

# Smarter Humans. Smarter Machines.

Insights from the Refinitiv 2019 Artificial Intelligence /  
Machine Learning Global Study

REFINITIV™  
DATA IS JUST  
THE BEGINNING



---

*“We’re on a mission at Refinitiv to enable smarter humans and smarter machines. We started our journey in artificial intelligence and machine learning more than a decade ago to provide the technology, analytics and real-time, intelligent data for competitive advantage. This survey confirms the important role AI and ML play in the transformation of financial services and can aid your organization on its technology course. In the end, data is just the beginning.”*



**David Craig**  
*Chief Executive Officer*  
Refinitiv

# Welcome

History will show the 21<sup>st</sup> century as the period when the world entered the fourth industrial revolution: the fusion of the physical, the digital and the biological, enabled by data. But data is just the beginning. Profound shifts, facilitated by technologies such as artificial intelligence and machine learning (AI/ML), robotics and the Internet of Things, are already underway.

Futurists warn that it is impossible to project what the world will look like in 2050 as it will be so different from what we know today. The Institute for the Future says that, over the next decade, emerging technologies will underpin the formation of new human-machine partnerships, which will help humans transcend their limitations, enhance daily activities and reset expectations for learning and work. The actualization of smarter humans, smarter machines.

## **Enabler of competitive advantage**

We predict that AI/ML will be the single greatest enabler of competitive advantage in the financial-services sector. In the last few years, there has been an explosion in the use of machine learning, led by applications for image processing, natural-language processing (NLP) and machine translation. Because these new capabilities are largely based on open-source libraries, and can be deployed relatively cheaply in the cloud, the barriers to entry have fallen dramatically. We expect a flurry of commercial and product innovation from organizations of all sizes. The benefits will extend well beyond automating rules-based repeatable tasks once done by humans.

Refinitiv's inaugural Artificial Intelligence/Machine Learning Survey of nearly 450 financial-institution leaders and data scientists reveals the degree to which machine learning is already becoming an integral part of running the business. Financial institutions have gone beyond experimenting with and testing machine learning, deploying it in key areas such as financial risk management, pre-trade analytics and portfolio optimization.

## **Data quality is essential**

The survey shows that the devil is in the data: data quality (meaning data of poor quality) is the biggest barrier to the adoption and deployment of machine learning. Unstructured data, as well as data from alternative sources, are increasingly important areas but need considerably more work before their insights are truly reliable. The adage ‘garbage in, garbage out’ has never been more pertinent. If data is the new oil, then much of it still needs a lot of refining and that’s a heavy lift for the consumers of data.

## **C-suite vs. data scientists**

Data scientists are tasked with creating the models and algorithms that will set their organizations apart from the competition. But there is a mismatch between the vision in the boardroom and the reality on the ground. C-level professionals believe it is important to be seen using the latest tools and techniques for competitive advantage and may be overstating the company’s actual adoption. Data scientists, on the other hand, are under pressure to deliver on the promise of machine learning but must navigate real organizational constraints. As with all surveys of this type, there is bound to be some sampling and human-based bias.

## **Geographic differences**

The study also reveals a disparity in how technologies are being adopted and used around the world. Financial institutions in North America (the United States and Canada) are the front-runners. Asian institutions are more advanced in some areas than those in Europe, such as in machine learning being core to business strategy and the projected growth in numbers of data scientists. However European organizations lead those in Asia in terms of having deployed machine learning.

This quite stark difference, again, could be due to human factors, but is more likely explained by the fact that much of the underlying capabilities were originated in North America—from the development of the first algorithm for random forests (developed by Tin Kam Ho at Bell Labs in 1995<sup>1</sup>), to the open sourcing of tools such as TensorFlow. TensorFlow<sup>2</sup> 2.0 was just released and it promises further democratization of the technology. But regional differences will likely diminish with time—open source knows no boundaries.

<sup>1</sup> Ho, Tin Kam (August 1995). “Random decision forests”. Proceedings of 3rd International Conference on Document Analysis and Recognition. 1: 278–282 vol.1. doi:10.1109/ICDAR.1995.598994. ISBN 978-0-8186-7128-9.

<sup>2</sup> <https://github.com/tensorflow/tensorflow/blob/07bb8ea2379bd459832b23951fb20ec47f3fdbd4/RELEASE.md>

## Buy side vs. sell side

Similarly, the study suggests that the buy side is ahead of the sell side. Perhaps this is more indicative of the hedge funds surveyed, where innovation and investment has traditionally been focused. Here, greater availability of advanced tools is likely to materially level the playing field—on both buy and sell sides.

Happy reading. I hope you find this as interesting as we do. Don't hesitate to reach out with your thoughts and comments.

*T Baker*



**Tim Baker, CFA**

*Global Head of Applied Innovation & Refinitiv Labs*  
Refinitiv





## Introduction

The Refinitiv inaugural Artificial Intelligence/Machine Learning<sup>3</sup> Survey took place in December 2018. We conducted 447 telephone interviews of 20-25 minute duration with data-science practitioners and C-level data-science decision-makers in financial institutions with annual revenue of more than \$1bn.

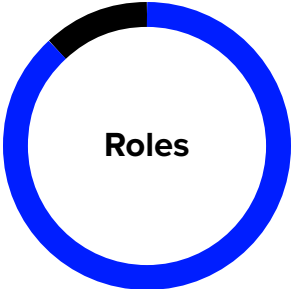
The survey was global, representing a wide variety of financial institutions and individuals in data science and related business-leadership roles. It focused on machine learning and included only participants from organizations that are currently using machine learning (98% of participants) or intend to in the future (2%).

<sup>3</sup> We defined machine learning to research participants as follows: machine-learning technologies enable the use of statistical techniques / models to give computer systems the ability to learn from data without being explicitly programmed.

The interviews broke down as shown in Chart 1.

CHART 1

## Breakdown of AI/ML Survey interviews



**394**  
**Data Scientists**

Data Scientists Example Roles:

- Data Analyst
- Data Engineer
- Data Scientist
- Head of AI
- Head of Data Science
- Head of ML
- Head of Natural Language Processing
- Innovation Manager
- Quant / Modeler / Statistician

**53**  
**C-Level**

C-Level Example Roles:

- Data Officer
- Information Officer
- Technology Officer

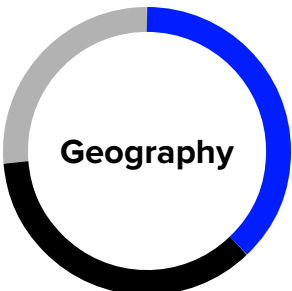


**320**  
**Sell Side**

- Commercial or Retail Bank (144)
- Investment Bank (107)
- Broker Dealer (66)
- Inter Broker Dealer (3)

**127**  
**Buy Side**

- Asset Mgt (87)
- Hedge Fund (23)
- Investment Advisory (17)



**170**  
**Asia Pacific**

- India (50)
- Singapore (49)
- Australia (28)
- New Zealand (20)
- Korea (13)
- China (10)

**161**  
**Europe**

- UK (71)
- France (49)
- Germany (41)

**116**  
**North America**

- US (59)
- Canada (57)

Source: Refinitiv AI/ML Survey

The data scientists we surveyed are responsible for at least three of the activities shown in Chart 2.

CHART 2

## Data scientist activities

### Developing models

(applying, testing and iterating statistical and machine-learning methods)



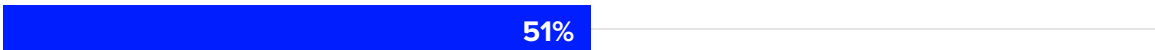
### Deploying models for operational use



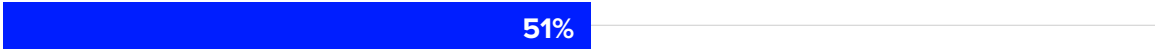
### Building reusable components to perform data-science work



### Preparing data for use in models



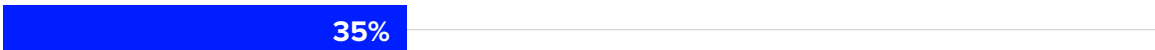
### Managing a team of data scientists



### Presenting findings and research about ML methods to business stakeholders



### Sourcing, evaluating or using new data from data vendors



Source: Refinitiv AI/ML Survey

Q: Within your role which of the following do you or are you responsible for?

Base: All respondents (447n)



# Key findings

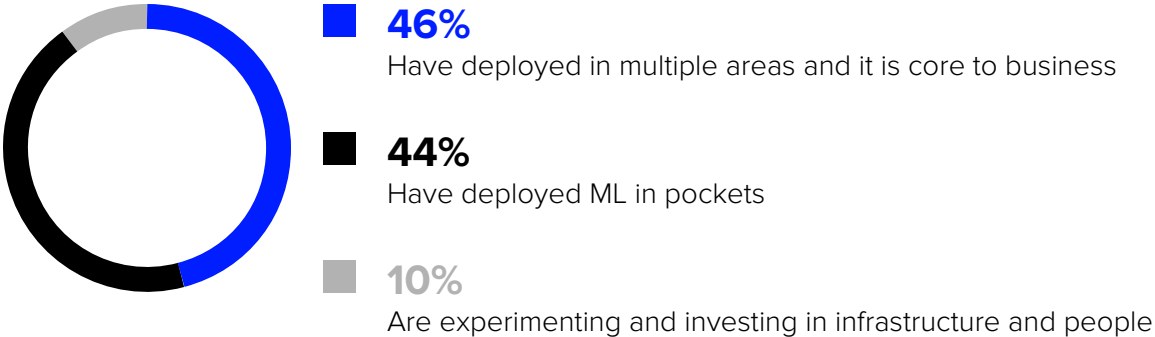
- 01** Financial institutions are further advanced in the deployment of machine learning than expected
- 02** Key applications of machine learning are in risk management, performance analysis and trading idea generation
- 03** Data quality is the primary barrier to machine-learning adoption
- 04** Alternative data sources are almost as widely used as market and company data
- 05** There is a desire to use market data and alternative data sources in conjunction with companies' own data, which requires the combination of disparate data sets
- 06** Advances in machine learning have facilitated the use of sources of unstructured data, such as text-based market news
- 07** C-level leaders are more likely than data scientists to say that machine learning is core to strategy
- 08** Data scientists are more likely to see data quality as a barrier to the adoption of ML than their c-level leaders
- 09** C-suite respondents plan to hire more data scientists in the future
- 10** Participants in North America and on the buy side of the market are more advanced in their use of machine learning, but may not sustain this advantage

# Machine learning already drives competitive advantage

It's not news that AI and machine learning are being tested and deployed in businesses of all types. But what is surprising is just how advanced the application of machine learning has become in the financial institutions that we interviewed. Our research shows that c-level business leaders have embraced these technologies as mainstream and are using them not just to improve productivity, save money or cut costs, but in key strategic areas including risk management and trading in order to differentiate themselves from competitors.

As shown in Chart 3, most organizations (90%) have gone beyond the experimental stage and have deployed machine learning, either in pockets or more fully across their businesses. Approximately three quarters say machine learning is a core component of their business strategy and they're making significant investments in it, as shown in Chart 4. This is striking because to have implemented machine learning requires an organization to have gone through a lengthy process involving hiring a team, sourcing the right data, building a model and back-testing and validating it to the degree that they are confident it can be used operationally.

CHART 3  
**Adoption of machine learning**



Source: Refinitiv AI/ML Survey  
Q: Which of the following describes the adoption of machine learning technologies / techniques to manage or analyze data or content within your organization?  
Base: 437n / C1 = You currently use



CHART 4

**The use of ML as a core component of business strategy**

- 9-10
- 7-8
- 4-6
- 1-3

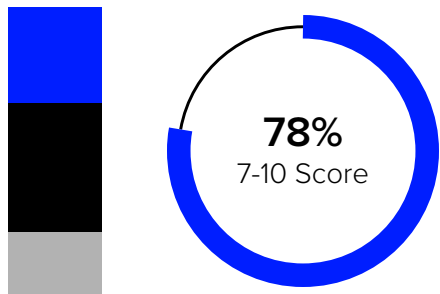
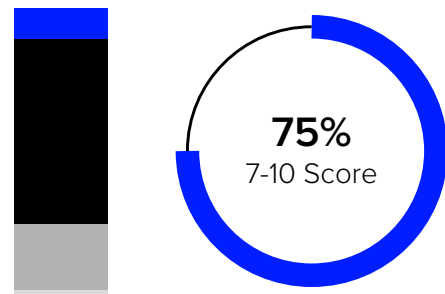


CHART 5

**Those that make a significant investment in ML**



Source: Refinitiv AI/ML Survey

Q: On a scale of 1 to 10, where 1 is strongly disagree and 10 is strongly agree, how much do you agree that the use of ML is a core component of your business strategy and that your organization makes a significant investment in ML?

Base: core component (446n); significant investment (447n)

# What's driving ML adoption

The reasons for ML adoption and its primary focus in financial organizations are quite telling. Current media articles point to saving money as the primary motivation for machine learning. *American Banker*, for instance, recently reported that AI is projected to cut costs and increase productivity to the tune of \$1 trillion for U.S. financial institutions.<sup>4</sup> But our research shows that back office efficiency is not the only reason these organizations invest in developing and deploying machine learning models (see Chart 6). Rather, the main driver is to make informed decisions with better quality data (60%). The desire to increase productivity and speed is second (48%), and cost-cutting checks in at 46% as the third most important reason.

Similarly, we expected to see the primary application of machine-learning in the automation of repetitive tasks. The survey, however, unearthed a different reality as the top applications are in business-critical areas such as risk avoidance (82%), trading and investment idea generation—aka seeking alpha (63%), and automation is in fourth place (61%), as shown in Chart 7. Data is core to the way financial-service businesses operate. Any innovation that makes better use of data, and enables data scientists to combine disparate sources of data in a meaningful fashion, offers the potential to gain competitive advantage.

CHART 6

## Adoption Drivers

**Extract better quality information**



**Increase productivity and speed in processes**



**Reduce costs**



**Extract more value from data**



CHART 7

## Applications for Machine Learning

**Risk management**



**Performance analysis & reporting**



**Trading investment idea generation**

*(alpha generation)*



**Automation**



Source: Refinitiv AI/ML Survey

Q: On a scale of 1 to 10, where 1 is not important at all and 10 is very important, what would you say are the most important factors driving adoption of new machine learning technologies / techniques in your organization / team?

Base: All respondents (447n)

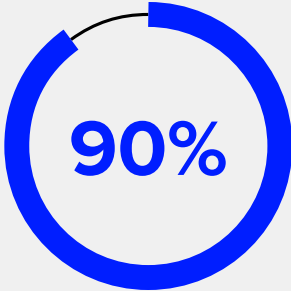
Q: Which of the following are focus areas in terms of applying ML techniques?

Base: All respondents (447n)

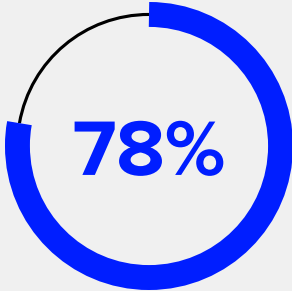
4 American Banker, February 6, 2019, <https://www.americanbanker.com/opinion/mergers-like-fiserv-first-data-are-a-red-flag-for-banks>

# Smarter humans.

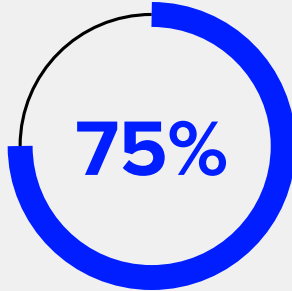
## Machine learning usage is increasing



deployed ML by one or more departments to manage or analyze content

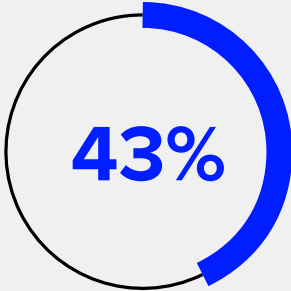


say ML is core component of biz strategy

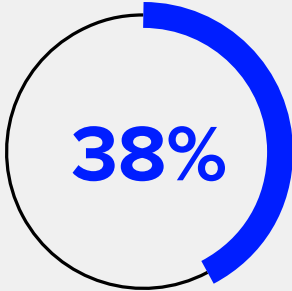


financial-services firms invest significant money in ML

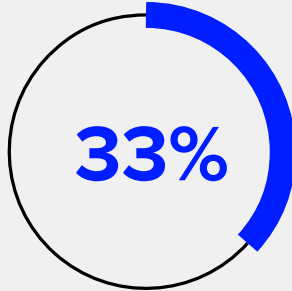
## Data quality is biggest ML challenge



say poor data quality is biggest ML barrier



report lack of data availability as 2<sup>nd</sup> highest barrier



find it hard to hire data-scientist talent

Source: Refinitiv AI/ML Survey

# C-suite vs. data scientists—different realities

## ML deployed and core to strategy



## Clear vision of use of ML tech



## Use market data in ML

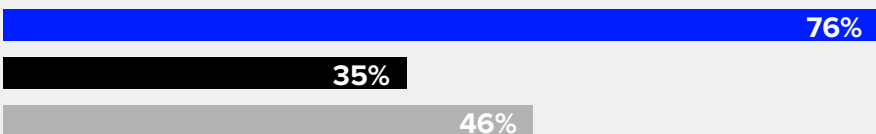


# North America leads in ML

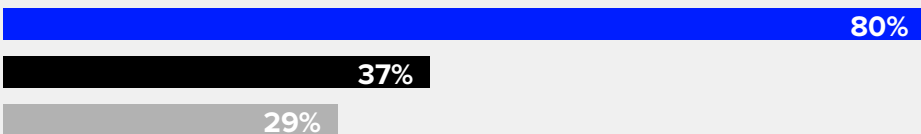
## Making significant ML investment



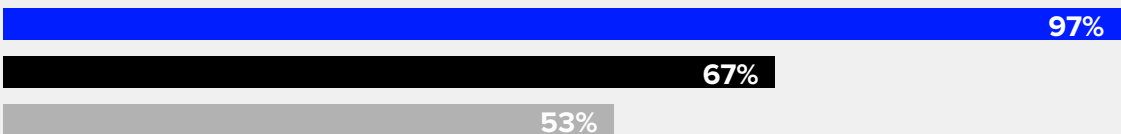
## No. data scientists will increase next year



## Adoption of ML deployed and core



## Using alternative data



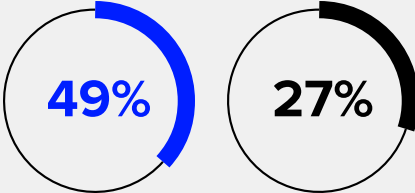
Source: Refinitiv AI/ML Survey

# Smarter machines.

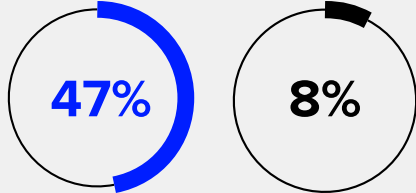
## Buy side vs. sell side

Buy  
Sell

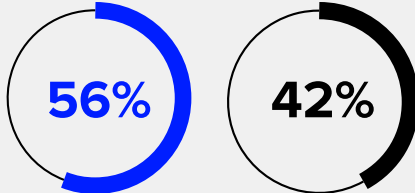
**ML core part of biz strategy**  
*(strongly agree, score: 9-10)*



**Invest significantly in ML**  
*(strongly agree, score: 9-10)*



**ML is deployed and core to biz**



## Main ML application: risk avoidance not automation or cost cutting



across all regions selected avoiding risk as their primary focus

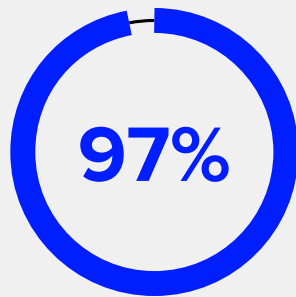


of the North American participants selected risk avoidance first

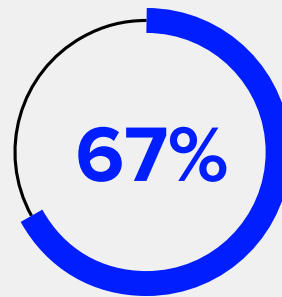
Source: Refinitiv AI/ML Survey

## ALT D (alternative data) growing in its ML significance

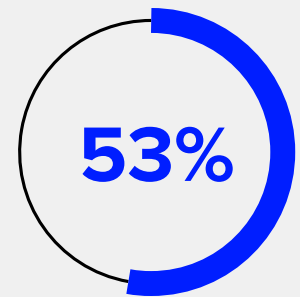
### Respondents who use alternative data



America



Europe



Asia



Alternative sources: web scraping, social media, credit-card data, geo-location and satellite imagery



Financial market and proprietary data = other sources used



Structured data preferred over unstructured

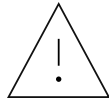




## Why this matters

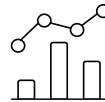
Machine learning is becoming increasingly important in financial services. Machine-driven models are being used in areas that are highly regulated, which can have a negative effect if mistakes are made. Machine learning enables you to rapidly model and test multiple trading strategies using different risk scenarios. It is important to ensure that the industry doesn't move from human-based errors to systematic machine-based ones. In the rush to adopt new technologies and techniques, we need to focus on the quality of the data that is feeding the models, and on understanding where biases can occur.

Risk management and trading have always been data-intensive areas in financial services and have employed individuals with sophisticated quantitative analytical skills. It therefore makes sense that these are now among the key applications for machine learning, as shown in Chart 7. They are the business areas where the challenges align with the people and their skills. The big opportunity for financial-service organizations comes from applying existing skills to different challenges and using new types of data. There are multiple use cases for machine learning in these application areas, including:



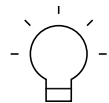
### **Risk management**

- Managing investment risk
- Customer on-boarding risk, such as Know Your Customer and Anti-Money Laundering
- Operational risk, such as looking at exposures across a range of asset classes



### **Performance analysis and management**

- Understanding which traders performed best and why
- Modeling and predicting how to make better trades



### **Trading and investment idea generation**

- Maximizing returns
- Finding new investment opportunities



### **Automation**

- Reducing costs through RPA (Robotic Process Automation)

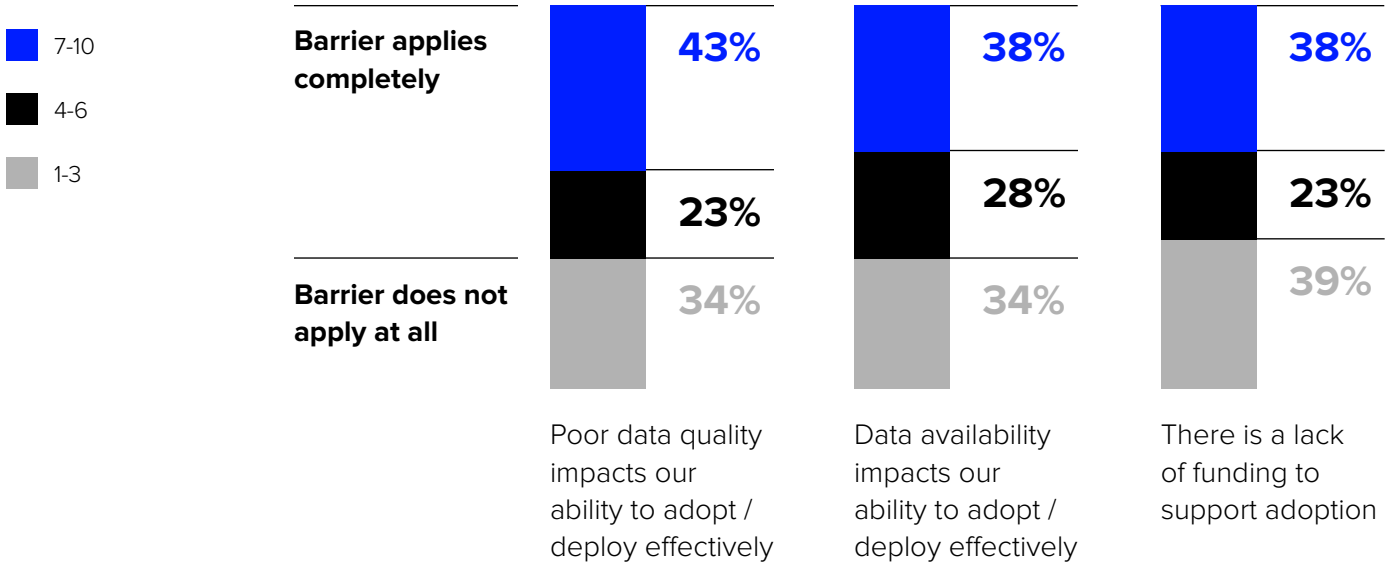
# Data is the biggest challenge

Machine learning involves running complex algorithms and processing massive financial-data sets. As financial institutions are actively using machine learning as part of their core processes, they will have an increasing need for more data, and especially that which can be readily accessed and is

curated, normalized and tagged. The adage ‘garbage in, garbage out’ is highly applicable; data quality is a core differentiator. Our research, as portrayed in Chart 8, shows that poor quality data is the biggest barrier to the adoption and implementation of machine-learning models in financial institutions.

CHART 8

## Barriers to Adopting and Deploying Machine Learning



Source: Refinitiv AI/ML Survey

Q. To what extent do you agree these are barriers to adopting new ML technologies / techniques in the organization, where 1 means ‘does not apply at all’ and 10 means ‘completely applies’

Base: All respondents (447n)

Additionally, Chart 9 shows that three of the top four challenges when using new data in models relate to data quality. The challenge of having sufficient capacity to manage data, which is third from the top, also relates indirectly to data quality in that poor quality data takes up data scientists' time.

At the recent AI and Data Science in Trading conference in New York, several presenters talked about how difficult it is to find data of the appropriate quality and that some groups can spend 80% - 90% of their time normalizing and cleaning it.

Our research participants put the figure for cleaning and normalizing data at around 30% of their time, regardless of geography or company type. Whichever end of the spectrum you're on, data in the financial services sector is likely to start off 'cleaner' than in other industries, as much financial data is structured and comes from sources that can demonstrate the provenance of the data over extended periods of time. Notwithstanding, even 30% is still far too much to have valuable (and expensive) resources tied up in relatively mundane tasks.

CHART 9

## Top Challenges when Working with New Data for Machine Learning

**Accurate information about the coverage, history and population of the data**

56%

**Identification of incomplete or corrupt records**

48%

**Capacity to manage the size and / or frequency of the data**

39%

**Cleaning and normalization of the data**

35%

Source: Refinitiv AI/ML Survey

Q: Thinking about the past few months, when working with new data used in ML models, what have been your major challenges?

Q: When working with a new data set, what percentage of time working with the data is required to clean and normalize it?

Base: All respondents (447n)

## Data's insidious problem

Data quality is paramount when performing any type of analysis, be that machine learning or otherwise. Data quality relates both to data that is missing and data that is wrong. Both are severely problematic and judgement is required on a case-by-case basis to determine the severity and impact of each quality issue. This takes time and effort, often away from addressing the core aspects of a project. Solutions require additional code and processes, increasing the complexity of the task. Similarly, software built to handle data quality can be difficult to design, particularly if quality issues are unknown at the start of a project. Additionally, an insidious problem in machine learning is when data quality is not identified and leads to incorrect inference and misunderstanding.

### Research participants said...

*“There is a massive volume of financial data, and diversity in the structure and volume means that managing this data is a big deal.”*

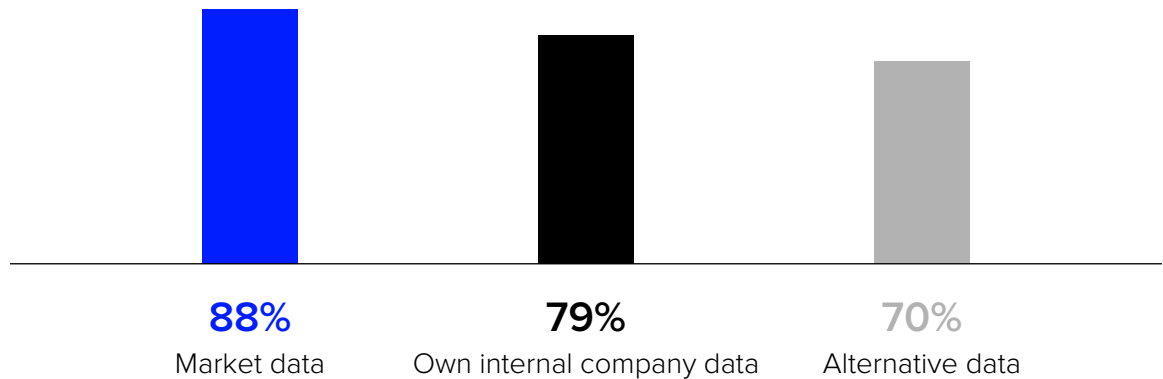
*“The most complicated task is to get data which is relevant, reliable and from a secure data source, which therefore has some statistical value.”*

*“One of the difficulties is that we are not getting defect-free data.”*

*“We always ask ‘which source of data will require the least effort to convert it into structured form?’”*

CHART 10

## Financial institutions are using all sources of data in machine learning



The type of data that companies are using is also a key issue. Chart 10 shows that financial-services professionals are using not just their own internal company data, but also market data and alternative sources in support of machine learning. Many are using all three, which implies that they are combining data in novel ways to gain insights that would have been hard to uncover using human analysts alone. Alternative data, though still less widely used than the other types, is likely used in conjunction with other sources as an overlay or to add context to market or internal company data.

Chart 11 shows a more detailed breakdown of the types of market data that financial-service institutions use in machine learning. The top category here is news (cited by 76% of those who use market data—around two thirds of participants as a whole), which is critical to developing context for much of machine learning. Structured data (often numeric) has long been the cornerstone of human-powered quantitative analysis in the financial-services sector, however news is text-based, unstructured data and is more challenging to prepare and employ in analytics. It must be structured via tagging, which needs to be done in a consistent fashion if it is to be combined with other data sources.

Source: Refinitiv AI/ML Survey

Q: Do you use or does someone in your organization use machine-learning technologies / techniques with financial-market data?

Q: Do you use or does someone in your organization use machine-learning technologies / techniques with your own company data?

Q: Do you use or does someone in your organization use machine-learning technologies / techniques with alternative data sets such as text, social media, specialist limited-access datasets or images?

Base: All respondents (447n)

Technology is enabling machine-learning approaches such as natural language processing (NLP) to bring structure to text-based data. Historically, NLP was a laborious process, but there are new open-source models and capabilities being released that make the process faster and more accurate. In addition, models from companies such as Facebook and Google have focused on bringing structure to the most challenging forms of unstructured data. Their growing success means that text-based analysis has become comparatively easier, more effective and more accurate.

This is good news for the financial-services industry as so much data is held in unstructured text-based documents—and that is where many opportunities lie. Machine learning will make it possible to extract insights from previously inaccessible data sources. However, although new capabilities have made the process easier, it is still highly skilled and requires a great deal of patience to train the models effectively, as well as a tolerance for error and deep knowledge about the market to extract meaning from the findings.

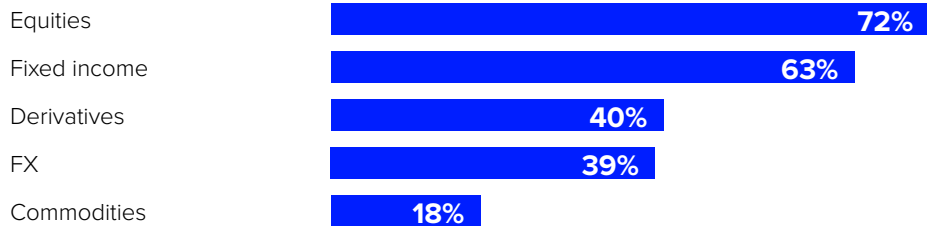
News, company financials and equity-pricing data are the market-data types currently most prevalent in machine learning (Chart 11). They reflect the diversity of the survey participants' roles and organization types. There are emerging data categories, such as ownership and commodities data, but are not currently as widely used (perhaps due to data availability, quality, ease of use and cost). These have potential to become more prevalent in the future.

CHART 11

## Types of Market Data Used in Machine Learning

### Structured Data

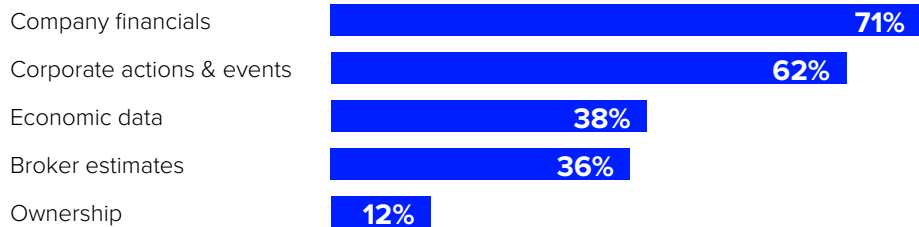
#### Pricing Information by Asset Class



#### Pricing Information by Frequency

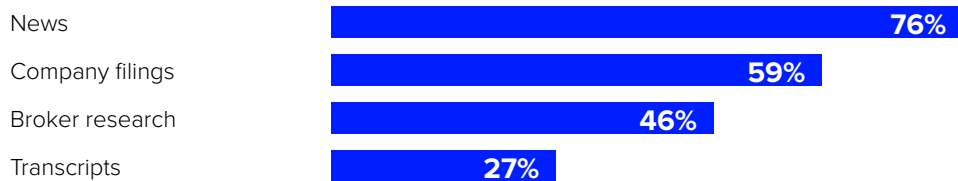


#### Company Data



### Unstructured Data

#### Textual Information



Source: Refinitiv AI/ML Survey

Q: To which of the following types of financial-market data do you commonly apply machine-learning models?

Base: All respondents using Market data (395n)



CHART 12

## Types of an Organization's Own Internal Data Used in Machine Learning

### Customer-profile information

84%

### Customer-usage data

67%

### Prospective-customer information

64%

### Customer-purchase data

60%

### Investment-performance data

59%

### Investment-risk data

51%

### Trading-execution data

50%

### Internal research & commentary

44%

### Research

37%

The majority of internal company-data sources relate to an organization's customers (Chart 12), are highly regulated and require secure handling to ensure privacy is protected. This is particularly important when combining customer data with data from other sources, which may change the nature of the information held about the clients.

Source: Refinitiv AI/ML Survey

Q: Which types of alternative data do you or the teams in your organization commonly work with?

Base: Alternative data (311n)

## Alternative data

Alternative data sources such as the Internet and in particular social media are becoming much more widely used. Chart 13 shows that 97% of those using alternative data use web scraping. This is approximately two thirds of all participants, which is a similar percentage to those using news data. Historically, it has been more difficult to use alternative data, as much of it is unstructured. Fortunately, however, new AI capabilities are making this possible.

What is critical in using machine learning is not just the type of data, but the ability to join together different data sets. The goal of machine learning is to find patterns and relationships between seemingly unconnected phenomena and use the insights to inform decision making. For example, there is currently a lot of excitement in financial services around how to use credit-card data. When this is viewed in conjunction with other data sources, it can be used to uncover trends in industries such as retail, hospitality and manufacturing. Similarly, text-based social media data has the potential to reveal patterns in customer sentiment.

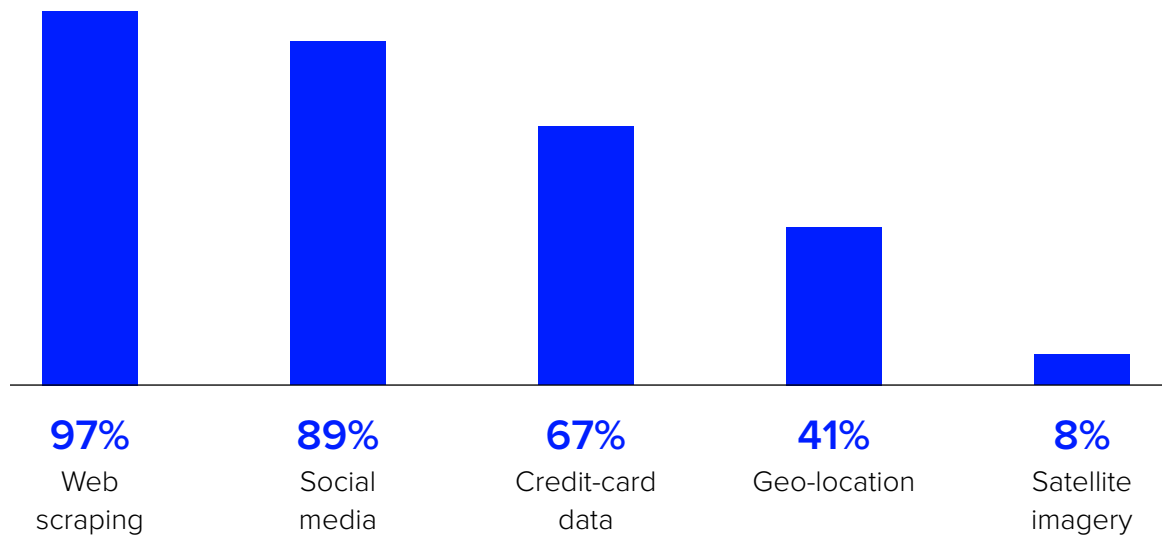
The challenge with much alternative data is that it is not well organized or consistently structured and needs a lot of work to clean it, deal with missing or incomplete records and, in the case of textual data, to apply tagging. Even when the data has been organized, it is still hard to detect repeatable signals. And even where there are stronger signals, it means it is more likely that they will also be detected by other rival organizations looking in the same area.

Another issue with unstructured data sets is consistency over time. Machine learning models need to be tested and deliver replicable results before they can be implemented in operational situations. Additionally, even when alternative unstructured data has been organized and tagged, it typically has not been done consistently for a long enough period of time to provide a high quality longitudinal data set.



CHART 13

## Types of Alternative Data Used in Machine Learning



Source: Refinitiv AI/ML Survey

Q: To which of the following types of company data do you commonly apply ML models?

Base: Company data (353n)

# Why this matters

The opportunity for financial institutions in machine learning is to tap into currently under-exploited sources of data and combine them with existing sources. Although machine-learning capabilities that enable these opportunities are developing rapidly, they are of no use if the data that feeds them is hard to find, poorly curated, dirty and inherently biased.

Biased data and robo-ethics are also growing concerns as ML technology advances, with scientists and researchers realizing that the machine can by matter of course take on the biases of the programmer unless there are checks and balances in place to screen for this. (IEEE explains there is an increasing awareness by researchers and non-researchers alike about the urgent need to understand the social impact and ethical implications of robot technology.<sup>5</sup>) It may not even be happening on a conscious level, but is an inherent consequence of the programming process and one of which data scientists need to be aware.

It is essential that the c-suite understands the challenges data scientists face and invest in the tools and resources for the successful use of ML technologies.

<sup>5</sup> <https://www.ieee-ras.org/robot-ethics>

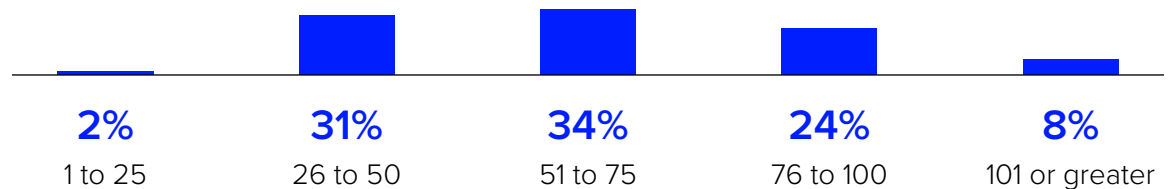
# The c-suite vs. data scientists: vision versus reality

Alongside the growth of machine learning is the growth in the size and number of data science teams in financial organizations. These advanced-degree scientists and engineers are essential for developing and training the models, organizing data, and back testing the algorithms. New roles are being created, such as 'Head of AI' in the same way that Chief Data Officer (CDO) roles appeared in recent years.

Our research shows that there are typically two or three data-science teams in each organization, with an average of 66 employees in each institution who have an element of data science in their job description (see Chart 14). This includes both data scientists and people in supporting roles. Half of the businesses we spoke to expect growth in the number of data scientists in their organization in the coming 12 months; no organizations predicted that they would have fewer data scientists.

CHART 14

## Number of Data Scientists or Employees Involved in Data Science



Source: Refinitiv AI/ML Survey

Q: How many data scientists or employees involved in data-science activities as part of their role are there in total in your company?

Base: All respondents (447n)

The survey comprised data scientists at the coalface working with data on a daily basis, and c-level professionals in roles such as CIO, CDO and CTO. It revealed there were several areas in which their views were not well aligned.

C-level leaders report higher levels of machine-learning adoption than those of data scientists (100% versus 89%). This highlights the degree to which the c-suite believes it is important to be seen using the latest tools and techniques for competitive advantage and may be overstating the company's actual adoption or projecting what is in development. Such sentiment is the job of board-level executives in selling the vision of the organization's future. The remaining 12%<sup>6</sup> of data scientists say they are still experimenting with and investing in machine learning.

This disconnect also shows up in the finding that 47% of c-level interviewees 'strongly agree' (score 9 or 10) that machine learning is a core component of business strategy, as compared to only 31% of data scientists. Similarly, 91% of c-level executives state there is a clear vision around the use of machine-learning technologies versus only 78% of data scientists reporting the same. Data scientists are in a difficult position. Everyone agrees that AI and machine learning are the future, but only data scientists are able to make it happen.

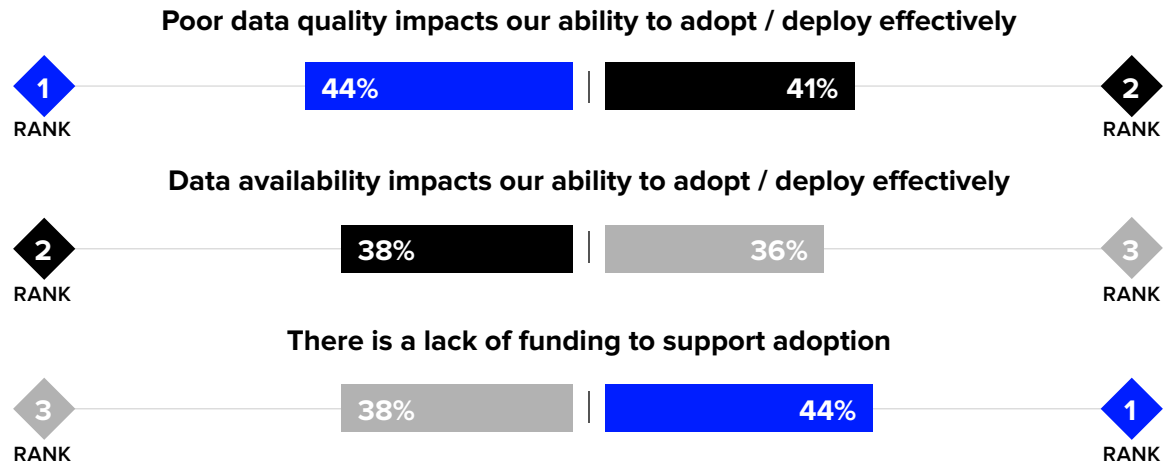
<sup>6</sup> Does not add up to 100% due to rounding (89% say they have adopted and 12% say they are still experimenting).

CHART 15

## C-Level vs. Data Scientist ML Perspectives

**Data Scientists**

**C-Level**



This c-suite/data-scientist disconnect also shows up when looking at how the different roles perceive the barriers to machine learning adoption and deployment, as shown in Chart 15. Participants in c-level roles are closer to the strategic reasons for adopting machine learning and are keen to push developments through, saying funding is their biggest challenge. Conversely, data scientists see more of the day-to-day operational issues. For the same reason, data scientists are less likely to say they are using all types of data, as shown in Chart 16.

Data scientists are more in touch with the reality of what is being done with machine learning on a daily basis. The discrepancy calls into question whether organizations are aligned around what needs to be done to adopt machine learning successfully. If sufficient focus isn't given to establishing clean data, the risk is that the models will fail.

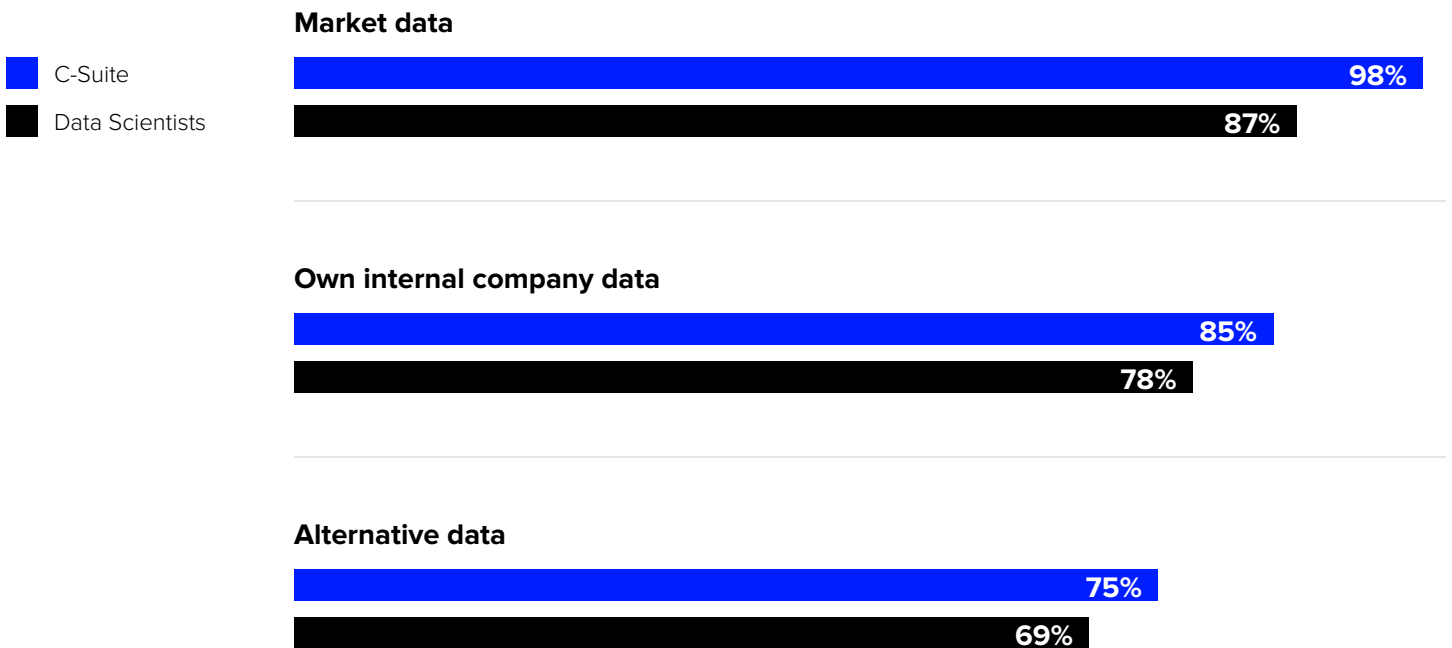
Source: Refinitiv AI/ML Survey

Q: To what extent do you agree these are barriers to adoption, using a scale of 1 to 10, where 1 means 'does not apply at all' and 10 means 'applies completely?'

Base: C-Level (53n); Data Scientists (394n). \*Percentages may not sum up to 100% because of rounding.

CHART 16

## Types of Data Used in Machine Learning, by Role



Source: Refinitiv AI/ML Survey

Q: Do you use or does someone in your organization use machine-learning technologies / techniques with financial-market data?

Q: Do you use or does someone in your organization use machine-learning technologies / techniques with your own company data?

Q: Do you use or does someone in your organization use machine-learning technologies / techniques with alternative data sets such as text, social media, specialist limited-access datasets or images?

Base: C-Levels (53n); Data Scientists (394n)



# Hype and cynicism

There is a lot of hype in the market around AI and associated technologies, and a certain amount of healthy cynicism about how companies in all industry sectors are actually using it. There have been several sayings that have been widely retweeted on this theme, such as, “If it is written in Python, it’s probably machine learning. If it is written in PowerPoint, it’s probably AI,”<sup>7</sup> and, “When you’re fundraising, it’s #AI; when you’re hiring, it’s #ML; when you’re implementing, it’s linear regression.”<sup>8</sup> We can draw a parallel with the hype around blockchain in recent years. However, although blockchain was a technology looking for an application, with machine learning, solutions are already in play in the market.

Research shows that organizations don’t all take the same approach to structuring their data-science teams and it appears that the jury is still out on which is most effective. There’s a clear correlation in the participants’ responses between being more advanced in deploying machine learning and having the work distributed more widely. As machine learning becomes more prevalent, companies tend to take the approach of having data scientists located in functional teams, rather than having a center of excellence for data science. Data scientists are an important component of the tech team and are most effective when connected to the departments and functions they’re supporting.

Chart 17 shows that work in applying machine learning is more commonly undertaken by just a few teams, rather than distributed throughout the organization. Data scientists are highly skilled, and hard to find and hire, so it is critically important to employ them in the most effective way.

CHART 17

## Machine Learning and the Structure of Organizations



Source: Refinitiv AI/ML Survey

Q: Is the work applying machine-learning techniques consolidated in a few teams or distributed across a number of teams?

Base: All respondents (447n)

<sup>7</sup> <https://twitter.com/matveloso?lang=en>

<sup>8</sup> Baron Schwartz, @xaprb



## Why this matters

The disconnect between the c-suite and data scientists highlights a key issue for financial-services organizations. How do you close the gap and ensure that those in the organization who could benefit the most from machine learning properly understand both the capabilities and the challenges? And how do you build and grow the analytics talent while staying connected to the business and aligning the organization around an effective machine-learning strategy? The c-suite needs to understand that machine learning is really machine training—the process requires clean, well-curated data, from a reliable source over a sufficient period of time to test, back test and validate models before they can be used.

# North America and the buy side lead—but not for long

In many areas of the survey, the financial services professionals from North America are more advanced than those in Europe and Asia. They are more likely to say machine learning is core to their business, that their organizations employ more data scientists, and more likely to use all types of data, including alternative sources. Similarly, they are significantly less likely to perceive any of the barriers to adoption that affect professionals in other regions.

While this may seem surprising to some, there are historical and structural reasons why North American professionals are currently ahead. And the expectation is that these reasons are quickly eroding. The biggest financial-services organizations, most sophisticated hedge funds and biggest technology companies have their headquarters in North America; these organizations tend to lead the way in technological developments. The innovation in AI and machine learning has also come largely from North America, out of universities such as Stanford, Berkeley and MIT.

## North America has almost universal use of all types of data

### North America

- 100%** Market data
- 98%** Company data
- 97%** Alternative data

Source: Refinitiv AI/ML Survey

Another reason for this current lead is that the financial market in North America is more homogeneous than in the rest of the world. With traders primarily focused on equities, there has always been an appetite for more data and analysis in order to gain an edge on the competition. In Europe and other markets, trading is more diverse, focusing more on foreign exchange and fixed income, as well as equities, and involves language differences and regulatory complexity.

## 100%

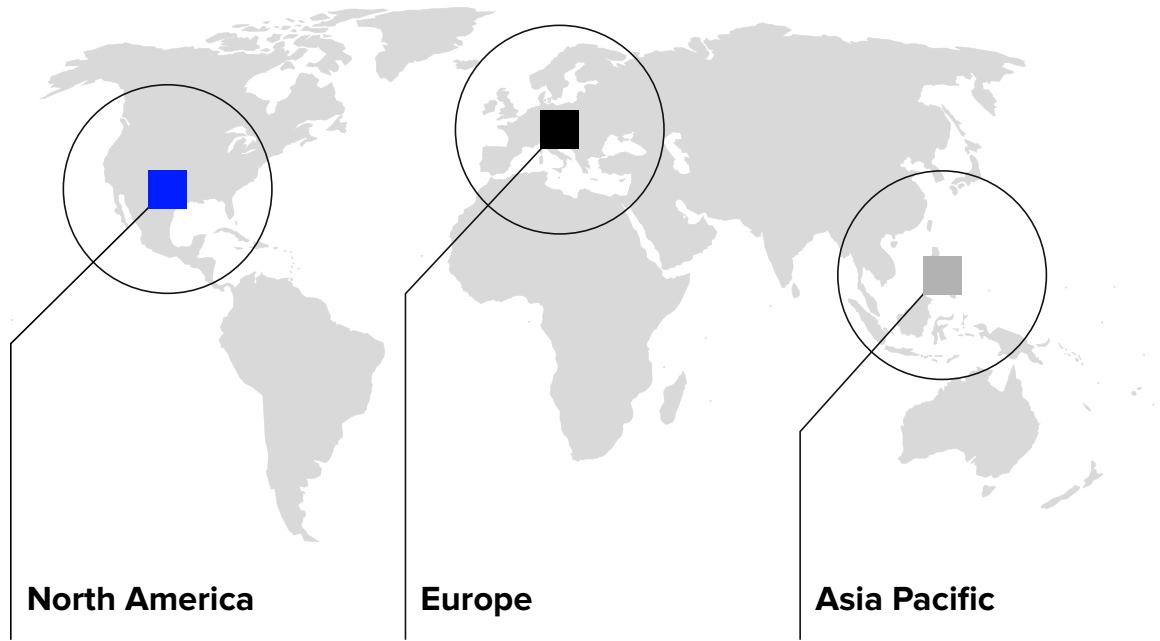
of North American respondents say their primary application for ML is **Risk Avoidance** (versus 82% in Europe and 69% in Asia)

## 95%

of North Americans list **Performance Analysis and Reporting** as the 2<sup>nd</sup> application for ML (versus 68% in Europe and 65% in Asia)

## 71%

in Europe say **Automation** is main ML application (versus 53% in North America and 56% in Asia)



**The use of ML is a core component of our business strategy**

*(Agree: Score 7-10)*



**We make significant investment in ML**

*(Agree: Score 7-10)*



**Number of data scientists in your company will increase in next 12 months**



**Adoption of ML: 'Deployed and it is core to business'**



**There is a clear vision around the use of ML technologies in the organization**

*(Strongly agree: Score 9-10)*



Source: Refinitiv AI/ML Survey

# Buy side vs. sell side

It is not a surprise, therefore, to find that the buy side leads the way over the sell side in this research. Professionals on the buy side are more likely to see machine learning as core to strategy, have machine learning distributed throughout the business and see significantly fewer barriers to adoption than those on the sell side. The buy side has long used data as an advantage and has employed quantitative analysts whose role it is to build models and algorithms to interrogate various sources of data to look for

patterns and signals that can inform buying decisions and gain a competitive edge. Simply put, the buy side has been better equipped to move ahead with machine learning, not only in the skills of the existing quantitative analysts, but also because of the buy side business model. The margins and commission are greater than on the sell side, linked to the potential returns for customers, and it is critical for the buy side to demonstrate to customers that they are doing everything they can to drive up return on investment.

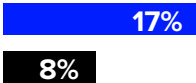
## The use of ML is a core component of our business strategy

*(Strongly Agree: Score 9-10)*



## The use of ML is a core component of our business strategy

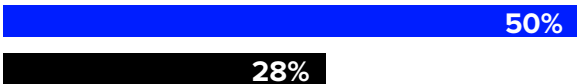
*(Strongly Agree: Score 9-10)*



## Adoption of ML: 'Deployed and it is core to business'

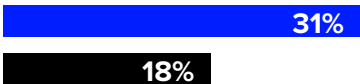


## Data quality: say the barrier does not apply at all



## There is a clear vision around the usage of ML technologies in the organization.

*(Strongly agree: Score 9-10)*



Source: Refinitiv AI/ML Survey



## Why this matters

What is interesting in this research is not the degree to which North America and the buy side lead, but the degree to which machine learning is now being used in the rest of the world, and on the sell side. Machine learning is a tool to give financial institutions competitive advantage, as is being witnessed on both the buy and sell sides, and will likely continue to grow in significance for both groups.

Machine learning is not just an industry fad. It is a tool that is being strategically used across the globe by a multitude of financial organizations in order to gain competitive advantage and better serve customers.

# Moving forward with machine learning

Whether your organization is on the buy side or sell side, located in Asia, Canada, Europe or the United States, and regardless of your role, the use of AI and machine learning are important parts of the financial-service industry's technological transformation. All indicators point to the fact that their inclusion in your operations will have profound effects.

It's no longer 'if' your organization should be utilizing artificial intelligence tools such as machine learning, it's 'to what degree' they should be applied. Companies that hesitate will ultimately find they're behind competitors and could have to scramble to maintain or acquire their edge.

It is essential to ensure you are working with a partner that provides a secure, scalable platform and that you have access to comprehensive, well-curated and clean data. Garbage in undoubtedly leads to garbage out. Ferraris don't run on water or crude oil.

The quest for quality sources of data to use in machine learning is a reality that's likely to continue into the foreseeable future. While it is important to be on the lookout for new data assets, this can sometimes be like searching for a needle in the haystack. It is equally if not more important to ensure the data you are using is of the highest quality possible.

For in the end, data is just the beginning.

# Data is the beginning of your AI/ML journey—and Refinitiv’s data is widely used for machine-learning initiatives

## News & Commentary

- Financial News
- Global and Domestic News
- Market Commentary & Sentiment
- Significant Developments
- Newsletters
- Video
- Commodities Research & Forecasts
- Specialist News like IFR

## Macro-economic Data

- Country Data
- Economic Indicators and Polls
- Industrial Activity

## Market Data and Pricing

- Benchmarks & Fixings
- Equities
- Fixed Income
- Foreign Exchange
- FX and Interest Rate Polls
- Commodities & Energy
- Cryptocurrencies
- Derivatives (Futures & Options)
- Evaluated Pricing
- Global Aggregates
- Indexes
- Loans Pricing

## Reference Data

- Index Constituents and Weightings
- Industry Classifications
- Security Identifiers
- Fixed Income Terms and Conditions

## Specialized Data

- Commodities Fundamentals, Pricing & Indices
- Deals & Transactions Intelligence
- Mutual Fund Data (Lipper)
- Quantitative Analytics and Models
- Private Equity Data

## Risk, Regulation & Compliance

- Financial Crime Prevention
- KYC as a Service
- Know Your Client (KYC)
- Risk Screening
- PEPs
- Sanctions
- World-Check
- Regulatory Compliance

## Company Data

- Broker Research & Content
- Business Classifications
- Credit (CDS)
- Company News
- Competitors
- Corporate Actions
- Debt & Syndicated Loans
- Entity Risk (Corporate Structures)
- ESG Data (Rankings and Ratings)
- Estimates
- Events & Transcripts
- Fundamentals
- Filings
- M&A
- Officers & Directors
- Ownership & Bond Holding
- Organization Data
- Private Company Data
- Shareholder Activism Intelligence
- StarMine® Scores
- Transactions
- Valuation



# Meet the Labs



Refinitiv's Innovation Labs are in London, New York, San Francisco and Singapore, alongside other regional tech hubs. You can supplement your organization's technological transformation with the Labs' team of talented data scientists and engineers, at the ready to help you tackle your AI/ML challenges and give your business a competitive edge.



**Tim Baker**  
*Global Head  
of Applied Innovation*



**Daniel Mattioli**  
*Head of Labs Engineering*



**Tim Gaumer**  
*Director, Innovation,  
Fundamental Research*



**Sanjna Parasrampur**  
*Head of Innovation Lab -  
Singapore*



**Geoffrey Horrell**  
*Head of Innovation Lab -  
London*



**Joe Rothermich**  
*Head of Innovation Lab -  
San Francisco*



**Daniel Lewington**  
*Head of Product Design*



**Amanda West**  
*Global Head  
of Innovation Enablement*

## About Refinitiv

Refinitiv is one of the world's largest providers of financial markets data and infrastructure, serving over 40,000 institutions in over 190 countries. We provide 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](https://www.refinitiv.com)

