LSEG LABS

The defining moment for data scientists

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING READY TO CHANGE FINANCE

Contents

Introduction

Data changes how we see the world and how we act. This research highlights how rapid acceleration in data-driven transformation, resulting from the Covid-19 shock, is really just a glimpse of what's yet to come.

Geoff Horrell
Head of Group Innovation
London Stock Exchange Group

Hanna Helin
Head of Technology Innovation and Ecosystems
London Stock Exchange Group

The future of financial services is distributed and tech-enabled, with data at its core. This in-depth research with global data practitioners shows us how artificial intelligence and machine learning (AI/ML) are advancing. Of the companies that took part in this research, 80% now apply deep learning in their business.

We see a new world view with exciting opportunities only the data-centric can seize. More companies than ever are relying on alternative sensor (41%), shipping (40%) and supply chain (46%) data to unpack abrupt trade and economic patterns.

However, the gap is growing between businesses capable of integrating high-quality data into production-grade AI/ML products and services, and those that aren't.

Nearly half (45%) of companies globally are accelerating away from the others with more AI/ML deployments, broader talent plans and clearer data and technology strategies, leaving 55% that are stuck in an experimental phase and at risk of falling further behind if they don't take the leap to scale.

We hope you find this report to be a useful read as you grow your own AI/ML strategy, move towards scale and overcome challenges along the way.

As a diversified global financial markets infrastructure and data provider following the acquisition of Refinitiv, LSEG (London Stock Exchange Group) is open to sharing its 300-year experience and deep knowledge to help businesses and economies around the world to innovate, advance and grow.

Ten key themes to explore

01

The time is now to make the leap and scale AI/ML

02

Finance firms move to integrate unstructured data into their models

03

Embracing alternative data ramps up data access and pipeline-building issues

04

Data quality and availability are still stubborn issues, but tech is becoming more of a barrier

05

Deep learning is now the favoured type of AI/ML

06

Diversity in data science sees role span model lifecycle and strategy

07

Data engineering and MLOps skills are essential for success

08

Model governance focus is on reducing costs and improving model quality

09

Stakeholder trust and ethics in AI/ML has a long way to go

10

North America loses its AI/ML lead as EMEA and Asia-Pacific catch-up

Research methodology

LSEG surveyed hundreds of data scientists working across the financial industry

The purpose

The AI/ML study is an annual independent survey providing a unique insight into the thinking and work of data leaders and practitioners in financial services. 2021 themes include:

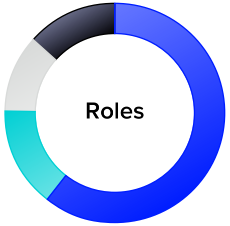
– Level of AI/ML adoption and use cases
– Data strategy, including which data is used
– The changing role and influence of data scientists
– How companies are hiring and cultivating talent
– Platforms and tools used by the data science community
– A comparison with previous years' findings

The data

The survey took place between 3 May and 16 June, 2021, based on 482 telephone interviews. Respondents include:

– Data scientists, quants, model governance professionals and C-level executives
– A combination of sell-side and buy-side firms with revenue in excess of $1 billion
– A mix of geographies across North America, EMEA and Asia Pacific

Profile of survey respondents

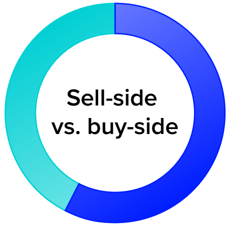


293
Data scientists
Data engineer
Data analyst
Data scientist
Head of AI
Head of data science

54
C-level
Data officer
Information officer
Technology officer

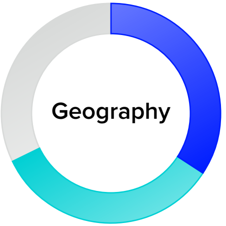
71
Quants
Quant analyst
Quant developer
Quant researcher
Algo trader

64
Model governance
Senior/model governance officer
Director of model governance
Senior manager, model risk governance



278
Sell-side
Commercial or retail bank (81)
Investment bank (111)
Broker-dealer (82)
Exchange (4)

204
Buy-side
Asset management (82)
Hedge fund (13)
Wealth management (49)
Private equity/venture capital (12)
Corporate commodities (13)
Insurance (32)
Pension funds (1)
Mutual funds (2)



165
Asia Pacific
India (17)
Singapore (29)
Australia (13)
China (43)
Japan (32)
Hong Kong (31)

155
North America
US (89)
Canada (66)

162
EMEA
UK (45)
France (43)
Germany (39)
Switzerland (18)
Netherlands (17)

CHAPTER 1

The defining moment: time to take the leap

Over half (55%) of companies are yet to deploy AI/ML across their business. What will enable them to make the big leap to scale?

- AI/ML adoption remained steady this year, with a similar number of companies deploying in multiple areas compared with 2020 and 2018
- Most use-case growth is in traditionally data-native departments, such as operations, risk and portfolio management
- Companies are eager to improve and personalise the customer experience, but are concerned over regulation
- Nailing end-to-end processes that deliver on AI/ML's promise remains a challenge, and budget concerns are rising
- Firms are increasingly confident about their AI/ML progress, becoming less concerned of competition

Figure 1: **AI/ML adoption remains steady**

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

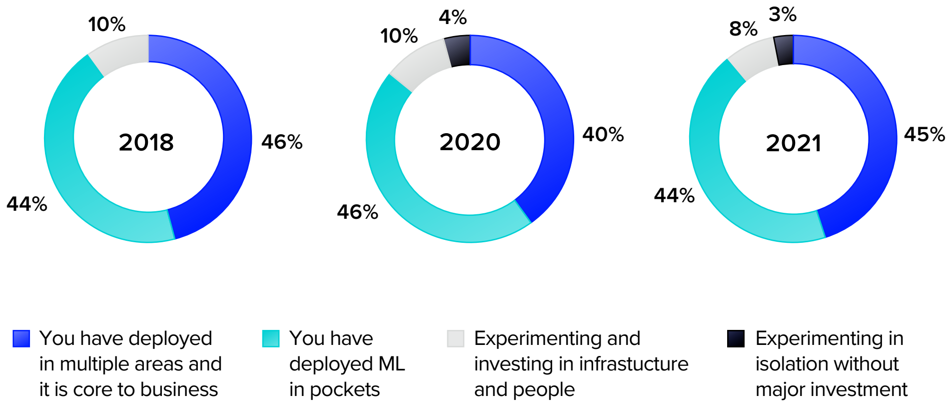
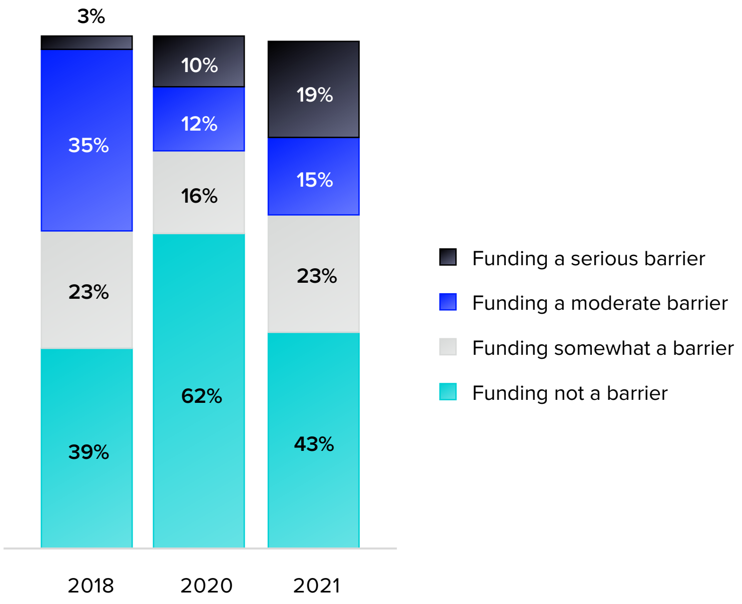


Figure 2: **Data scientists feel the need for increased budget**

To what extent do you agree funding is a barrier to adopting new AI/ML technologies/techniques in the organization, where 1 means 'does not apply at all' and 10 means 'completely applies'?



What are data scientists achieving?

The diversity in use-cases that data scientists are working remains high. Compared with last year, AI/ML techniques are being applied at similar levels to a wide range use-cases, with increases most notable in operational risk, reporting and compliance, and portfolio management. These are all business areas that have traditionally relied on data-centric employees, such as quants and senior analytics professionals.

A similar picture emerges when looking at where data scientists are seeing revenue gains or cost savings. Companies are more easily extracting value in business segments that were predisposed to adopt AI/ML. Generally, respondents report similar levels of cost savings and revenue gains – and most are reported in traditionally data-heavy parts of the business.

Figure 3: **Top use-cases and gains are in data-native departments**

Which of the following are focus areas in terms of applying AI/ML techniques?

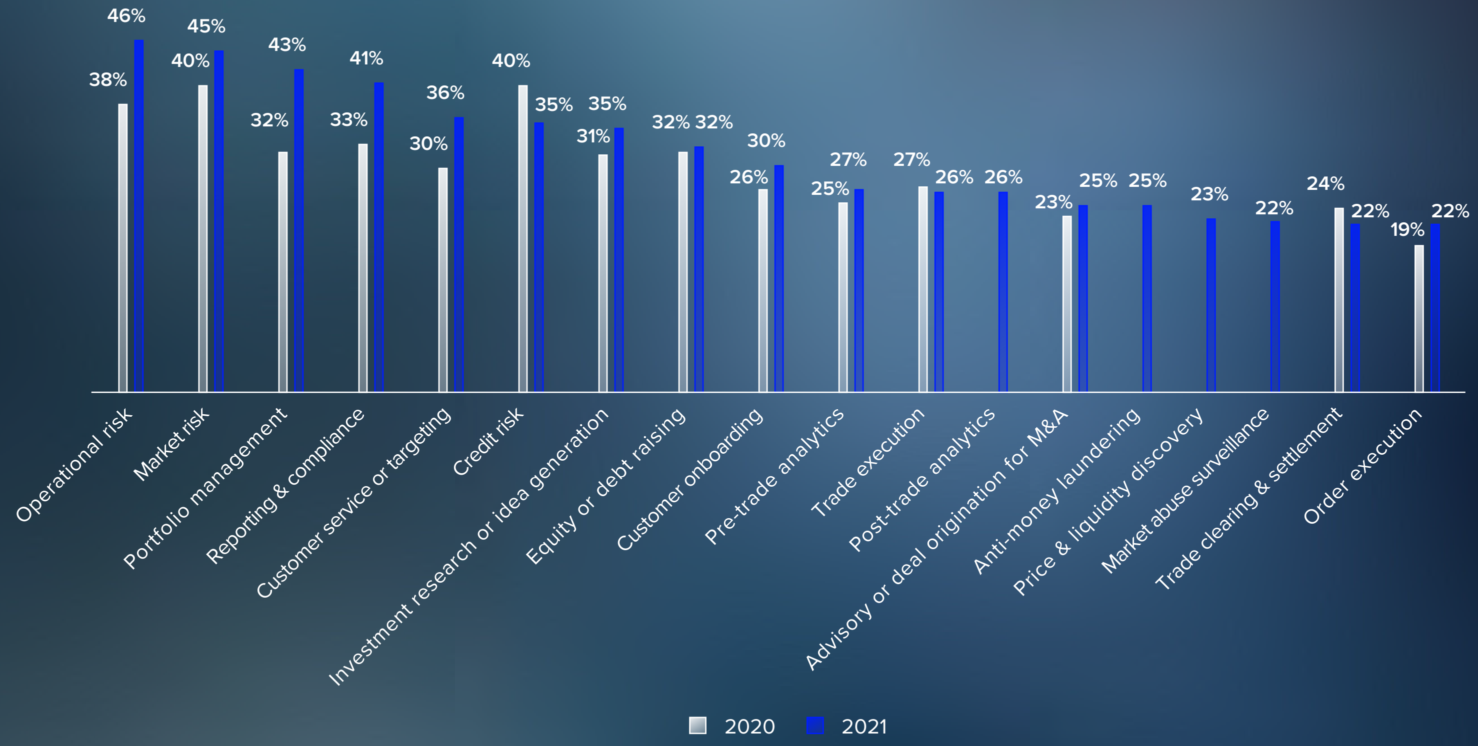
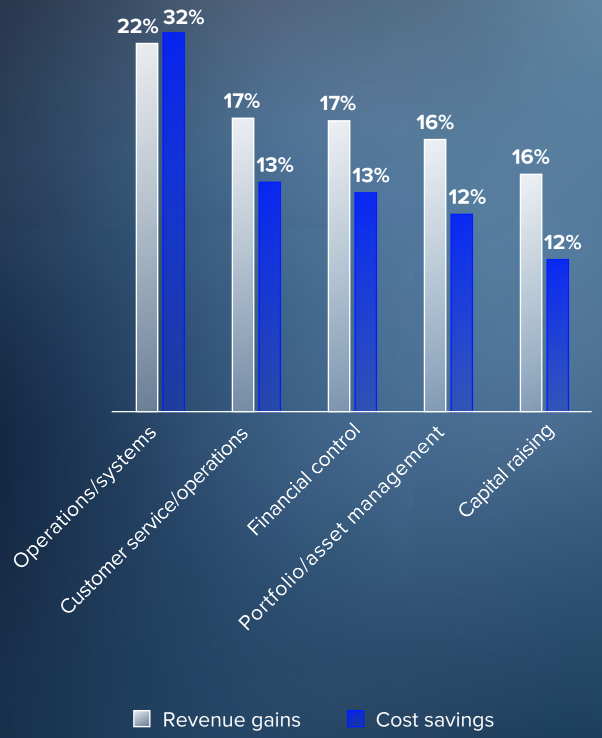


Figure 4: **Data-native departments are reaping the rewards**

In which areas of the organisation, if any, have you seen revenue increases and costs savings from AI/ML?



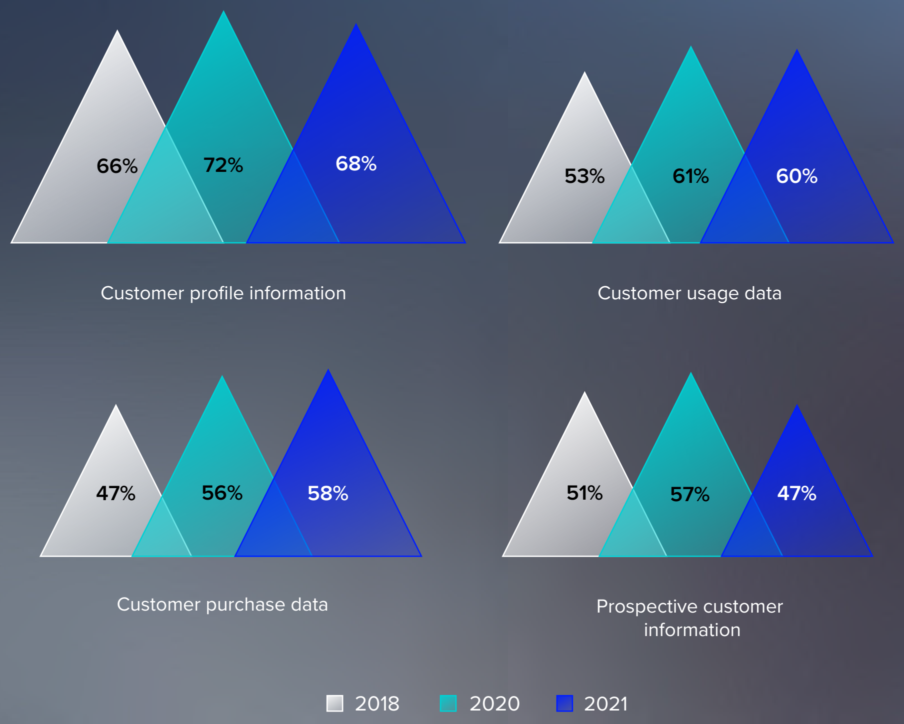
Improving the customer experience

Consumer applications have influenced expectations for enterprise services, and financial services clients now expect an improved and personalised experience. Accordingly, 53% of survey respondents cite improving the customer experience as a top priority for the coming two years.

Having seen AI/ML pay off in some use-cases, such as operational risk mitigation, the industry sees an opportunity. However, use of customer data is flat or falling, as data privacy regulations (such as GDPR) remain a challenge. In addition, leaping towards customised solutions concerns strategic decision-making: what exactly is the type of personalised experience a company wishes to solve for?

Figure 5: **Customer experience a top priority but customer data is flat or falling**

Which of the following types of company data do you commonly apply in AI/ML models?



> We are seeing more and more financial institutions using chat bots. These bots are used to interact with customers without human intervention, but in the future these bots will be able to collect massive amounts of data based on customer behaviour and habits, and learn or study customer behaviour.

Head of data science operations,
Canadian brokerage firm

AI/ML remains a top priority

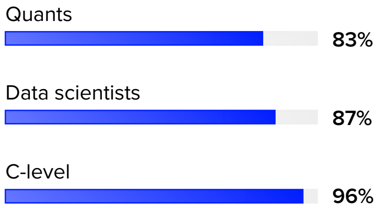
All participant groups report their company intends to continue investing in AI/ML. Evidence throughout the survey suggests this goes beyond lip service: team sizes are growing, use of data is expanding, and companies are investing in tools and technology.

C-level participants are overwhelmingly committed. Their awareness of future spending plans (to which data scientists or quants may not be privy) helps explain the perception gap as one between forward thinking and current execution.

The challenge of building robust pipelines for ever-growing data types is a second contributing factor to this perception gap. It helps explain why quants and data scientists, who have long worked with data in different, more established frameworks, are less bullish than their leaders.

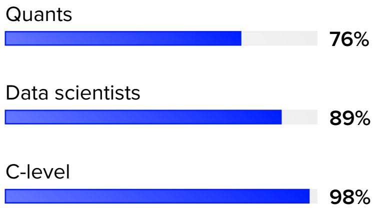
C-level participants are overwhelmingly committed, but quants and data scientists are less bullish.

Figure 6:



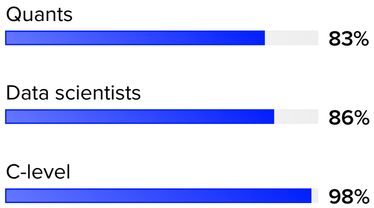
% who think their business has a formal process in place to evaluate AI/ML

Figure 7:



% who think AI/ML is a core component of their business strategy

Figure 8:



% who think their business invests significantly in AI/ML

Most companies believe they're ahead of the pack

Despite the various challenges data scientists face, many feel a strong sense of pride in their frontier work – and in how it positions and advances their company. On average, 63% of respondents see their company as either an AI/ML industry leader or challenger, while only 2% consider their company an industry laggard. Companies employing many data scientists see themselves even more favourably.

In practice, few companies have yet to emerge as leaders or challengers, especially with 44% still deploying AI/ML in pockets (**Figure 1, page 7**). This suggests a degree of dissonance between how advanced companies are relative to average industry progress.

Notably, data scientists are feeling more confident about their companies' competitive advantage compared to their colleagues. In a recent LSEG study with bankers, portfolio managers and analysts, only 50% saw their company as either a market leader or a challenger.

Confidence levels may also explain why staying ahead of the competition appears a markedly lower priority this year. Only 43% cite it as priority, down from 57% last year.

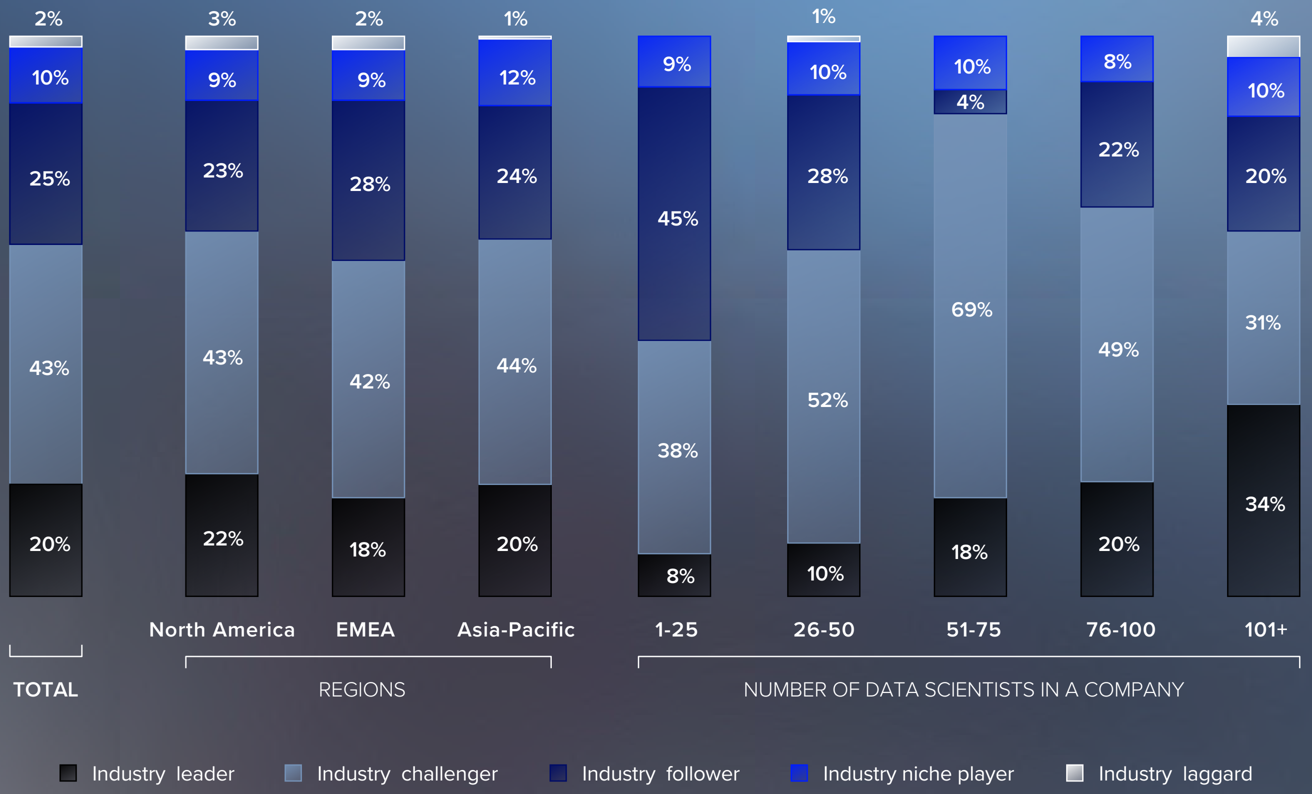
> Most of our business areas have integrated advanced AI processes and data management processes.
>
> **Chief data officer**
> German bank

> As this (AI) is still an unexplored field, most companies are now moving towards full-scale implementation.
>
> **Head of data science and machine learning**
> Japanese insurance company

> The adoption is uneven across companies. The gap between early adopters and us is only widening.
>
> **Quantitative researcher**
> Dutch investment bank

Figure 9: **Data scientists feel they are beating the competition**

How would you categorise your organisation?



Key takeaways

1. Remember that end-to-end processes underpin achievable AI/ML visions

Ensure processes designed to deliver AI/ML products are carefully considered, aligned to a clear strategy and supported by the requisite AI/ML talent, who can help guide decisions being made across the company.

2. Embed design thinking into AI/ML projects for the best customer experience

Understanding how your customers interact with the output of AI/ML systems should guide how your models are built and optimised. Explainability, interpretability and trust in models will be critical as AI/ML scales and becomes more sophisticated.

3. Understand whether you are really an industry leader

Compare your team and your wider business to a respected analytic maturity model, like Robert L. Grossman's 'Framework for evaluating the analytic maturity of an organisation'. Comparing both team and business will give you a truer benchmark, as your team may be doing advanced work, but the wider company may not yet be an AI/ML-driven organisation utilising the technology across the business.

CHAPTER 2

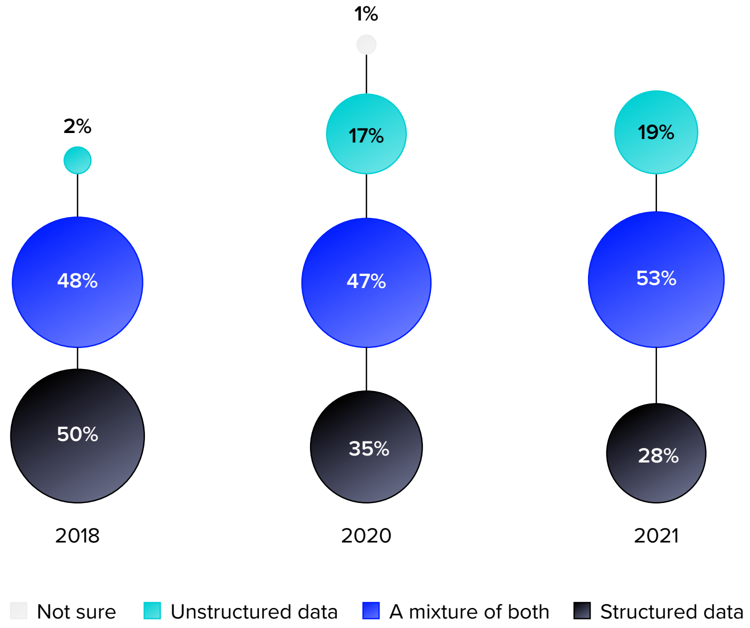
Data-hungry: firms diversify their data

The industry's appetite for data is only growing, but managing and connecting data assets is a major challenge

- More companies are now working to integrate unstructured data into their models, and fewer work exclusively with structured data
- Poor data quality and lack of data availability remain significant barriers
- Firms are turning to alternative data to supplement financial data
- Companies are spending more time annotating data. Where data quality is an issue, a substantial majority are also using synthetic data
- Most firms lack processes to manage data assets, leading to multiple issues around connecting and aligning different data sets

Figure 10: **More companies now rely on structured and unstructured data**

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



The data sets firms are using

Adoption levels of financial data sets are steady, both for quantitative data sets and standardised textual data. What comes across strongly this year is that companies are turning to less standardised data sets to build a competitive advantage. This is evidenced by use patterns emerging for alternative data. In 2018, 30% of companies didn’t use alternative data; this year it was only 1%.

Compared with last year, standardised credit card data became much less attractive (dropping to 18% from 51%). By contrast, use of specialised data sets – such as app installs, Web traffic and job listings – has leapt. The abrupt change in trade and movement data due to Covid-19 has led to a surge of interest in tracking data, ranging from supply chain and shipping to sensor data and satellite imagery.

Figure 11: **Use of financial data is stable**

Which of the following types of financial markets data does your company commonly apply in AI/ML models?

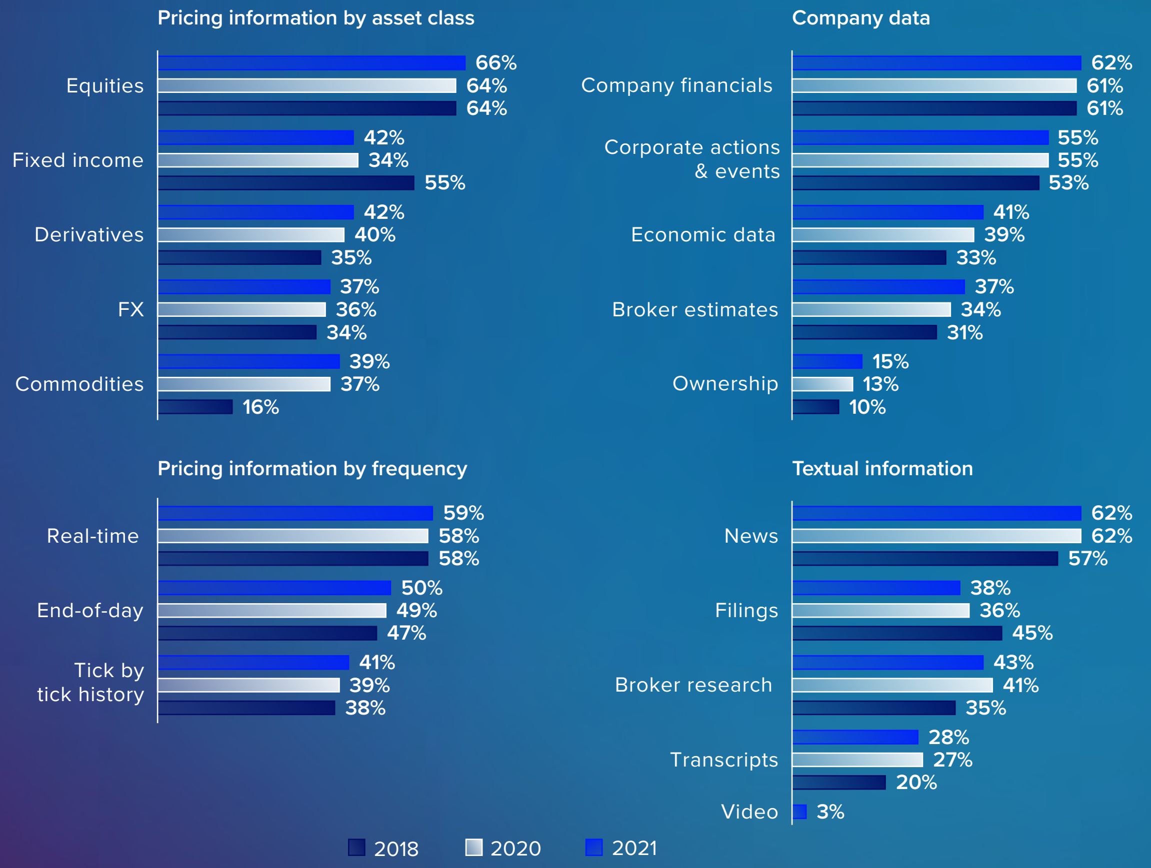
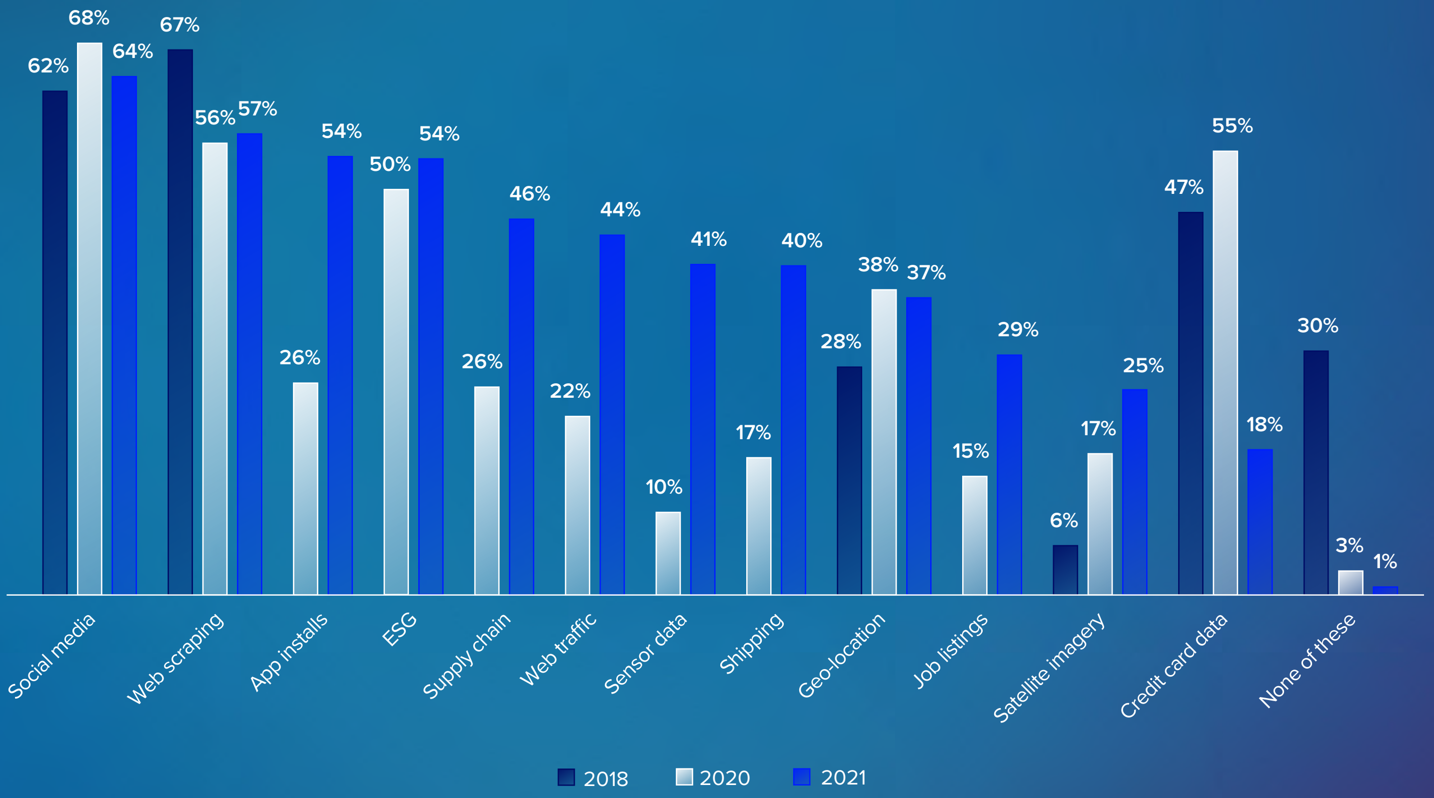
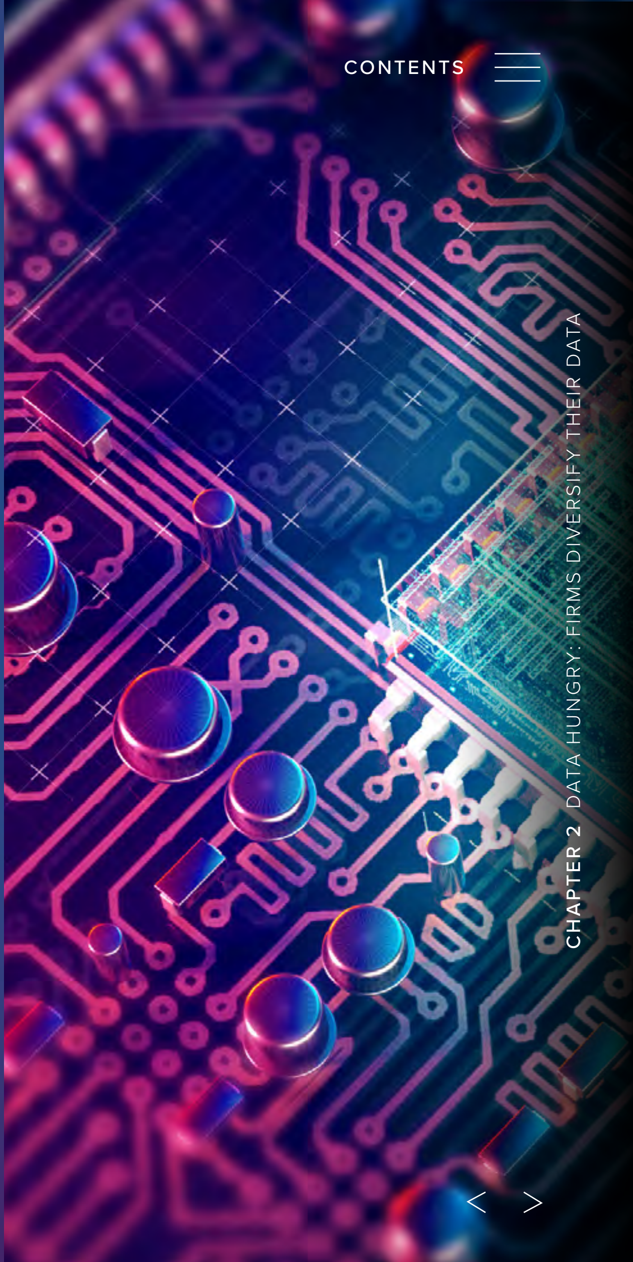


Figure 12: **While use of alternative data is growing**

Which of these types of alternative data does your company use in AI/ML models?





Data annotation and synthetic data

Answering a question about barriers to AI/ML adoption, 40% of the survey's respondents cite poor data quality and 43% lack of data availability. These dual challenges are leading many in the financial sector to seek two types of creative solutions.

One, companies have increased their use of data annotation. 39% of respondents say creating or obtaining sufficient annotation is a key challenge, up from 30% last year. This increase may also be directly related to more companies using unstructured data sets and NLP.

Two, 84% of those citing data availability issues say they use synthetic data. The exceptional circumstances instigated by the pandemic played a part in the push to synthetic data, which was already becoming more prevalent with newer techniques and better hardware utilising more data.

Companies needed to adjust models trained on pre-Covid-19 data to account for post-pandemic implications, but also to attempt to predict new outlier scenarios to test models on.

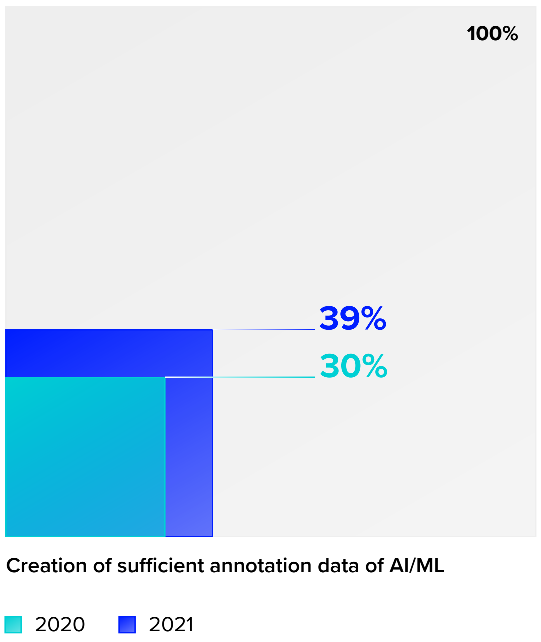
These two tactics suggest companies are moving from plug-and-play to extracting more subjective qualities of the data, such as sentiment or intent.

Figure 13: **Data quality and data availability pose dual barriers**

To what extent do you agree these are barriers to adopting new AI/ML technologies/techniques in the organisation, where 1 means 'does not apply at all' and 10 means 'completely applies'?

Figure 14: **Data annotation becomes a pressing challenge**

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



Barriers to expanding use of data

Embracing new data sets leads to a spike in challenges around accessing, using and handling data.

Linking and reconciling different data sets is an acute problem. Its sharp rise coincides with companies using more data sets and adding non-standardised data to the mix.

Managing data ownership is also proving to be a challenge, leading to issues around data access.

Permission management for data is a concern for over a third (34%) of respondents, doubling from 17% last year.

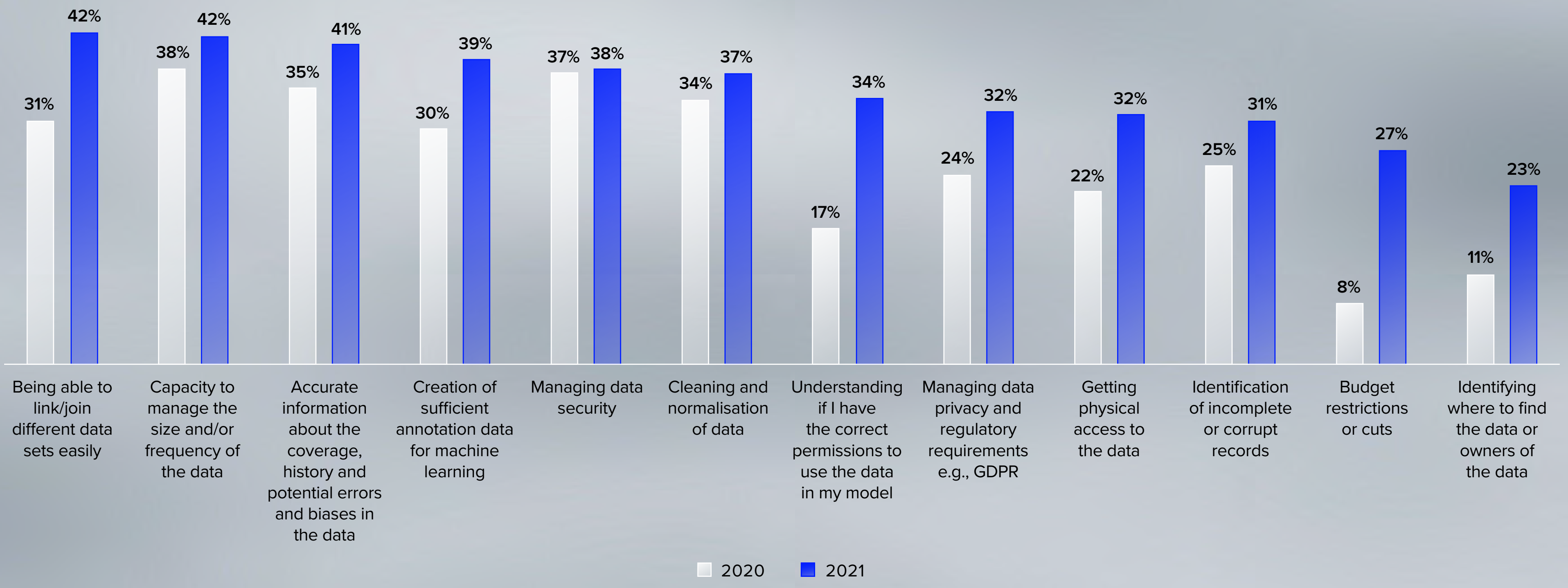
Since models have built in assumptions on the data records they process, incomplete or corrupt data can derail an entire AI/ML pipeline.

As 27% cited budget restrictions or cuts this year – compared with only 8% last year – this finding suggests companies are underinvesting in processes to scale the amount of data sets they use.

> “Companies will have to think about how they manage and store data. With the use of AI/ML everything is becoming bigger. Bigger data sets, bigger computers and bigger neural networks. It will be difficult to move to the next stage if we don’t train our systems to manage data well.
>
> **Quantitative analyst**
> Japanese investment bank

Figure 15: **Challenges of working with new data are becoming more acute**

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



Key takeaways

1. Let’s stop complaining about data quality and recognise it as a discipline

The perennial issue of ‘data quality’ can be misleading, as systematic data problems will always exist. One person’s ‘quality’ issue can also be another’s pattern or insight. The complexity involved in delivering reliable data for AI/ML is a valuable discipline in its own right - from sourcing data with trusted legal agreements and specifying API formats and standards, to employing people with deep expertise who can own and manage sourcing relationships and the data sets day to day.

2. Not all data annotation tasks are created equal

Depending on your data set, annotation can range from a simple to fiendishly complex task that could add weeks, sometimes months, to a project. This is especially true in fields such as NLP, where the complexity and subjectivity of annotation can be even more costly and time-consuming. Teams should be prepared to factor in annotation time to projects and also consider exploring the active learning landscape, which could see users iteratively teach models to label data.

3. Don’t underestimate the ability to link data

Linking data is crucial to finding insight in new or alternative data sets, especially if you’re linking new and fundamental data that is unlikely to offer value in isolation. Linking should also be factored into the cost of onboarding new data, as data science teams often have to revisit traditional data sets to ensure they are easy to access and link. The linking space is one to watch, as open identifiers, like Refinitiv’s PermID, begin to offer advanced search and discovery across company documents and data.

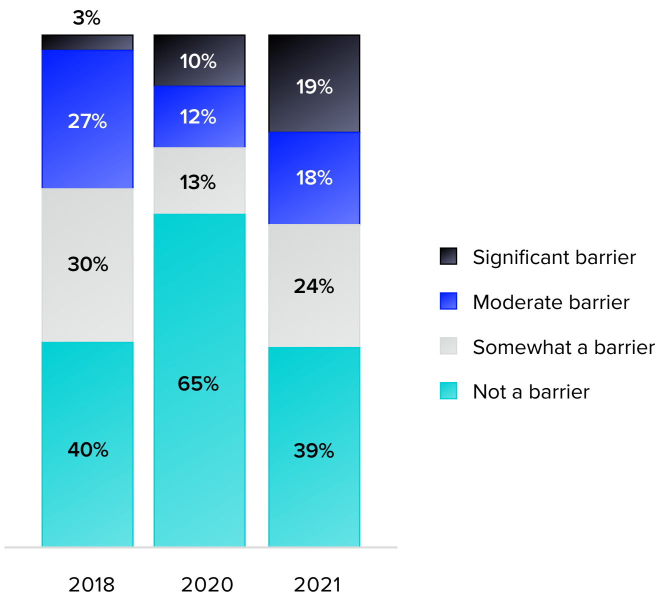
CHAPTER 3 Technology choices: tools and processes changing rapidly

In a complex, fragmented and rapidly changing technology landscape, firms are weighing up which core competencies to build and how they partner with external vendors

- Financial services firms are migrating to cloud infrastructure – and are favouring the market's largest players
- There's growing buy-in for outsourced AI/ML services, with demand greatest for third-party vendors that integrate with internal systems
- NLP has become a core focus area and many firms choose to use APIs by the large cloud providers
- Financial services no longer prefer supervised learning. Deep learning dominates, with reinforcement and unsupervised learning growing rapidly

Figure 16: **Technology is becoming more of a barrier**

To what extent do you agree technology is a barrier to adopting new AI/ML technologies/techniques in the organisation, where 1 means 'does not apply at all' and 10 means 'completely applies'?



Building in-house capabilities

The global growth in AI/ML adoption presents financial services companies with a multitude of choices around whether to build capabilities in-house or rely on external applications. This year, only 37% of companies reported building the majority of their capabilities in-house.

However, buying full solutions from a vendor was still a minority choice. This year, our findings show most companies prefer to integrate third-party solutions into their internal applications, and nearly half choose to use open source components. Strikingly, firms with over 100 data scientists were more likely (36%) to buy a full solution from an external vendor, while those with less than 25 data scientists were nearly certain (82%) to opt for third-party solutions.

Automation and work elimination are also coming into sharper focus, with 16% of respondents citing them as major trends for the next two years. Two forces are at play. On the one hand, data scientists now have machine learning operations (MLOps) technology like Amazon Sagemaker, meaning they rely less on technology teams. It's a massive win for data scientists, allowing them to be much more involved in the build cycle to make projects faster and more aligned to production.

On the other hand, companies see an opportunity to improve financial performance, by reducing fixed costs in business units requiring a high degree of repetitive work.

Figure 17: **Companies are now more likely to outsource tech**

When your organisation is looking for new capabilities to support your models, what proportion is built within the organisation?

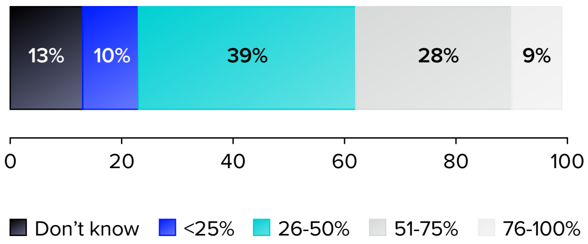
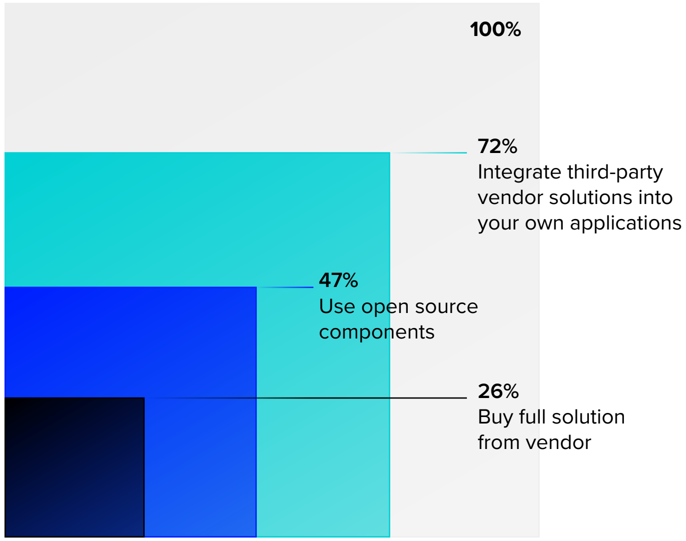


Figure 18: **Most companies integrate third-party solutions**

Which of these ways do you typically use external solutions?



Choice of AI/ML cloud providers

Data science as a service (DSaaS) is going mainstream. Cloud providers and hardware specialists are enabling companies to build quicker and maintain workflows. And the major providers are only growing more popular, to the point that only 1% of respondents say their company doesn't use external compute resources.

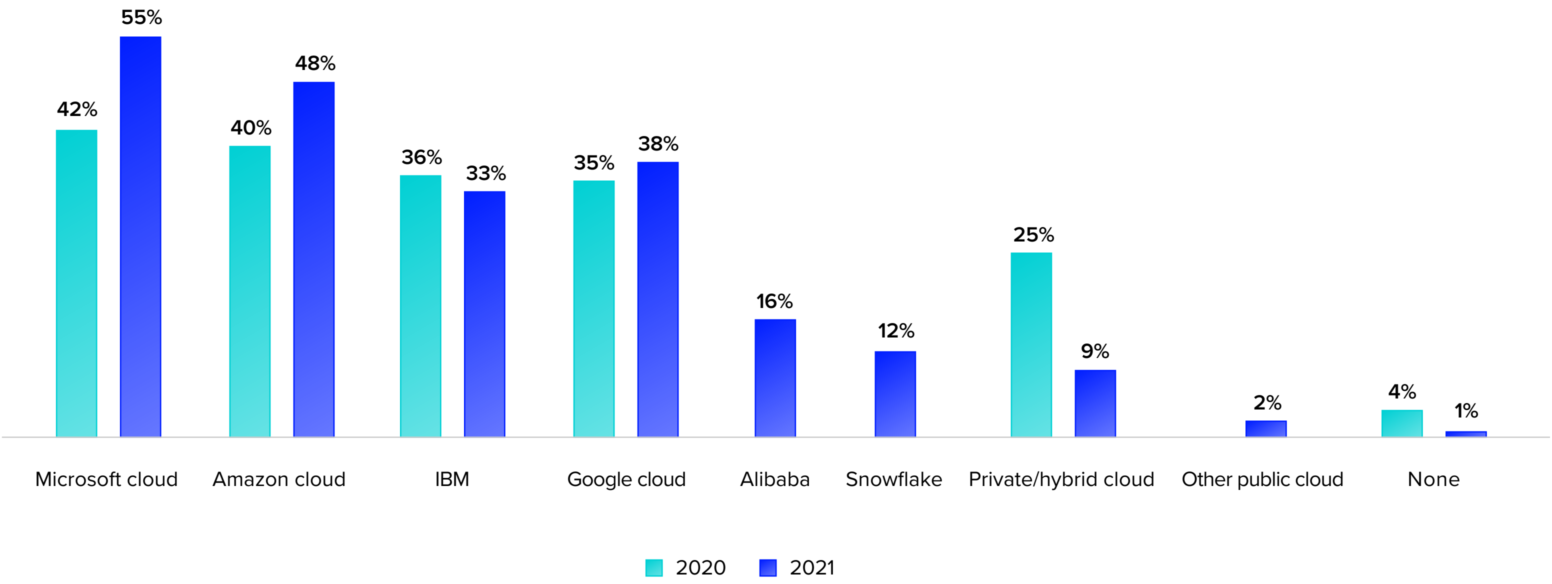
This year, far more companies report using Microsoft and Amazon's cloud services, and this growth comes mostly at the expense of private or hybrid cloud providers.

> For 2021-2022, I see more application of cloud computing to machine learning algorithms. Machine learning tasks can be time-consuming and with cloud computing these tasks can be sped up significantly.

Data scientist
Japanese Investment Bank

Figure 19: **Cloud providers – Microsoft and Amazon make significant gains**

Which, if any, cloud providers do you use to run models?



NLP is going mainstream

With growing appetite for unstructured data, this year we see NLP move front and centre to become a core focus area.

Companies also rely heavily on cloud providers for off-the-shelf text analysis. Notably for each of the three leading providers – Microsoft, Google and Amazon – respondents were far more likely to use their NLP services when already using the provider for cloud services.

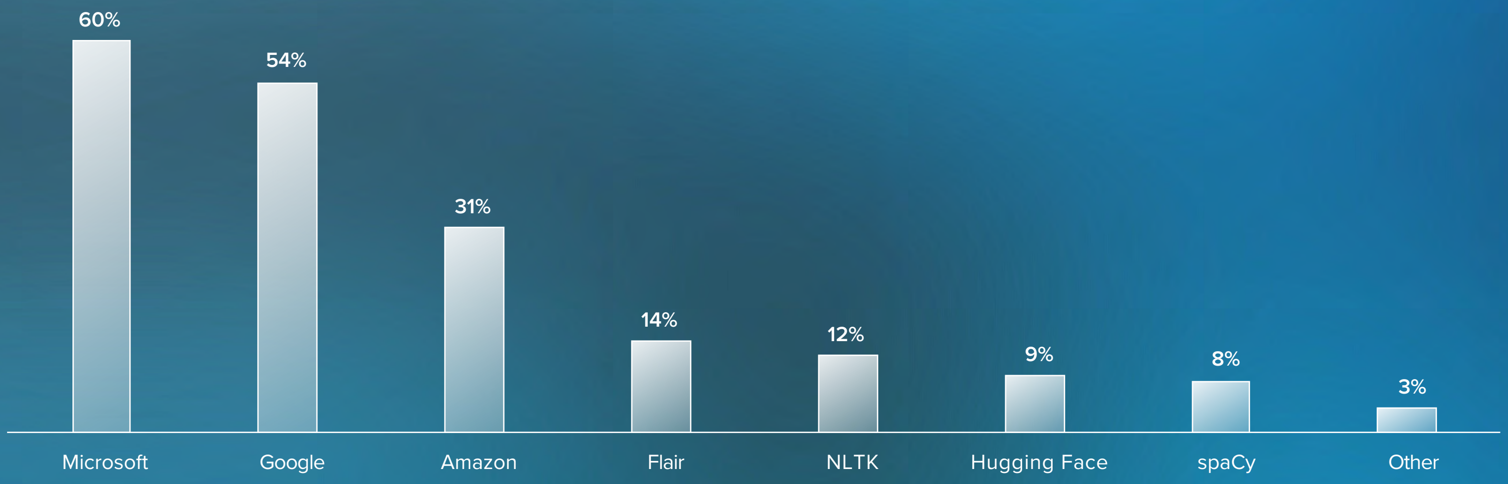
At the same time, companies are now far more willing to bring open source into core parts of their business. Our survey finds Flair, NLTK and spaCy are the leading open source libraries used to analyse and process unstructured data.

Nearly one in 10 respondents cite using Hugging Face, an open source tool that enables companies to experiment with tweaking large, pretrained models that require significantly less annotated data than training a model from scratch.

> “We are focusing on data mining and NLP to improve productivity and user experience.
>
> This will also help us improve our revenue and reduce costs drastically.
>
> **Head of digital innovation**
> Singaporean bank

Figure 20: **Microsoft and Google lead with tools for text**

What tools do you use to analyse text?



A push into deep learning

The commercialisation of deep learning, both through third-party APIs and off-the-shelf services from cloud providers, have led to a robust growth in adoption. Our previous surveys showed supervised learning dominating AI/ML applications, but this is no longer true.

Firms are also experimenting with other types of learning, with the biggest growth in adoption seen for reinforcement learning and unsupervised learning.

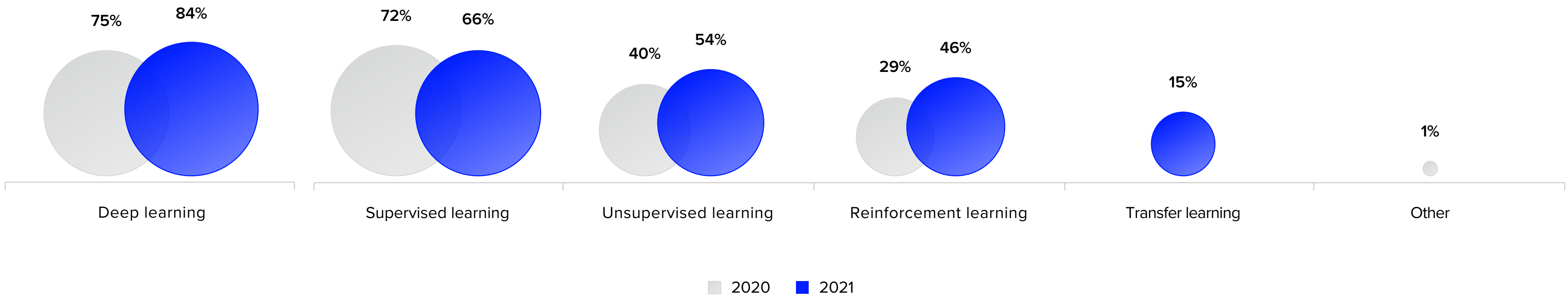
A not insignificant 15% of respondents cite using transfer learning. The pace of adoption of the technique is noteworthy, given its very recent evolution from a niche academic branch of AI/ML to a commercial technique. The increase in adoption is likely to be linked to the rise of NLP, where models are built on a general corpus and then can be fine-tuned with transfer learning to a specific area.

Survey respondents also expect these shifts to continue into 2022, with one in six forecasting a change in AI/ML learning techniques, mostly into greater adoption of deep learning.

> “The emerging trend that I see right now is the use of deep learning that will enhance our development, performance and simulation models.
>
> **Manager of model risk and governance**
> Chinese hedge fund

Figure 21: **Types of learning have become more diverse and advanced**

Which types of AI/ML do you use?



Key takeaways

1. Understand what your team needs versus what tools provide

The tooling landscape is changing fast and teams can easily spend money on something that doesn't solve for their needs. Existing tools also continue to rapidly evolve, improve and grow user bases. Understanding your team's unique needs and challenges is paramount to choosing the right tools.

2. Think modular, stay flexible

Data science tools are built to solve specific problems, with no offering likely to provide an entire end-to-end solution. Thinking modular about how to use each tool effectively can help build a mindset of being multi-platform and counter any fears of being locked-in to technology. More platforms are supporting this approach with interoperable and transparent data, analytics and technology. With the technology landscape being so volatile and fast-moving, an imperfect 'modular' decision that advances your learning and projects is often better than no decision at all.

3. Don't underestimate the learning involved in new and niche tech

New technologies are increasingly sophisticated, as seen in the rise of deep learning, reinforcement learning and NLP. This is driving data scientists to work more closely with software engineers and develop more technical skills. However, each new tool comes with a learning curve. Choosing a toolkit that is widely adopted can increase the talent pool and resources available to deploy and solve problems and not lock you in to talent or technology. Niche tools, while offering advantages in differentiation, can potentially be challenging to maintain over long periods of time due to a lack of skilled and specialist talent.

CHAPTER

4 Battle for brains: acquiring talent, building skills

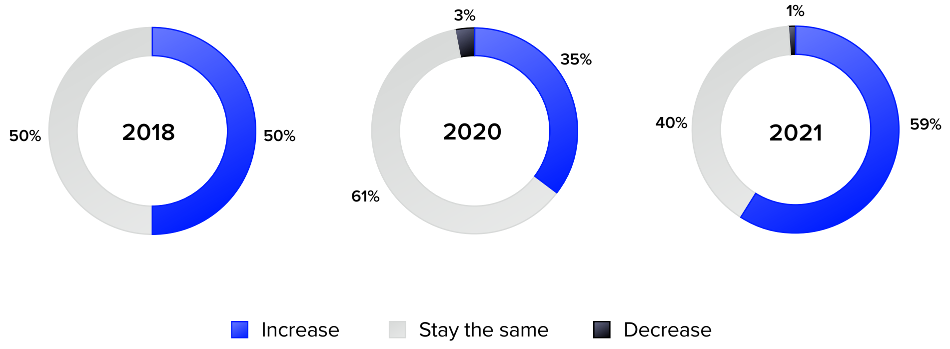
Companies are expanding their data science teams, while becoming more specific about the skills they need

Hiring expectations have risen sharply this year. While in 2020 hiring expectations were tempered amid great uncertainty, this year 59% of respondents expect their company to increase the number of data scientists employed.

- Companies are turning to a hybrid model for their data science teams: a core centralised team and many smaller ones embedded within business units
- Data scientists' responsibilities are increasing, from building and deploying models to selling the business case and influencing internal strategy
- Companies are eager to hire more specialist roles, especially data engineers and data architects
- Employees with domain expertise are keen to build their coding skills

Figure 22: **Appetite for data scientists is at an all-time high**

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?



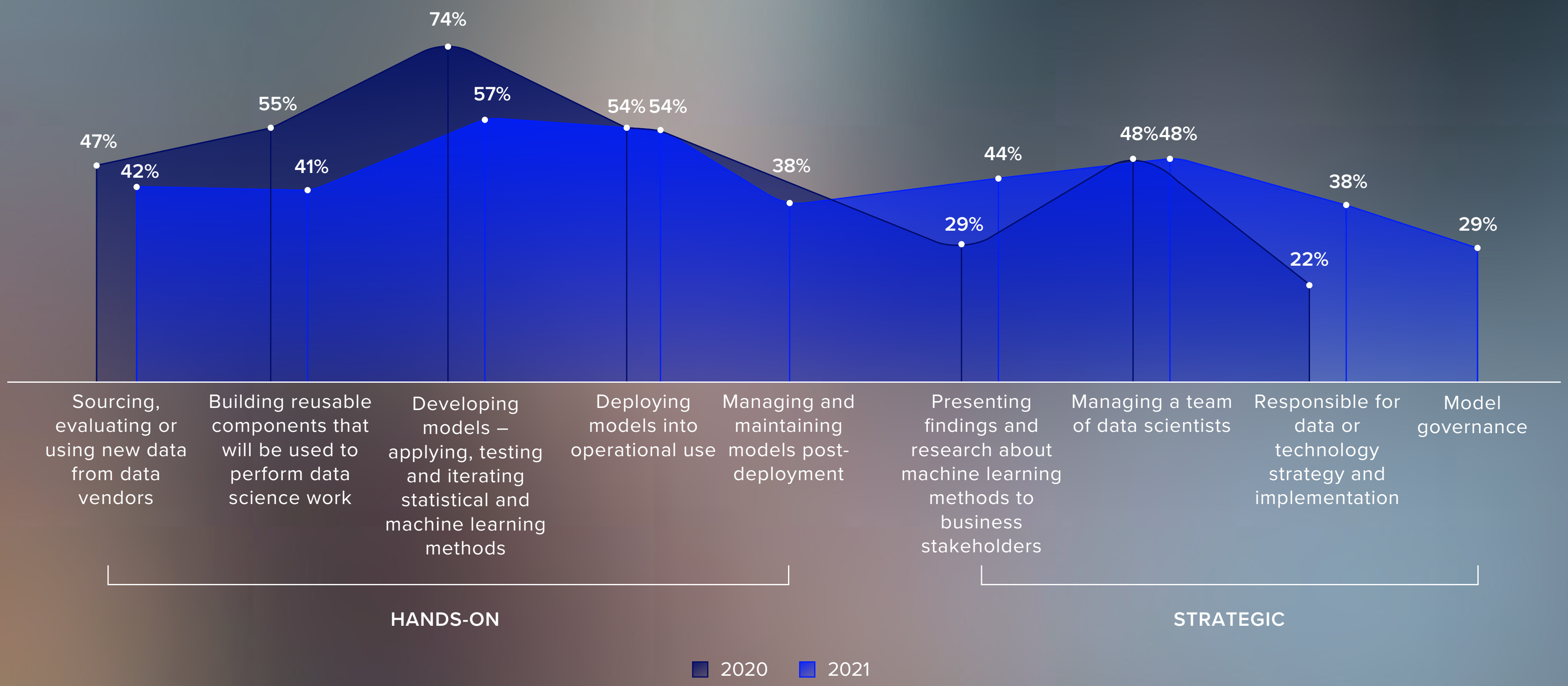
The role of the data scientist is in flux

This year, fewer data scientists were engaged with building models or components. Instead, more were engaged in strategic activities, including presenting findings to management and overseeing model governance.

This duality points to the different directions in which the industry's data scientists are currently stretched: both executing on AI/ML strategy and helping to shape it. Data scientists are acting as internal resources on breaking down business problems into an AI/ML setting, on setting up the end-to-end flows and on how to deliver solutions at scale.

Figure 23: **Data scientists are busy shaping AI/ML strategy**

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



Centralised or distributed? Both

As data scientists take on more strategic responsibility, their dispersion within their company changes. In previous years data scientists were either in a centralised group or embedded within business units, whereas this year's survey shows the industry adopting a hybrid model, where some capabilities are centralised, while others apply their skills within a specific knowledge domain.

This is also evidenced by the number of teams, which has grown over the years. For the first time, most respondents reported their company having five or more data science teams within the business, with the average number of teams jumping from seven last year to 10 this year.

Figure 24: **Many more data science teams**

How many data science teams are there in your company?

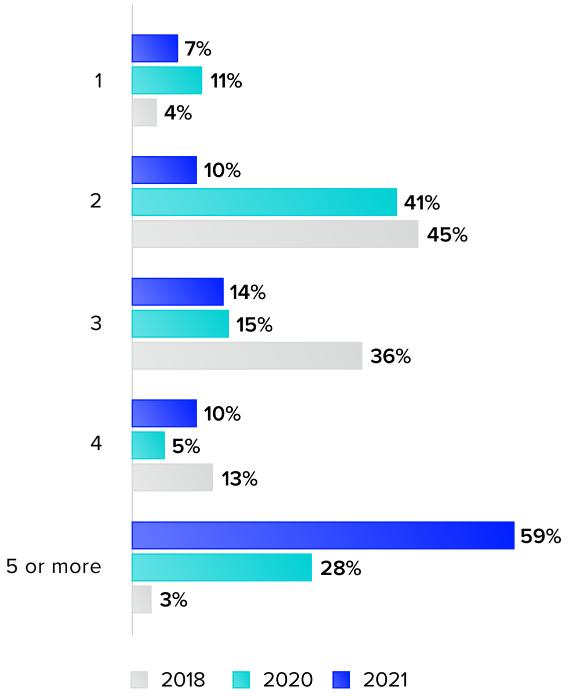
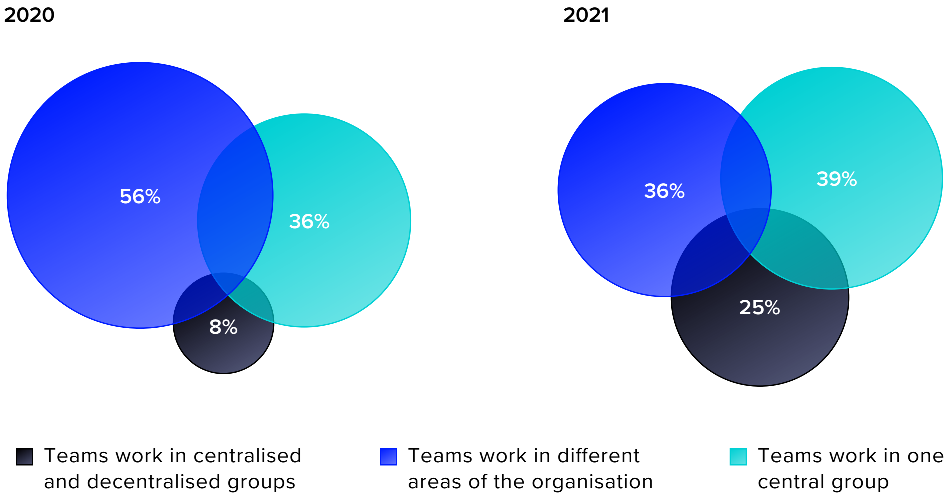


Figure 25: **Operating in a hybrid model**

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



Hiring versus retraining

The most in-demand roles mirror the data availability challenges we report in Chapter 2, with data engineers and data architects most needed to design, build and manage data pipelines. Data scientists increasingly occupy more strategic positions, serving as a bridge between understanding AI/ML and the business domain.

MLOps skills are also in demand. As more companies scale AI/ML into production, serving and managing data models and monitoring performance is becoming a critical specialisation.

Also, as financial services firms realise financial know-how isn't easily transferrable, they're supporting employees' efforts to build coding skills. In a separate LSEG survey of analysts, bankers and portfolio managers, the majority of both senior and junior employees are seeing coding skills becoming relevant to their current role. With increased competition over data science talent and the push toward integrating AI/ML into strategic decision-making, we expect this trend to continue.

Figure 26: **Plugging skills gap through hiring**

Which specialist AI/ML roles do you anticipate will grow the most over the next one to two years?

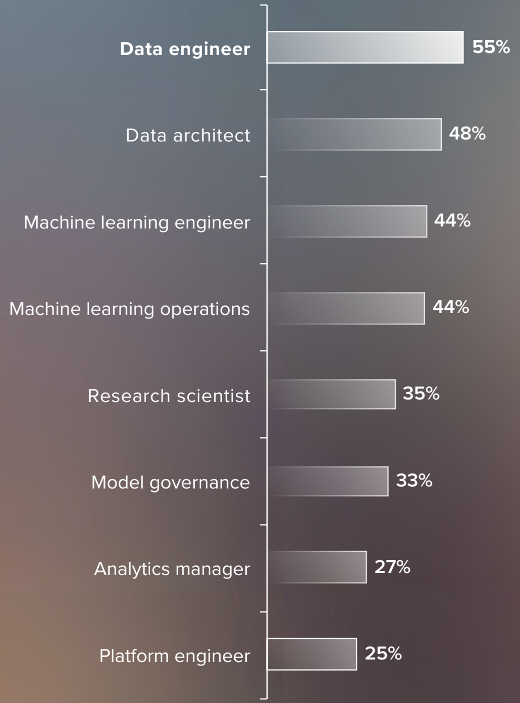
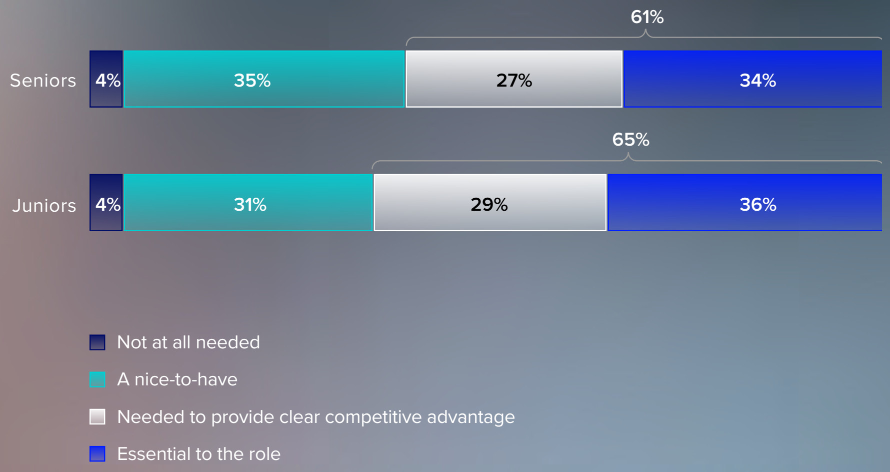


Figure 27: **Coding is seen as a must-have skill**

To what extent do you think the ability to use programming is needed for the role you currently have in two years' time?



Workspace without limits (2021), Refinitiv

Key takeaways

1. Embrace data scientist specialisation to be effective

Data scientists can progress through multiple career paths and will need a variety of managers, support networks and opportunities depending on their roles. Some data scientists are now supporting project management and strategy by acting as technical advisors. Others are more involved in software engineering – with new tools allowing data scientists to contribute to MLOps and close the gap between research and production.

2. Expect retraining barriers to lower

The skillsets in DevOps, DataOps and MLOps are increasingly overlapping and could reduce the barrier of retraining. Furthermore, these disciplines can be learned and put into practice on the job without formal qualifications. For example, software engineers who are good at building systems are likely to be able to transfer their skills to MLOps. Some cloud platforms are also merging these 'Ops' capabilities together and upskilling teams as a by-product, with many providers also offering hands-on training and crash courses in AI/ML, like Google.

3. Make sure your centralised data science groups are seen as active enablers

A hybrid model with centralised and distributed data science capabilities poses a risk to the distributed teams' morale if the central team is seen to be doing all the 'fun stuff'. Centralised teams should be seen as enablers for the rest of the organisation. The team should think strategically and provide guidance and support by investigating potential technology offerings, suggesting best practices for building and deploying models, supporting teams on their AI/ML approaches and investigating novel approaches.

CHAPTER 5

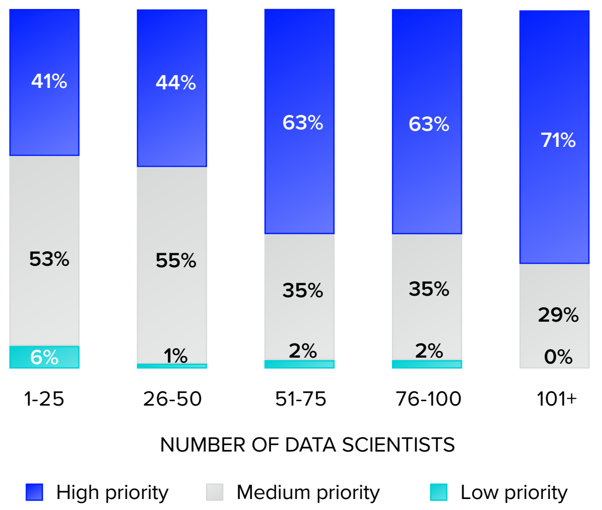
Trustworthy AI: the rise of model governance

Greater AI/ML adoption and fear of regulation are pressing financial services firms to build and implement model governance frameworks

- As more investment flows into AI/ML and companies grow their data science teams and improve their processes, model governance becomes more important
- With growing model inventories, model governance is seen as a way to reduce costs and improve model quality
- Stakeholder trust in AI/ML remains an ongoing challenge, with efforts currently focused on stakeholder engagement and communicating benefits
- Moving from experimentation to production, model governance is more likely to be implemented for statistical robustness rather than due to ethical concerns
- Today, ethics is less of a driver for model governance, but some respondents expect this to change as AI becomes more pervasive and ethics, bias and the issue of explainability are impossible to ignore

Figure 28: **Model governance becomes more important as companies grow data science teams**

To what extent is model governance a priority in your organisation?



Model governance as a sign of maturity

Model governance comes into sharper focus when teams move AI/ML projects from experimental testing into production. This year, 38% of respondents report managing models post-deployment, with 87% of companies already managing an inventory of models.

Our survey shows the drivers for governance appear to be less about meeting ethical standards and more to do with mitigating regulatory risk. Regulations such as GDPR have been shaping how companies manage data sets with personal information, while firms are anticipating more AI/ML regulation. In Europe, for example, the European Commission has already proposed new rules and actions for excellence and trust in Artificial Intelligence. Once adopted, the new legal framework will be directly applicable throughout the European Union.

Concerns of bias in model outputs raise questions about how data is sourced and pipelines are built, and which new risks they present to the business. Data scientists need to pay close attention to this area and ensure they have the support of legal and risk colleagues to spot and manage any risks.

Figure 29: **Most companies have a model inventory**

Do you have an established way to identify models? Do you have organisation-wide model definition? Does the organisation maintain an inventory of models?

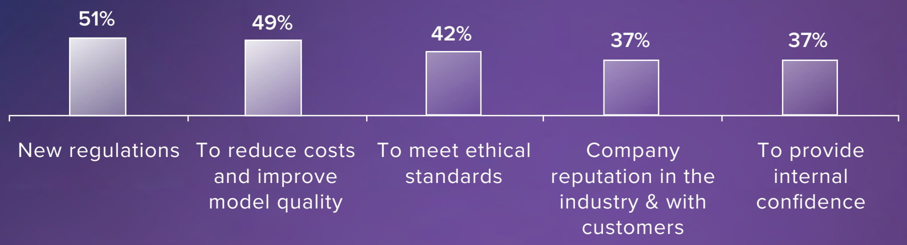
86%
Established way to identify models

87%
Organisation-wide model definition

87%
Firm has inventory of models

Figure 30: **Model governance is driven by regulation and improved model quality**

What is most likely to drive investment in model governance in the next year?



How companies build internal trust in AI/ML

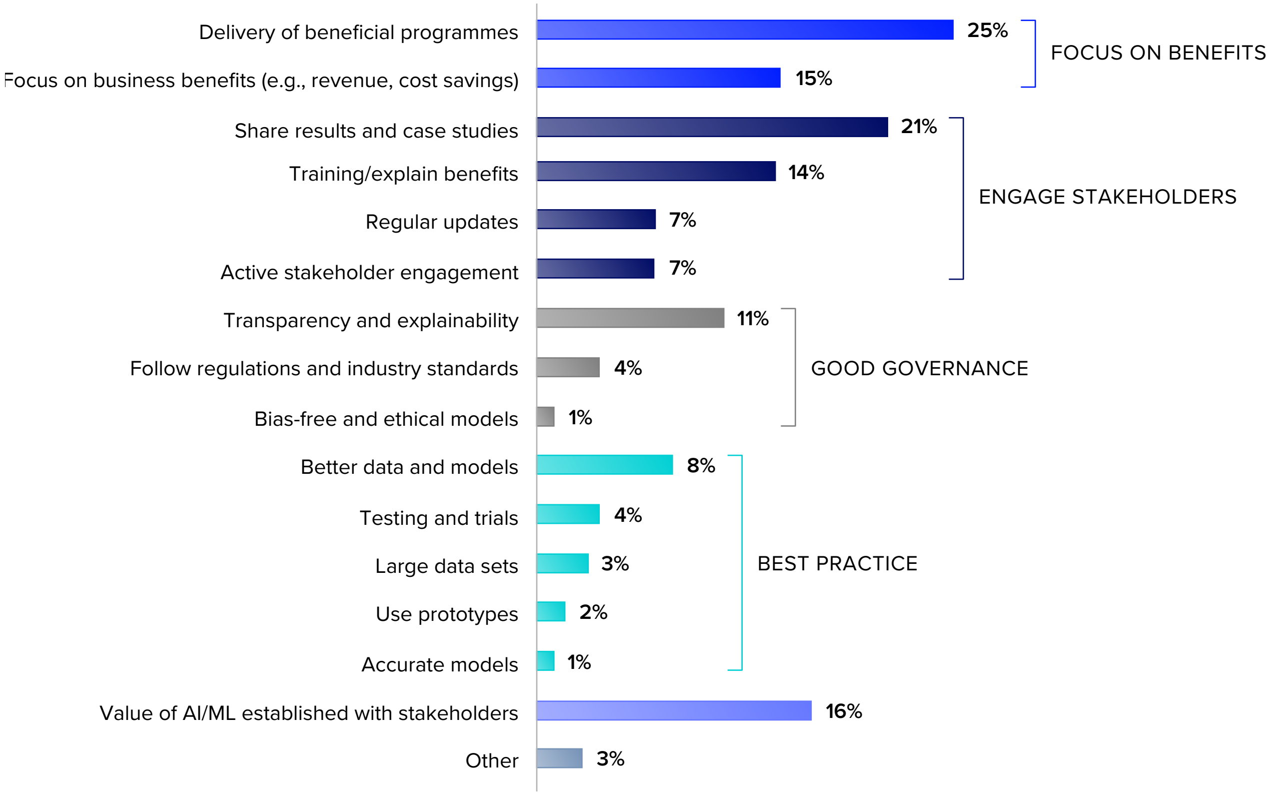
This year, only 16% of respondents believe the value of AI/ML has been established with stakeholders. Building internal support is still work in progress.

The results show the preferred way for building trust in models is engaging with stakeholders, mostly by focusing on the beneficial outcomes of the models themselves. This suggests some business scepticism regarding cost versus benefit of AI/ML, which may not be helped by nascent regulation. Introducing industry regulation and legal frameworks is likely to help build trust in AI/ML, thereby increasing usage and adoption.

Our survey shows data scientists are increasingly responsible for communicating the strategic value of AI/ML to the wider business (see **Figure 23, page 38**). However, while deep learning is becoming the most popular AI/ML framework, only 11% of respondents use transparency and explainability to manage stakeholder trust, but just a concerning 1% use bias-free/ethical models and accuracy. This surprising finding underlines the need for data scientists to have a firm grasp of the business domain and data sets in order to identify the risks of implemented AI/ML.

Figure 3: **Trust in AI/ML is being driven by stakeholder engagement and a focus on outcomes**

How does your organisation manage stakeholder trust and confidence in models that use AI/ML?



> With these technologies booming we will see new regulations surfacing.
> I think these will be more consistent and stable.
>
> **Senior director of model governance**
> French broker

Types of governance being implemented

Financial industry firms acknowledge they need to implement some form of model governance. Only one in six companies surveyed has yet to implement any governance changes.

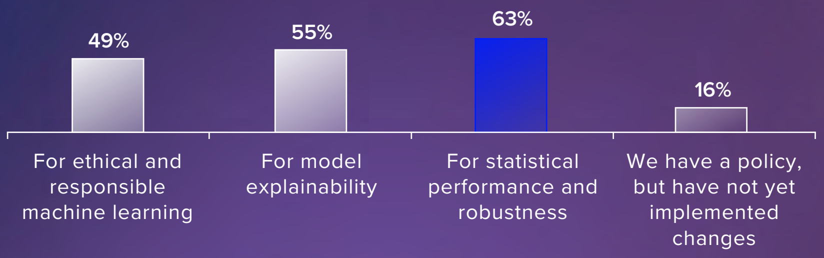
While half 49% of respondents cite ethical AI/ML as a desired outcome, more firms are implementing model governance with the hopes of generating better statistical performance and better explainability.

This year, only 49% of model governance respondents cite ethical machine learning as stated purpose of a framework in place. However, ethical and bias issues are likely to emerge from the move into less-explainable AI/ML frameworks (particularly deep learning, see **Figure 21, page 33**) and the adoption of bias-laden data (particularly unstructured data, see **Figure 10, page 17**).

Our survey shows the increasing appeal of geolocation, satellite, credit cards, text and voice data that all contain substantial biases, which could be further reinforced once integrated into decision-making models.

Figure 32: **More companies implement model governance for statistical robustness**

In which of these ways have you implemented model governance for AI/ML?



Key takeaways

1. Don’t limit model governance’s scope

The data insights in this report suggest most respondents see model governance as the creation, storage and version control of documentation for AI/ML code and models. While this view is not incorrect, it is a limited view of the number of ways model governance can both advance innovation and proactively limit potential damage to a company.

2. Raise the profile of AI/ML and model governance with a cross-divisional working group

A working group including representatives from risk, technology, data, procurement, legal, information security, government relations, regulatory and compliance can provide oversight of firm-wide AI capabilities and applications, and engage key stakeholders on model governance.

3. Govern the models you buy as well as the ones you build

Chapter 3 on ‘technology choices’ shows how companies have never relied more on their cloud and other service providers. Teams should ensure that their robust model validation practices and model policy extend to the models and analytics they buy as well as build.

Regional AI/ML trends

Asia Pacific and EMEA are catching up fast with North America

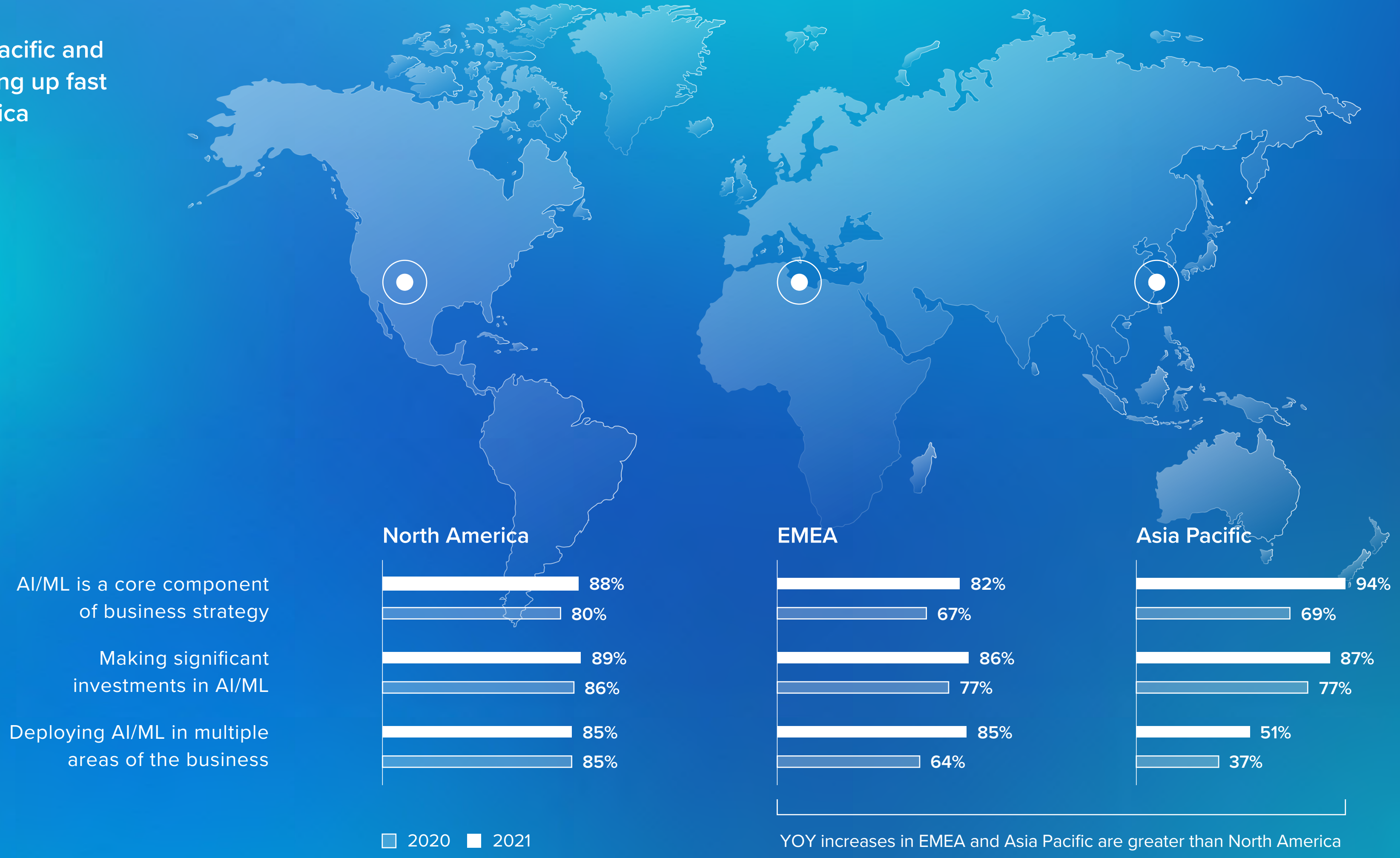
AI/ML adoption levels into multiple areas of the business have increased in Asia-Pacific. This year, 51% of respondents at Asia Pacific-based companies say their company deploys AI/ML in multiple areas of the business, compared with 37% last year.

Crucially, companies in both Asia Pacific and EMEA perceive AI/ML as a core component to business strategy far more than was the case last year. In EMEA, 82% of respondents believe this to be the case, up from 67% last year. In Asia Pacific, 94% say AI/ML is core to the business strategy, up from 69% last year.

This strategic commitment is also backed up by rising levels of investment. This year's survey shows, both for Asia Pacific and EMEA, a 10% and 9% rise respectively in the number of respondents who agree their companies significantly invest in AI/ML.

Companies in EMEA and Asia Pacific are also rapidly evolving their processes to match organisational patterns in North America. For example, companies are now on par with having their AI/ML decision-makers distributed across the business.

Figure 33: **Asia Pacific and EMEA are catching up fast with North America**



Technology glossary

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.

Data engineering: The infrastructure built to receive, store and prepare data for use in machine learning. Akin to pipe-building, data engineering controls the flow of data from raw input into final outcomes.

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.

Model governance: The processes put in place by a company to control access to its models and asses their quality. It also includes policies concerning models’ implementation and performance tracking.

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 provides inputs as labelled training data, as the basis for predicting the classification of unlabelled data.

Unsupervised learning: Machine learning approach that looks for patterns in unlabelled 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.

Transfer learning: Machine learning approach that adjusts large pre-trained models using a relatively small data set of labelled data.

Solving the most challenging problems in financial services. Redefining the art of the possible

At **LSEG Labs** we collaborate with customers around the world to redefine the art of the possible by solving the most challenging problems in financial services.

We are the in-house fintech at LSEG (London Stock Exchange Group), a diversified global financial markets infrastructure and data provider.

As a Group we are dedicated, open-access partners with a commitment to excellence in delivering the services our customers expect from us.

With extensive experience, deep knowledge and worldwide presence across financial markets, we enable businesses and economies around the world to fund innovation, manage risk and create jobs.

It's how we've contributed to supporting the financial stability and growth of communities and economies globally for more than 300 years.

Discover how open makes more possible.

About LSEG

LSEG (London Stock Exchange Group) is more than a diversified global financial markets infrastructure and data business. We are dedicated, open-access partners with a commitment to excellence in delivering the services our customers expect from us. With extensive experience, deep knowledge and worldwide presence across financial markets, we enable businesses and economies around the world to fund innovation, manage risk and create jobs. It's how we've contributed to supporting the financial stability and growth of communities and economies globally for more than 300 years.

Discover more at lseg.com

LSG1396454/9-21