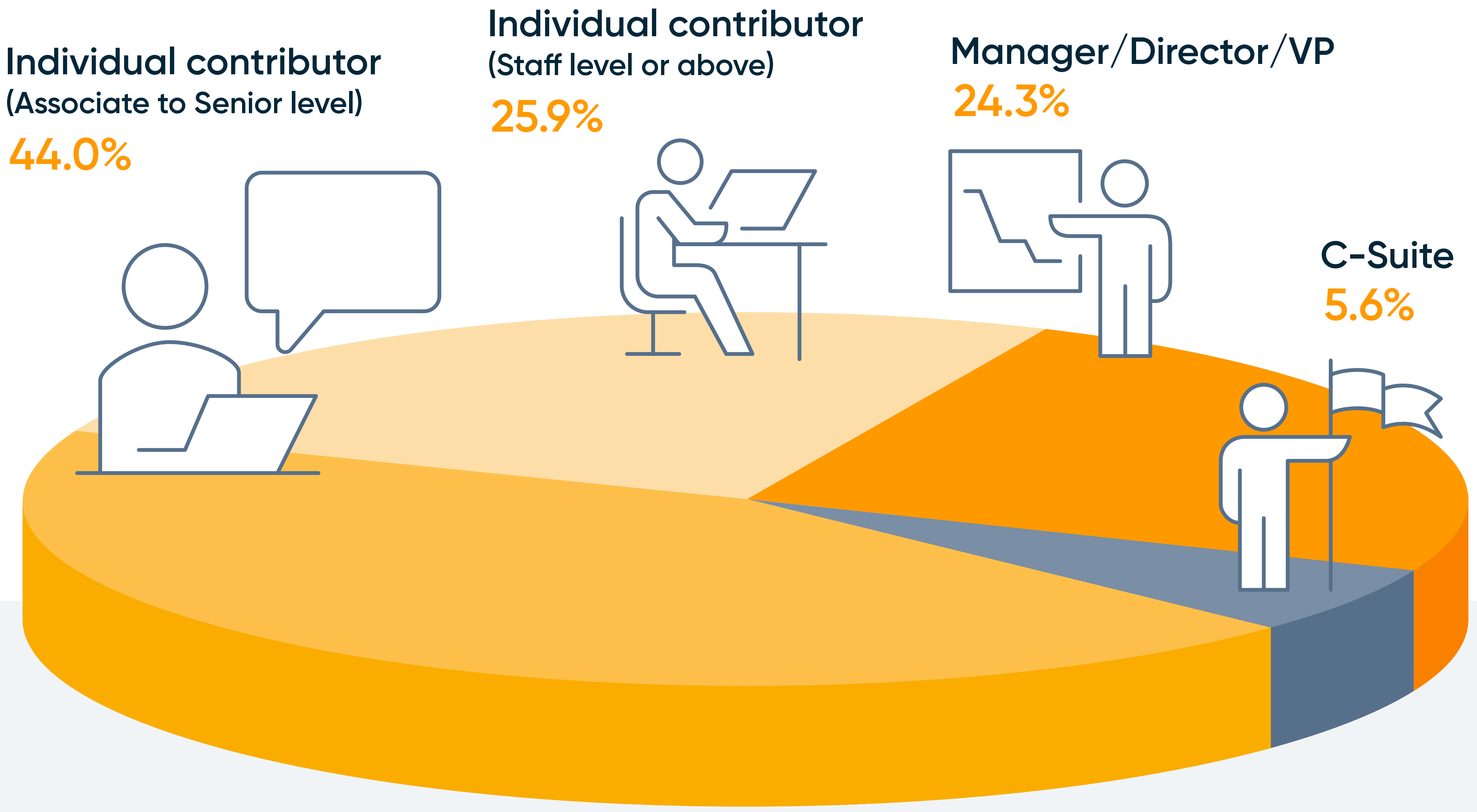
Company Size



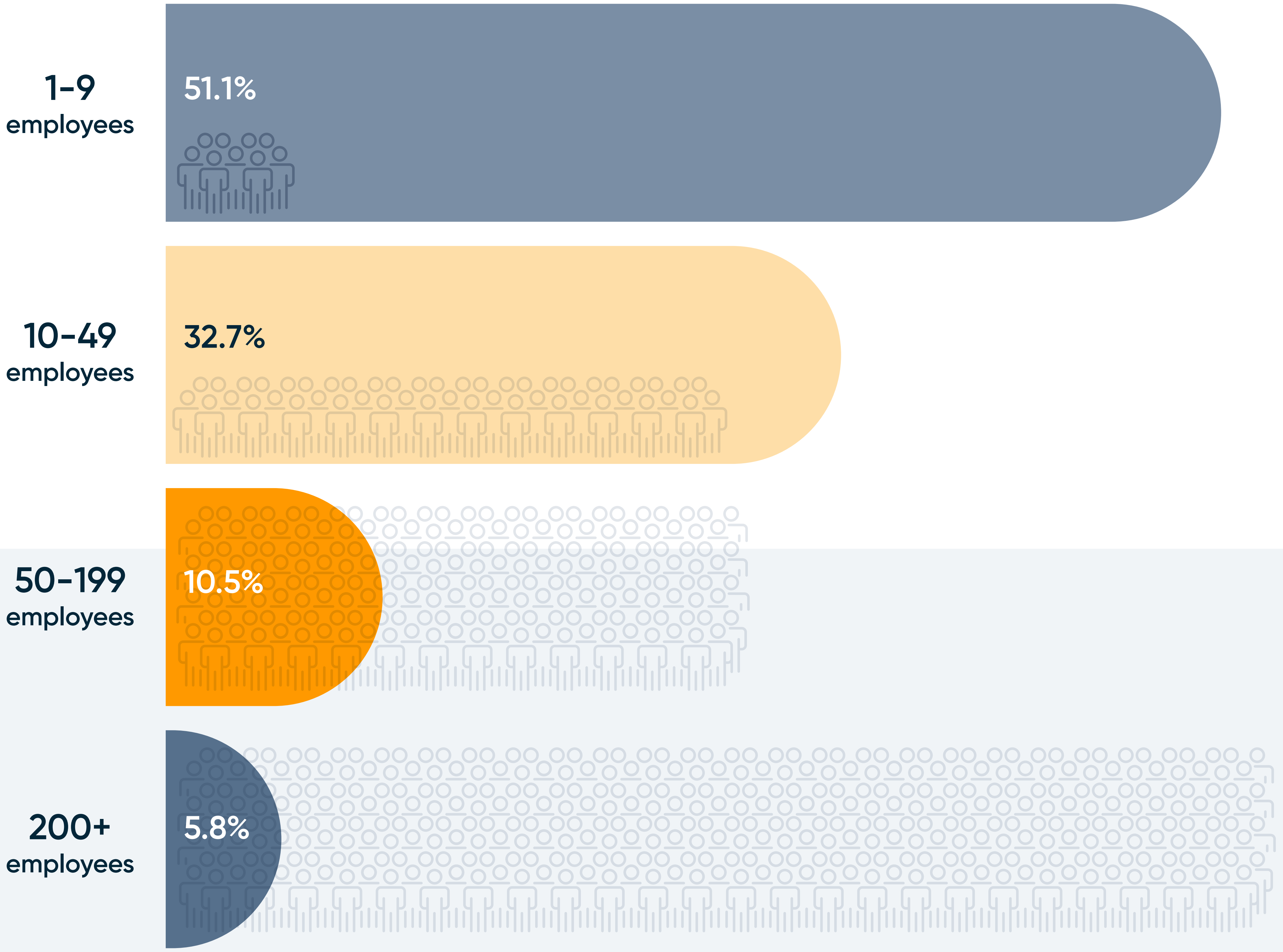
Role



Seniority



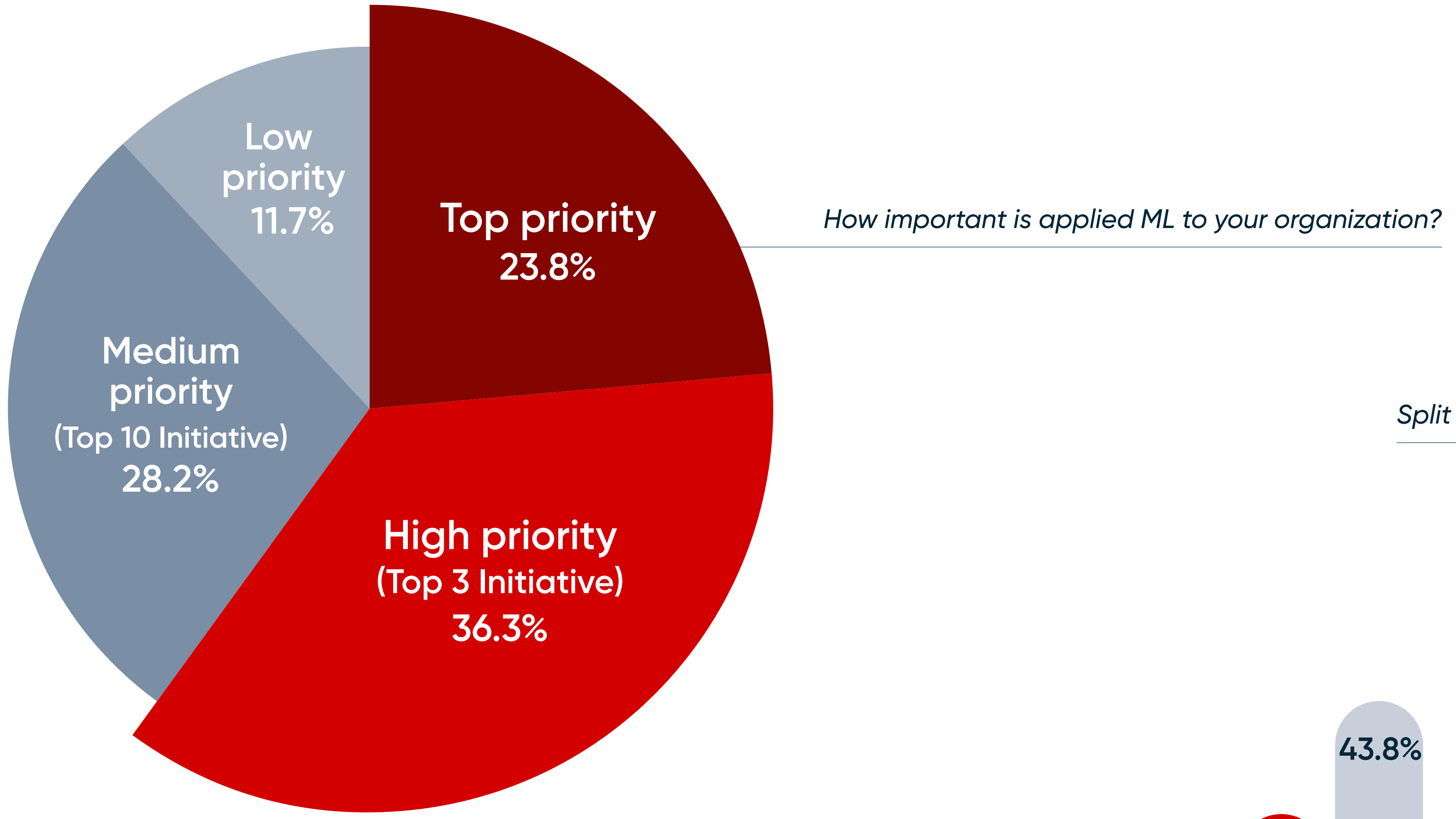
Data Team Size

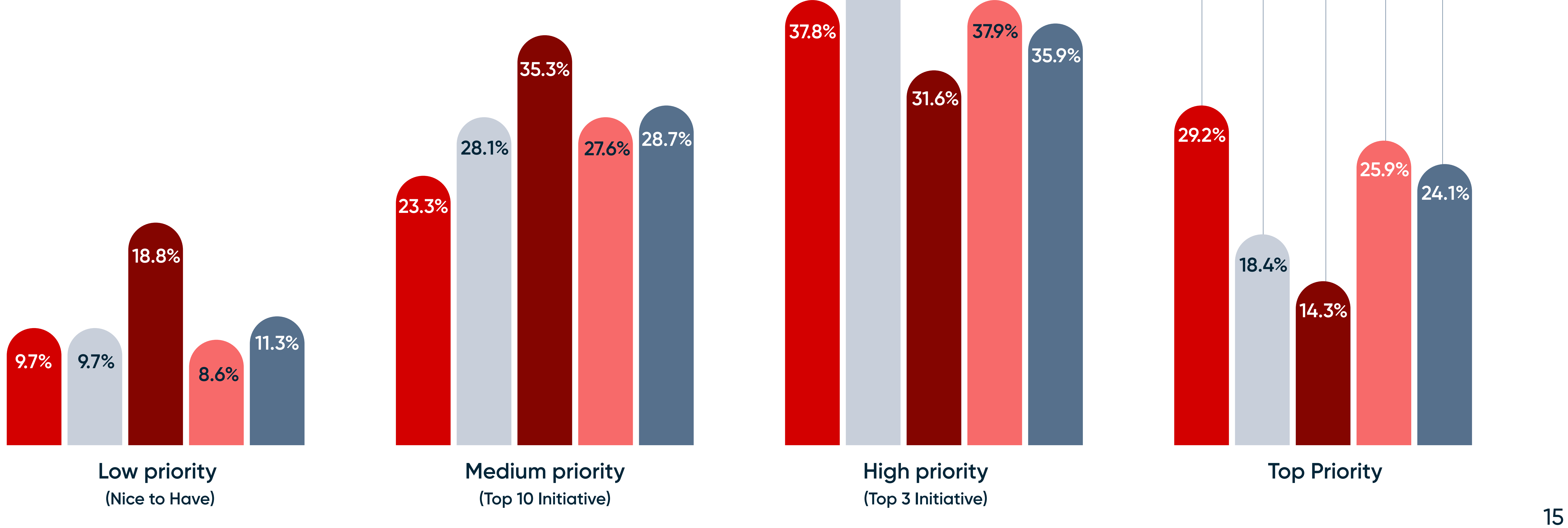


Applied ML is a top priority for organizations

Highlights

- **Over of 60% survey respondents identified applied ML as a top 3 company initiative.** Among these, 23.8% of respondents identified applied ML as *the* top company initiative.
- **The Software & Internet industry places the highest emphasis on applied ML**, with 29.2% of respondents identifying it as the top priority and 67.0% naming it as a top 3 priority.





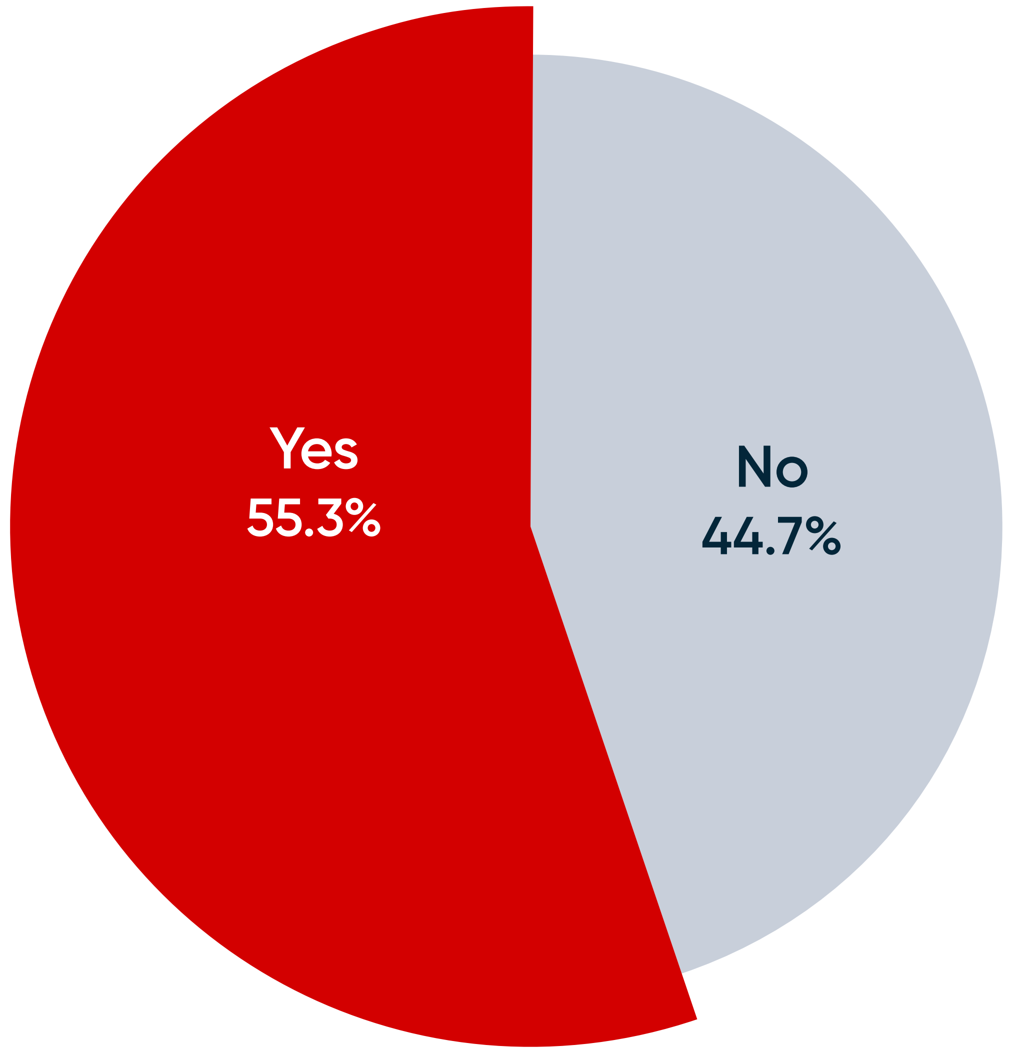
Applied ML is a top priority for organizations

ANALYSIS

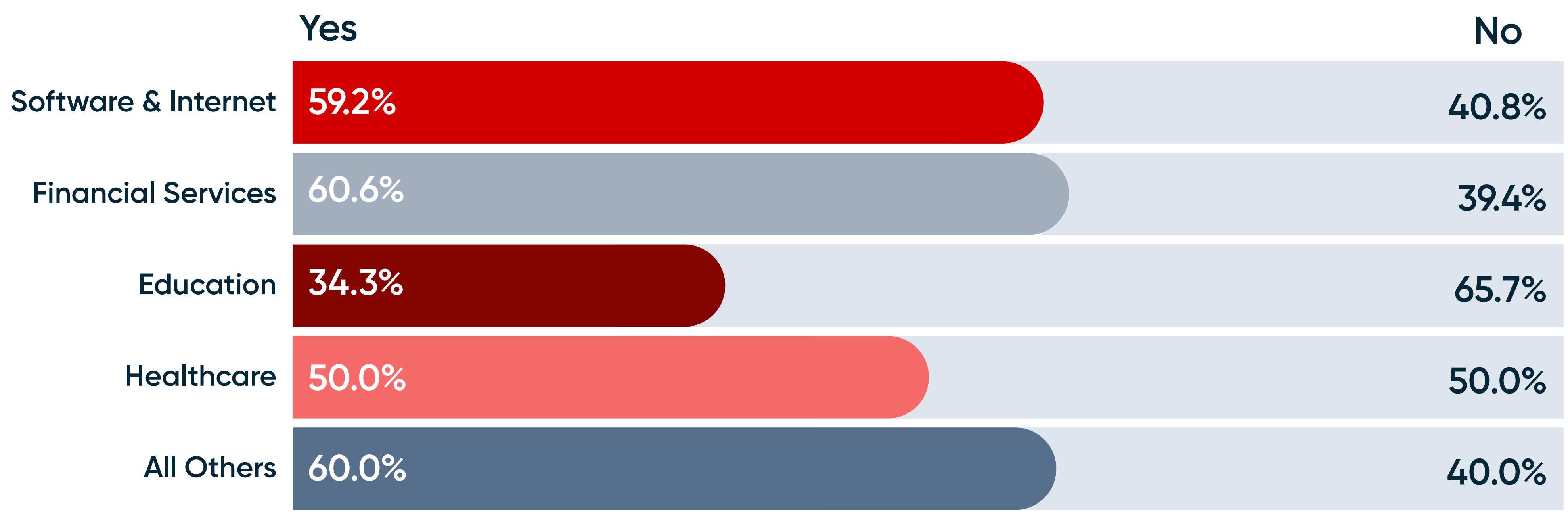
Highlights: Central Platform Teams

- 55.3% of respondents shared their organizations have a **central ML platform team**
- The presence of a central ML platform team is **correlated with the size of the organization**—for companies with over 1,000 employees, an average of 64.3% have central ML platform teams, compared to 45.3% of companies with fewer than 250 employees

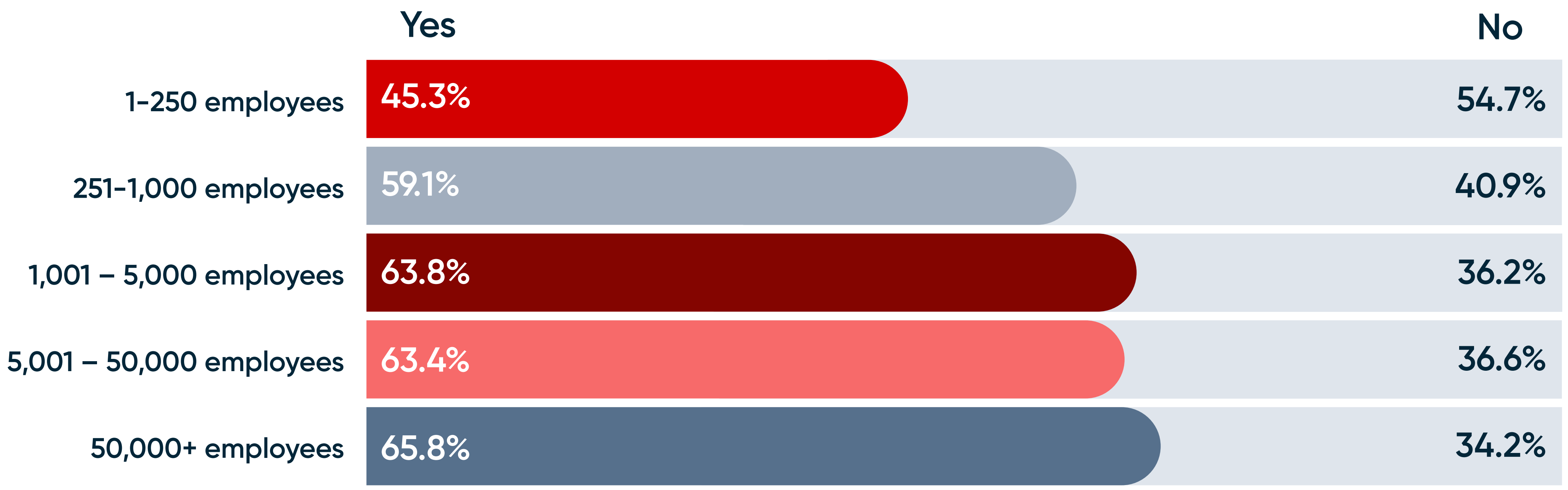
Do you have a central ML platform team responsible for building and maintaining your ML infrastructure?



Split by industry



Split by company size



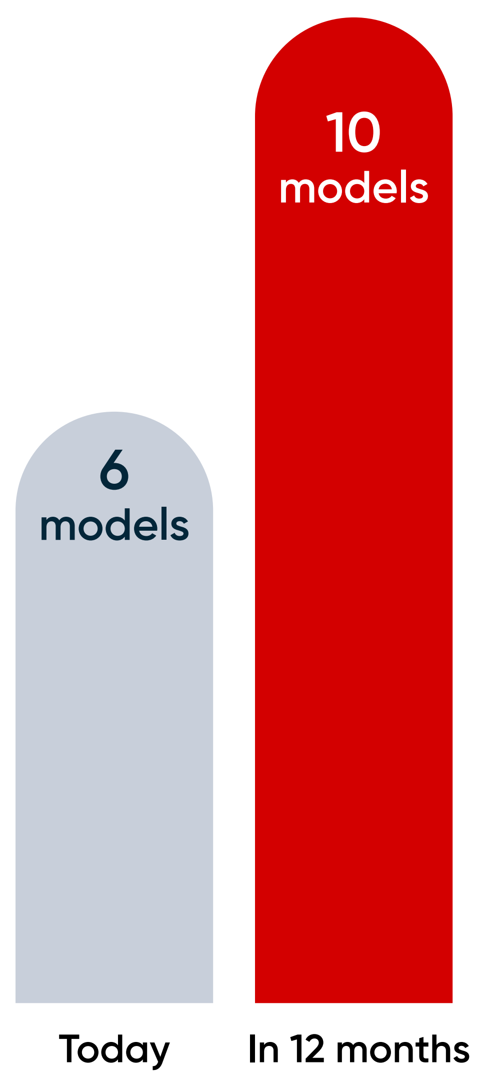
ANALYSIS

Applied ML is a top priority for organizations

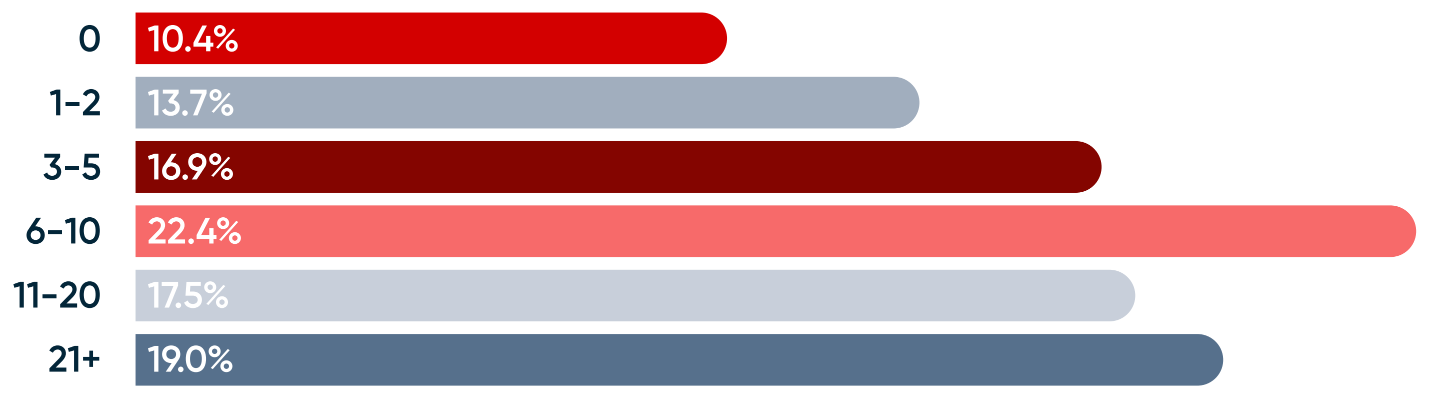
Highlights: ML Models in Production

- Half of survey participants indicated their organizations have 6 or more models in production, and companies are planning on deploying more models quickly. Based on survey responses, the projected median for **the number of models in production in the next 12 months will increase from 6 to 10**
- **The Financial Services industry is leading** other industries with the number of models in production today, as well as the number of models expected to be in production within the next 12 months
- As expected, company size correlates with the number of models in production today— i.e., **the largest companies have the most models in production (median: 12 models)**, while the smallest ones have the fewest number of models in production (median: 4 models)
- The survey also found that the number of models in production increases with the size of data team. However, even small teams manage to have a significant number of models in production, with **a median of 4 models in production for the smallest teams (<10 team members)**

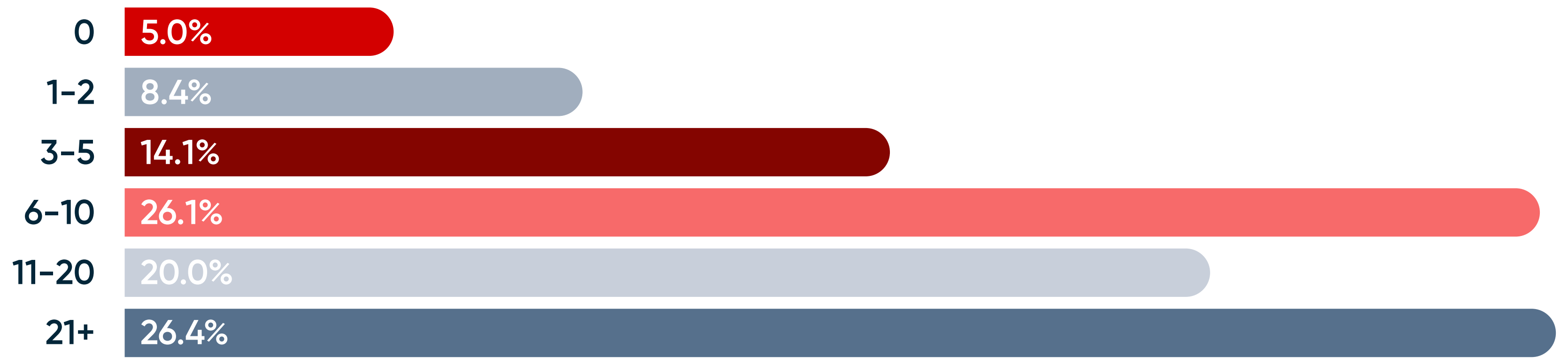
Median number of models in production



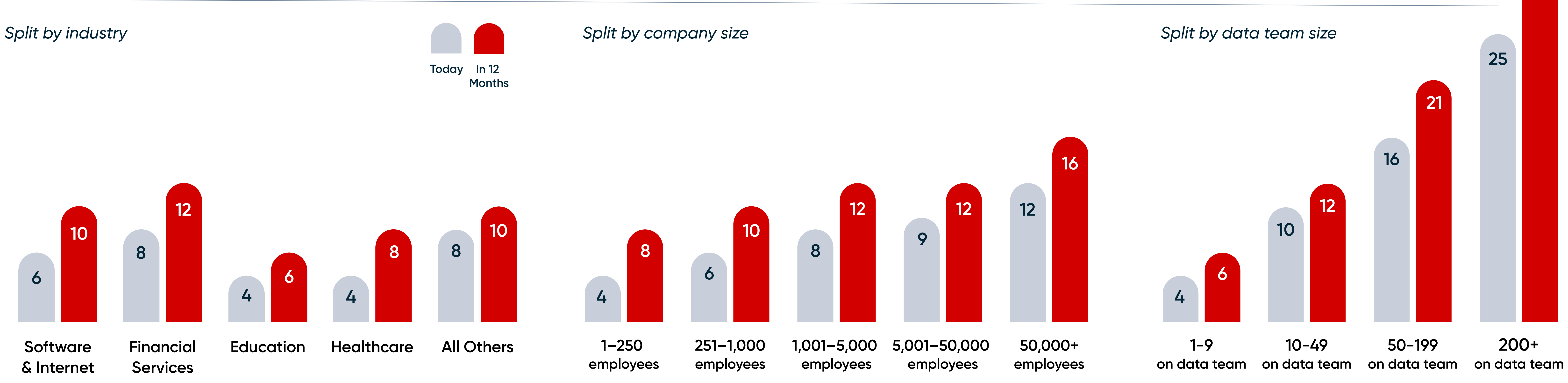
How many ML models do you have in production today?



How many models do you plan to have in 12 months?



Median number of models in production today vs. 12 months



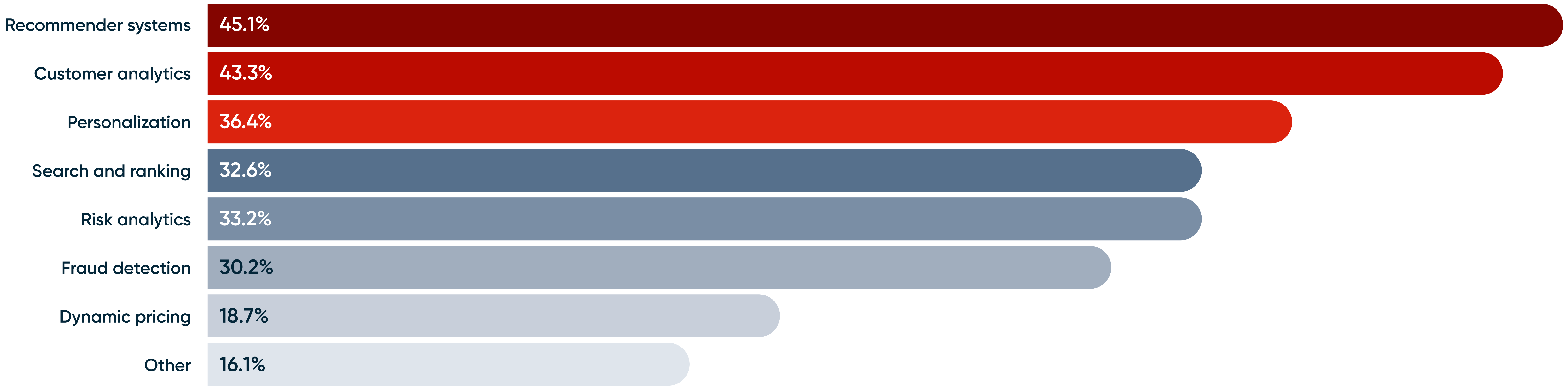
ANALYSIS

Applied ML is a top priority for organizations

Highlights

- The top 3 use cases are **recommender systems (45.1% of respondents), customer analytics (43.3%), and personalization (36.4%)**

Use cases: Which types of use cases are your ML models used for?



Real-Time ML

About Real-Time ML

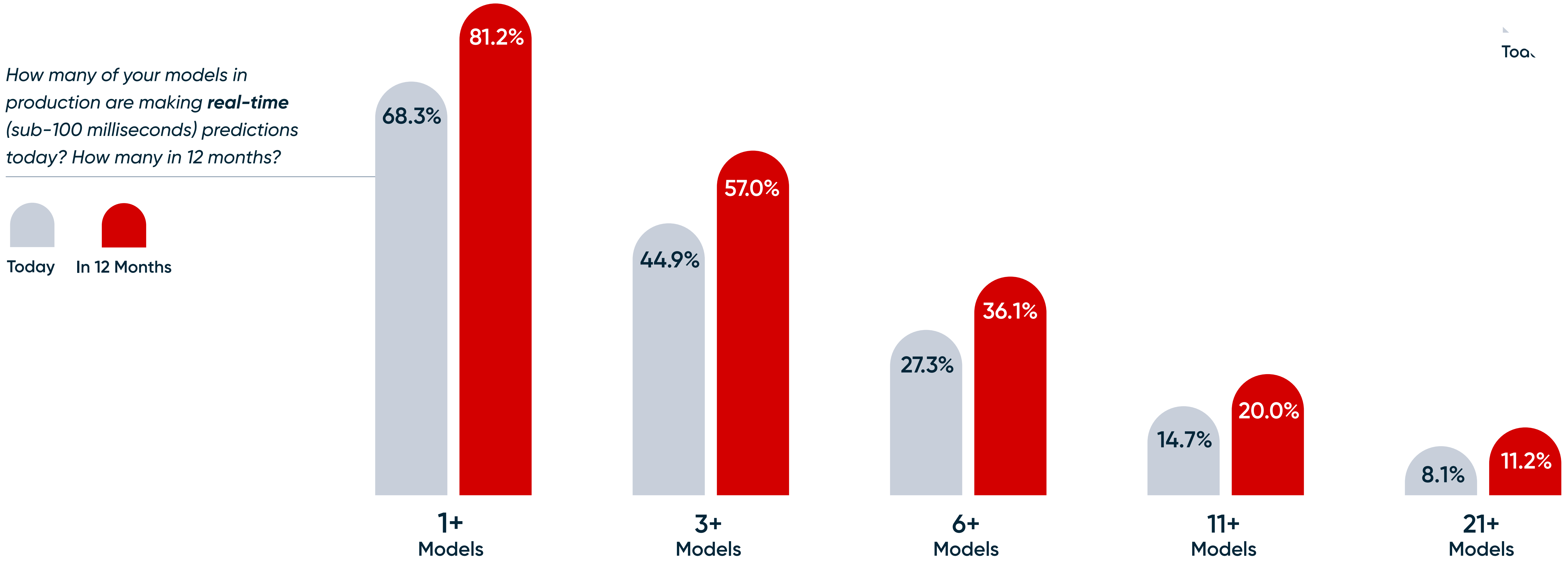
Real-time ML is a subset of applied ML use cases. Real-time ML consists of running ML models in production to make real-time predictions and using those predictions to power production applications, with no human in the loop. These use cases are typically more advanced and complicated than batch ML use cases.

Real-time ML models can be powered by batch data only, but can also use streaming or real-time data to increase the accuracy of predictions by incorporating the freshest information available. (*Read more about real-time ML.*)

Highlights

- Real-time ML already has significant adoption: **68.3% of respondents say their companies already have at least one real-time (i.e., sub-100 milliseconds) ML model in production**, while 14.7% have more than 10[1]
- Survey responses indicate that **the above number will likely increase to 81.2% in 12 months.**
- The **Software & Internet industry** has the highest rate of real-time ML adoption today.

*How many of your models in production are making **real-time** (sub-100 milliseconds) predictions today? How many in 12 months?*



How many of your models in production are making real-time (sub-100 milliseconds) predictions today? Split by industry

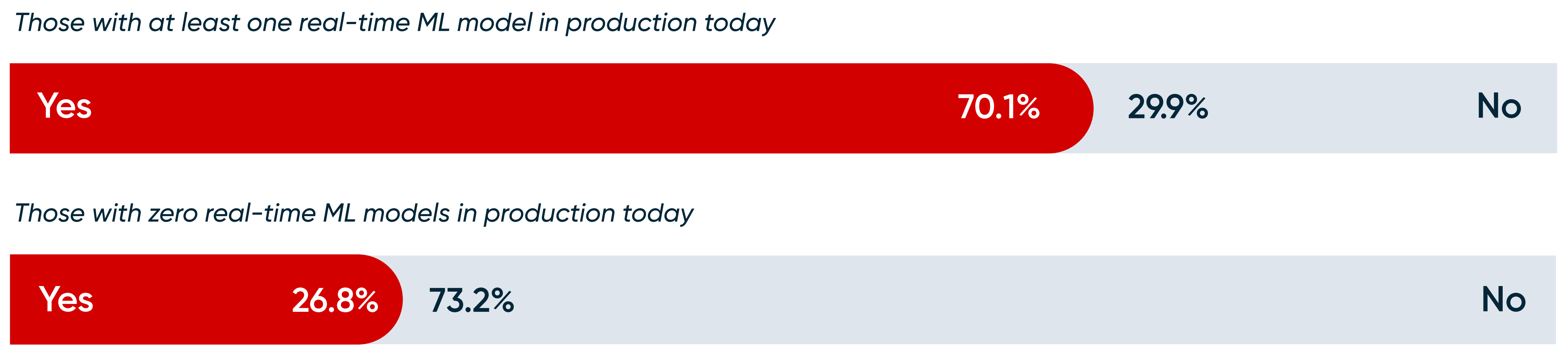
[1] *This number is one of the biggest surprises uncovered in this survey, and it's a big step forward compared to where the industry was even just a year ago. Keep in mind, however, that this number reflects mainly early real-time experiments on the path to full ML adoption rather than a full-blown real-time ML deployment. Even so, 68.3% is a very promising number, and even more so when looking at 12-month out prospects.*

Real-Time ML

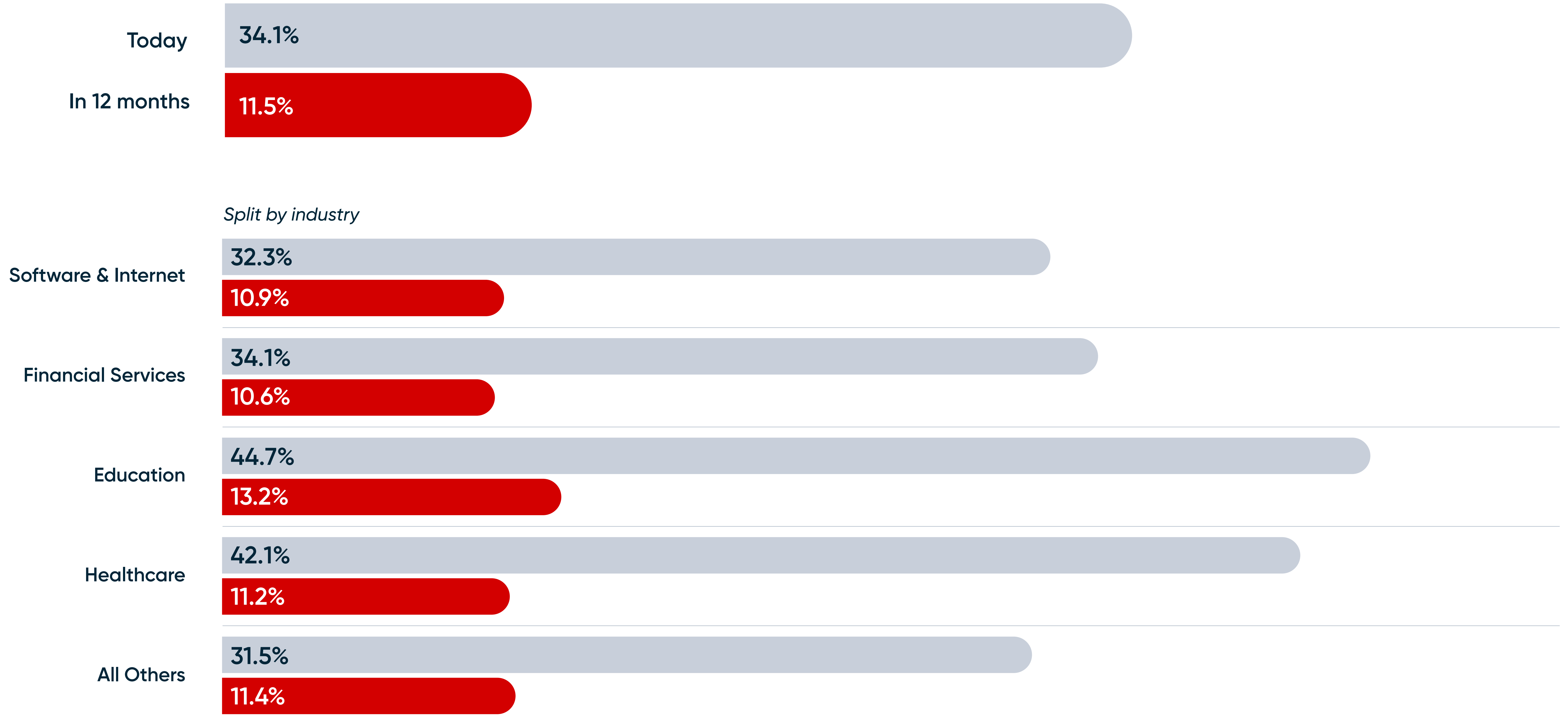
Highlights

- The existence of a central ML platform team is highly correlated with real-time ML adoption. **For organizations with at least one real-time ML model in production, 70.1% have a central ML platform team** (for organizations with no real-time ML models, only 26.8% have a central ML platform team.)
- The number of companies using only batch data to power their models (as opposed to streaming or real-time data) is expected to fall from 34.1% to 11.5% in the next 12 months.

Do you have a central ML platform team responsible for building and maintaining your ML infrastructure?



% of companies using only batch data today vs. 12 months from now (Other options were streaming and real-time data)



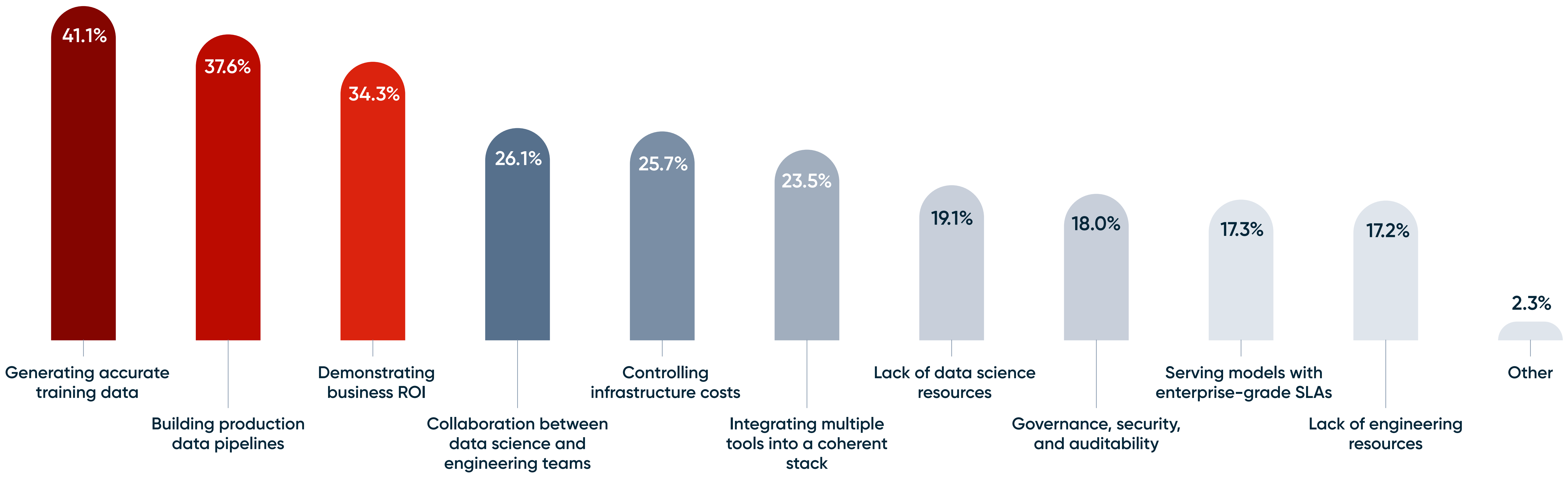
Challenges on the road to applied ML

Highlights

The top 3 challenges encountered when deploying new models to production are:

- Generating accurate training data (41.1%)
- Building production data pipelines (37.6%)
- Demonstrating business ROI (34.3%)

What are the 3 biggest challenges encountered when deploying new models to production?



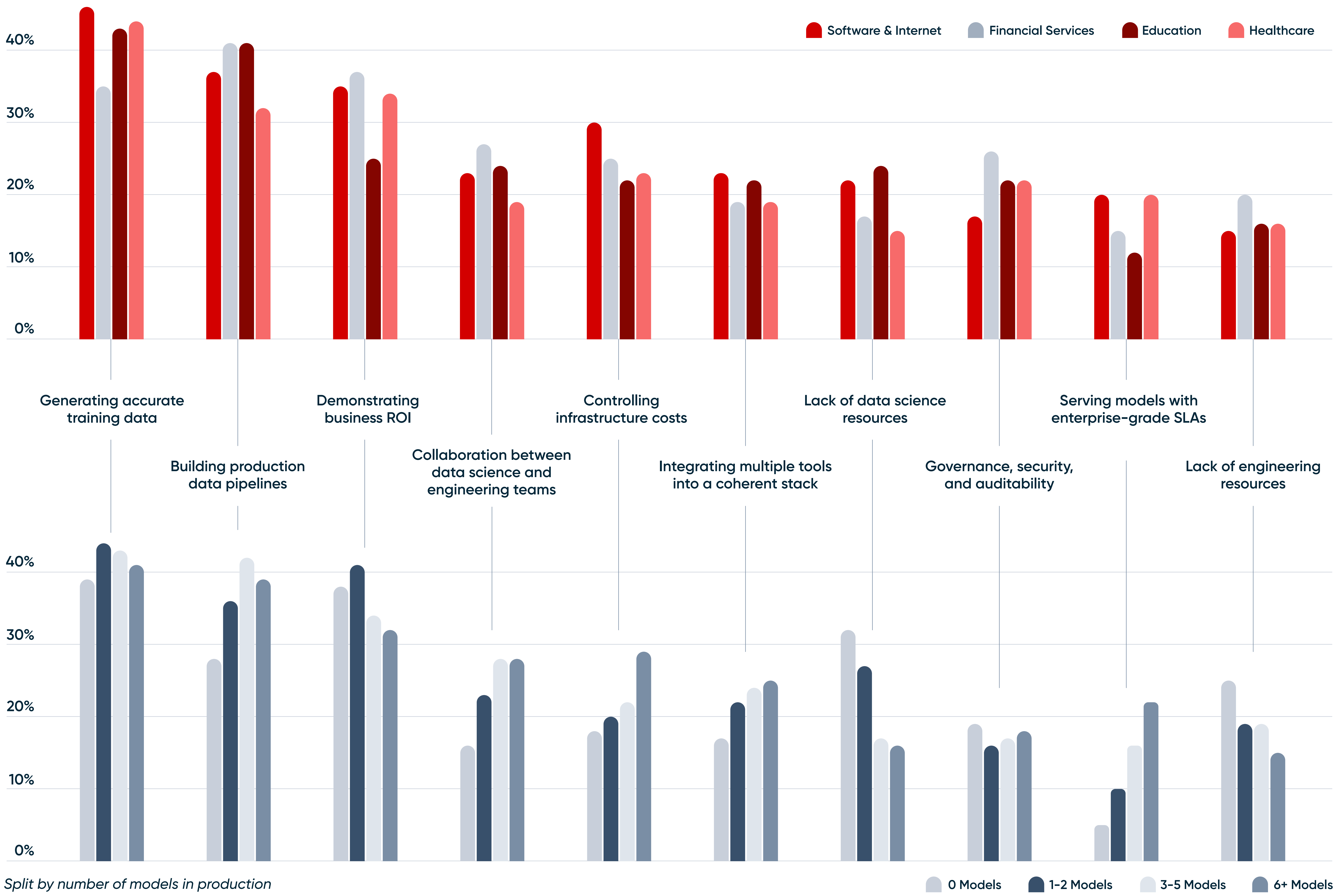
Challenges on the road to applied ML

Highlights

The number of ML models in production are correlated with different concerns / pains:

- Companies with more ML models in production are more concerned about controlling infrastructure costs and having effective enterprise-grade SLAs
- Companies that are just getting started with applied ML (defined as having 0-2 models in production) are more concerned about the lack of data science and engineering resources

What are the 3 biggest challenges encountered when deploying new models to production? Split by industry



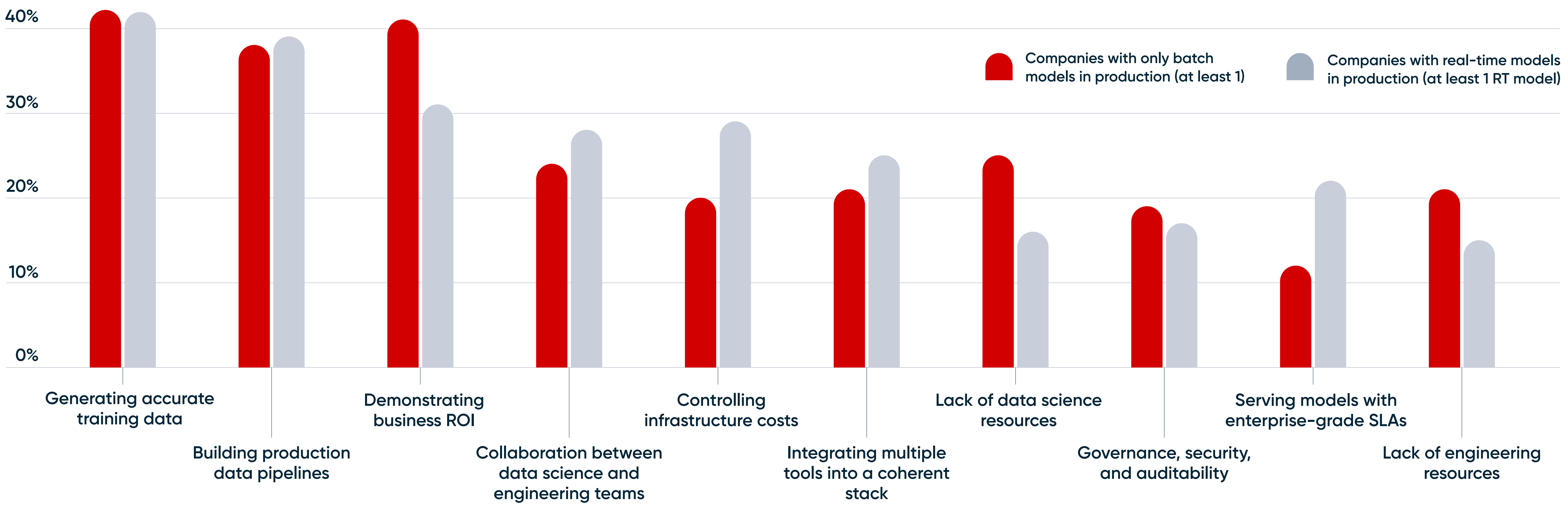
Split by number of models in production

ANALYSIS

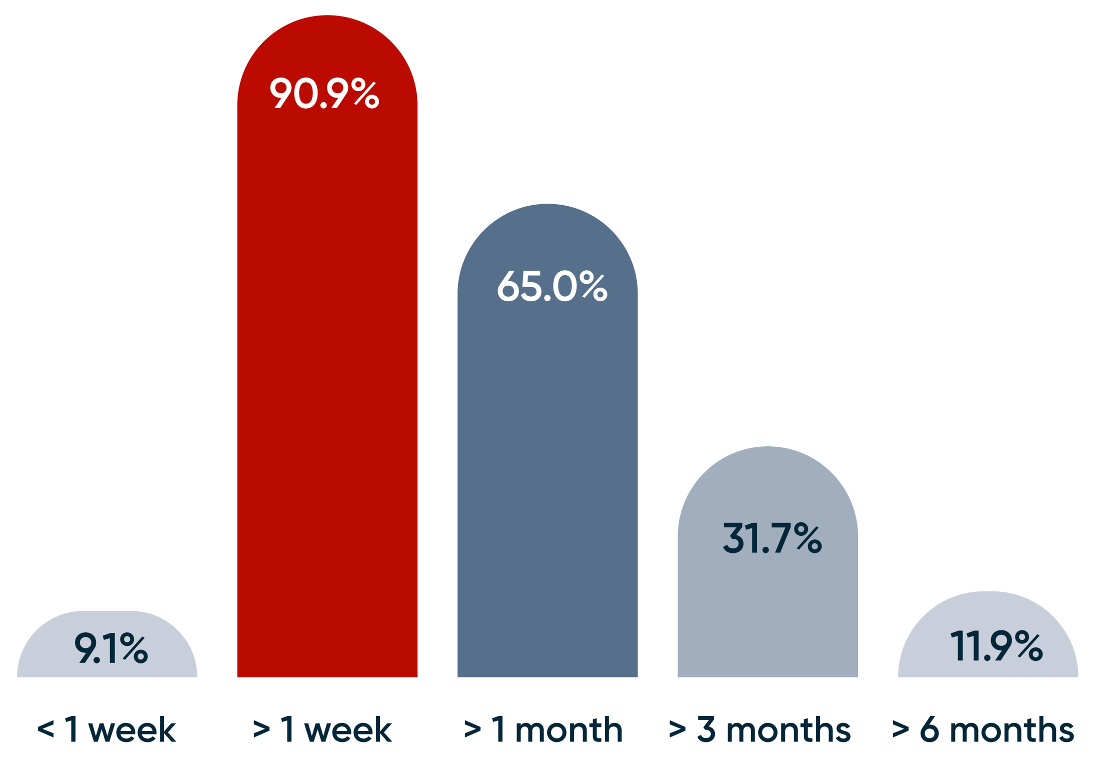
Challenges on the road to applied ML

- Respondents who shared that their companies have only batch models in production also shared that they struggle more with simpler organizational problems, such as demonstrating business ROI (41.5%) and lack of engineering and data science resources (21.5% and 24.8%, respectively)
- Meanwhile, respondents who shared that their companies have real-time models in production struggle more with "advanced" challenges, such as collaboration between engineering and data science teams (28.0%) and serving models with enterprise SLAs (21.5%)
- Deploying a new model to production is a long process (>1 month for 65.0% of respondents and >3 months for 31.7%)
- Companies in the Software & Internet industry deploy models to production faster than average.

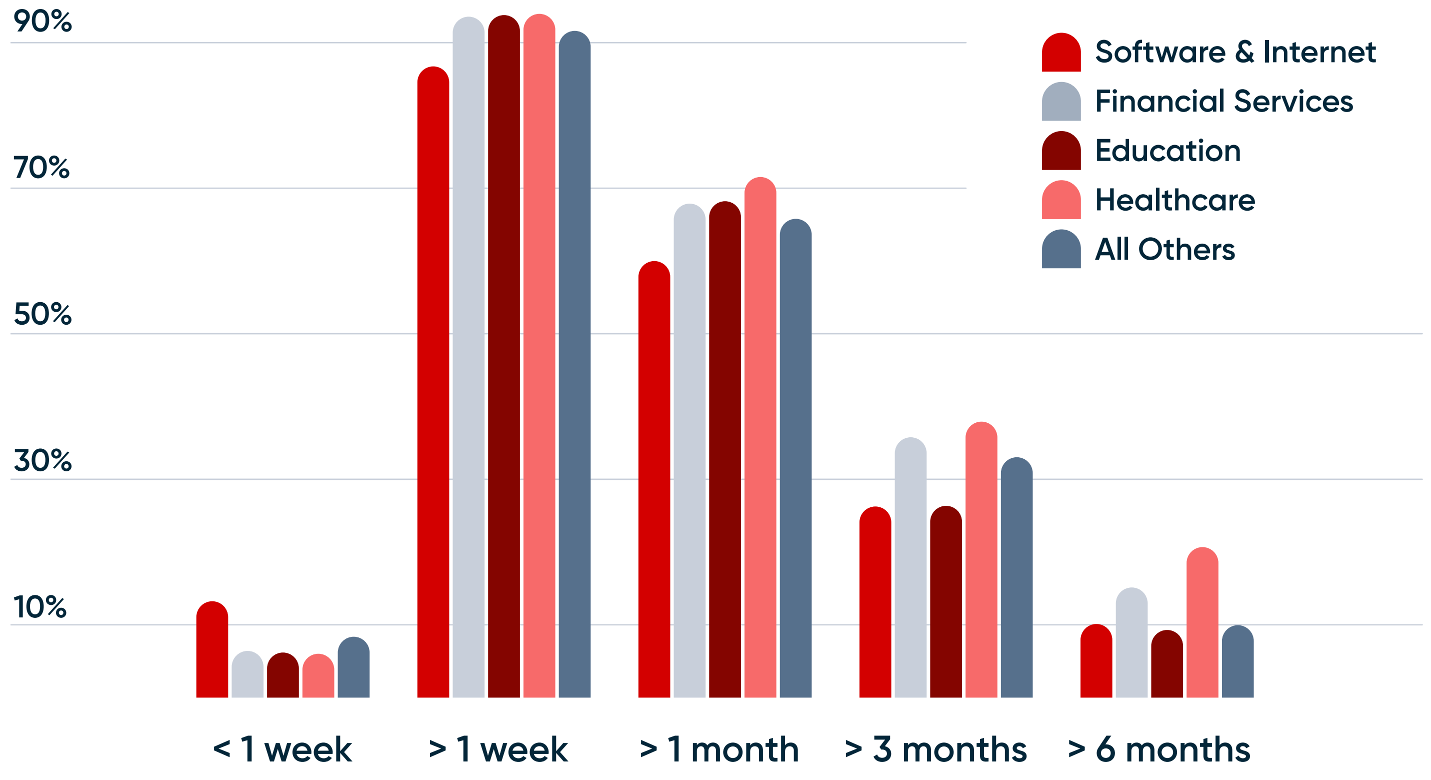
What are the 3 biggest challenges encountered when deploying new models to production? Split by batch vs. real-time



How long does it take to deploy a new model from idea to production?



How long does it take to deploy a new model from idea to production? Split by industry



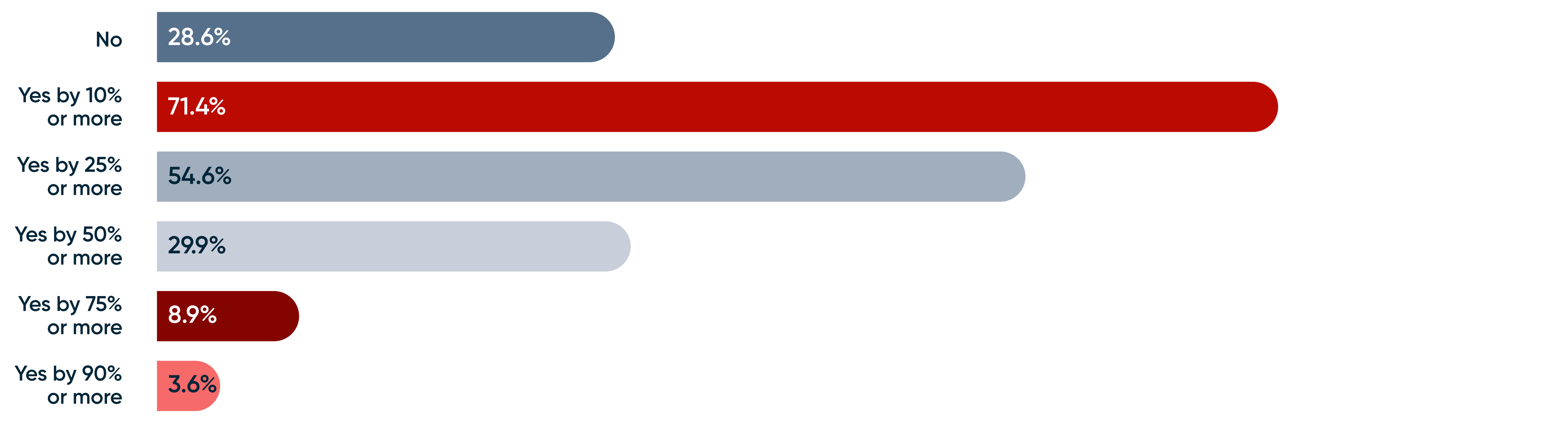
ANALYSIS

Challenges on the road to applied ML

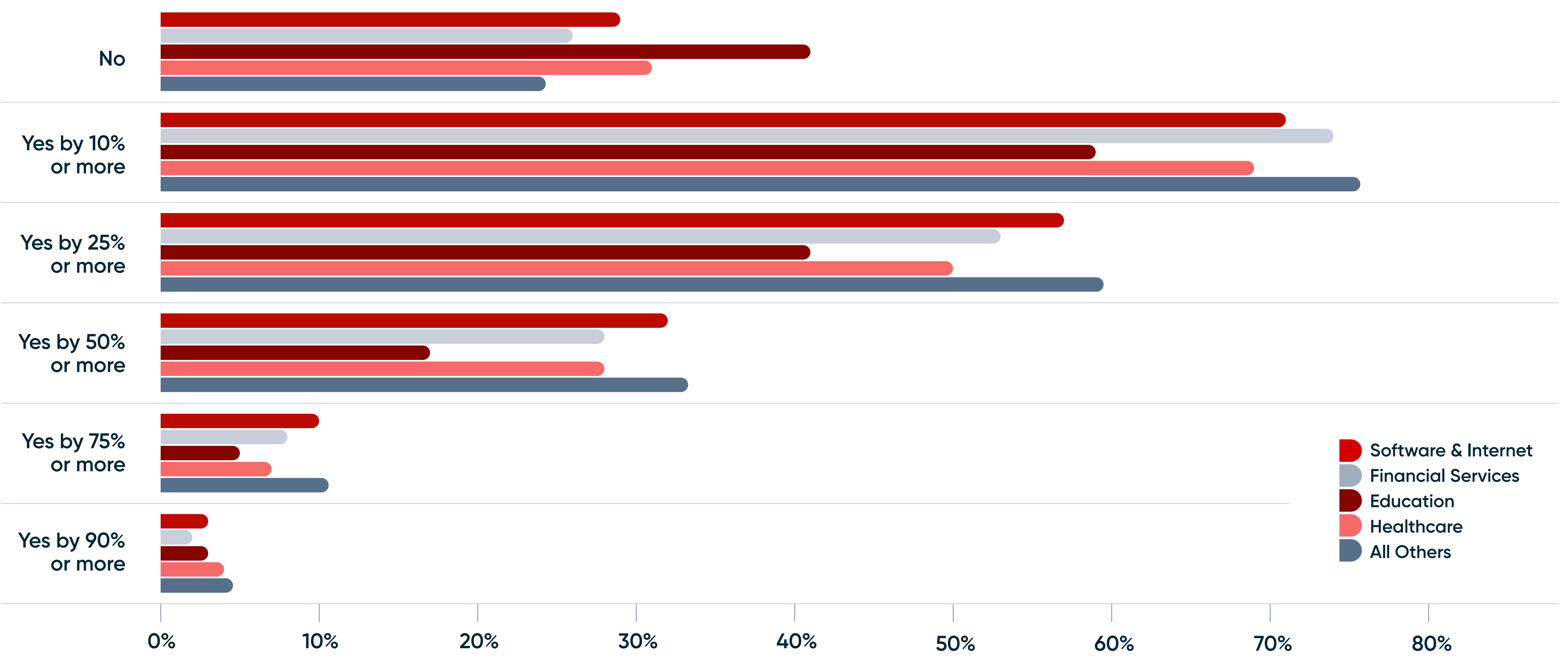
Highlights

- 71.4% of respondents shared that their companies aim to improve deployment time by at least 10% in the next 12 months.

*Does your company aim to **reduce model deployment time** within the next 12 months? If so, by how much?*



*Does your company aim to **reduce model deployment time** within the next 12 months? If so, by how much? Split by industry*

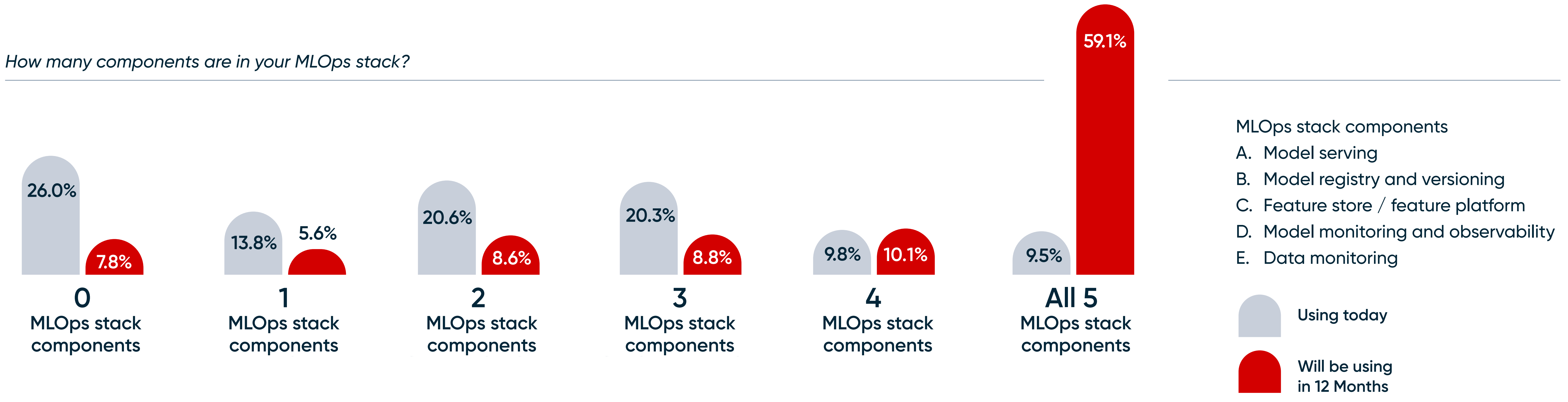


MLOps Stack

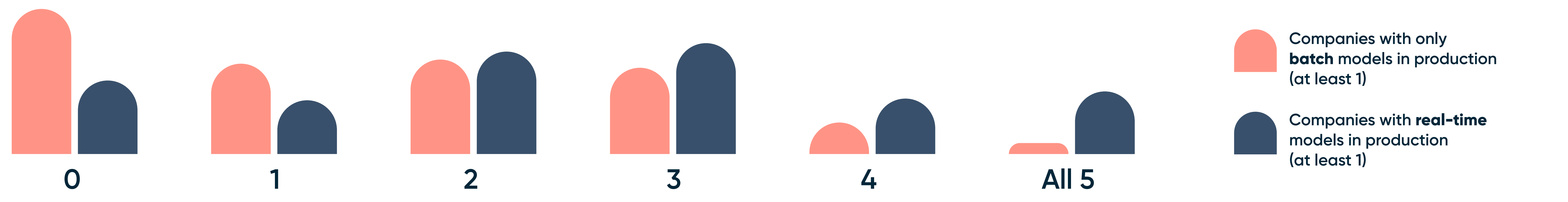
Highlights

- Companies are investing heavily in machine learning, and a number of them are planning to shore up their MLOps stack to solve for these challenges.
  - 39.6% of survey participants shared that their companies have adopted 3 or more components out of the 5 major ones—and this number is expected to almost double (to 78.0%) by the end of 2023
  - Only 9.5% of participants shared that their companies have adopted all 5 major components at the time of the survey, but 59.1% of participants expect to adopt all 5 in the next 12 months
- Companies with real-time ML models in production are more advanced in terms of MLOps adoption

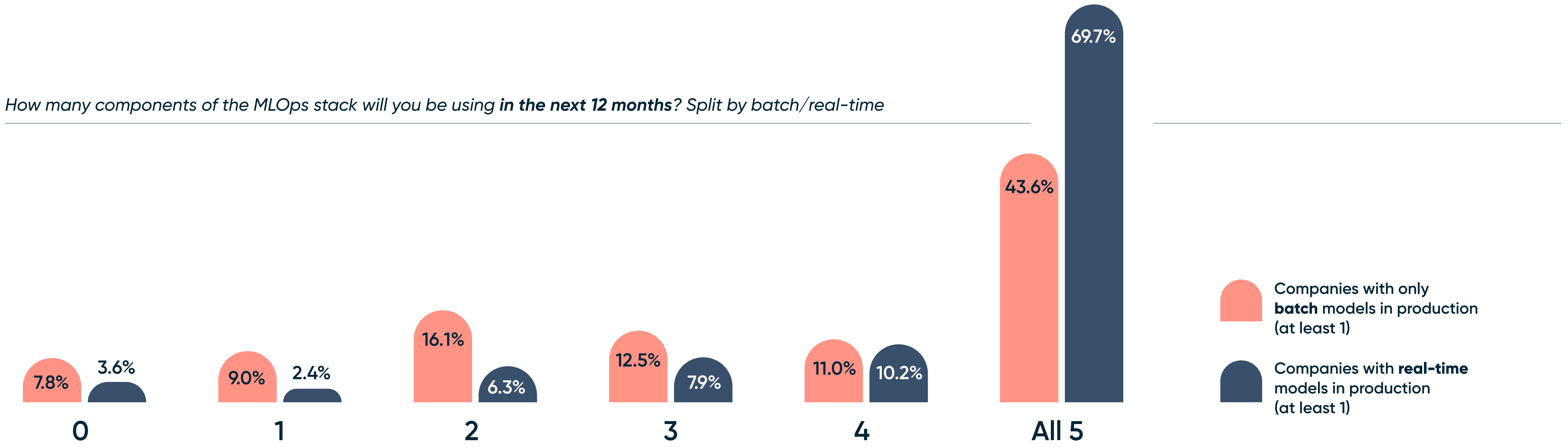
How many components are in your MLOps stack?



How many components are in your MLOps stack? Split by batch/real-time



*How many components of the MLOps stack will you be using **in the next 12 months**? Split by batch/real-time*

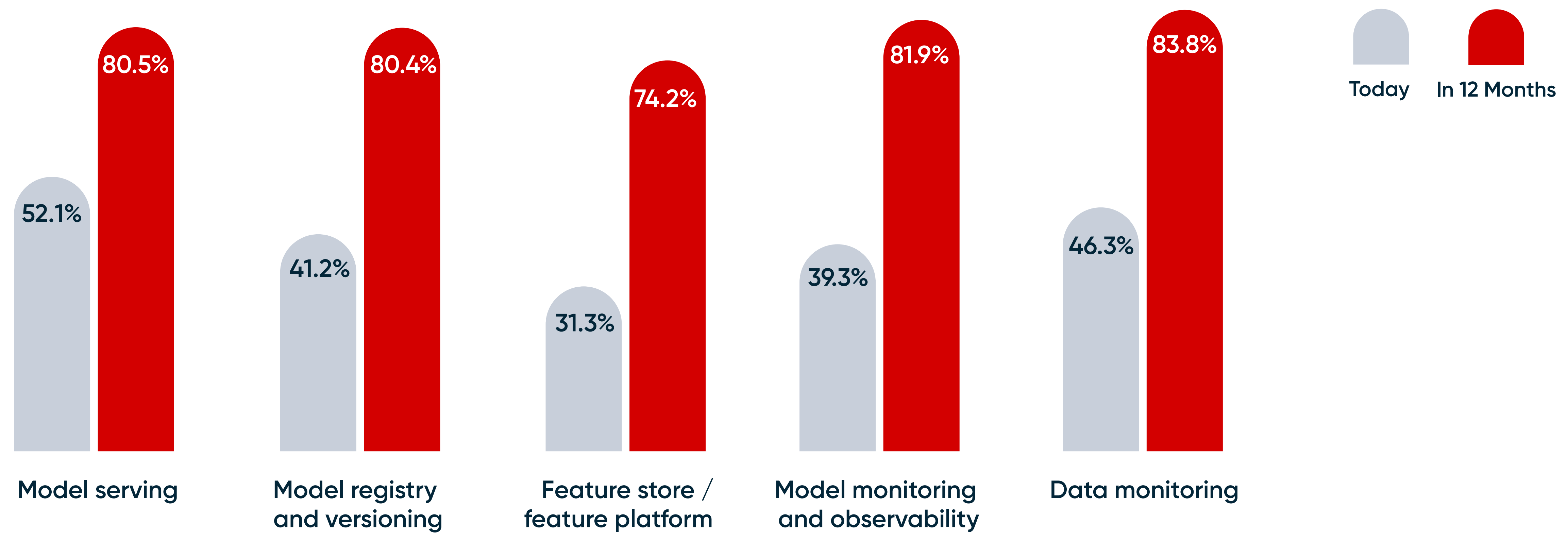


MLOps Stack

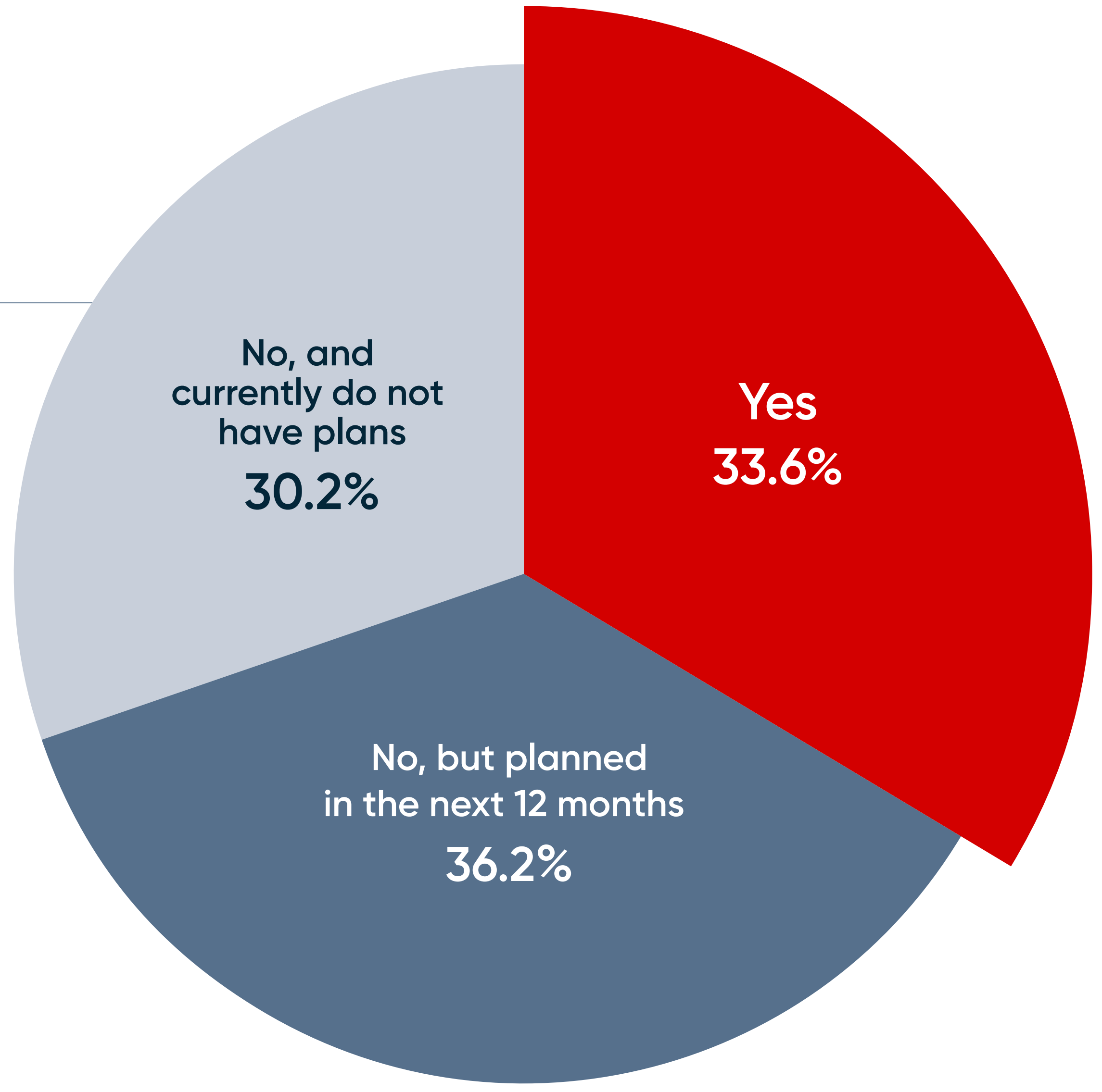
Highlights

- The Feature Store / Feature Platform and Monitoring & Observability components will see the largest increases (~43 percentage points increase for both) in adoption in the next 12 months
- Nearly 70% of respondents say they either have or plan to have a central MLOps platform in the next 12 months

Which components of the MLOps stack does your company use today or plan to adopt within 12 months?

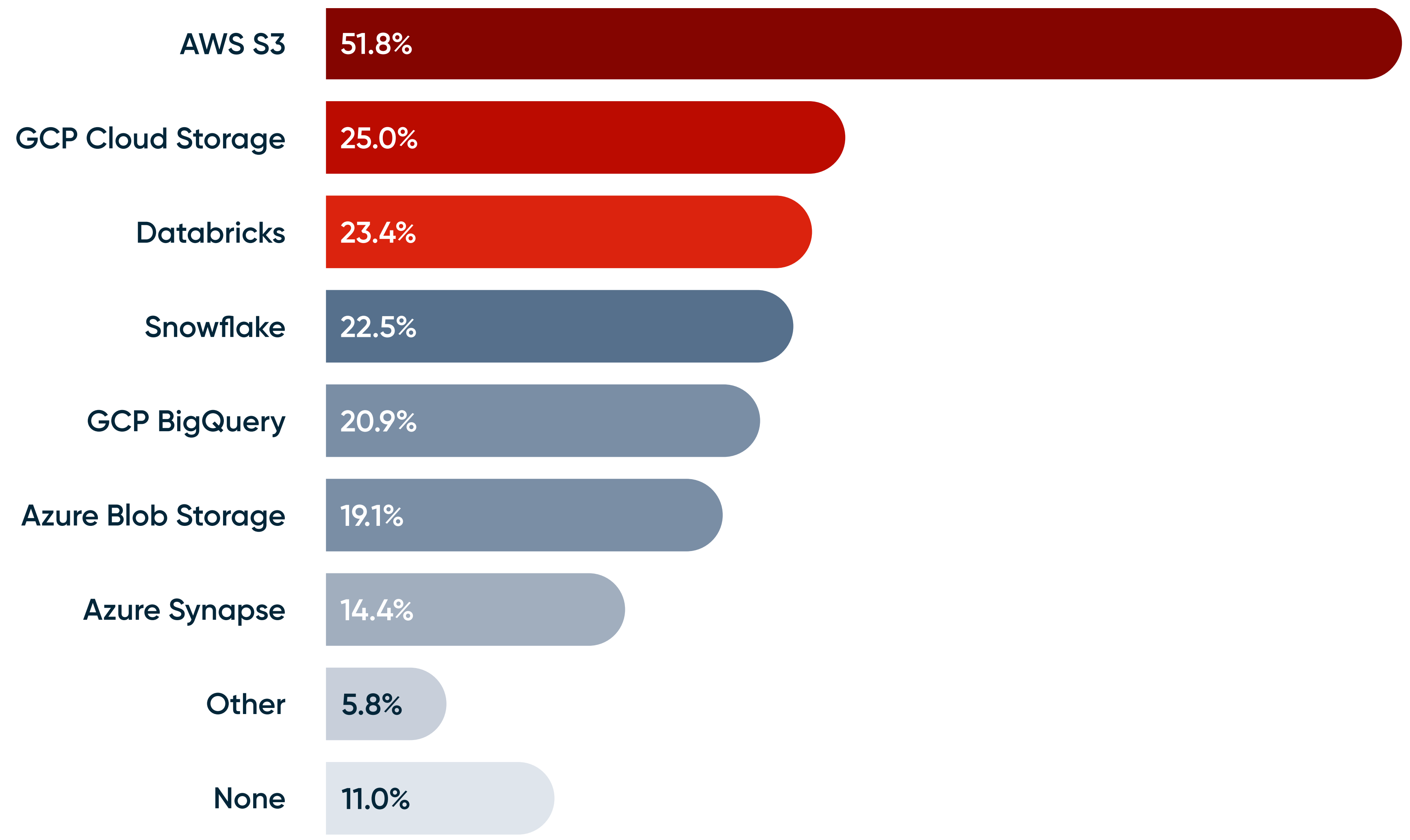


Does your company have a central MLOps platform shared by all your data science and ML engineering teams?

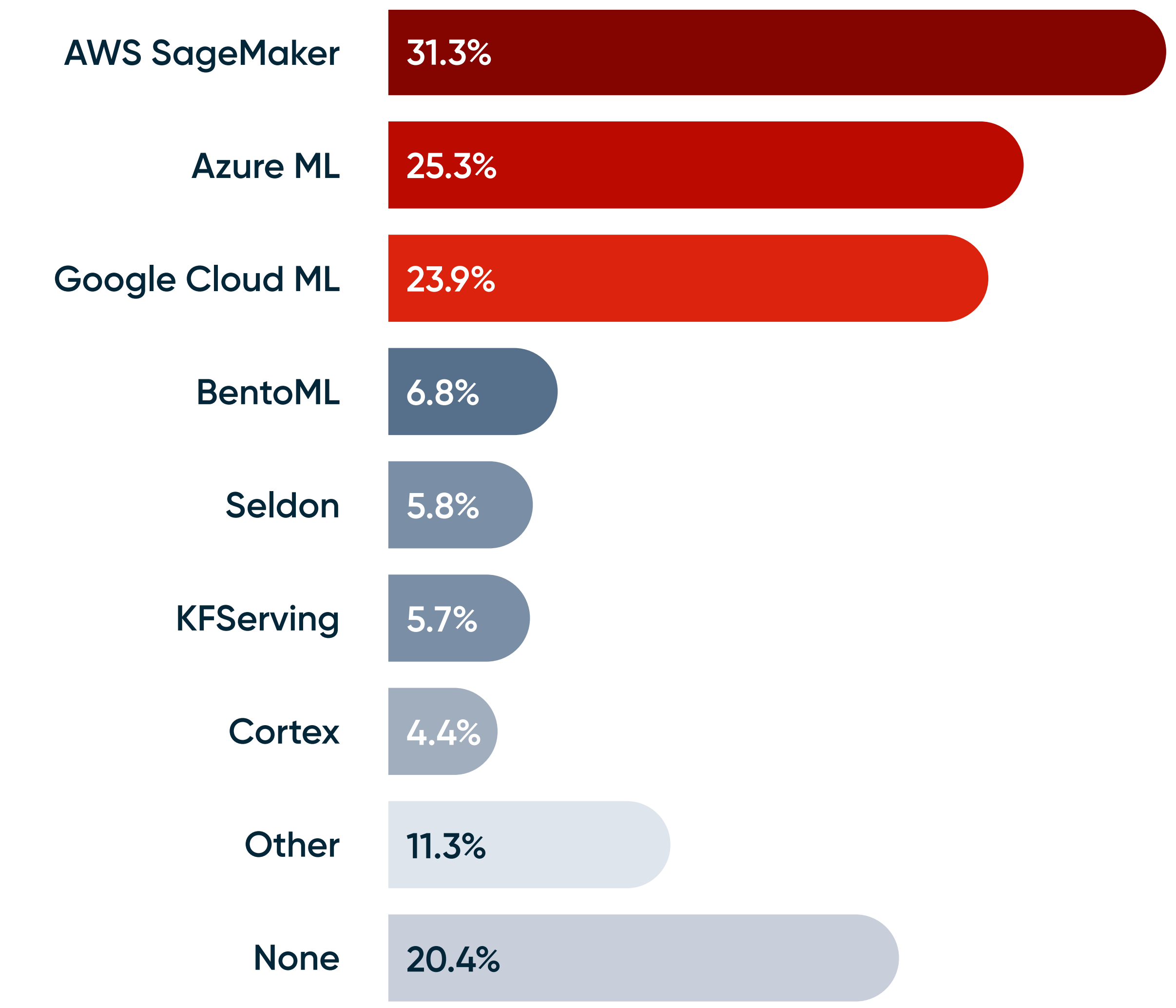


MLOps Stack

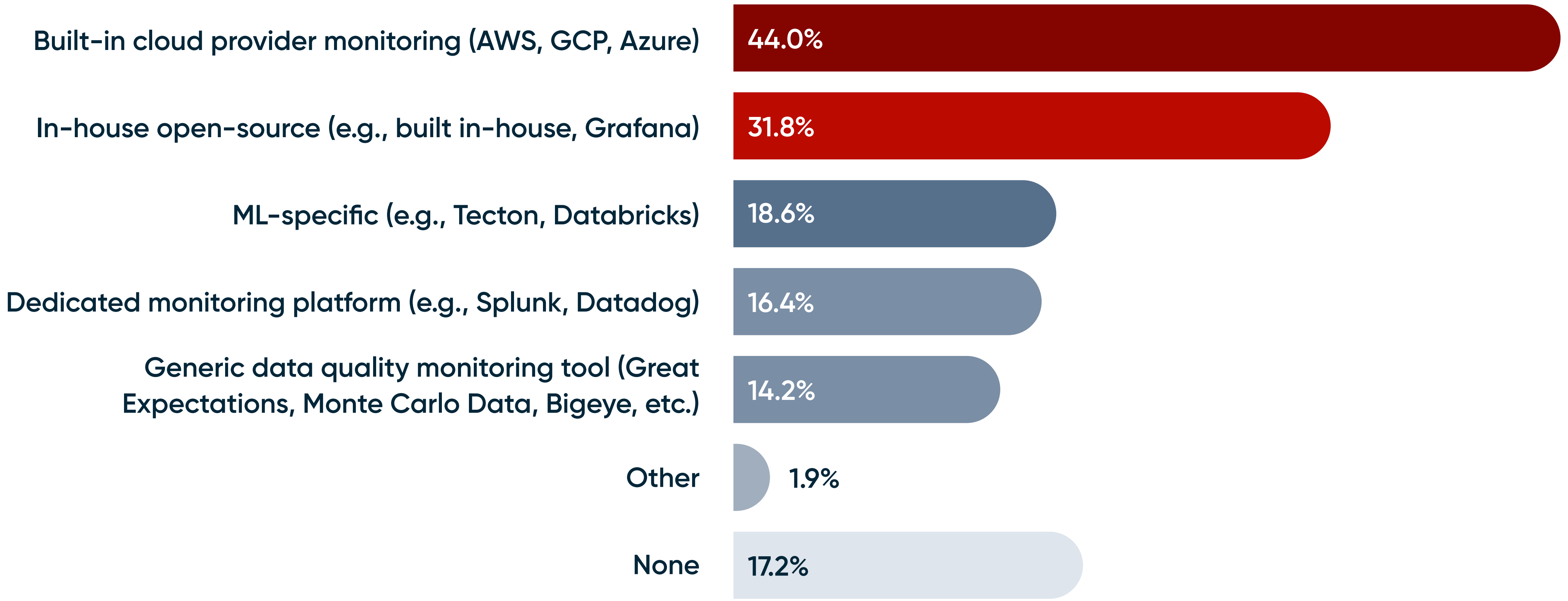
Which enterprise data platforms does your company use?



What is your company using for model serving?



Which tool is your company using to monitor ML-related data?



Thanks for reading!

This report has zeroed in on the current state of applied machine learning, the challenges faced by organizations, and their goals for the future. The adoption of applied machine learning in various industries is on the rise, and companies are leveraging it for a wide range of use cases. Despite the challenges in generating accurate training data, building production data pipelines, and demonstrating business ROI, survey results show companies are committed to building out their MLOps tech stack to improve their ML capabilities.

As companies continue to invest in machine learning and work to overcome the challenges, the future of applied ML is shaping up to benefit both the builders behind the scenes and the end users of those applications. Organizations are expected to increase the number of models that make batch-driven and real-time-driven predictions to make better decisions and provide improved customer experiences. As a result, companies who move faster toward real-time ML use cases stand to gain a competitive advantage in the market.

By sharing the insights from this survey, we hope to inform and guide data teams on the evolving landscape of applied machine learning and contribute to its growth and success.

ADDITIONAL RESOURCES

Blog Post: What Is Real-Time Machine Learning?

Blog Post: What Is Online / Offline Skew in Machine Learning?

Blog Post: A Practical Guide to Building an Online Recommendation System

Blog Post: Why Building Real-Time Data Pipelines Is So Hard

Financial Services Customer Case Study: Tide

Meal Kit Customer Case Study: HelloFresh

Healthcare Customer Case Study: Vital

Whitepaper: The Buyer's Guide to Evaluating Feature Stores & Feature Platforms

Video: Designing & Scaling FanDuel's ML Platform: Best Practices & Lessons Learned

Video: How to Make the Jump From Batch to Real-Time Machine Learning

Accelerate your journey to applied machine learning. Request a demo today.

tecton.ai