DATA-FIRST THINKING

THE RISE OF THE DATA SCIENTIST:

Machine learning models for the future

REFINITIV®

DATA IS JUST
THE BEGINNING®

INTRODUCTION

COVID-19 is the arch accelerator. In under six months we have experienced a level of technological change in financial markets that would otherwise have taken a decade to play out.

While continually improving computing power is driving the rise of electronic trading, our more virtual "new normal" is obliging financial institutions to embrace the cloud; fast track digital collaboration and explore the future with ideas as radical as augmented reality headsets to mimic the trading floor.

The need to build a more virtual markets infrastructure is spurring new technology at an incredible pace and across our industry, players are doubling down on their investments in data science and machine learning to gain an edge. Four fifths of those who took part in this research are making significant investments in artificial intelligence and machine learning and almost as many are making it a core component of their business strategy.

This is a revolutionary moment in financial technology and one that is quickly widening the gap between the "haves" and the "have nots". Those businesses that can best harness data and emerging data science techniques – and deploy them at scale – are stretching their advantage. But their success is about much more than investment dollars alone.

As this report shows, financial players (big and small) must find a delicate balance between talent, technology, leadership and business culture – all underpinned by quality data. From that foundation our industry's brilliant data scientists are delivering solutions we would scarcely have believed a year ago. Unleashed by this pandemic, the technological revolution sweeping our industry will only grow stronger.

David Craig
Chief Executive Officer
Refinitiv

CONTENTS

FOREWORD

Amanda West
Global Head of
Refinitiv Labs,
Refinitiv

Welcome to Refinitiv's Second Annual Artificial Intelligence/Machine Learning (AI/ML) Report.

If there is one thing to take away from the global emerging tech trends in 2020, it should be that scaled AI/ML is the new normal in financial services. Natural language processing (NLP) has become mainstream through a focus on creating value from unstructured data.

McKinsey & Company recently estimated that AI technologies could deliver up to $1 trillion in incremental value annually to global banking. In the same report, they warn that "banks that fail to make AI central to their core strategy and operations will risk being overtaken by competition and deserted by their customers."

AI/ML models are maturing, companies are striving to deploy ever more sophisticated techniques at scale, such as deep learning, and firms are beginning to execute the rapid innovation cycles enabled by next-generation AI/ML technology.

Investment and deployment trends are positive. **Our 2020 survey uncovers a plethora of AI/ML use cases in financial services, from customer onboarding (26%), personalized services and targeting (30%), to trade execution (27%) and managing market risk (40%).**

Previous challenges relating to investment levels, technology choices and access to talent have diminished, providing a robust foundation to implement AI/ML models at scale.

The growth of the data science community – both in size and influence – in financial services is a key driving force behind these developments. The number of teams within firms has seen an exponential increase of over 260% since 2018, and data scientists are now more likely to be found working across different business units, rather than as part of a support function in technology.

This year's research also reveals that AI/ML models and their associated insights are increasingly powered by different data sources. **The number of firms that only use unstructured data has shot up from 2% in 2018 to 17% in 2020, and only 3% of the firms surveyed report that they do not use alternative data sources, down from 30% in 2018.**

However, even as firms diversify their data sources, quality and accessibility challenges continue to disrupt their AI/ML strategies. Data remains both a fundamental prerequisite to success and a critical barrier to successful adoption.

If firms are going to genuinely benefit from the speed, agility and value of an 'AI-first' vision, they need high-quality, trustworthy data that can be easily accessed, ingested and manipulated for the variety of financial use cases they are progressing.

Tackling the fallout from COVID-19 has only made implementing data-driven strategies, which give current and accurate insights we can trust, even more important.

At Refinitiv we appreciate all of this is easier said than done. Over the past two years we have been expanding and scaling the impact of the work we do in Refinitiv Labs, fueling our own global innovation team. Our teams combine Refinitiv data, emerging technologies such as NLP, with close customer collaboration that fuels informed decision-making at speed.

Throughout this report, we offer practical advice, based on our experience developing and deploying AI/ML and deep learning models, and share updates on the new prototypes built by our teams in London, New York and Singapore.

I hope you find this an insightful and beneficial read.

DEFINITIONS

Artificial intelligence (AI): Machines performing cognitive functions we associate with humans, such as perceiving, learning and problem-solving.

Machine learning (ML): Machine learning algorithms are mathematical models based on sample data, known as “training data”, in order to make predictions or decisions without being explicitly programmed to do so.

Data science: A multi-disciplinary field that uses scientific methods, statistical data analysis, computer science and domain expertise to generate data insights and build machine learning models.

Deep learning: A family of machine learning approaches that uses artificial neural networks, with several layers that abstract the data so that features can be identified and complex classification tasks can be performed.

Natural language processing (NLP): NLP is a field dedicated to the harnessing of human language in programmatic ways, using linguistics, computer science and machine learning.

Supervised learning: Machine learning approach that provided inputs as labeled training data as the basis for predicting the classification of unlabeled data.

Unsupervised learning: Machine learning approach that looks for patterns in unlabeled data and with minimal human supervision.

Reinforcement learning: Machine learning approach that trains models to select outcomes through a continuous reinforcement loop such as “trial and error”.

METHODOLOGY

Objectives

The 2020 AI/ML survey was designed to examine the current market landscape for AI and ML, including:

- Level of AI/ML adoption in financial services, including key use cases, triggers and barriers to adoption
- What shapes financial firms' AI/ML strategy, and the impact of COVID-19
- Companies' data strategy, including how they search for data, make decisions about what to purchase, and what content they actually use
- The changing role and influence of data scientists in financial firms
- Platforms and tools used by the data science community
- A comparison with the findings of our 2018 AI/ML survey

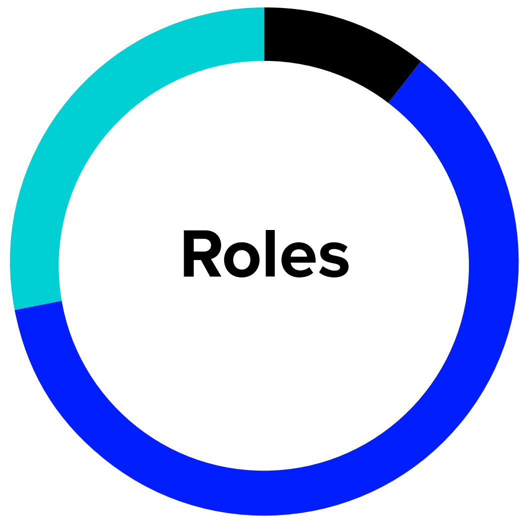
Approach[1]

The survey took place between June 29 and August 14, 2020, based on 423 telephone interviews.

Respondents included:

- Data scientists, quants, technology and data decision-makers
- A combination of sell-side and buy-side firms with revenue in excess of $1 billion
- A mix of geographies across the Americas, EMEA and the Asia-Pacific regions
- Quants and wealth management as new job and organizational types added to the 2018 sample

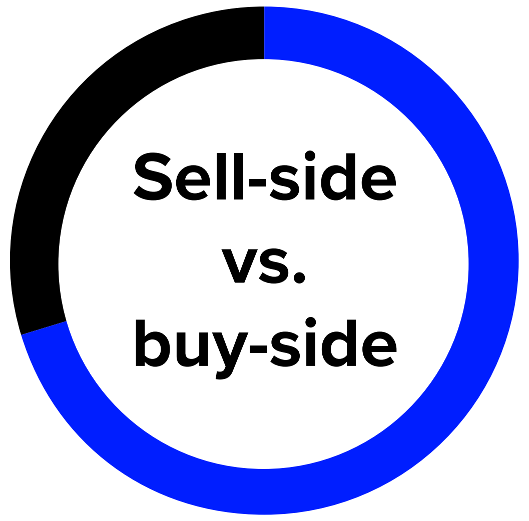
Breakdown of survey respondents



■ **260**
Data scientists
Data engineer
Data analyst
Head of innovation
Head of AI
Innovation manager

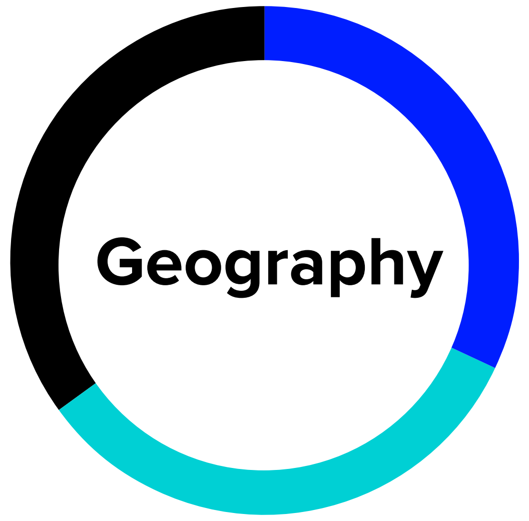
■ **118**
Quants
Quant analyst
Quant developer
Quant researcher

■ **45**
C-level
Data officer
Information officer
Technology officer



■ **298**
Sell-side
Commercial or retail bank (136)
Investment bank (87)
Broker-dealer (72)
Exchange (3)

■ **125**
Buy-side
Asset manager (78)
Hedge fund (10)
Wealth management (30)
Private equity/venture capital (7)



■ **147**
Asia-Pacific
India (30)
Singapore (20)
Australia (20)
New Zealand (15)
Korea (14)
China (23)
Japan (25)

■ **140**
EMEA
UK (49)
France (34)
Germany (31)
Switzerland (26)

■ **136**
Americas
U.S. (82)
Canada (54)

[1] Percentages in the charts drawn from the survey results may not sum up to 100% because of rounding.

KEY FINDINGS

1 Firms scale AI/ML capabilities across multiple business areas

2 As AI/ML models mature, more data scientists are recruited and deployed across firms

3 Data scientists evolve from a supportive function to driving strategy

4 Advances deploying models to production make AI/ML more of a reality than “hype”

5 Firms turn to deep learning, with implications for hardware, cost optimization and AI/ML explainability

6 Natural language processing (NLP) unlocks value in unstructured data

7 Data quality and availability are considered the biggest barriers to AI/ML adoption, as talent, technology and funding issues fade

8 Firms’ AI/ML models will only be as good as their data strategies

9 COVID-19 upset AI/ML models, and is set to drive up investment in firms with high levels of AI/ML maturity

10 AI/ML models need alternative data to be ready for more black swans

STRATEGY: FIRMS SCALE AI/ML ACROSS MULTIPLE BUSINESS UNITS

AI/ML is clearly maturing

Nearly three-quarters (72%) of respondents in our AI/ML survey say that it is a core component of their business strategy, and 80% claim they are making significant investments in the technology.

Figure 1.1: Level of AI/ML adoption

On a scale of 1 to 10, where 1 is strongly disagree and 10 is strongly agree, how much do you agree?
Base: all respondents (423)

72% state AI/ML is a **core component of their business strategy**

80% state they are making **significant investments in AI/ML** technologies/techniques

70% agree that **decision making** about AI/ML takes place across **multiple parts of the business**

Source: AI/ML survey, August 2020

AI/ML use cases become more diverse

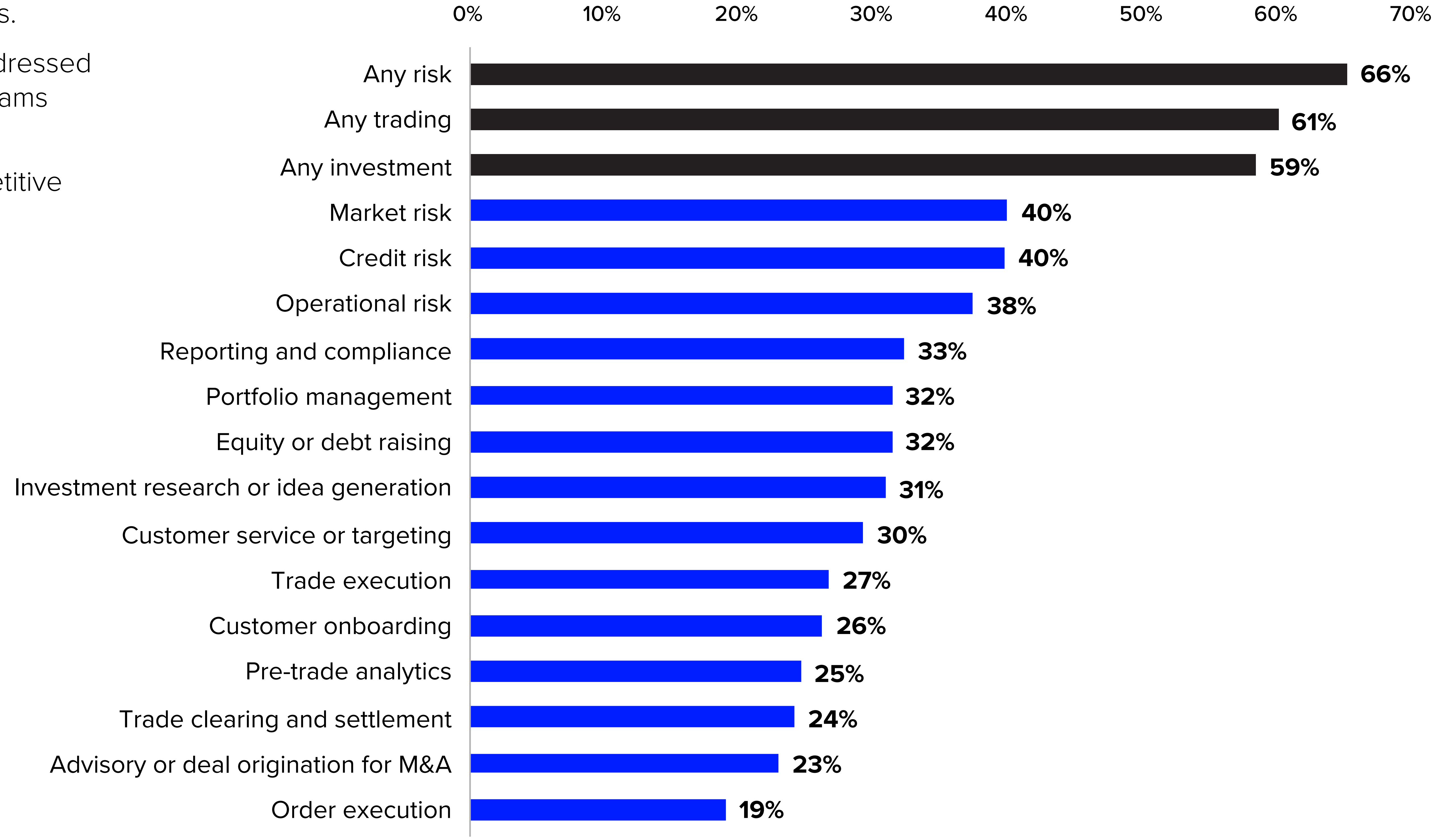
It is also clear that data-driven business strategies need AI/ML at scale. This year's survey shows that technology has advanced to become a horizontal capability available across the business.

This is visible in the plethora of business use cases now addressed by AI/ML, but also in the way that AI/ML and data science teams are structured – 64% of teams now work in business units.

Business units will increasingly rely on AI/ML to drive competitive advantage and manage risk, rather than an afterthought the technology team implements to increase efficiency.

Figure 1.2: Key areas of focus for AI/ML deployment

Which of the following are focus areas in terms of applying AI/ML techniques?
Base: all respondents (423)



Any risk = credit risk; market risk; operational risk
Any trading = trade execution; pre-trade analytics; trade clearing and settlement; order execution
Any investment = investment research or idea generation; portfolio management; advisory or deal origination for M&A

Source: AI/ML survey, August 2020

Sell-side set to scale faster than buy-side

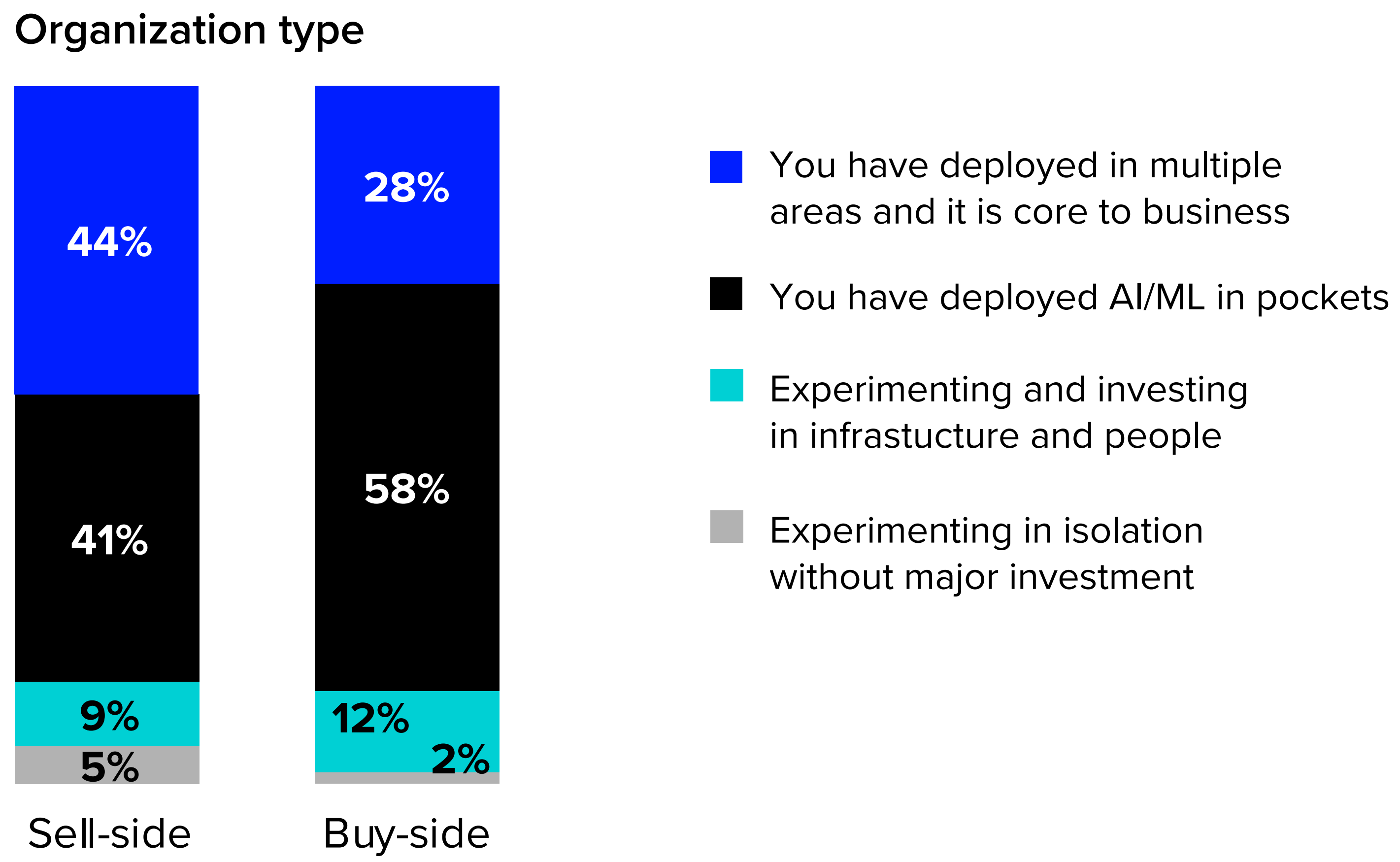
The results of our 2020 survey show the number of teams using AI/ML is increasing significantly. The sell-side is using AI/ML in multiple business areas, and the buy-side is deploying AI/ML into pockets of the business.

Figure 1.3: Level of AI/ML adoption: Sell-side vs. buy-side

Which of the following describes the adoption of AI/ML technologies/techniques to manage or analyze data or content within your organization?

Base: sell-side: 281; buy-side: 120

**Percentages may not sum up to 100% because of rounding.*

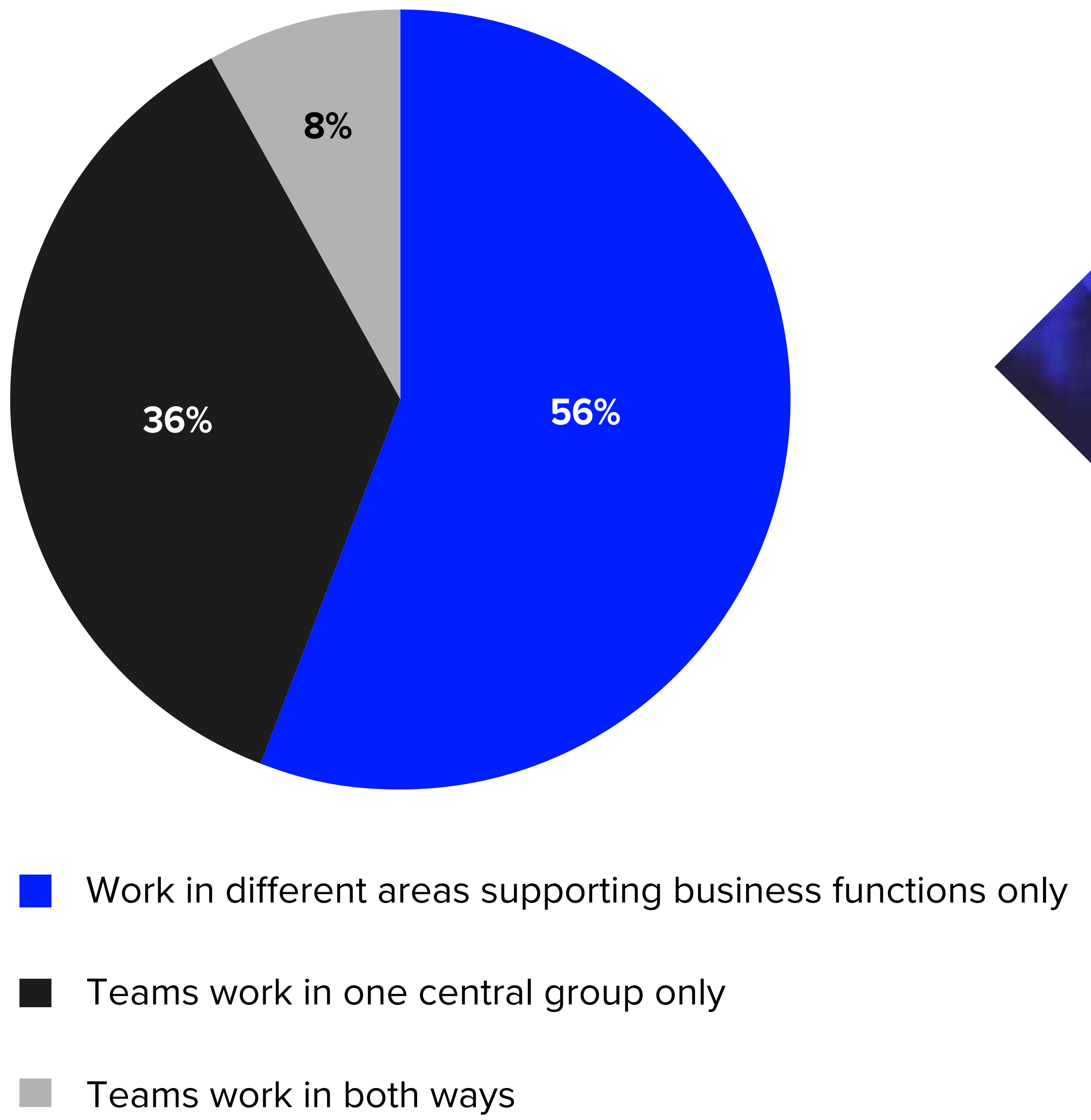


Source: AI/ML survey, August 2020

Figure 1.4: Where do data scientists work?

In which of these ways do data science teams work in your organization?

Base: all respondents (423)



Source: AI/ML survey, August 2020

Looking at who uses what, there are more firms on the sell-side focusing on risk, compliance and trade execution, compared to the buy-side.

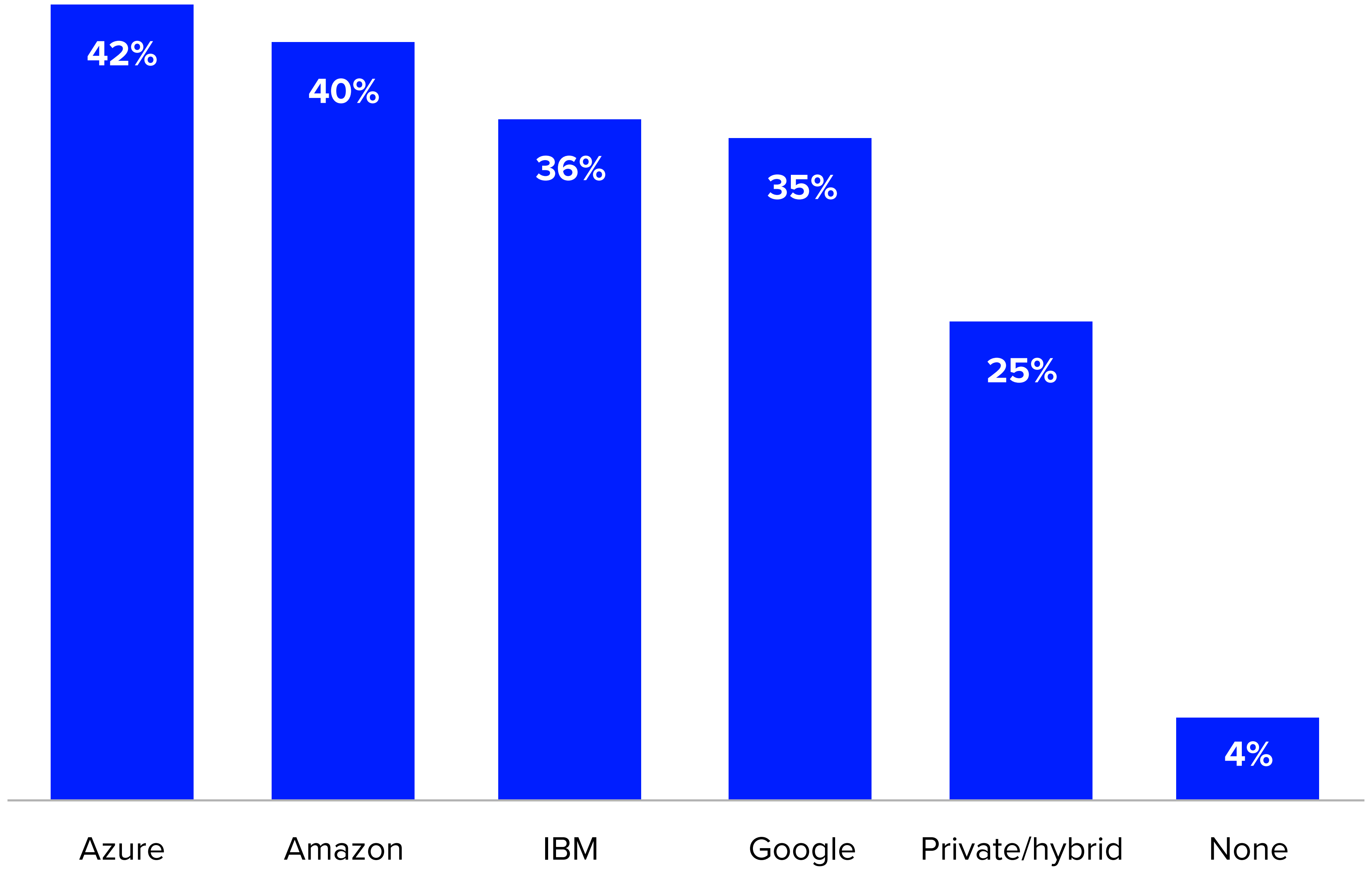
Financial firms that have implemented one or two use cases are less likely to focus on risk and more likely to deploy AI/ML to customer onboarding and targeting, equity and debt raising and portfolio management.

Scaling AI/ML makes high-performance computing essential

To ensure the resilience of AI/ML and the ability to scale, most financial services firms use more than one cloud provider to run their models. Amazon and Microsoft® Azure are favorites in the Americas, while Google and Azure are somewhat more popular among sell-side organizations compared to the buy-side. Where there are more than six use cases, Amazon and Azure have become even more widespread.

Figure 1.5: Cloud providers used to run AI/ML models

Which, if any, cloud providers do you use to run models?
Base: all respondents (423)]



Source: AI/ML survey, August 2020

Cloud providers are increasing their computing power to meet the requirements of AI/ML. AWS and Microsoft are adopting AI-optimized hardware, such as Graphcore, that increases resilience and computing power.

Migration to the cloud accelerated in the wake of COVID-19, highlighting the limitations of existing systems and the need for agile, cloud-based solutions.

> **"Cloud providers and hardware specialists have developed next generation compute to meet the demanding requirements of AI/ML. Google and Nvidia, as well as new entrants such as Graphcore and Cerebras, offer AI-optimized hardware to enable the training of large, high-performing machine learning models."**

Laura Sartenaer
Strategy and Partnerships Manager
Refinitiv Labs, EMEA

KEY TAKEAWAYS

Scale effectively by keeping the end goal in mind

- Make sure you have an end-to-end deployment path for AI/ML projects, then begin to iterate with simple models. These may not even contain AI/ML, but you want to ensure that you can deploy basic analytics before spending budget on computing costs
- Expect to re-train and re-deploy multiple times, continuously checking and refining as things change
- Having flexible infrastructure and a test environment that allows for this continuous iteration is critical

Gauge your AI/ML maturity before scaling in the cloud

- If your AI/ML maturity is low, start small with on-premises machines, or with packaged AI/ML services from cloud providers. Move to build on top of industry advances, such as Google's pre-trained BERT language models, rather than create from scratch. Iterating existing models can save you from spending a significant proportion of your research budget on pre-training and compute
- As your maturity increases, consider partnering with your cloud provider to develop a cost-optimized approach, i.e., what hardware will you use for training vs. running models? As you scale your training, you will want to use different services

Enable data scientists to learn about the industry they are solving for

- Partnering with domain experts allows data scientists to understand business goals, error tolerance and relevant data
- For example, when Refinitiv Labs was building Project SentiMine, a discoverability tool for unstructured data, data scientists collaborated with target research analyst users on data annotation
- This partnership ensured SentiMine generated accurate sentiment signals for equity performance across 110 themes covering accounting, business drivers, valuation, economics, management change, key risks and ESG

TALENT: DATA SCIENTISTS ADOPT A MORE STRATEGIC ROLE

Data scientists have transitioned from developing and implementing AI/ML models at the request of the business, to influencing the technology and data strategies required to achieve business objectives

"There is very little that business users can understand about what we are doing, and it is very logical that something you don't understand, you don't believe in. While investment was initially low and the biggest challenge when we started, it is no longer an issue," explains a senior data and analytics professional at a credit services company in India.

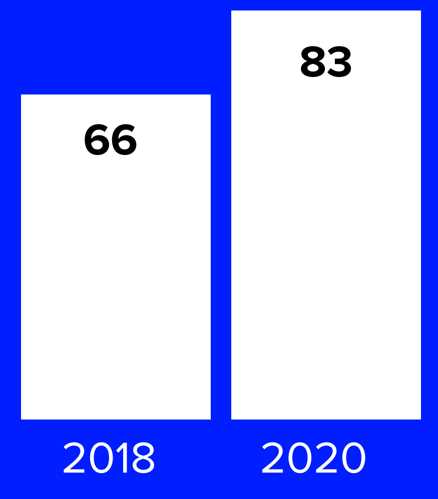
The number of data scientists employed influences AI/ML adoption, which is highest in firms with 26 to 50 data scientists.

The importance of data scientists to the success of AI/ML strategies is also evidenced in this year's research. The average number of data science roles in each financial firm increased by 26%, from 66 in 2018 to 83 in 2020. Looking forward, 35% of the firms surveyed expect the number of data scientists hired to grow over the next 12 months.

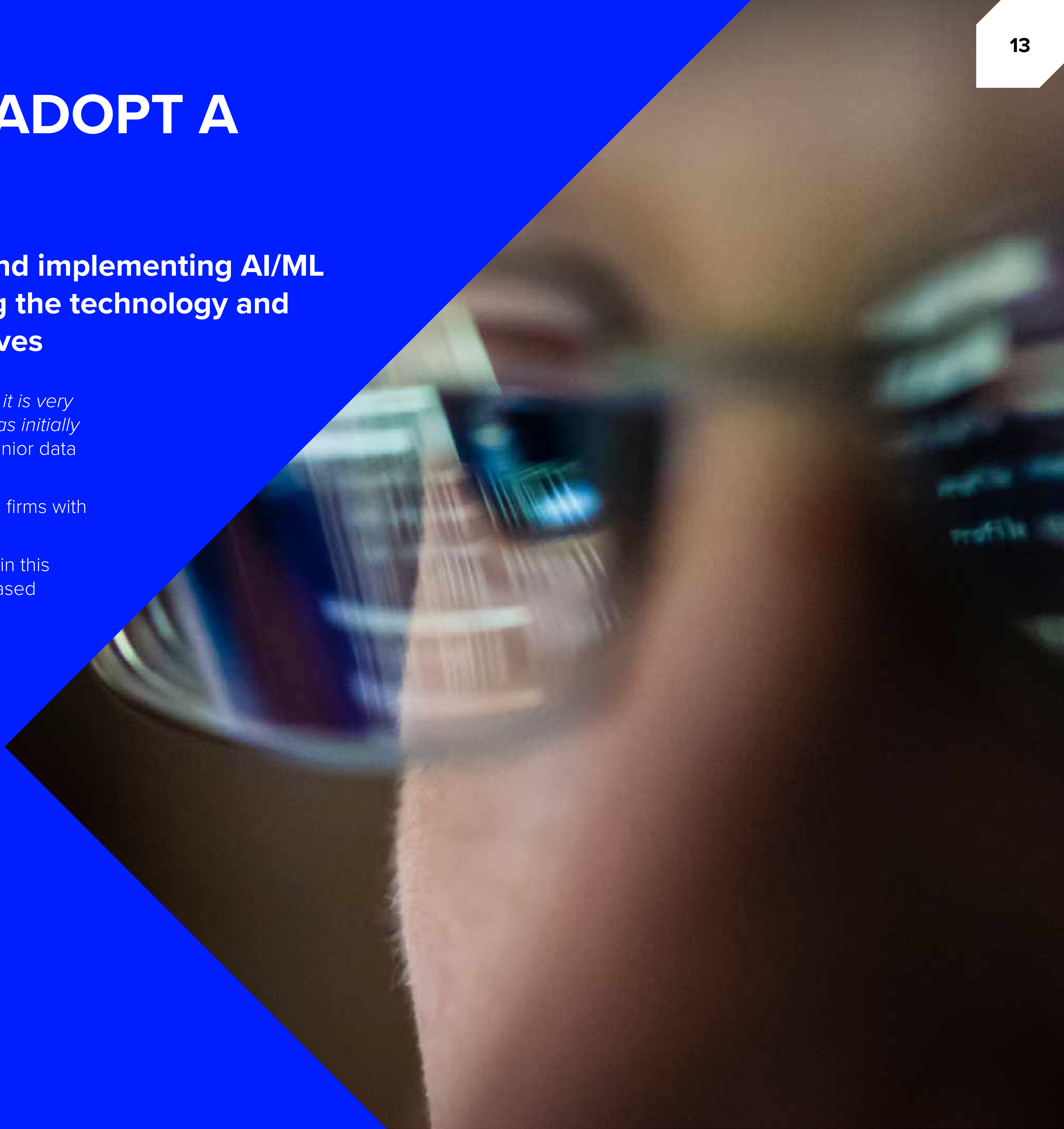
Figure 2.1: Average number of data science roles per company

How many employees are involved in data science activities as part of their role in your company?

Base: all respondents (2018: 447; 2020: 397)



Source: AI/ML survey, December 2018; August 2020

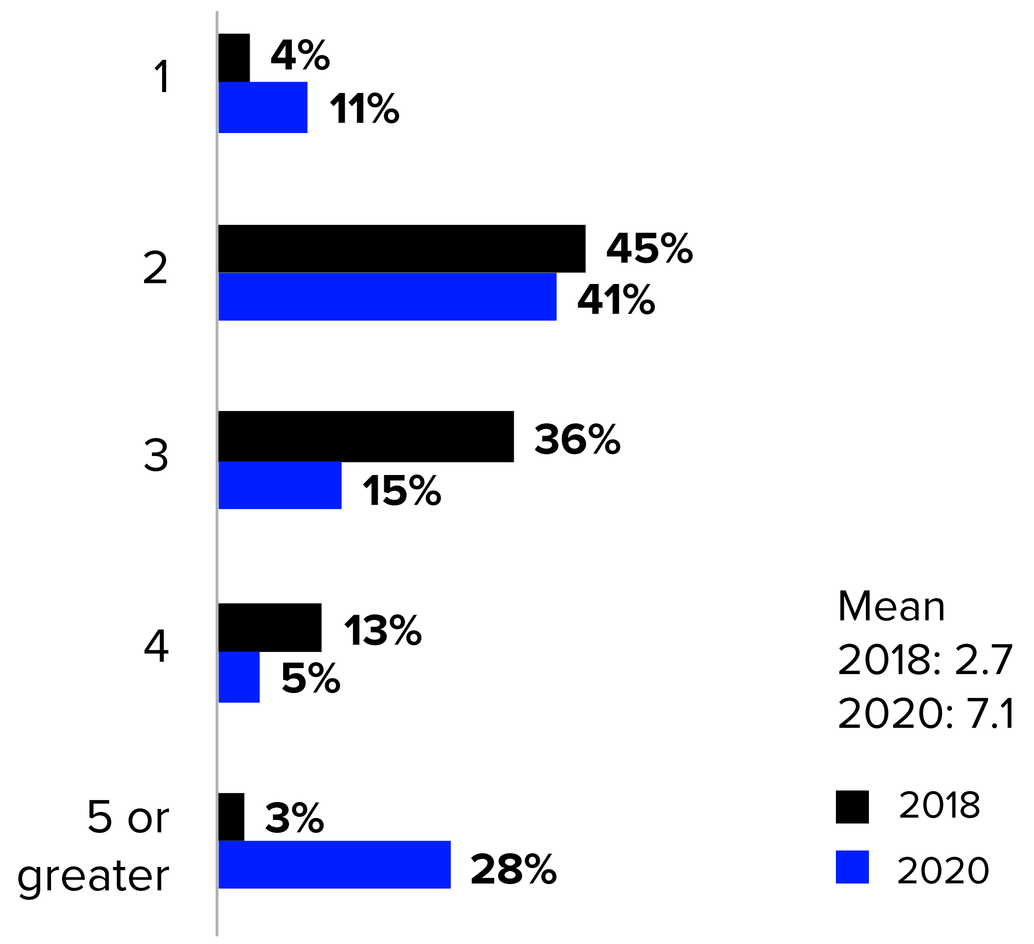


A surge in data science teams and roles

Figure 2.2: Number of data science teams per company

How many data science teams are there in your company?

Base: all respondents (2018: 447; 2020: 400)

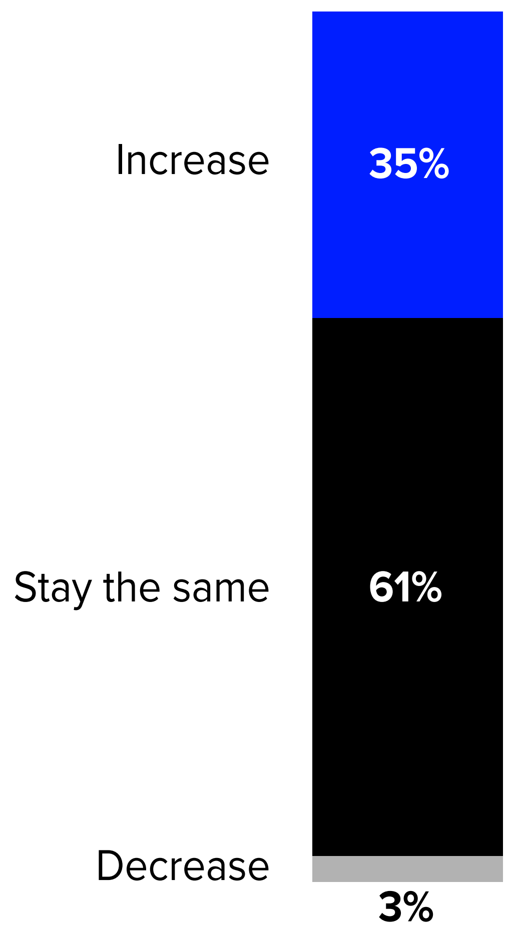


Source: AI/ML survey, December 2018; August 2020

Figure 2.3: Change in number of data science employees in the next 12 months

To the best of your knowledge, will the number of data science roles in your company increase, decrease or stay the same in the next 12 months?

Base: all respondents (423)



Source: AI/ML survey, December 2018; August 2020

> **“People strategy is very important. You need people who have the right skills, who have deep engineering skills, who have the right business knowledge, and who can stitch everything together.”**

Senior data scientist at an international investment bank in Singapore

Who makes the data decisions?

This year's survey also highlights that data scientists are strategically important to data procurement strategies, as they know what business value different data sets can be deliver.

Data scientists are key influencers in deciding which data sets to trial, although decisions on buying data are more likely to fall into the domain of procurement.

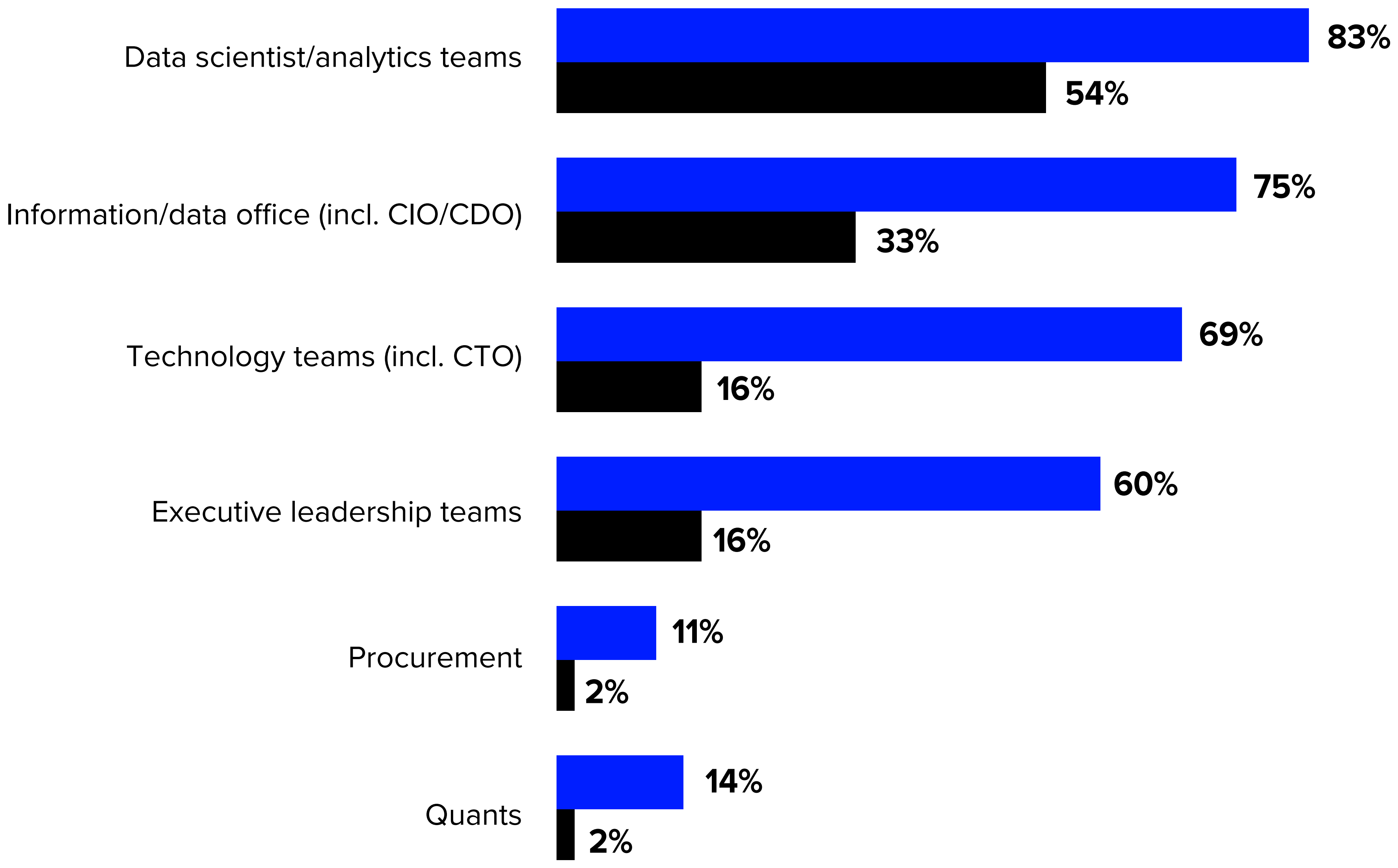
"For traditional machine learning, over the next year we're going to continue to see great advancement in automation," shares a chief data scientist at a U.S. university. *"There will still be a role for data scientists doing feature engineering, working with data, so a lot of work is going to be in data curation and serving as problem solvers. In terms of model training and deployment, I see tools getting better and better."*

Data scientists' increasing influence on data procurement

Figure 2.4: Decision makers for trialing data

In your organization, which teams/functions are involved in the decision/make the final decision to trial data sets you would use with machine learning?

Base: all respondents (423)

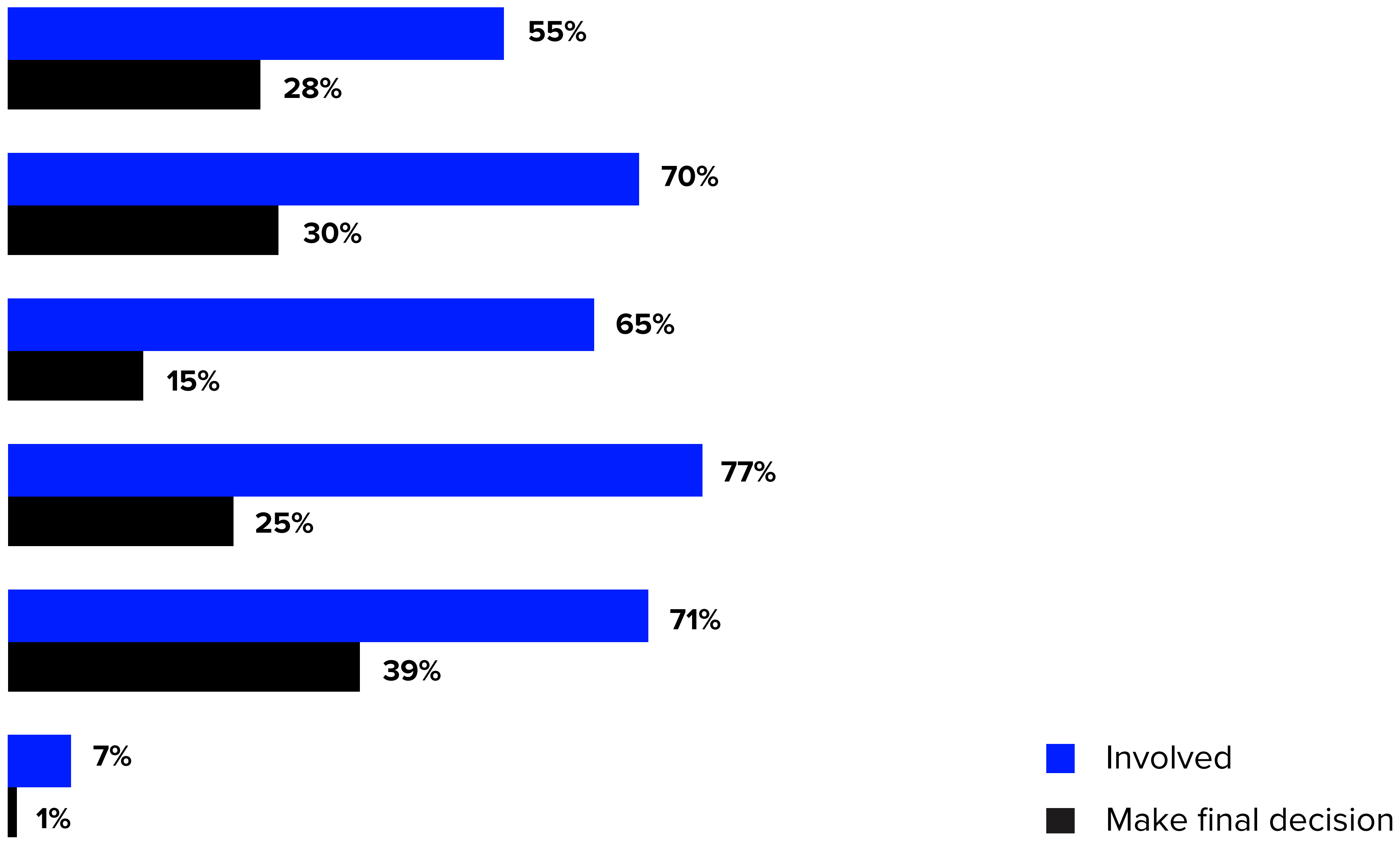


Source: AI/ML survey, August 2020

Figure 2.5: Decision makers for buying data

In your organization, which teams/functions are involved in the decision/make the final decision to buy data sets you would use with machine learning?

Base: all respondents (423)



Source: AI/ML survey, August 2020

KEY TAKEAWAYS

Build stakeholder trust by explaining probability and risk

- Ensure your business partners understand the consequences and trade-offs of AI/ML – namely that either a) precision or b) recall will have to be prioritized for the project
- For example, you would not want to miss any data being screened for anti-money laundering (AML) (recall), but will have to sacrifice precision in the process
- On the other hand, you could be looking for precise signals screening news data, and sacrifice recalling every single relevant news story

Always let data scientists try before someone else buys

When a data scientist assesses pricing data, as an example, they will examine:

- The data when the market is open and outside of trading hours
- Whether the data overwrites
- If data attributes are classified
- Whether attributes have useful meanings that can be interpreted (Was it a trade? Normal, opening or closing auction?)
- How is missing data handled (Is it 'N/A' or classified?)
- How data changes over time, and much more

These are factors critical to the success of AI/ML, and features that only data scientists would be looking for.

Check out these communities for data scientists in finance

Here are some useful groups and events in finance, recommended by the team at Refinitiv Labs:

London Quant Group

AI and Data Science in Trading

The Python Quants

Refinitiv Developer Community

RE.WORK AI and Deep Learning events

INQUIRE

Data Council

Data Science SG

TECH: AI/ML IS MORE REALITY THAN "HYPE"

In 2018, 75% of firms were making significant investments in AI/ML technologies and techniques. Based on our 2020 findings, companies continue to invest heavily in AI/ML, and we are also seeing the results of this funding.

Firms turn to deep learning

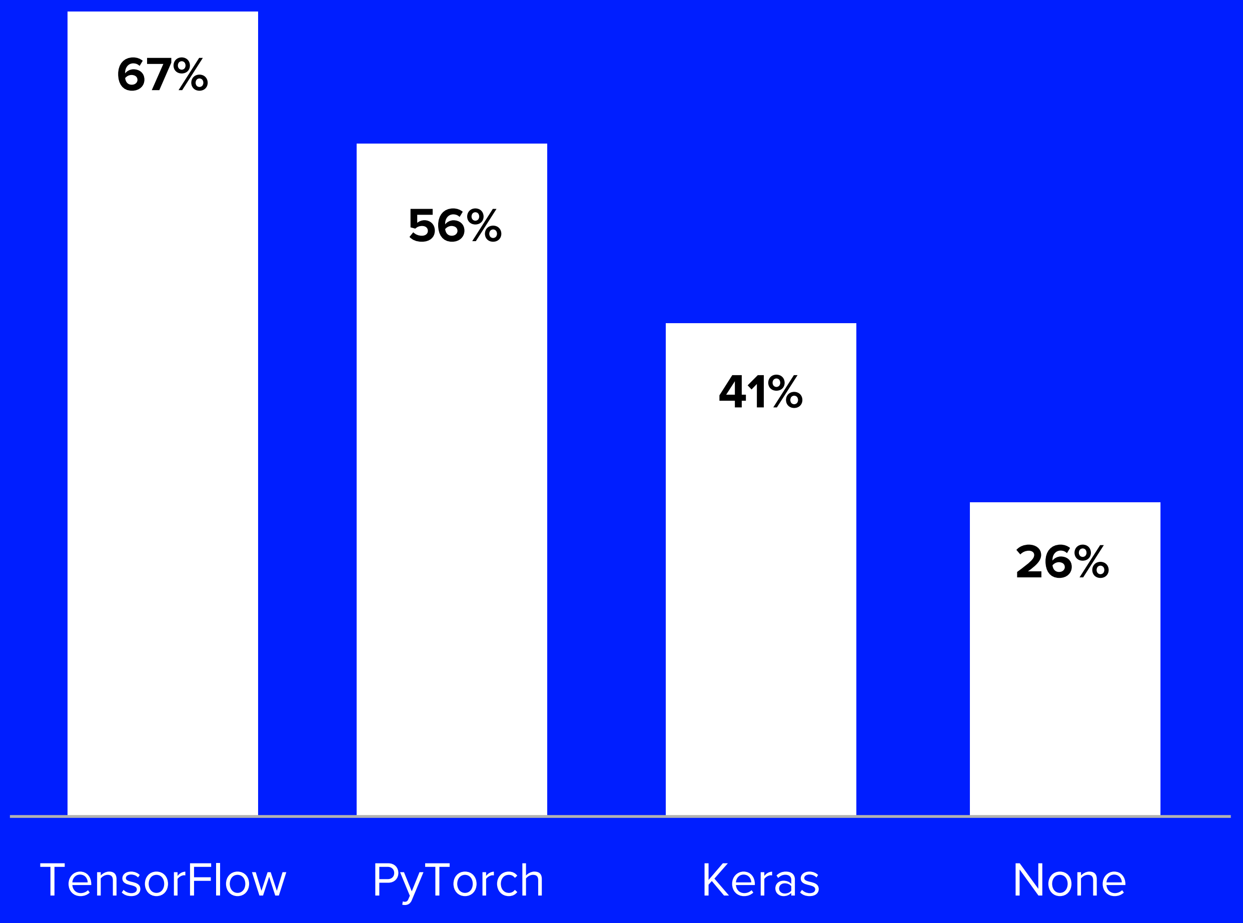
A striking result of this year's research is that 75% of firms are using deep learning. This is a major and unexpected technology advance, considering that deep learning has previously been seen as niche, expensive and academic.

This finding is backed by a surge in the application of leading deep learning frameworks, including TensorFlow from Google and PyTorch from Facebook. TensorFlow is used more by the buy-side than the sell-side, with commercial banks showing the greatest uptake.

Figure 3.1: Deep learning frameworks

Which deep learning framework are you using?

Base: respondents using deep learning (316)



Source: AI/ML survey, August 2020

“We use TensorFlow and PyTorch in Refinitiv Labs to build AI/ML models, so it is not surprising to see their growing popularity in the data science community.

What is surprising is the percentage of firms using deep learning, which takes time and resources to get right, but can deliver exceptional results.”

Geoff Horrell
Head of Refinitiv Labs, EMEA

As progress is made from AI/ML to deep learning, it becomes ever more critical to address concerns around data quality. Deep learning is gaining traction to derive insights from large and diverse unstructured data sets such as text, voice and video, and its more widespread deployment also comes with implications for hardware, cost optimization and AI/ML explainability.

“I think the ethics of using AI/ML models without human intervention will be a big challenge. That is where explainability comes into play,” shares a data scientist at an insurance firm in the UK. *“It should be a requirement to have all models audited or be able to understand why they’re making the predictions or why they’re doing what they’re doing, to understand the in-between layers.”*

Natural language processing set to unlock value in unstructured data

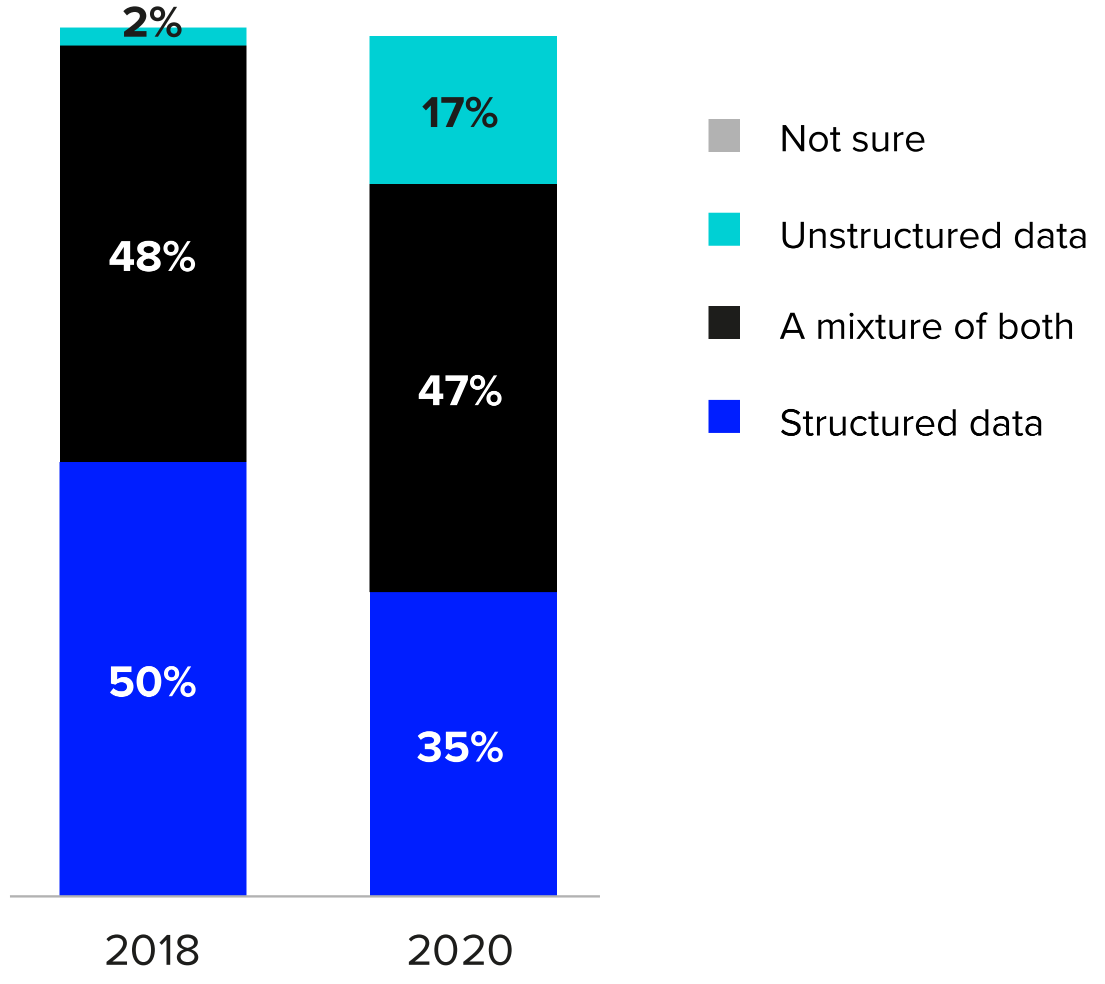
Management of unstructured data, such as text, images or certain types of alternative data, is a challenge for many firms today, but AI/ML technology advancements have enabled data scientists to drive more business value from these types of data sources.

The use of unstructured data has increased, with 17% of firms saying they only use this type of data, up from just 2% in 2018. This growth reflects the power of AI/ML models fed with unstructured data to provide new signals to the business and revenue to the company.

Figure 3.2: Structured vs. unstructured data usage

Do you personally work with structure data, unstructured data or a mixture of both?

Base: all respondents (2018: 447; 2020: 423)



Source: AI/ML survey, December 2018; August 2020

> **“We have a couple of data scientists who want to start tracking weather patterns and bringing in weather data on top of just news data to come up with alternative trading strategies that could help drive revenue for the firm.”**
>
> Managing director in IT at an asset management firm in the U.S.

Natural language processing (NLP) is a subfield of AI that programs computers to process and analyze large amounts of text and voice data. NLP is increasingly seen as a viable technique to extract value and gain insights from large volumes of unstructured data.

As 80 to 90% of all data in the digital “universe” is unstructured, NLP can help generate new insights, including the ability to conduct sentiment or risk analysis, recommendation engines and automated alerting.

However, to deliver precision, NLP use cases must be domain specific. A model built on news data will not transfer to transcripts. Data science teams have a difficult balance to strike – they must solve for specific use cases, build accurate models, handle data inputs, manage training and communicate their work to business users.

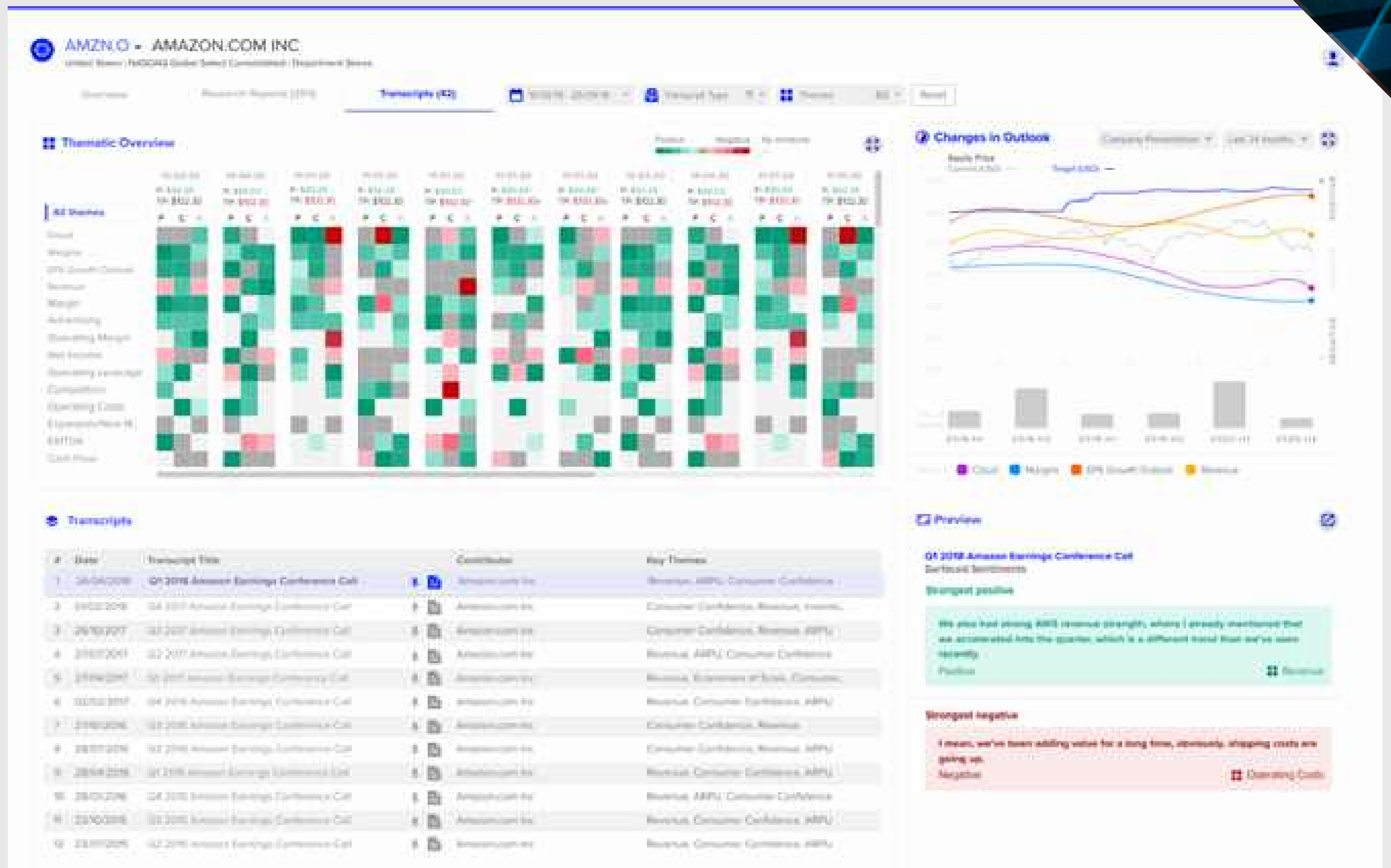
Project SentiMine: Surfacing equity performance themes in unstructured content

A recent NLP success story is the SentiMine prototype built by the Refinitiv Labs team in Singapore.

By combining deep learning, sentiment analysis and NLP, SentiMine surfaces equity performance themes and contrarian views from thousands of research reports and earnings call transcripts, in real time.

This innovation will save buy-side firms millions of staff hours combing through unstructured text to drive asset management and investment decisions.

Refinitiv Labs is working to deploy SentiMine’s unique insights in Refinitiv Eikon next year.



#	Date	Transcript Title	Contributor	Key Themes
1	26/04/2018	Q1 2018 Amazon Earnings Conference Call	Amazon.com Inc	Revenue, ARPU, Consumer Confidence
2	[illegible]	Q4 2017 Amazon Earnings Conference Call	Amazon.com Inc	Consumer Confidence, Revenue, Investm...
3	26/10/2017	Q3 2017 Amazon Earnings Conference Call	Amazon.com Inc	Consumer Confidence, Revenue, ARPU
4	27/07/2017	Q2 2017 Amazon Earnings Conference Call	Amazon.com Inc	Revenue, ARPU, Consumer Confidence
5	27/04/2017	Q1 2017 Amazon Earnings Conference Call	Amazon.com Inc	Revenue, Economies of Scale, Consumer...
6	02/02/2017	Q4 2016 Amazon Earnings Conference Call	Amazon.com Inc	Revenue, Consumer Confidence, ARPU
7	27/10/2016	Q3 2016 Amazon Earnings Conference Call	Amazon.com Inc	Consumer Confidence, Revenue
8	28/07/2016	Q2 2016 Amazon Earnings Conference Call	Amazon.com Inc	Consumer Confidence, Revenue, ARPU
9	28/04/2016	Q1 2016 Amazon Earnings Conference Call	Amazon.com Inc	Revenue, Consumer Confidence, ARPU
10	28/01/2016	Q4 2015 Amazon Earnings Conference Call	Amazon.com Inc	Revenue, ARPU, Consumer Confidence
11	22/10/2015	Q3 2015 Amazon Earnings Conference Call	Amazon.com Inc	Revenue, Consumer Confidence, ARPU
12	23/07/2015	Q2 2015 Amazon Earnings Conference Call	Amazon.com Inc	Revenue, Consumer Confidence, ARPU

Q1 2018 Amazon Earnings Conference Call
Surfaced Sentiments

Strongest positive

We also had strong AWS revenue strength, where I already mentioned that we accelerated into the quarter, which is a different trend than we've seen recently.
Positive — Revenue

Strongest negative

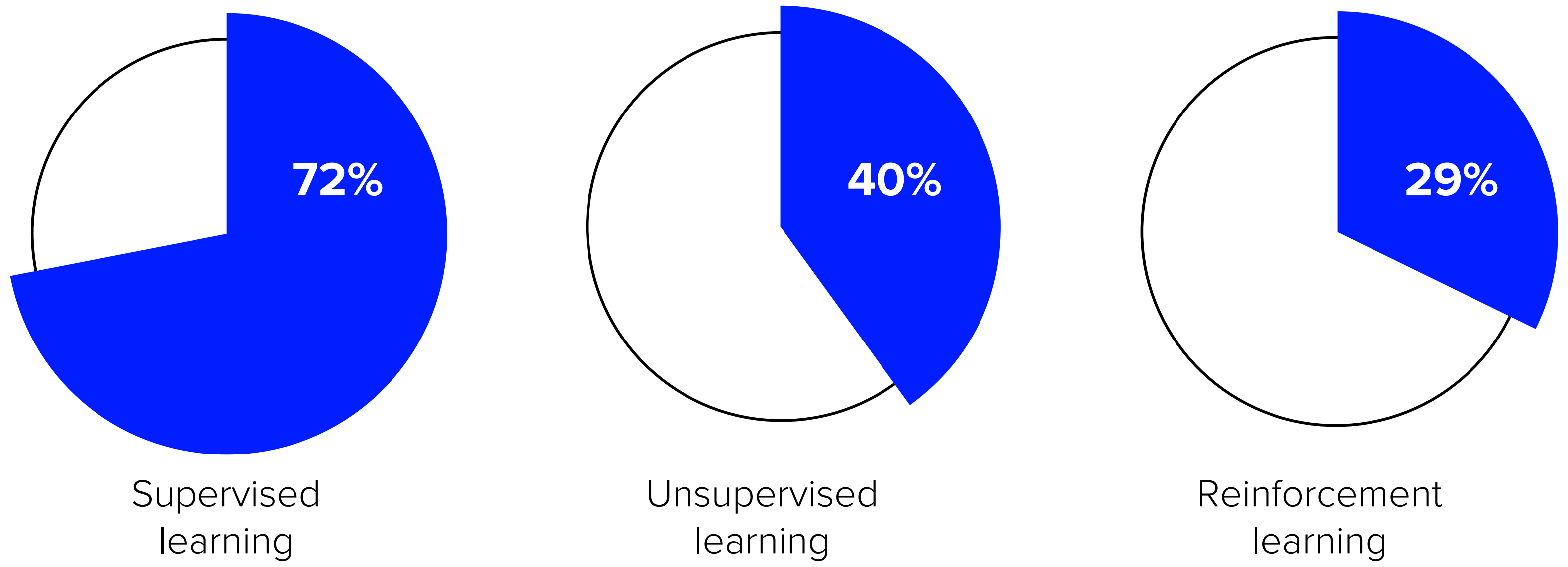
I mean, we've been adding value for a long time, obviously, shipping costs are going up.
Negative — Operating Costs

Increased focus on unsupervised learning – but supervised reigns supreme

In terms of AI/ML tasks, supervised learning is most widely applied, ahead of unsupervised learning and reinforcement learning, which does not use a training data set and makes sequential decisions based on learning from its experience.

Figure 3.3: Types of AI/ML

Which types of machine learning do you use?
Base: all respondents (423)



Source: AI/ML survey, August 2020

> **"Setting up AI/ML capabilities has become less of a problem, as talent and investment that were previously difficult to secure are more easily available this year. The maturity curve has moved up since our 2018 survey and we are now putting plans into production rather than creating capabilities."**

Sanjna Parasrampuria
Head of Refinitiv Labs, Asia

The Americas are most likely to use supervised learning, while the types of methods applied increase in parallel with the number of use cases deployed.

65% of financial firms use one type of AI/ML

28% of financial firms use two types of AI/ML

7% of financial firms use three types of AI/ML

Working with different types of AI/ML methods makes for more robust, cross-technique corroborated insights. Models can be layered to discover more insights, and if one fails during an unforeseen event, another can make sure the financial firm is not left exposed.

Which types of machine learning do you use?
Base: all respondents (423)
Source: AI/ML survey, August 2020

KEY TAKEAWAYS

Keep an eye on tomorrow's data science platforms

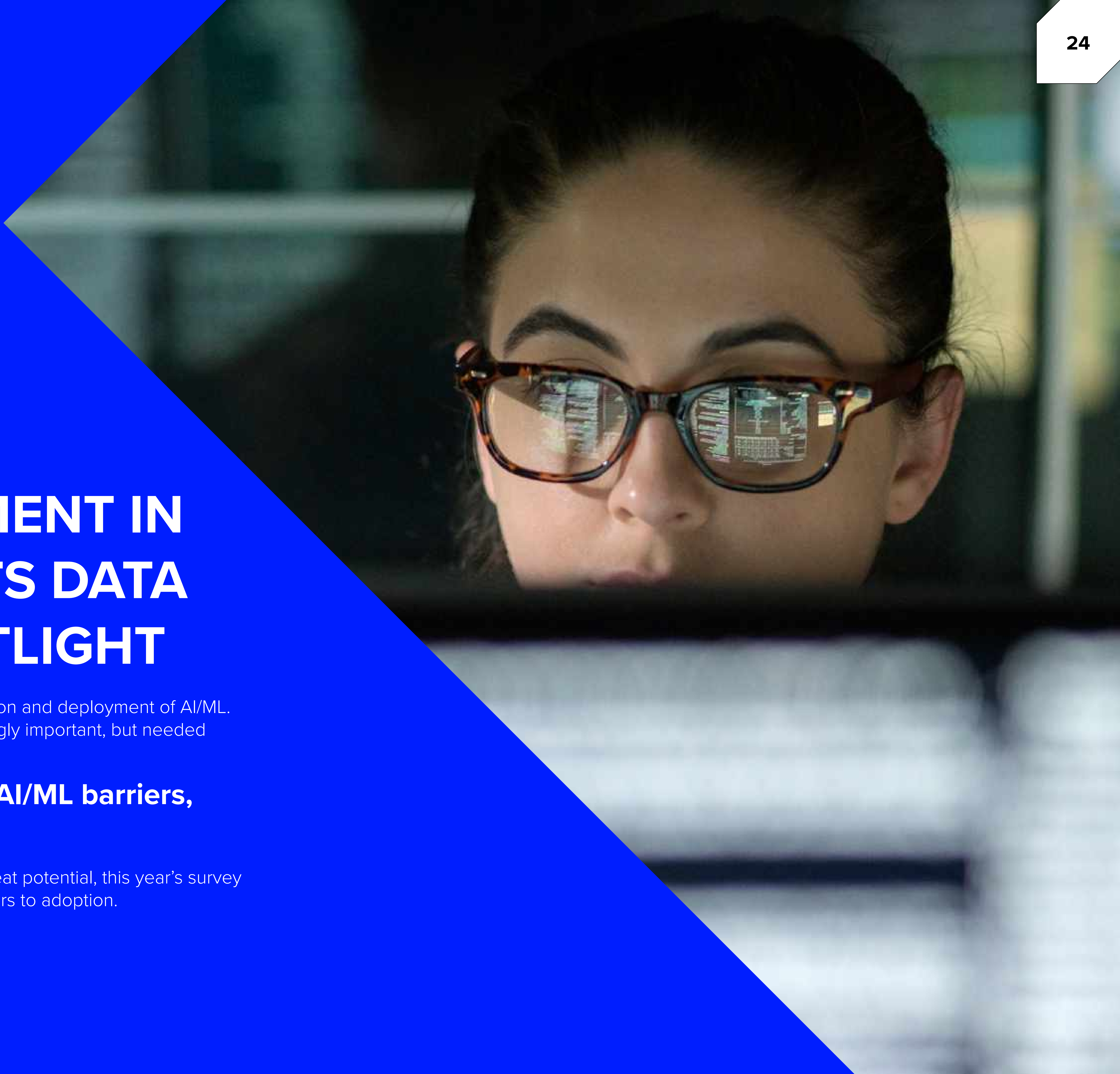
- Data science platforms that offer AutoML and facilitate the work of citizen data scientists could become commonplace in 2021. Citizen data scientists are able to perform simple or moderately sophisticated analytical tasks but their primary job function is outside the field of data science
- A good example is the H20.ai platform, which analyzes data, runs a wide range of iterations, and suggests a model and which parameters perform best
- Platforms will certainly help firms trying to scale citizen data scientist capabilities, by identifying a good AI/ML tool for the job at hand. However, they won't assist with refining models, or building them from scratch

Use NLP to extract value from data you already have

- There have been significant developments in NLP that are helping to extract insights from human language
- If you are new to the discipline, find the critical information that you need to extract with NLP, which is often entities such as companies, organizations and people, and start pulling out this key information to create a structure
- These elements will help you connect to other data sets and ensure your analysis (i.e., sentiment) relates back to specific entities

Explainable AI is going to be a key challenge in 2021

- As models become more complex and handle larger and more versatile data sets, a whole new "explainability" field is emerging to understand how AI/ML predictions and decisions are made
- Explainable AI seeks to understand, for example, what models are sensitive to, whether they use meaningful features for classification, and which weak points need improvement
- Those interested in this fast-emerging field should look up the explainable AI method LIME (Local Interpretable Model-agnostic Explanations) and exBERT – a visual analysis of transformer models



CHALLENGES: INVESTMENT IN TECH AND TALENT PUTS DATA STRATEGY IN THE SPOTLIGHT

In our 2018 survey, poor quality data was the biggest barrier to the adoption and deployment of AI/ML. Unstructured data, as well as data from alternative sources, was increasingly important, but needed more work before insights were truly reliable.

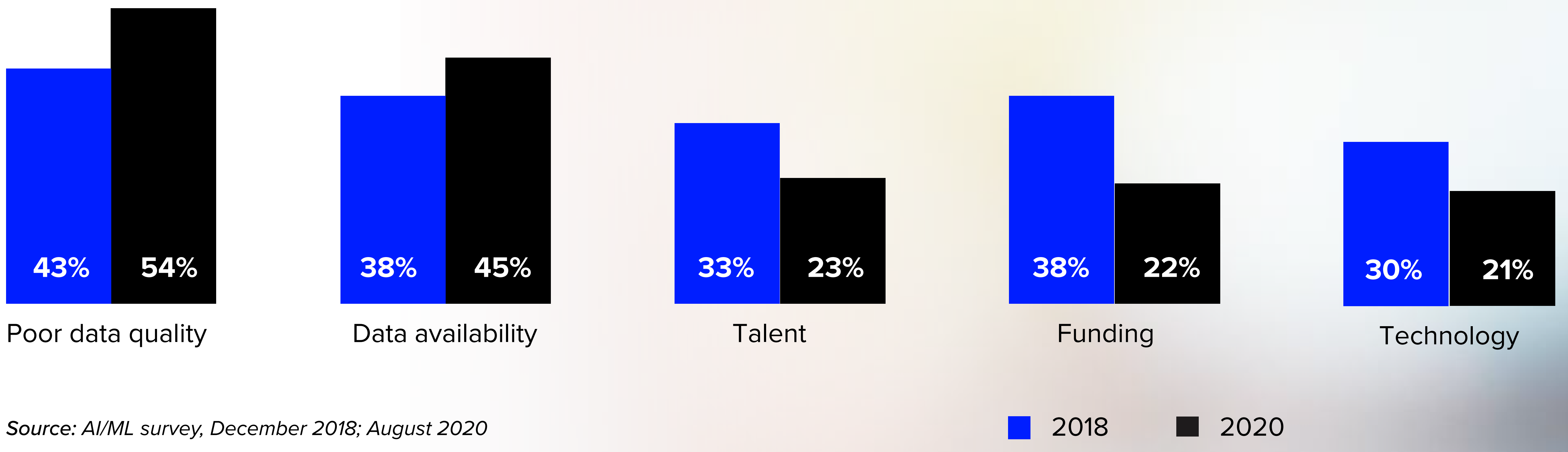
Data quality and availability are the biggest AI/ML barriers, as tech, talent and funding issues fade

While AI/ML projects using high-quality and accessible data sets show great potential, this year's survey highlighted that poor data quality and availability continue to rise as barriers to adoption.

Figure 4.1: Barriers to AI/ML adoption 2018 vs. 2020

To what extent do you agree these are barriers to adopting new AI/ML technologies/techniques in the organization, where 1 means "does not apply at alll" and 10 means "completely applies". Percentages describe those who answered with a score between 7 and 10.

Base: all respondents (2018: 447; 2020: 420)



Source: AI/ML survey, December 2018; August 2020

While data quality remains the biggest challenge, issues related to talent, funding and technology seem to be fading. This may be a result of countless citizen data scientist applications and dedicated AI/ML model tools that cater to data scientist workflows around data management, model management and model training.

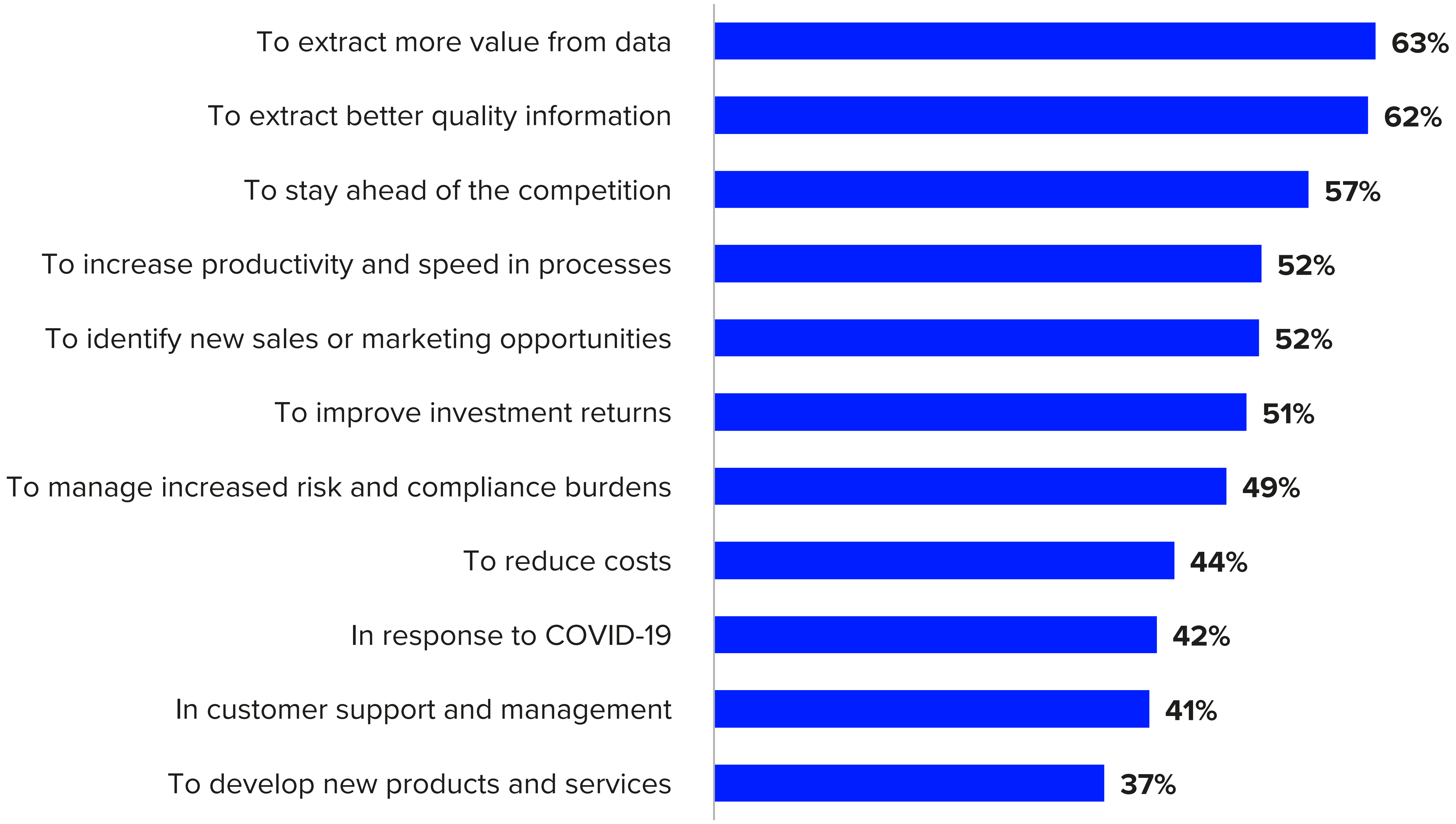
These innovations have been spearheaded by fast-growing start-ups like Databricks, Dataiku and DataRobot, and key cloud providers such as Amazon SageMaker, Microsoft Azure Machine Learning and Google Colab.

Considering a two-year horizon, our 2020 survey shows that the most important drivers of AI/ML will be extracting more value from data, extracting better quality information and staying ahead of the competition.

The days of "any AI is better than no AI" are over, giving firms with a data-driven AI/ML strategy the edge.

Figure 4.2: Which factors will become more important in the next one to two years?

Which factors will become more important in the next one to two years?
Base: All respondents (423)



Source: AI/ML survey, August 2020

“Scaling AI/ML requires an integrated data and technology strategy. The vast majority of the firms we interviewed for this year’s survey have data strategy as their top priority, which is a reassuring sign that firms are realizing the importance of the balance between technology, talent and data investment.”

Geoff Horrell
Head of Refinitiv Labs, EMEA

An AI/ML model is only as good as the data strategy supporting it

Given the barriers outlined by the survey, and as AI/ML models mature, firms are increasingly focusing on data quality and accessibility rather than just the technology investments that were key findings in our 2018 survey.

The change in focus is reflected in emerging use cases like alpha generation and trade execution, which enter a whole new level of interaction with the market and, as a result, rely heavily on high-quality and third-party data.

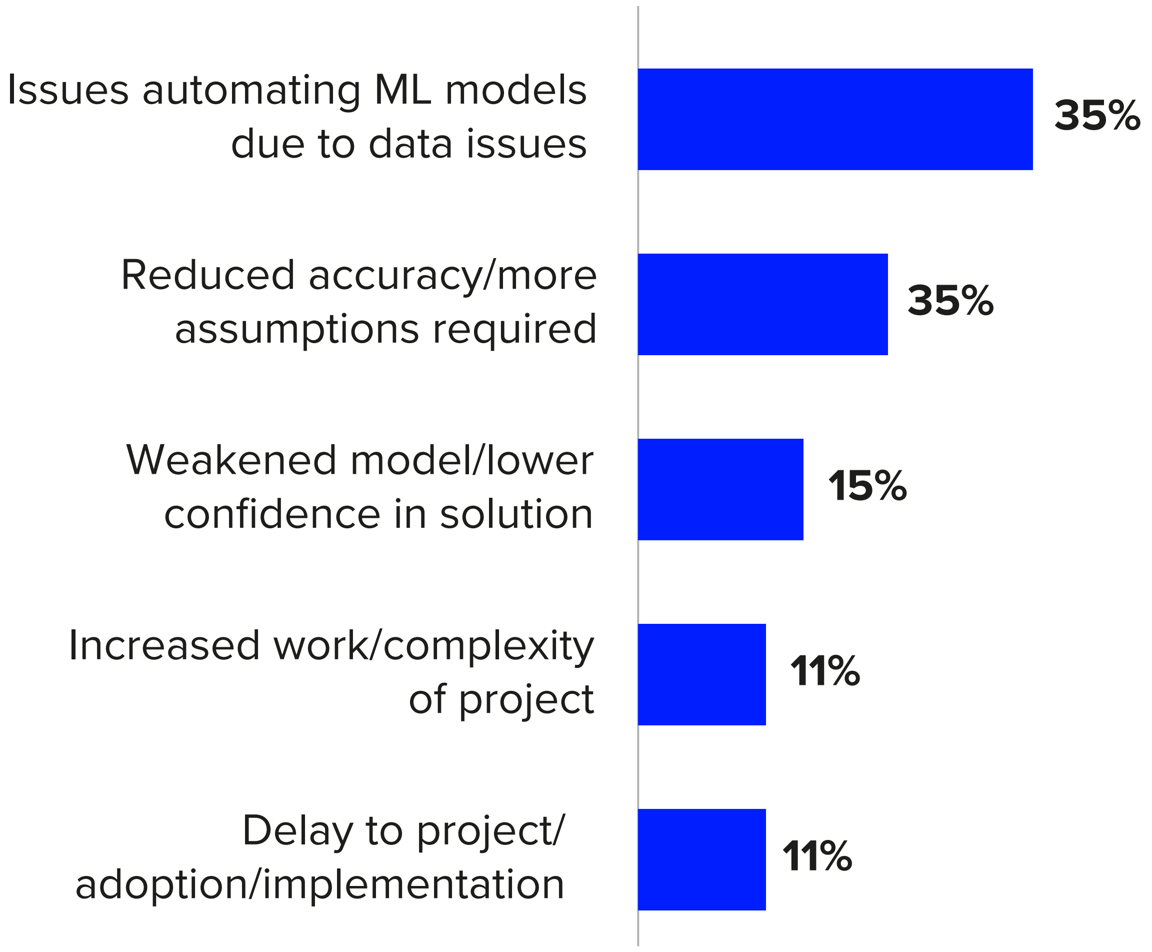
For example, to find unique investment opportunities, and to compete with the rest of the investment world in the process, you need:

- Unique data with enough history to prove your strategy works
- Diverse data sources to create new data combinations that your competitors will struggle to replicate
- The ability to join and link disparate data sets that may not be used in the market today

Figure 4.3: The impact of data quality

Can you describe instances where poor data quality has impacted your ability to deploy machine learning effectively? What data sets were you working with?

Base: all respondents who had a problem or problems with poor quality data (244)

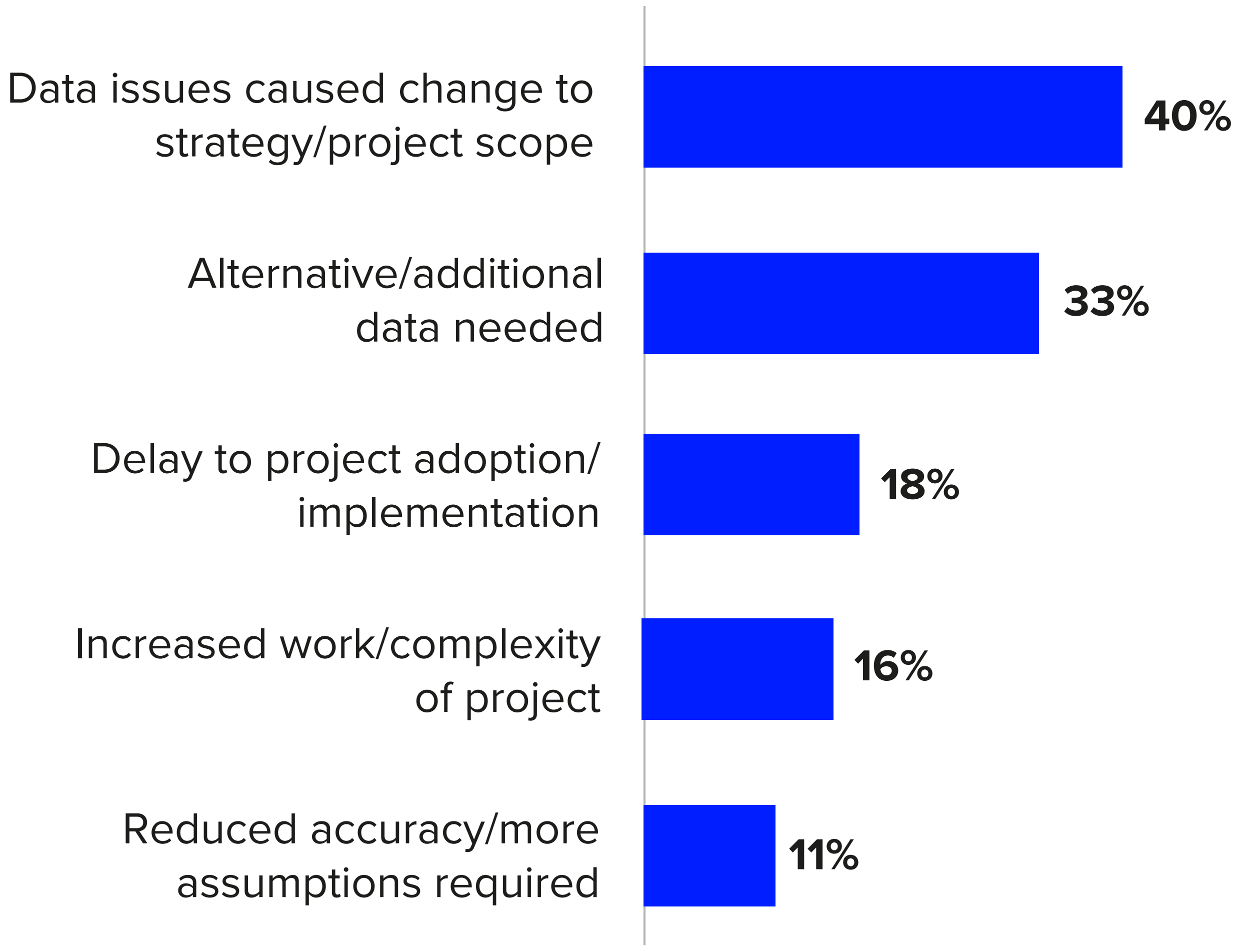


Source: AI/ML survey, August 2020

Figure 4.4: The impact of data availability

Can you describe instances where data availability has impacted your ability to deploy machine learning effectively? What data sets were you looking for?

Base: all respondents with a data availability issue (209)

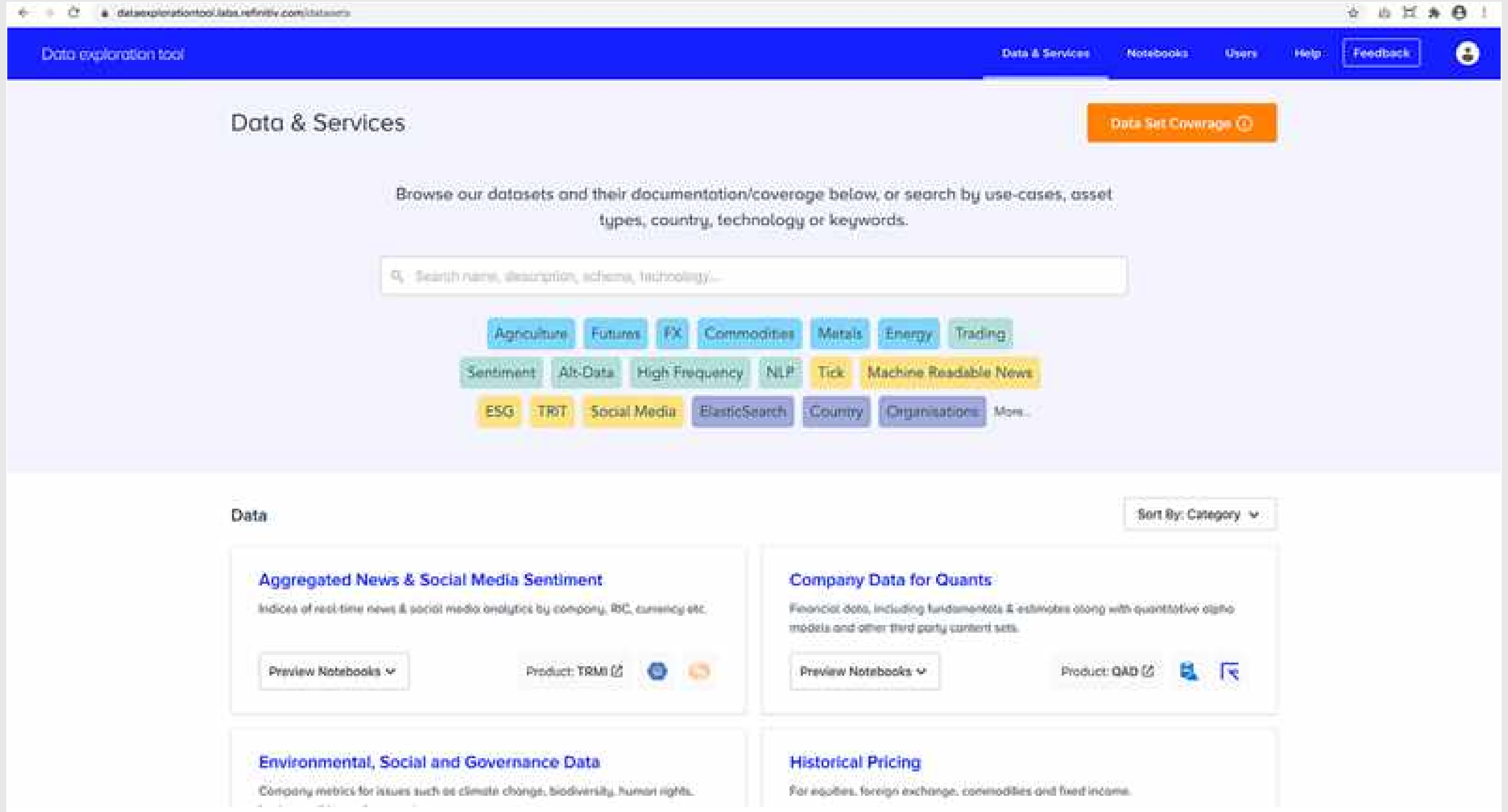


Source: AI/ML survey, August 2020

Meeting data scientists’ need for speed – and quality

Refinitiv Labs is working in collaboration with data practitioners to solve the challenge of AI/ML data quality and availability with solutions such as the data exploration tool.

The new tool gives data scientists free, easy and intuitive access to sample Refinitiv data sets and notebooks, so they can discover, explore and validate Refinitiv’s production-grade data, and ultimately, build and deploy AI/ML models faster.



Data & Services

Browse our datasets and their documentation/coverage below, or search by use-cases, asset types, country, technology or keywords.

Data

Aggregated News & Social Media Sentiment

Indices of real-time news & social media analytics by company, RIC, currency etc.

Company Data for Quants

Financial data, including fundamentals & estimates along with quantitative alpha models and other third party content sets.

Environmental, Social and Governance Data

Company metrics for issues such as climate change, biodiversity, human rights,

Historical Pricing

For equities, foreign exchange, commodities and fixed income.

KEY TAKEAWAYS

Trustworthy, high-quality data underpins the strategy of every AI-first firm. Here are five recommended steps from Refinitiv Labs to ensure your data is clean and machine-ready*:

Classify

- What type of data do you have – text, numerical, time series, dates, images, sound?
- What compliance, residency or usage rights are associated with the data?
- What is the ultimate source of the data and what steps has it gone through to get to you?
- Does your data contain sensitive, confidential or personally identifiable information (PII) that needs handling separately?

Link

- Does the data have a schema that explains fields, tables and indicates the keys to link it?
- Are the identifiers available? What are they for – countries, locations, publishers, companies, objects?
- Which other data sets are essential to link with to ensure the data makes sense?

Enrich

- Can you turn strings into numerical values, dates and meta-data for programmatic use?
- Have you removed outliers, handled missing values and tracked down reasons for the bad data?
- Can you embed meaningful labels into the data and mark your corrections so that others can track changes?

Analyze

- Have you grouped and profiled the data by real-world attributes – month, year, time, location, publisher, metadata, language?
- Have you visualized the data and shared it with others to assist in understanding it?
- Can outliers, patterns or groupings be understood and explained?

Normalize

- Do you need to change the format for processing – CSV, DataFrame, Parquet file?
- Does the data need to be broken into chunks or test set versions for processing?
- Can your data be packaged and stored as a feature, so you can test and reuse it free of any dependencies?

**Caveat: following the steps is not always a linear process.*

ACCELERATION: COVID-19 UPSETS MODELS AND DRIVES UP AI/ML INVESTMENT

One fascinating transformation from 2018 to 2020 is how tried and tested models that use trusted data failed in the systemic economic shifts caused by the coronavirus pandemic.

This change triggered an interest in using alternative data to both increase signal accuracy and generate a competitive edge as firms reset their strategies.

Before COVID-19, the effort and risk required to test an alternative data set was more difficult to justify, given that existing methods were still functioning, and there was no proof that delta would be outsized by using a new data source.

AI/ML models need to be ready for more black swans

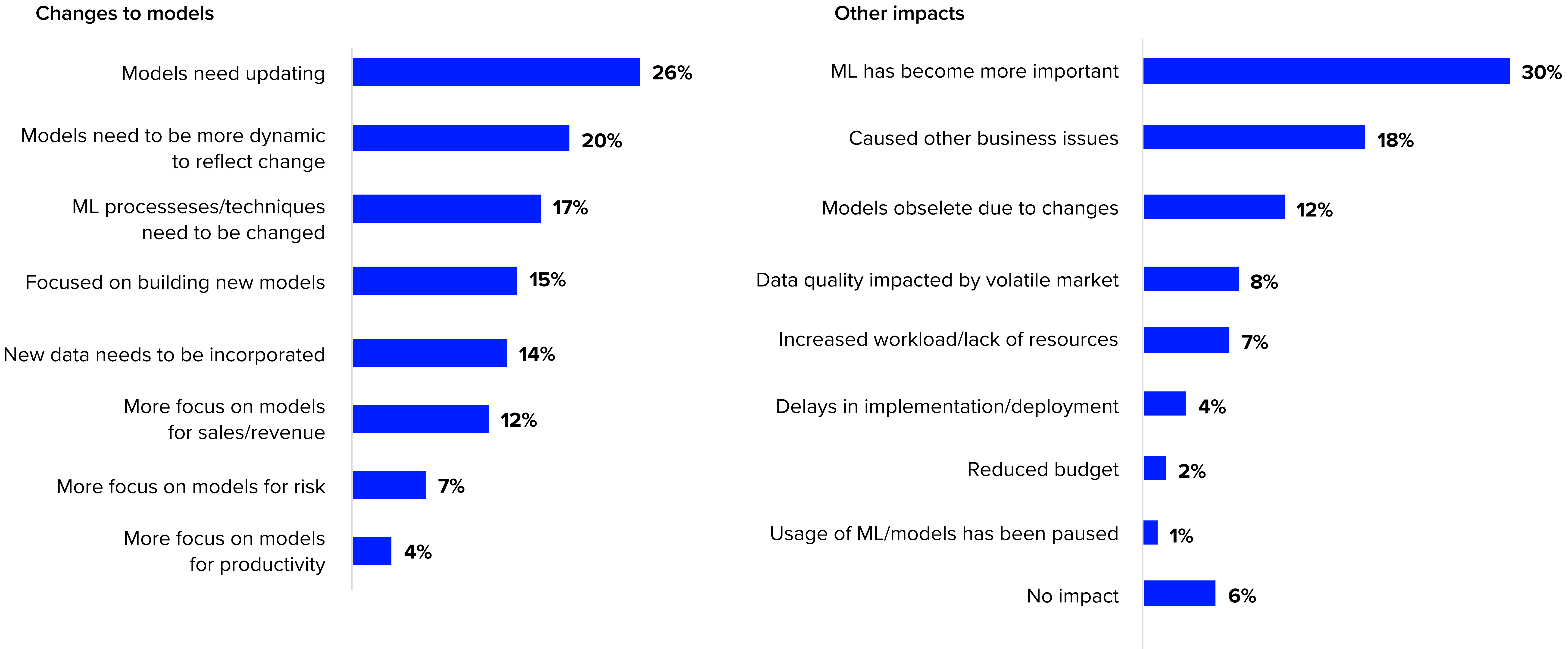
During the second quarter of 2020, models underperformed, and alternative data came into play as a provider of real-time data to derive insights for immediate action.

Our 2020 survey shows 72% of firms' models were negatively impacted by COVID-19. Some 12% of firms declared their models obsolete, and 15% are building new ones. The main problem was the lack of agility to quickly adapt and include new data sets in models as circumstances changed.

> **"Our models depend on data quality and a small mistake can lead to a major disaster in the system. Hence, we now need a large amount of quality data sets for the development of new models."**
>
> Data scientist at a commercial bank in Germany

Figure 5.1: The COVID-19 impact on AI/ML models

What has the impact of COVID-19 been on use of machine learning in your organization and on the data science community in general?
Base: all respondents (423)



Source: AI/ML survey, August 2020

A comparison between 2018 and 2020 shows the percentage of financial firms not using alternative data at all has plummeted from 30% to 3%.

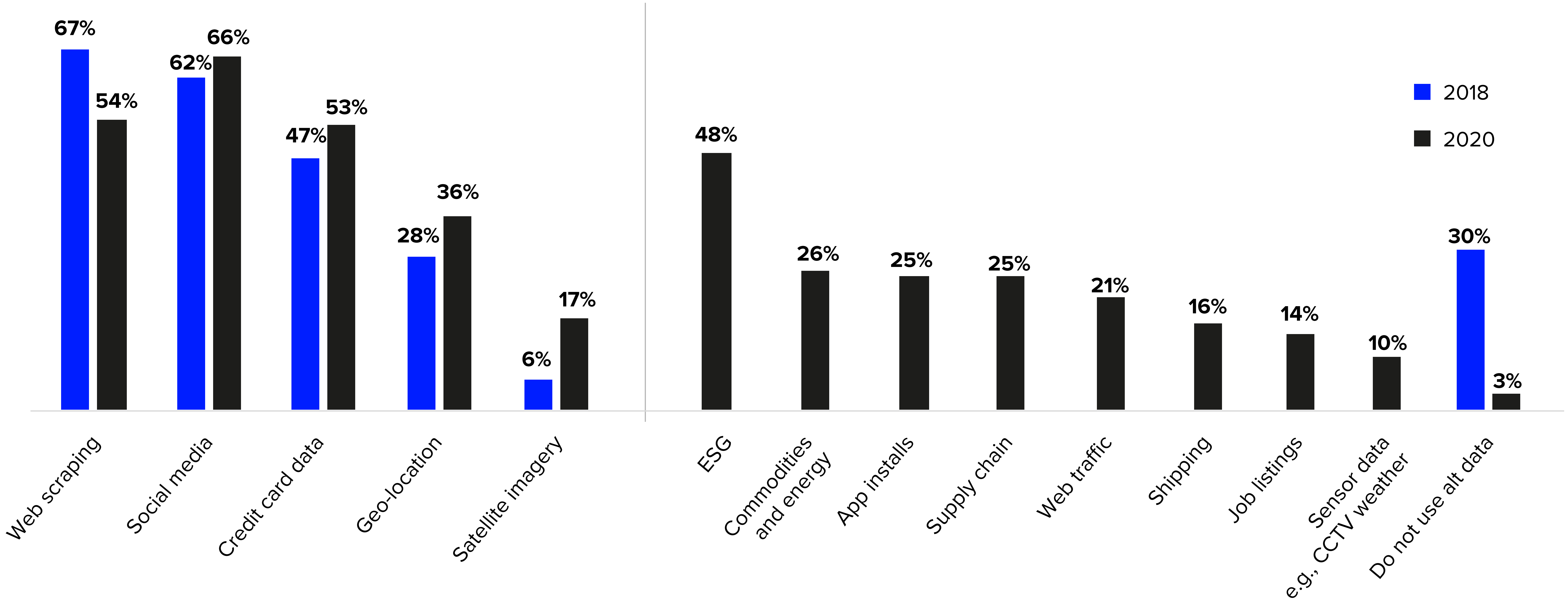
Similarly to 2018, social media, Web scraping and credit card data are the most frequently used categories of alternative data.

They are matched in 2020, and are likely to be overtaken going forward, by escalating interest in ESG data, a new category in the 2020 survey.

Figure 5.2: Alternative data usage in AI/ML

Which of these types of alternative data does your company use in machine learning models?

Base: all respondents (2018: 447; 2020: 423)]



Source: AI/ML survey, December 2018; August 2020

Refinitiv partners with alternative data firm Battlefin

Refinitiv's strategic partnership with alternative data firm BattleFin combines high-quality Refinitiv fundamental data with alternative data, such as app installs, Web traffic, geospatial data and employment data among many subsets, in an easy-to-use experience.

It allows the global investment community to develop and test differentiated ideas, optimize portfolios, manage risks and seek alpha, free from the burden of sourcing quality data.

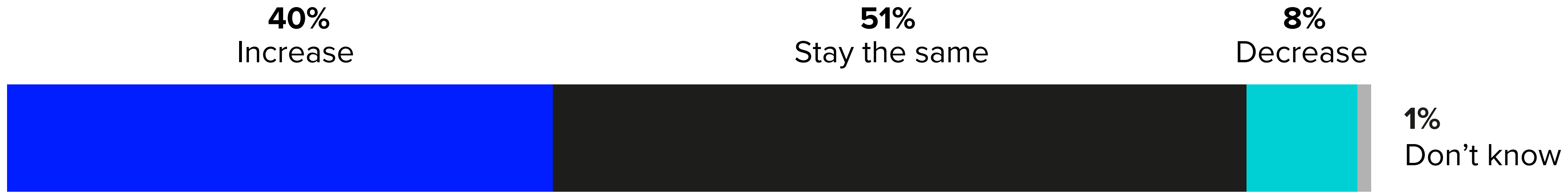
Will different investment levels post-COVID-19 create AI/ML "haves and have nots?"

COVID-19 has made AI/ML more important at 30% of the financial services firms surveyed, and is driving up investment. Our 2020 research shows 40% of firms expect to increase investment in AI/ML as a result of COVID-19.

Figure 5.3: Investment in AI/ML post-COVID-19

Will investment in machine learning increase, decrease or stay the same as a consequence of COVID-19?

Base: all respondents (423)

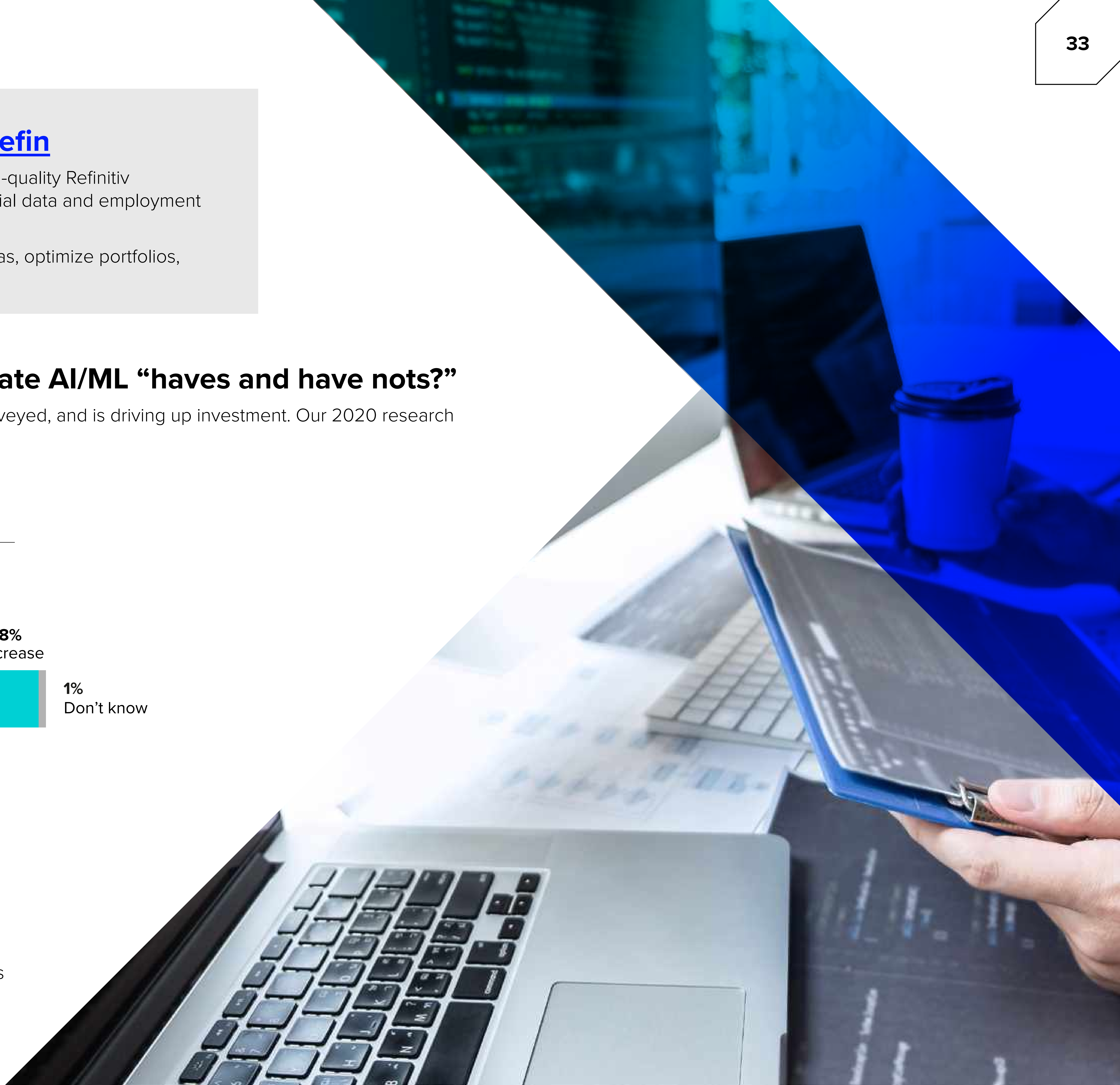


Source: AI/ML survey, August 2020

Diving deeper, companies that have already invested more heavily in AI/ML have a greater expectation of higher than average investment over the next 12 months.

Similarly, 50% of companies that employ more than 26 data scientists anticipate an increase in investment. Where the number of use cases is higher, companies also expect more investment.

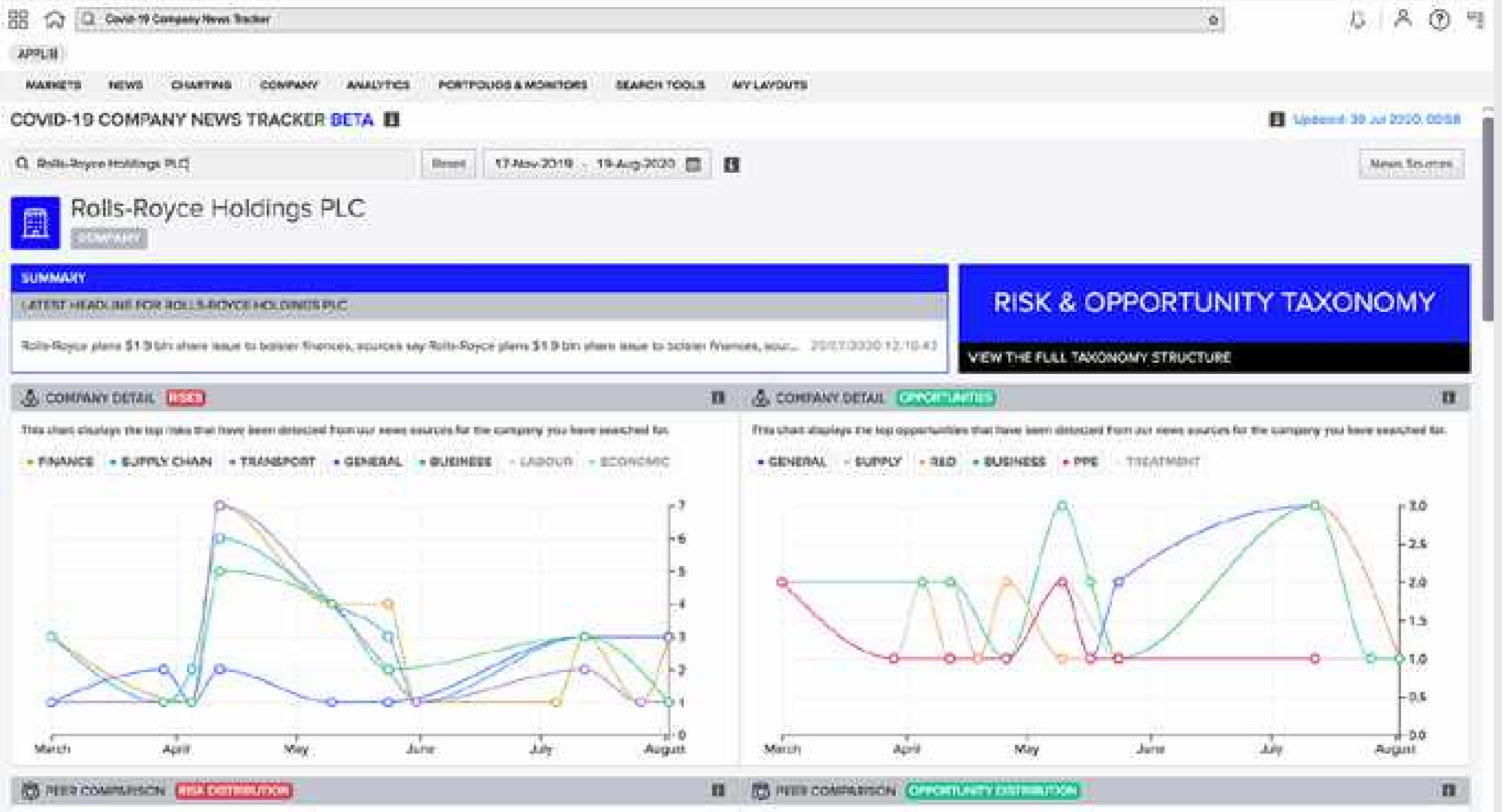
Where does this leave less mature firms with fewer data scientists and AI/ML use cases? One possibility is that COVID-19 could widen the gap between financial firms that are more and less invested in AI/ML.



COVID-19 Company News Tracker

Refinitiv Labs responded quickly to the challenges of the pandemic by building the COVID-19 Company News Tracker to help economists, traders and investors uncover new risks and opportunities across different companies, industries and supply chains.

The app applies AI/ML and Google's open source NLP model, BERT, to Refinitiv data sets including Machine Readable News and Company Fundamentals. The COVID-19 Company News Tracker is available in the Refinitiv Eikon Macro Vitals App.



Rolls-Royce plans $1.9 bln share issue to bolster finances, sources say Rolls-Royce plans $1.9 bln share issue to bolster finances, sour... 20/07/2020 12:10:43

This chart displays the top risks that have been detected from our news sources for the company you have searched for.

This chart displays the top opportunities that have been detected from our news sources for the company you have searched for.

KEY TAKEAWAYS

If your firm isn’t investing heavily in ML

- Begin by using products that already have AI/ML baked in, such as Machine Readable News. You don’t have to start from scratch and can build on top of pre-existing analytics
- For example, if a news article is flagged as high-risk, you can automatically generate an alert linked to your firm’s risk profile, internal processes and people

Ensure your models are trained for future disruptive events

- Create simulations to test robustness with extensive data history and backtesting. For example, if every financial crash is a data point, you will need to historically back-test the global financial crisis and the Dot-com bubble
- You can also change models based on risk appetite and an assessment of the environment. Switching to a low-risk model for trading may mean only making liquid investments, holding more cash daily and trying to reduce your exposure

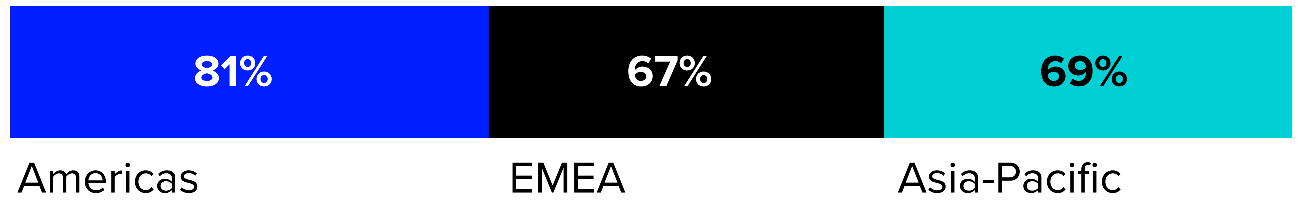
Combine alternative data with traditional data sets to test their reliability

- Linking is critical to understanding whether alternative data is a good or bad predictor. For example, the Refinitiv Starmine® models revealed that Web scraping data is good at predicting upside, but bad at predicting downside
- Comparing alternative data to fundamental data benchmarks, such as Estimates data in Starmine’s case, allows firms to gauge the reliability of their alternative data

REGIONAL AI/ML TRENDS

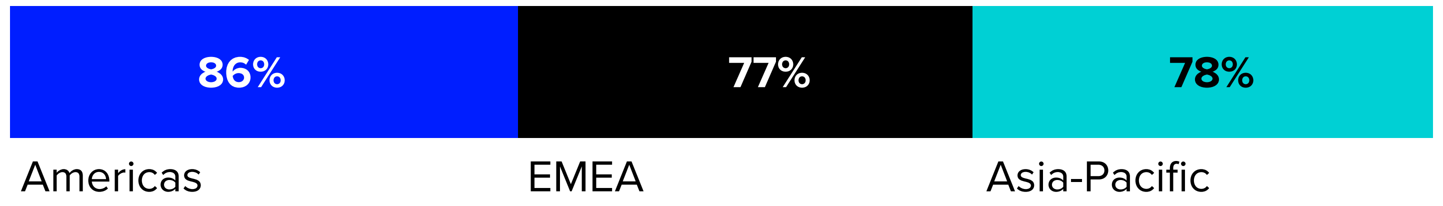
AI/ML is a core component of our business strategy (Score 7 to 10)

On a scale of 1 to 10, where 1 is strongly disagree and 10 is strongly agree, how much do you agree?
Base: Americas (135); EMEA (140); Asia (147)



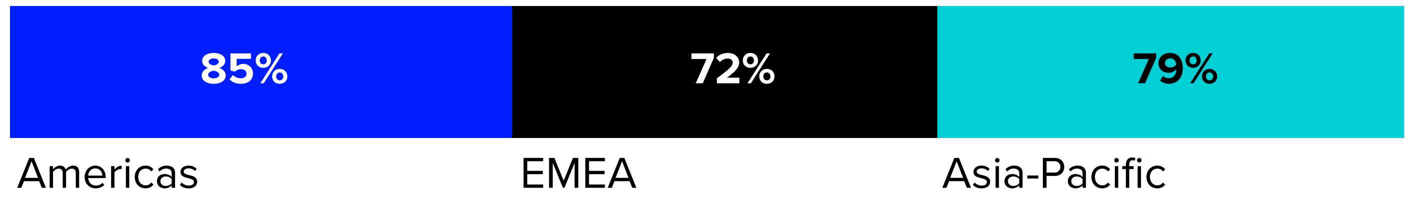
We make significant investment in AI/ML (Score 7 to 10)

On a scale of 1 to 10, where 1 is strongly disagree and 10 is strongly agree, how much do you agree?
Base: Americas (135); EMEA (140); Asia (147)



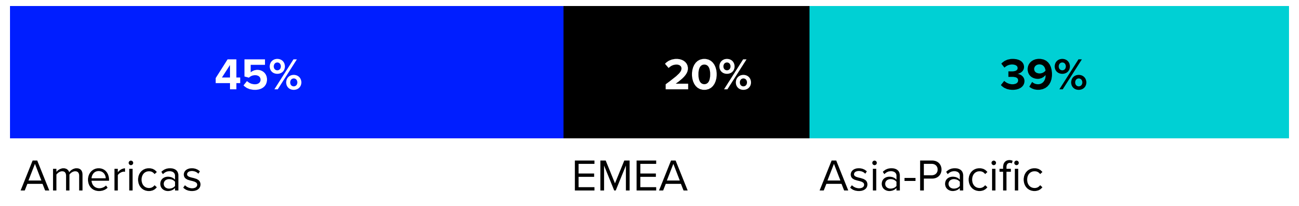
We have a clear vision around usage of AI/ML technologies (Score 7 to 10)

On a scale of 1 to 10, where 1 is strongly disagree and 10 is strongly agree, how much do you agree?
Base: Americas 135; EMEA 140



We anticipate an increase in the number of data scientists per region

To the best of your knowledge, will the number of data science roles in your company increase, decrease or stay the same in the next 12 months?
Base: Americas (135); EMEA (140); Asia (147)



Source: AI/ML survey, August 2020

Top four AI/ML drivers per region

AMERICAS

1. To extract better quality information
2. To extract more value from data
3. To stay ahead of the competition
4. To improve investment returns

EMEA

1. To extract better quality information
2. To extract more value from data
3. To stay ahead of the competition
4. To identify new sales or marketing opportunities

ASIA-PACIFIC

1. To extract more value from data
2. To reduce costs
3. To extract better quality information
4. To stay ahead of the competition

CONCLUSION AND PREDICTIONS

AI/ML is a horizontal capability

This year’s survey demonstrates growing maturity in AI/ML, and the ability to scale across business units. Additional use cases will have a head start if foundations, such as cloud deployment and investment in technology and teams, have been implemented.

AI/ML data strategies are now more important than technology strategies

This requires new approaches to data sourcing, data management, data governance and data quality. Having confidence in high-quality, production-grade data for AI/ML will enable firms to use more sophisticated techniques, like deep learning and NLP, to extract new value from existing and untapped data and new data combinations.

MLOps puts scale into production

Moving forward, the next target is ML Operations (MLOps), which will scale AI/ML for the enterprise. Lots of data science teams currently find themselves using PowerPoint® to present new insights, but MLOps will soon help these teams drive real change by operationalizing AI/ML models and replacing manual steps for data preparation and model evaluation, among others, with an automatic pipeline.

Financial data scientists and ML engineers will drive change

We predict financial data scientists will drive this strategic change. The year 2021 could also see the rise of a new role – the MLOps Engineer – as robust data pipelines handling petabytes of data become critical. With data at the heart of AI/ML, the role of the data scientist will evolve.

The ability to discover more unique and significant content will mean greater responsibility in the business, which will continue to increase as the business better understands the capabilities of its data. Then, perhaps, data scientists will join business and banking teams at the tipping point of modeling the future of finance.

```
    mirror_mod.use_x = False
    mirror_mod.use_y = True
    mirror_mod.use_z = False
elif _operation == "MIRROR_Z
    mirror_mod.use_x = False
    mirror_mod.use_y = False
    mirror_mod.use_z = True

    #selection at the end -a
mirror_ob.select= 1
modifier_ob.select=1
bpy.context.scene.objects.ac
print("Selected" + str(modif
```

PRODUCTION-GRADE DATA FOR AI-FIRST FIRMS

Trusted data to help you scale AI/ML and deep learning

Refinitiv’s production-grade data is carefully aggregated, cleaned, normalized and managed to help you train models that achieve quality and precision.

9.6 million

peak message rates distributed per second to financial markets for up to 90 million instruments

8.3 million

private companies covered

+1 million

research documents contributed per quarter from over 1,300 research contributors

68,000

public companies covered – 99% of global market cap

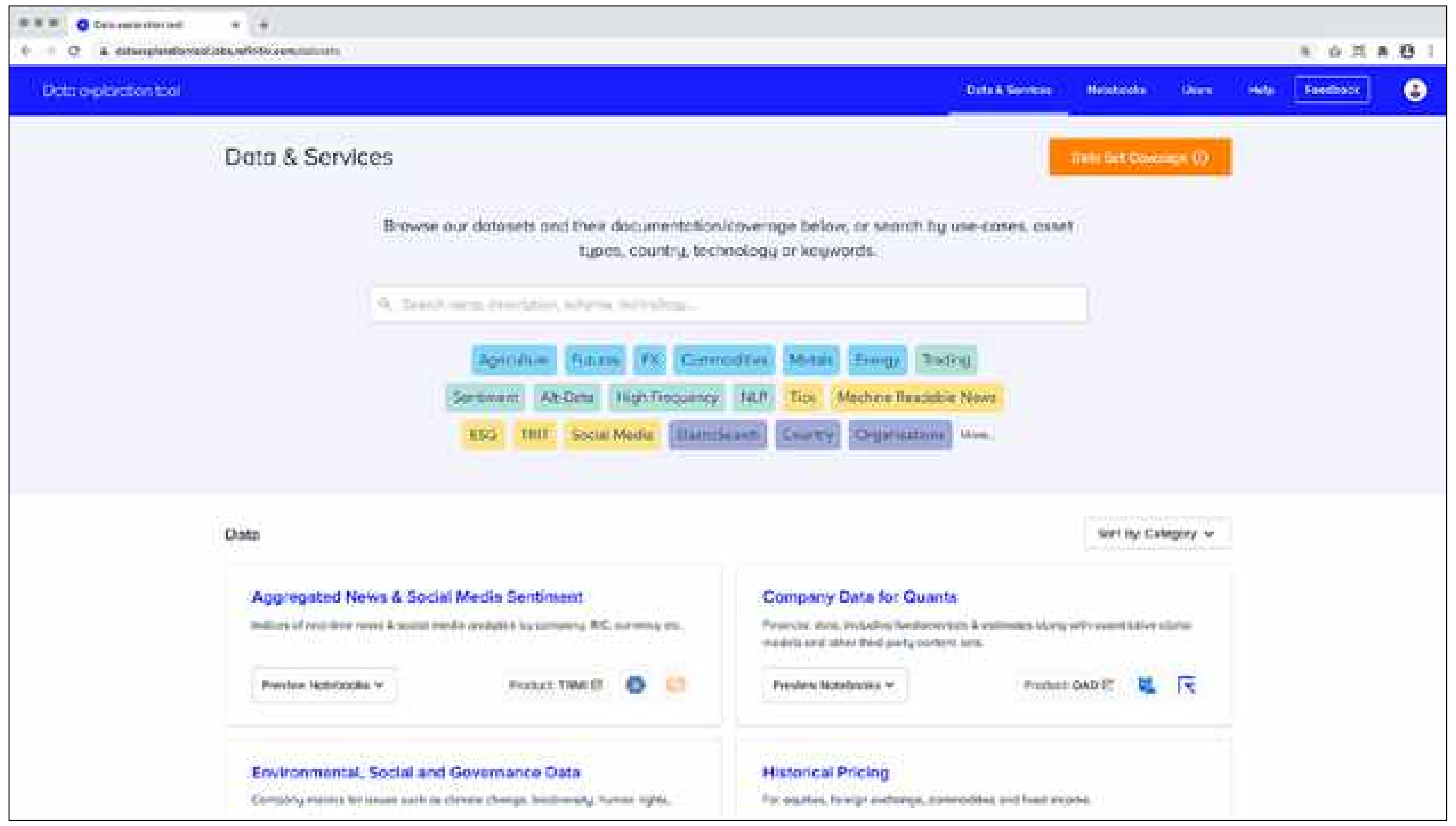
Reuters news

from 2,500 journalists and 110 bureaus in 72 countries

+1,200

content partners and over 1,000 outbound partners

EXPLORE FOR FREE TODAY



Data & Services

Browse our datasets and their documentation/coverage below, or search by use-cases, asset types, country, technology or keywords.

Data

Aggregated News & Social Media Sentiment

Indices of real-time news & social media analytics by company, ESG, currency etc.

Company Data for Quants

Financial data, including fundamentals & estimates along with quantitative alpha models and other third party content sets.

Environmental, Social and Governance Data

Company metrics for issues such as climate change, biodiversity, human rights...

Historical Pricing

For equities, foreign exchange, commodities and fixed income.

Data experimentation with fewer hurdles

Refinitiv’s data exploration tool gives data scientists, quants and developers free, easy and intuitive access to sample Refinitiv data sets and notebooks.

Refinitiv is one of the world's largest providers of financial markets data and infrastructure, serving over 40,000 institutions in approximately 190 countries. It provides leading data and insights, trading platforms, and open data and technology platforms that connect a thriving global financial markets community – driving performance in trading, investment, wealth management, regulatory compliance, market data management, enterprise risk and fighting financial crime.

Visit **refinitiv.com**

@Refinitiv Refinitiv

RE1152855/10-20

REFINITIV®

DATA IS JUST
THE BEGINNING®