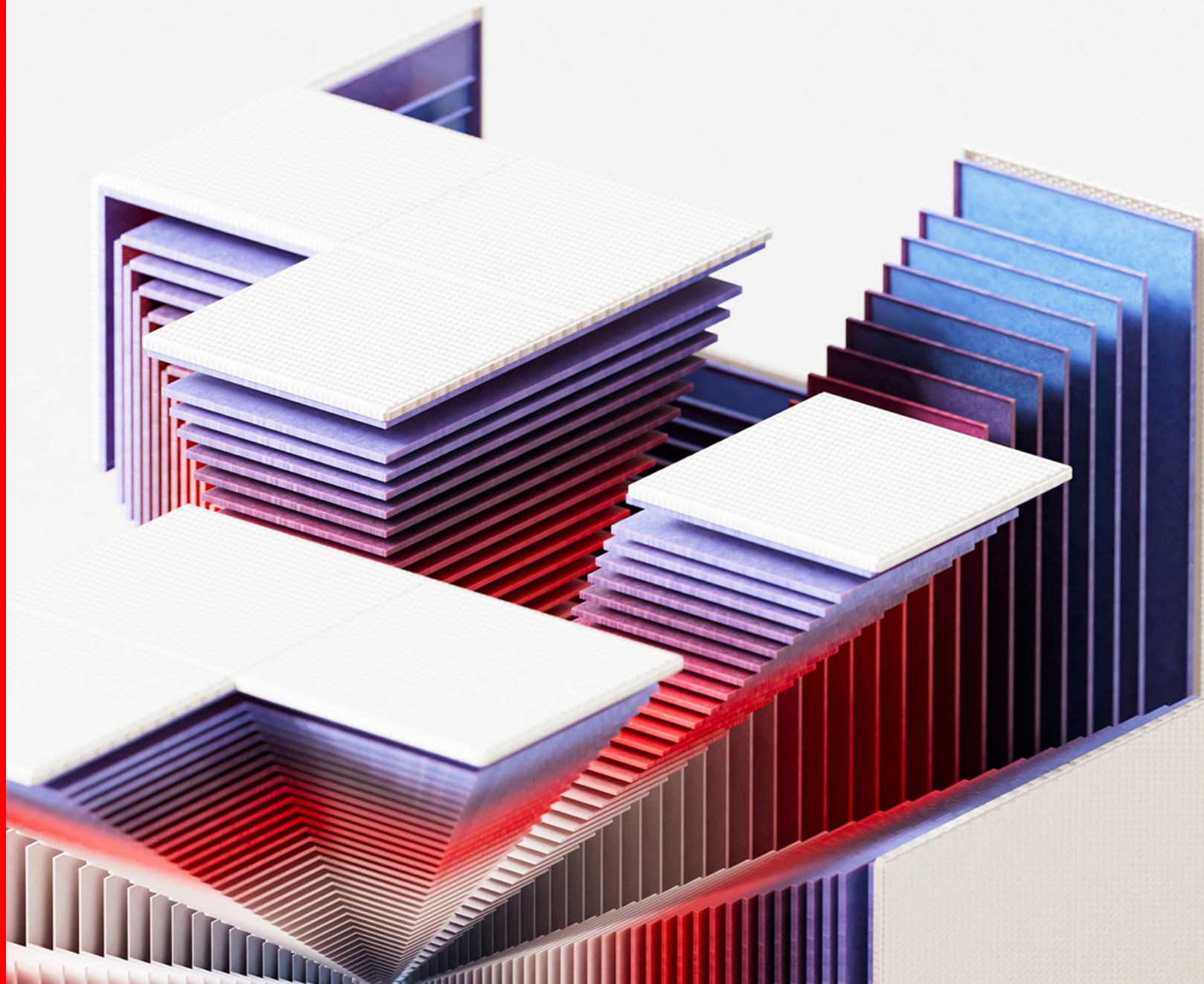
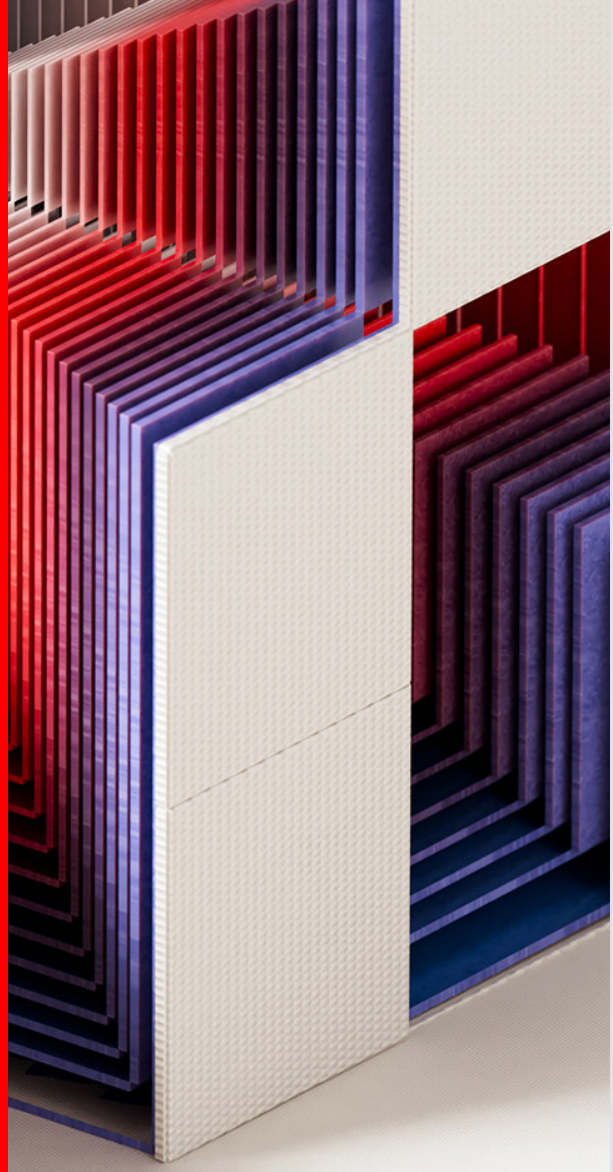
Hitachi Vantara

State of Data Infrastructure Global Report 2024

How AI is Shifting Data's Foundation





Introduction

As AI advances at breakneck speed, pushing traditional data infrastructure to its limits, businesses and their customers demand more from technology. How are IT leaders keeping up with this relentless pace?

To understand the pressures and opportunities businesses now face, Hitachi Vantara interviewed 1,200 IT decision makers in large organizations from 15 countries. Their observations, warnings and examples illustrate how important data infrastructure is for business and IT leaders who want to take full advantage of the wealth of information stored within their company.

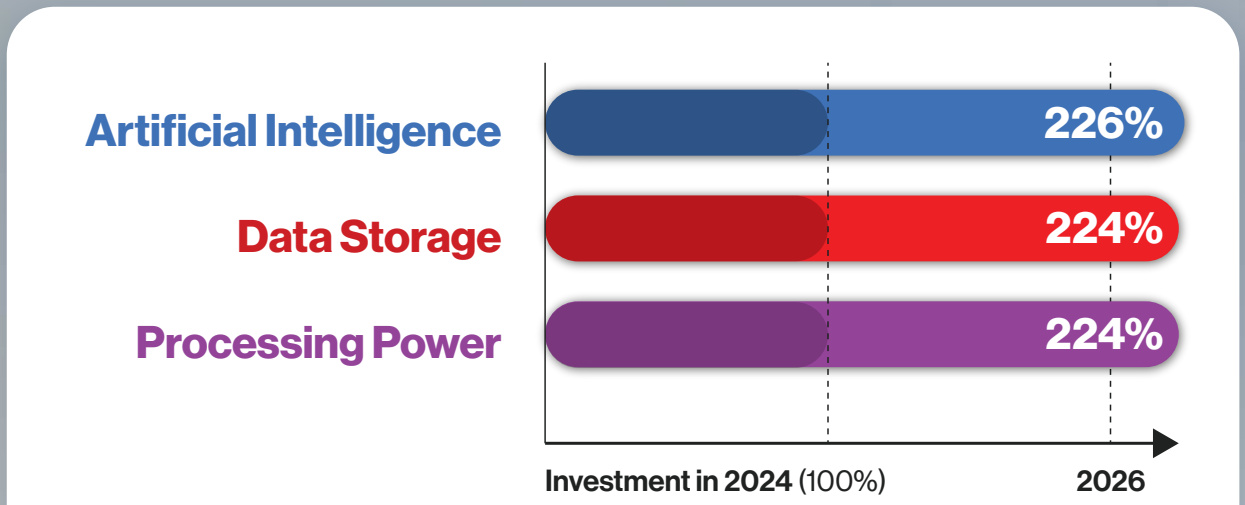
Last year, IT leaders expected data storage to double within two years[1]. Already, their needs have almost tripled as the average large organization now holds 150 petabytes (Pb) of data. By the end of 2026, they expect to be storing over 300Pb – equivalent to storing every film made worldwide since 1950[2] nearly 200 times over, in 4k.

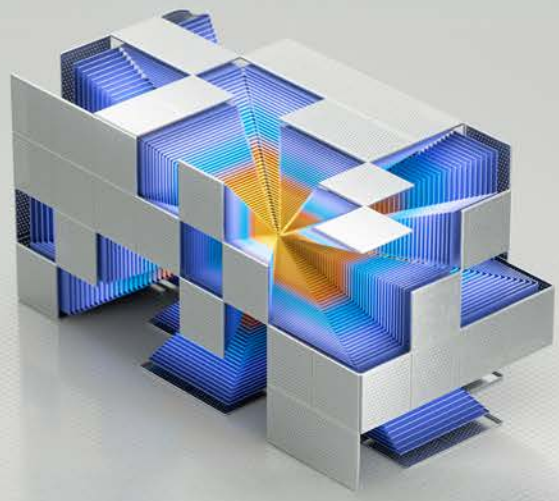
We interviewed...

1,200 IT decision makers + **15** Countries

> **“Companies are not prepared for what comes next. The data explosion. Many need to completely rearchitect their infrastructure.”**
>
> **CTO**
> Global telecoms company
> Germany

Expected Growth in Investments Over the Next Two Years

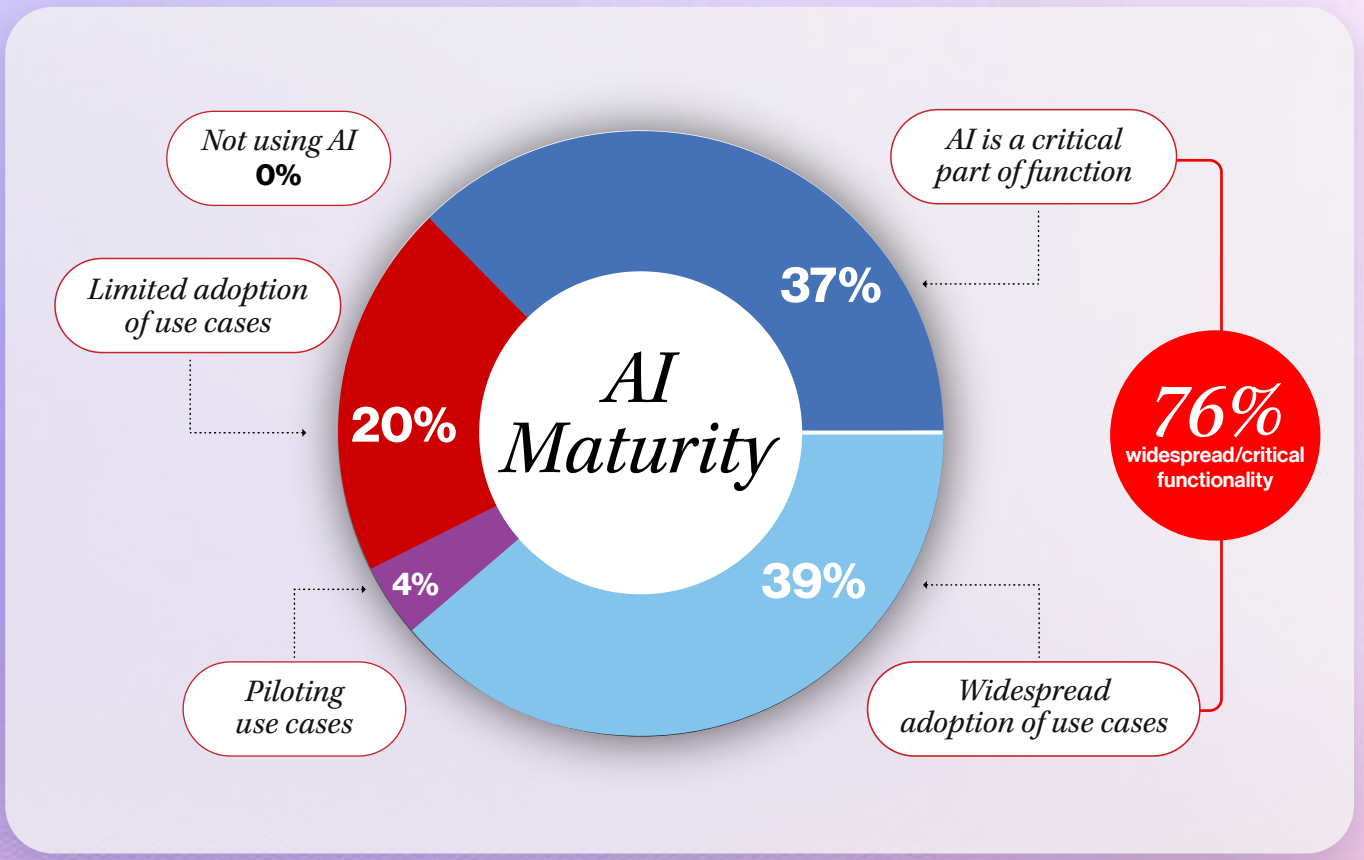




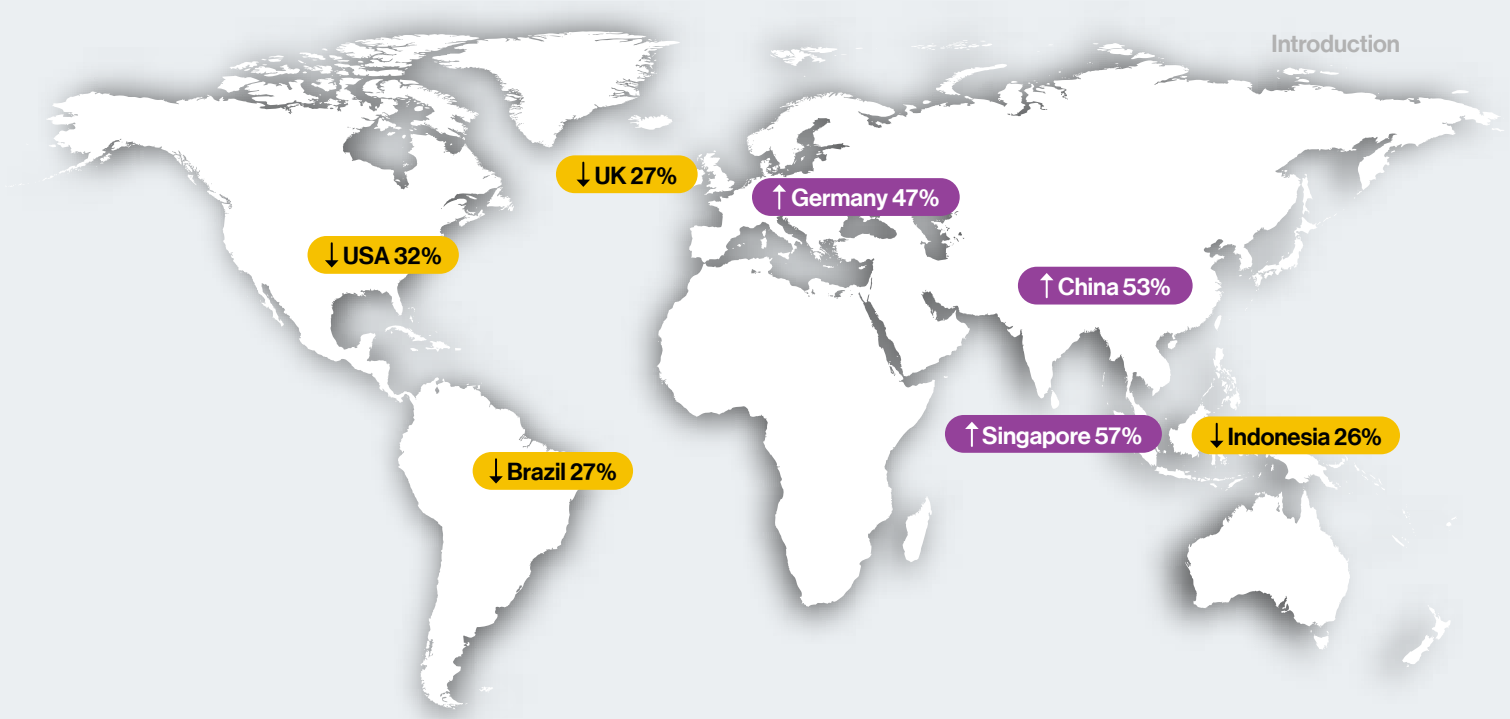
The increase may look steep, but our research indicates estimates are likely conservative. IT leaders are planning for investments in AI, data storage and processing power to more than double before the end of 2026.

GenAI has captured stakeholders' imaginations, from CEOs to customers. The groundbreaking successes of AI pioneers have sparked a surge of investment and expectation. And the opportunities are real. A recent study commissioned by Google Cloud showed that 86% of AI early adopters gained an average of 6% in revenue[3].

AI Adoption Among Large Organizations



Nobody wants to be left behind. Every company in our study has adopted AI in some capacity. **Three quarters (76%) have progressed beyond limited adoption of AI use cases**, and a third (37%) say AI is already critical to their business.



Adoption is highest in Singapore (57%) and China (53%). In Europe, Germany is leading (47%). Adoption is slower in the US (32%), UK (27%), Brazil (27%) and Indonesia (26%).

IT leaders face a future with more significant rewards, but also more risks than ever. The current rate of growth is placing a strain on organizations' critical assets. IT leaders are concerned about limited resources like quality data (37%), skilled workers (31%), data storage (31%) and processing power (28%).

Successful transformation depends on making informed, strategic decisions about data infrastructure. A robust foundation is critical for scaling effectively to get the most out of AI without sacrificing data quality, security or sustainability management.

37%

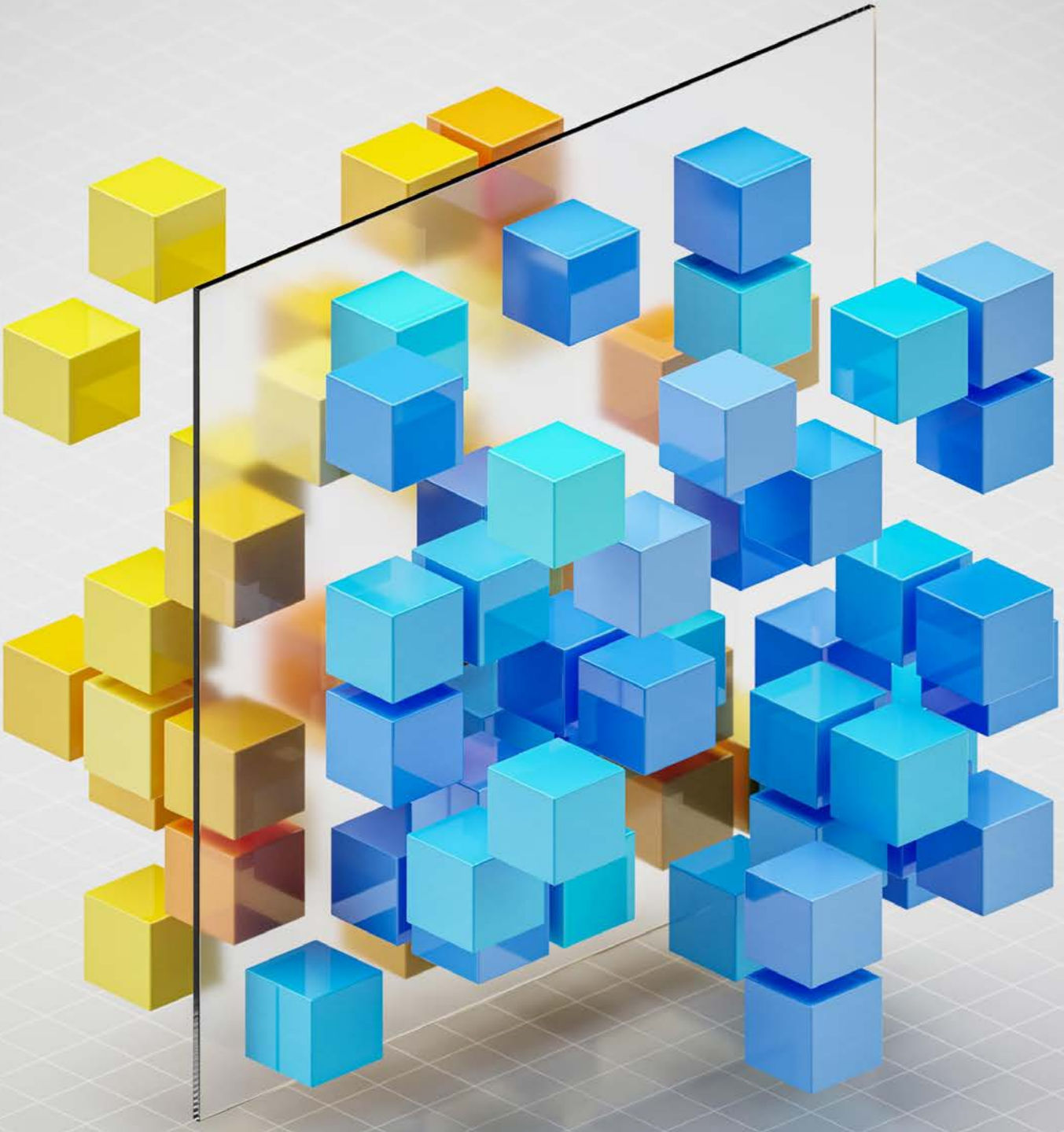
include data quality in their top 3 functions

> **To make AI a core strategic advantage, companies must lay the groundwork now with a strong data collection strategy. The complexity and cost of infrastructure will keep rising, especially in cloud environments. This demands proactive decisions around security, GDPR compliance and data privacy. Yet, few organizations are truly prepared with a comprehensive data strategy."**

Srihari Udugani
VP of Technology Innovation & Operations
Borderless Access, India

Table of Contents

01
The Great Data Quality Divide



IT Leaders Say Data Quality Is King, but Many Are Ignoring It

One of the most surprising takeaways from our research is the gap between how many IT leaders say data quality is essential for implementing new technologies like GenAI and how few prioritize data quality in their actions.

IT decision makers know successful AI implementation relies on quality data. Four out of ten (38%) point to data quality as one of the most important factors. In India, IT leaders were especially focused on data quality (58%). No wonder then, that high quality training data is the second highest concern IT leaders have when implementing AI (37%), after keeping the data secure (38%).

Although IT leaders say they are seeing successes with AI between 76% of the time (using free models) and 85% of the time (when partnering with global systems integrators), the bar for AI accuracy is low. **IT leaders say the results are accurate just 42% of the time.** About one-fifth of the time (21%), they see the models "hallucinating." Currently only 36% of IT leaders trust their AI outputs more than half of the time. Although in some industries like manufacturing the figure is higher (42%).

> **Before using AI, we were not really making data quality a discipline; that was a shift because we had to step back and say, 'What's really important to us here and how accurate do we want the outputs of AI to be?' I think that was a wake-up call."**

Jim Rutt
Chief Information Officer
Dana Foundation, USA

In many cases, this quality issue is due to the fact that many companies are missing the steps required to protect the quality they say is so crucial. Only 38% are enhancing training data quality to explain their AI outputs. A quarter (24%) don't even review the datasets they use to train AI for quality. A third (37%) don't tag their data, and for another sixth (18%), more than half the data they store is *dark data*. In the UK, where AI adoption is slower, 56% of leaders said more than half their data is dark.

With three-quarters (76%) of organizations rolling out AI for widespread or critical functions, accuracy must improve quickly if AI implementations are going to live up to their long-term hopes.

All of this begs the question:

If data quality is crucial for a successful AI implementation, **why are so many organizations ignoring the fundamental measures that provide clean data?**

IT Leaders Are Focused On Security and R&D Rather Than ROI and Sustainability

75% of IT leaders are improving AI applications through feedback loops and continuous learning, but building quality datasets needs to be an early part of that journey. Currently, 25% do not monitor AI accuracy in real-time, and only 28% monitor model accuracy to explain their outputs, so the models are unlikely to improve.

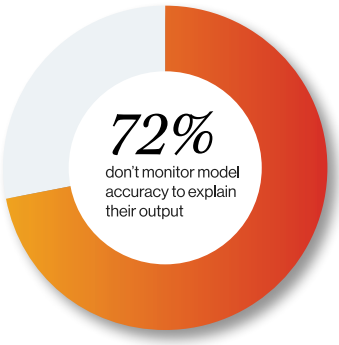
Even if data quality isn't good enough to guarantee accurate results from AI, it's a less urgent issue than security for many IT leaders. They are wary of AI being used to attack the organization from outside, as well as AI being misused, or introducing vulnerabilities within the organization. They are also concerned about AI itself being let loose on their data systems, leading to unforeseen consequences and security risks.

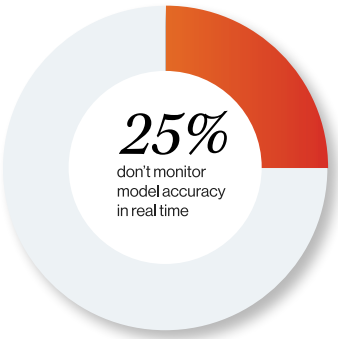
Seven out of ten organizations (69%) say they are treating AI like R&D, expecting ROI in the longer term (at least one to two years away), rather than immediately. Speed (41%) and cost (41%) are currently more pressing concerns than ROI (37%), as IT leaders seek to quickly iterate through and improve on AI implementations. Given the high energy costs of AI training and experimentation, it appears that sustainability is an afterthought for many organizations, too.

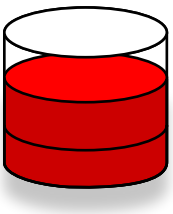
While data storage requirements are increasing at an unprecedented rate, the intricacy of the storage systems and the complexity of the data itself is also increasing. Three-quarters (76%) of IT leaders said more than half the data their organization now stores is unstructured.

As a result, providing high-quality data to train and run the AI models is challenging. Meanwhile, CEOs and customers see the technology as a must-have and are pushing IT leaders to implement AI regardless of data quality issues.

Unfortunately, the cost of experimenting with poor quality data and ignoring sustainability may be higher than IT leaders realize.







76%
of IT leaders say **more than half** of the data they store is unstructured

02

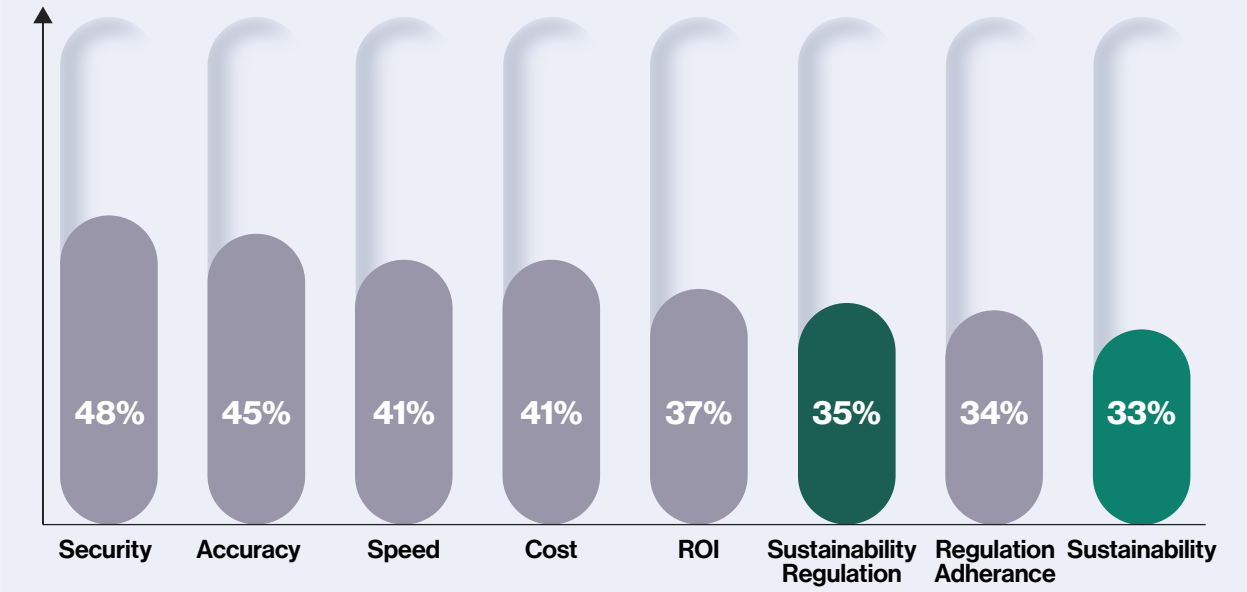
Where's Sustainability in the AI Conversation?

Profit Erosion: The Price of Ignoring Sustainability

Technologies like GenAI are creating revolutionary demands on our resources. Tech giants including Microsoft, Amazon and Google are turning to alternative sources of energy,– even nuclear [4]– to overcome electricity shortages and make their data centers more energy efficient.

Even so, sustainability ranks **last** among IT leaders' priorities for successfully implementing AI in most markets. Just a third (33%) say impact on the environment is among their top three priorities, slightly below adhering to sustainability legislation (35%).

Organizations' Priorities for Successfully Implementing AI (Included in Top 3)



With consumption likely to increase, responsible organizations must consider sustainability as part of their long-term strategy. This is slightly more prevelant in Italy (45%), the UK (44%) and the finance industry, where sustainability is one of the top three priorities of successful AI implementation.

65% of large organizations are focused on developing bigger, more general LLMs (large language models) rather than smaller specialized models. This despite large-scale models being greedier than regular models to train, consuming up to 100 times more power.[5]

Beyond training, the energy it takes to perform an AI-enhanced Google search query is 10 times as much as a regular search query[6].

Organizations that prioritize sustainability will need to limit their use of AI. More efficient, calculated usage is not only good for the planet but also for profits, since lower energy consumption translates into lower utility costs.

"It's not just the data quality, it's the higher order problems like sustainability that need to be addressed. A huge amount of stress is expected to be placed on the grid."

Simon Ninan
SVP of Business Strategy
Hitachi Vantara

65%
of large organizations are focused on developing bigger LLMs

Sustainability Mandate: Act Now or Face Regulations Later

Three out of five (60%) IT leaders are focused on getting their technology right before addressing ethical concerns rather than having ethical AI frameworks and guidelines in place first. Three in ten (29%) say that sustainability is not even a consideration for their organization's AI strategy.

With AI sustainability regulations in the pipeline[7] , ignoring sustainability is short term thinking. In the long run, infrastructure built without sustainability in mind will likely need rebuilding to adhere to future regulatory requirements.

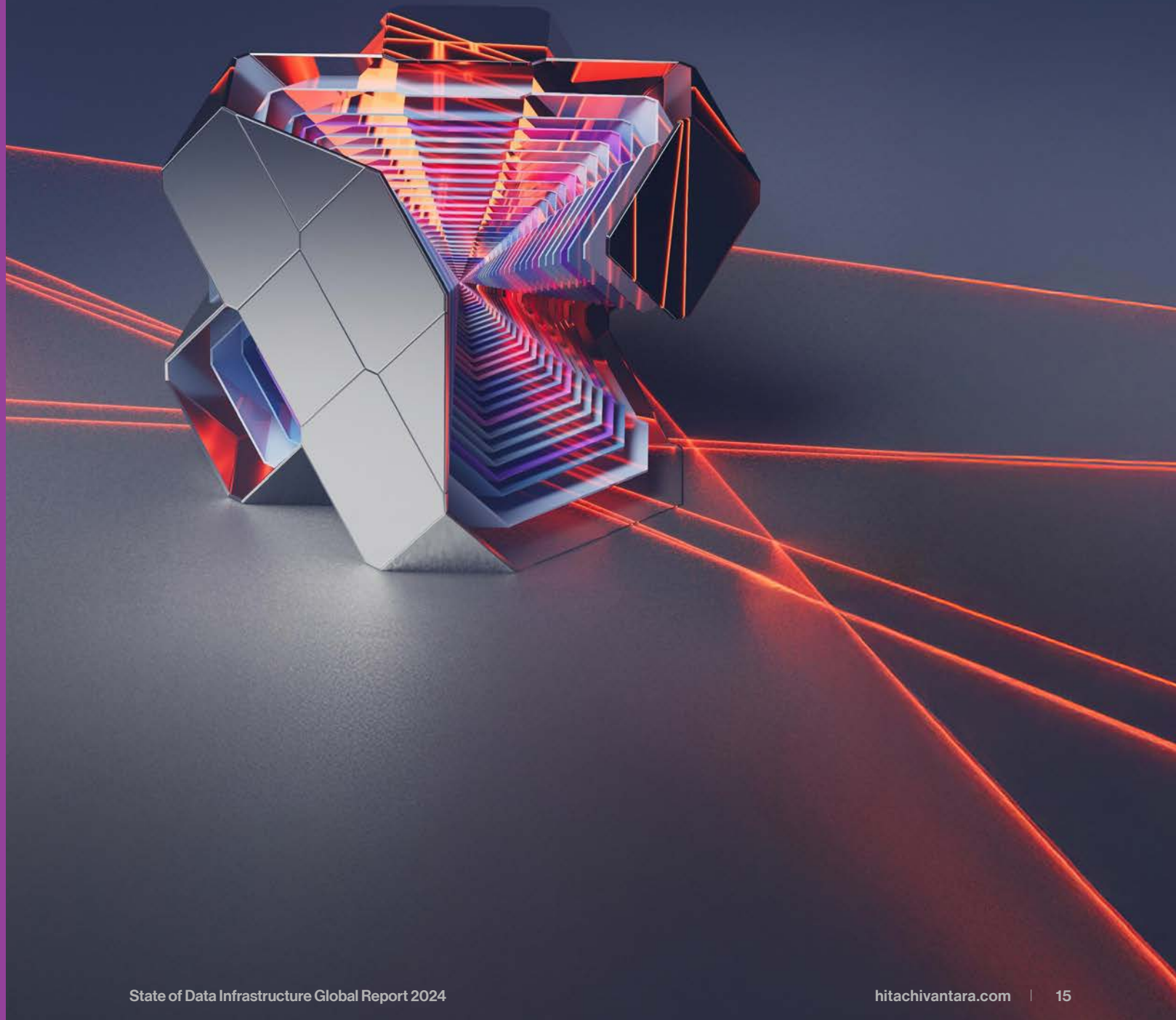
In the United States, mandatory sustainability reporting is expected to be enforced for large public companies in the fiscal year 2025 and for all public companies by 2026/2027[4]. However, a current lack of clear guidance on AI regulation has led many leaders to deprioritize sustainable AI. A third (34%) say having no standard available to inform what sustainable AI means is a barrier to implementing sustainable and responsible AI.

> **We needed to be more efficient and save energy; our industry is very expensive in terms of energy. We needed to start from scratch and migrate everything onto a new platform and close down all the existing old legacy stuff."**
>
> **CTO**
> Global telecoms company
> Germany

Only
2 in 5
IT leaders are addressing AI ethics before diving into technology

03

AI Risks Rise. Data Governance Holds the Line.



Securing Data is Currently Job #1

Data security is IT leaders' number one priority for successful AI implementation (48%). No wonder, when the stakes are so high. Three-quarters (74%) say if they lost their data through a mistake or an attack, the results would be catastrophic to their business – more so in sectors with highly sensitive data like healthcare (88%) and finance (84%).

In 2023, leaders were most concerned about internal risks from their own employees. Now, a data breach from an AI-enabled cyber attack is the most common concern (41%).

Three-quarters (73%) of IT leaders are very confident AI will benefit hackers more than cybersecurity defenders. It's an arms race between AI-enabled bad actors and AI-enabled cyber defenders.

74%

say if they lost data through a mistake/attack, the results would be **catastrophic to their business**

> **"We're highly focused on data security because of AI. Although essentially AI security is data security, it can have a lot of unintended consequences."**

Jim Rutt
Chief Information Officer
Dana Foundation, USA

Catching AI with AI

Last week, our security team identified a guy selling generative AI to respond correctly to 99.9% of all CAPTCHAs. So, CAPTCHA is not a solution anymore.

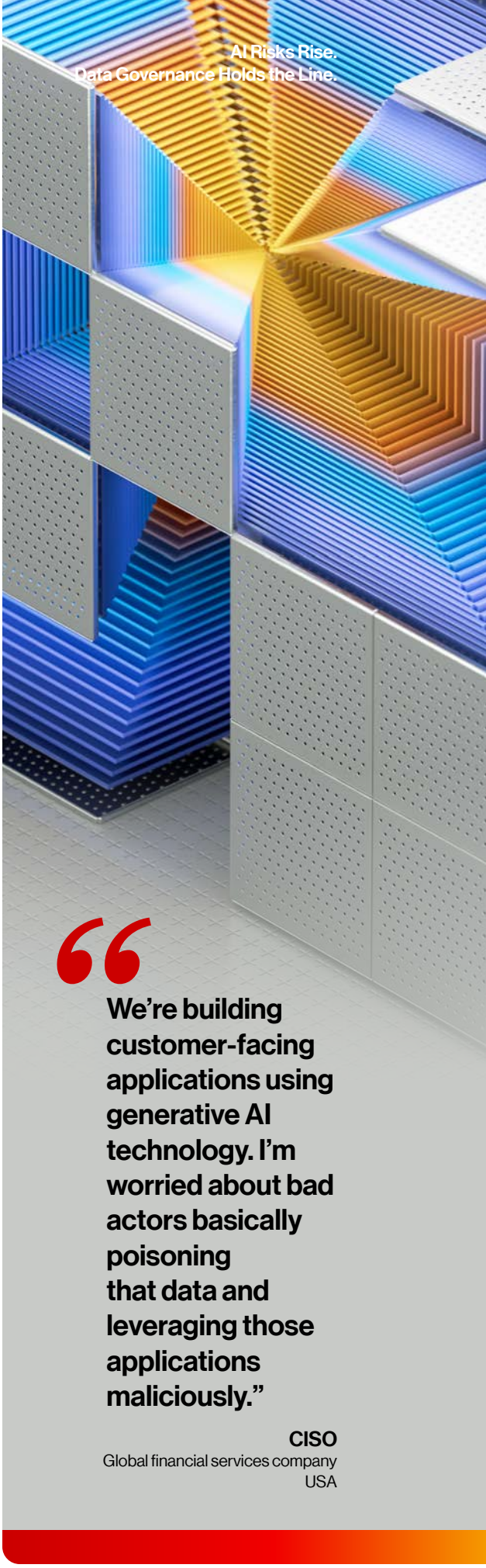
Now, our security problem is knowing if it's a human using our website, or a robot.

Right now, our solution is forcing our users to show us their face, but that can be faked. We're working on how we can prepare for that in the next few years.

We can't just keep evolving our current solutions.

I think the only way to combat AI-powered fraud in the future will be using GenAI against gen AI.

– José Dantas, CTO, Voltz, Brazil

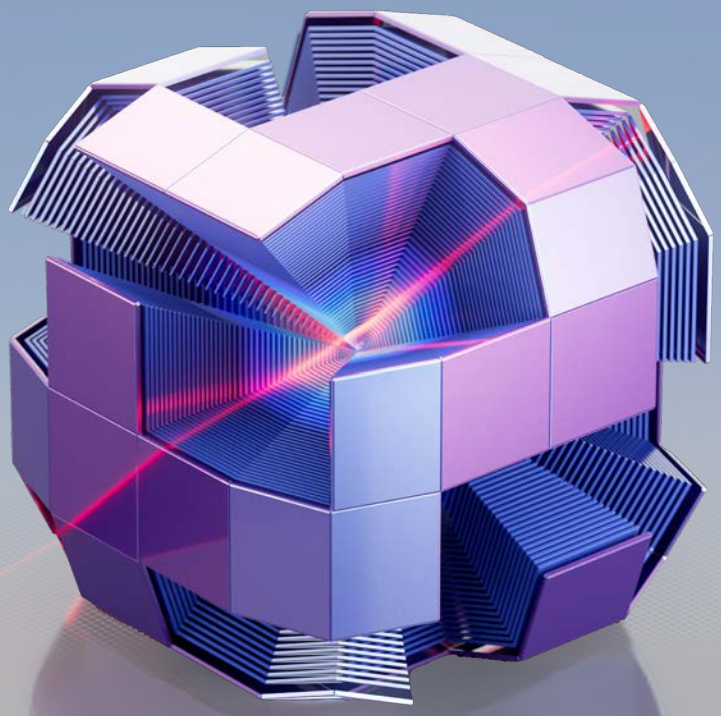


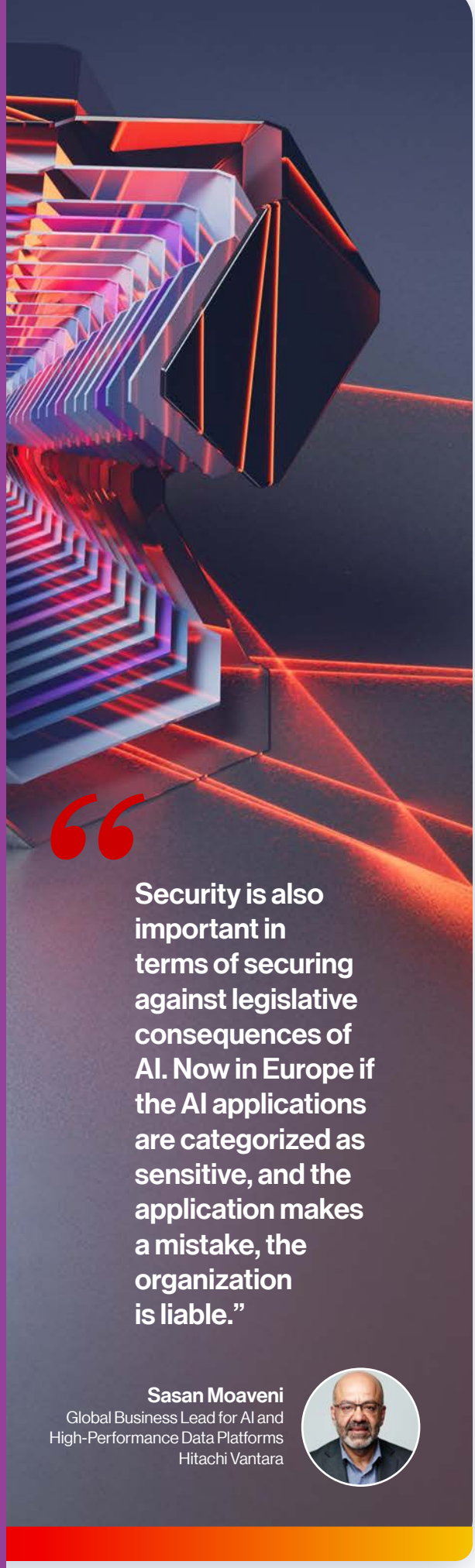
"We're building customer-facing applications using generative AI technology. I'm worried about bad actors basically poisoning that data and leveraging those applications maliciously."

CISO
Global financial services company
USA

Top Security Concerns of IT Leaders

Cybersecurity defenders can't only focus on external threats. While employee mistakes leading to a data breach was the top concern in 2023, now IT leaders are more concerned that AI itself may initiate security issues. **35% worry that they won't be able to recover data lost due to AI making a mistake, and globally, 30% fear that AI might accidentally cause a data breach**. In China, that rises to four in ten (40%).





In addition, AI tools give internal users access to more data and power than ever before. 31% of IT leaders rank recovering from a data breach caused by an employee's mistake among their top three concerns.

Generative AI is creating a whole new layer of security issues, as it crawls through dark data, and data sources that IT leaders might not have full control over. A poor dataset can taint data, but perhaps more concerning is good data that contains personal, sensitive or proprietary information. Securing AI models and the data that they train on requires tight control, visibility and the option to rewind model training to remove compromised data without starting from scratch if a problem is found.

"Security is also important in terms of securing against legislative consequences of AI. Now in Europe if the AI applications are categorized as sensitive, and the application makes a mistake, the organization is liable."

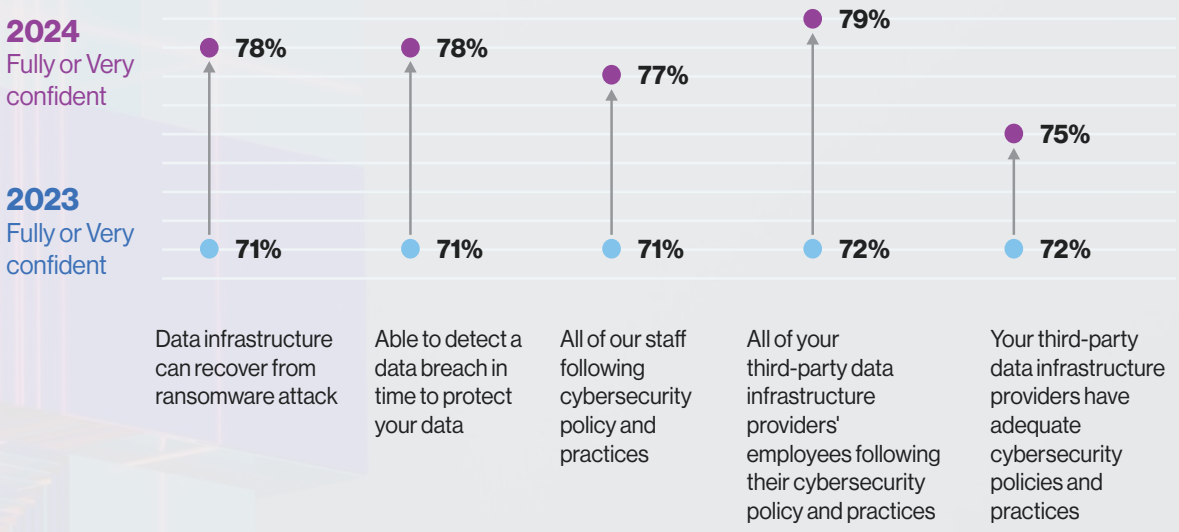
Sasan Moaveni
Global Business Lead for AI and High-Performance Data Platforms
Hitachi Vantara

31%

of IT leaders rank recovering from an employee-caused data breach among their top three concerns

Cyber Defenses Are Improving

Security Confidence Has Improved Since 2023



Despite increasing security threats, IT leaders feel more confident than they were a year ago in three critical areas:

1. Their ability to detect a data breach in time to protect their data or recover from a ransomware attack (both moved from 71% in 2023 to 78% in 2024).
2. Their staff are following cybersecurity protocols (from 71% to 77%) – particularly in finance (88%) and healthcare sectors (87%).
3. Their third-party data infrastructure providers have adequate controls (from 72% to 75%), and the third-party providers' employees follow their respective cybersecurity protocols (from 72% to 79%).

Forward-thinking organizations are stepping up their defenses internally and externally against human and AI risks.

04
Building AI for a Hybrid World

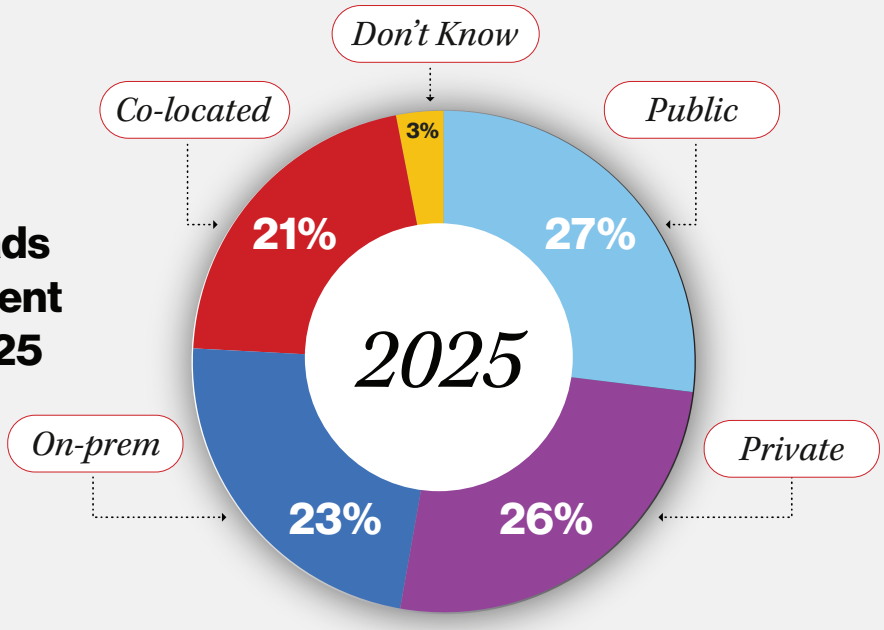
Hybrid Infrastructure Is Here to Stay

In 2023, our study showed although business leaders favor cloud when it comes to enabling digital transformation, they believe on-prem solutions are more secure (44% vs. 32%). In 2025, 23% of their data is expected to still be stored on-premises.

Mission-critical, personal and sensitive data is often stored in a private cloud or on-premises. More organizations push operational data out to hybrid or public cloud. However, with so many systems, there's a lot of overlap, and IT leaders must deal with data on multiple platforms. 98% use more than one storage platform for their data, and over half (57%) store data on all four types of platform: on-premises, private cloud, hybrid cloud, and public cloud. This multiplicity of platforms is particularly common in the UK (78%), China (76%) and North America (66%).

For many, hybrid cloud offers a balance between public cloud providers' future-proof, scalable services and the privacy and customization of a private cloud environment. Implemented correctly, hybrid cloud provides a single window into a flexible flow of data and services, allowing data quality and connectivity across the organization.

Percentage of Where Data Center Workloads Will be on Different Platforms in 2025



Compounding Success

My first problem was, how can I give the customer credit and get that money back?

So, the first AI produced a behavior score by performing tests using different channels and analyzing the customer's behavior. Which channel gets better response rates?

The information I had for this first AI was not that important. It was basic data and very incomplete. Even so, we achieved repayment rates of about 80%.

And the data that we collected from the first AI. Good quality data. We injected that into the second AI I created back then in 2022.

This was a credit score to separate the good payers from the ones who will never pay you back.

So, after that, the mix of these two AI, I jumped from 80% to 97% success in getting that money back and 35% less cost, because it's all automated, and it's more responsive.

– José Dantas, CTO, Voltz, Brazil

35%

in savings achieved due to automation and responsiveness

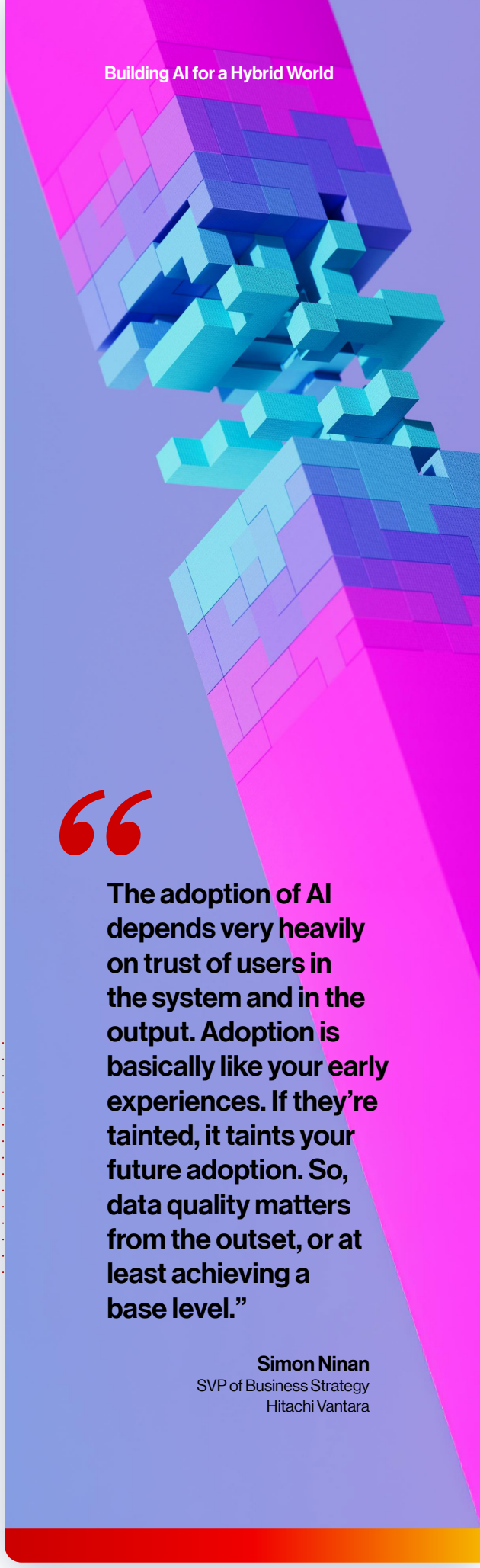
Fracturing Trust: The High Cost of Bad Data

Data infrastructure allows for the data quality that AI needs to succeed. It reduces replication in active data, while archiving data for protection. It balances the need for distribution with the need for speedy access on the data periphery. It automates and standardizes the ingesting and cleansing processes that data should undergo, and a good data infrastructure makes data visible, so IT leaders know what they are working with, and what is missing.

AI doesn't just need quality data, but it needs the right data. That critical company knowledge which IT leaders don't want to put in the cloud, is the very data that AI needs to provide the most benefit to an organization. IT leaders need hybrid cloud solutions that reach across cloud and on-premises solutions to get the right data to the right people at the right time.

Training AI models on poor-quality or incomplete data taints the model itself. It's like a skilled chef who is perfecting a recipe. If the ingredients are rotten, mislabeled or missing entirely, the chef will create a dish that doesn't taste at all as intended. Similarly, an AI model, no matter how sophisticated, will struggle to produce accurate results if trained on incorrect, incomplete or biased data.

Eventually, poorly performing models will poison the trust of users and investors. If they can't rely on the results of AI, they may reject the models entirely, and winning back their trust may be impossible.



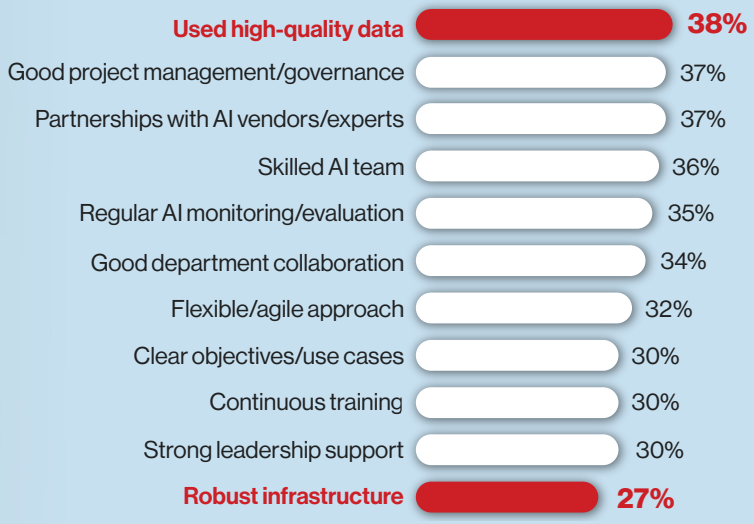
> **"The adoption of AI depends very heavily on trust of users in the system and in the output. Adoption is basically like your early experiences. If they're tainted, it taints your future adoption. So, data quality matters from the outset, or at least achieving a base level."**
>
> **Simon Ninan**
> SVP of Business Strategy
> Hitachi Vantara

Companies that don't provide AI models with known, precise and representative training data cannot ensure effective AI performance. **It's a recipe for disaster.**

The AI Rush Exposes a Blind Spot: Data Infrastructure

What's surprising is the disconnect between data quality and data infrastructure. **IT leaders rate data quality as the number one factor (38%) in successful AI projects but rate robust infrastructure as the least important factor (27%)**.

Top Factors for Why AI Projects Have Been Successful



As our data and observations have shown, many IT leaders are being forced to implement AI before their data infrastructure is ready to handle it. They are embarking on a journey of experimentation, hoping to stumble on ways to recoup their investments.

While investigation with new technologies is necessary, experimentation should happen within controlled conditions. Currently, only 5% indicate they are using sandboxes to test their AI experiments before implementing them. 70% are implementing AI testing, and improving as they go. This approach risks poisoning AI models, destroying users' trust in AI as a tool, and opening the door to new security vulnerabilities.

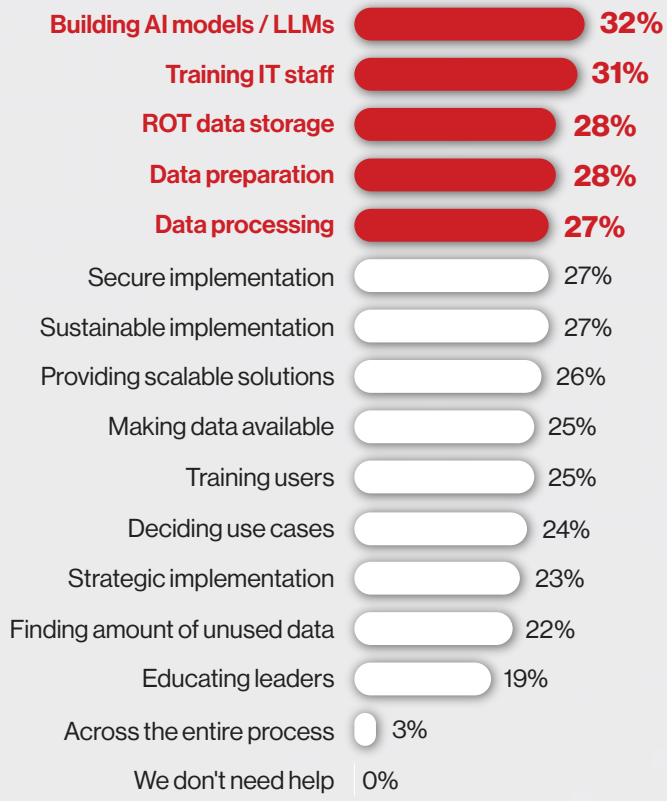
05
The Call for Targeted Expertise

Leveraging AI All the Way up the Value Stack

Our research shows that as technology evolves, almost everybody feels the need for outside help. Of the more than 1,200 IT leaders we surveyed, only one claimed not to need third-party help as they implemented AI.

However, very few felt that they needed help everywhere, just 3%. **The key for IT leaders is to identify and fill the gaps in skills, personnel and infrastructure that could prevent them from implementing AI successfully**.

Areas IT Leaders Say They Need Help Implementing AI



Hardware must be secure, available 100% of the time, and efficient to satisfy the coming sustainability requirements. A quarter (26%) of IT leaders say they need help creating scalable solutions.

Data storage and processing solutions need to bring data close to the users while still operating with security and sustainability in mind. 28% of IT leaders say they need help with ROT data storage, another 28% need help with data preparation and 27% with data processing. These essential steps allow organizations to scale efficiently and ensure that accurate, complete data is available to train and use AI models with confidence. Organizations in Mexico are leading the way, being more likely to have training data version control (81% vs. 67% globally) and tagging for virtualization (69% vs. 62% globally). 45% of IT leaders working with external partners to implement AI said the main benefits included future-proof structures and frameworks. This future-proofing was the most cited benefit in both the US (54%) and Australia (63%).

Software that is secure against malicious attacks and internal mistakes is key. It can also add a layer of data virtualization, which makes an organization's data even more accessible for AI and staff who need access. Considering LLMs and AI models, a third (32%) of IT leaders say they need outside help to build them effectively.

Skilled staff are essential to successful AI implementation, but many are learning to deal with these new technologies on the fly. 48% of IT leaders say they are building the skills to successfully implement AI through experimentation, and 35% say they are self-taught. Two-thirds (64%) of IT leaders say they are hiring the skills they need.

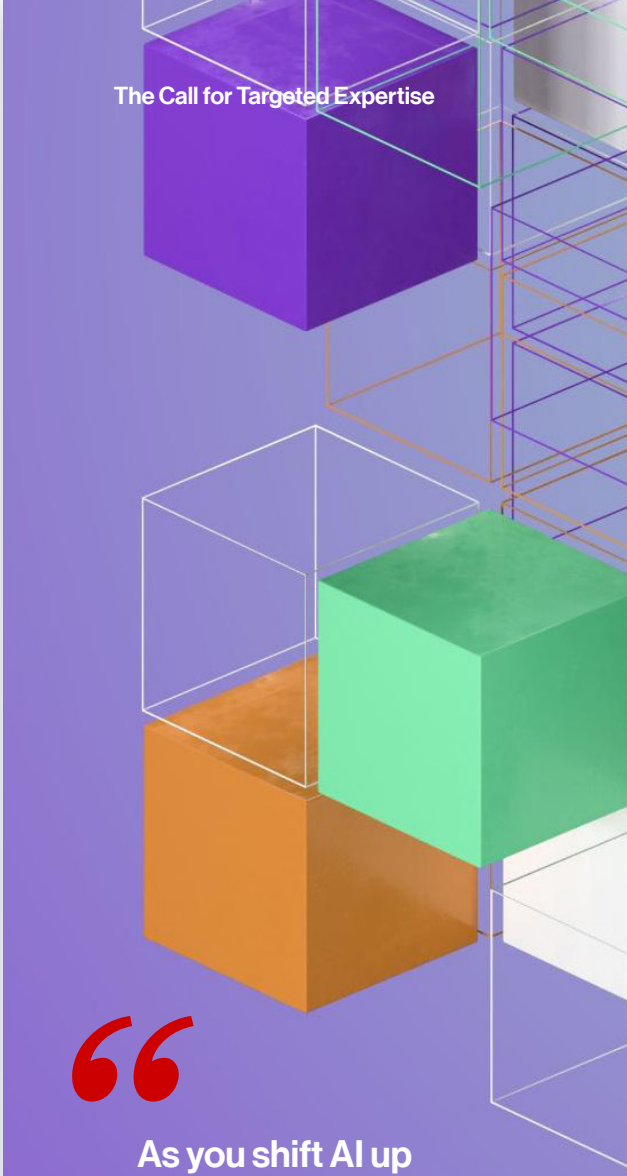
Where have you gained the skills to implement AI?

64%
Hired staff with AI expertise

61%
Consulted external AI experts

48%
Experimented

35%
Self-taught

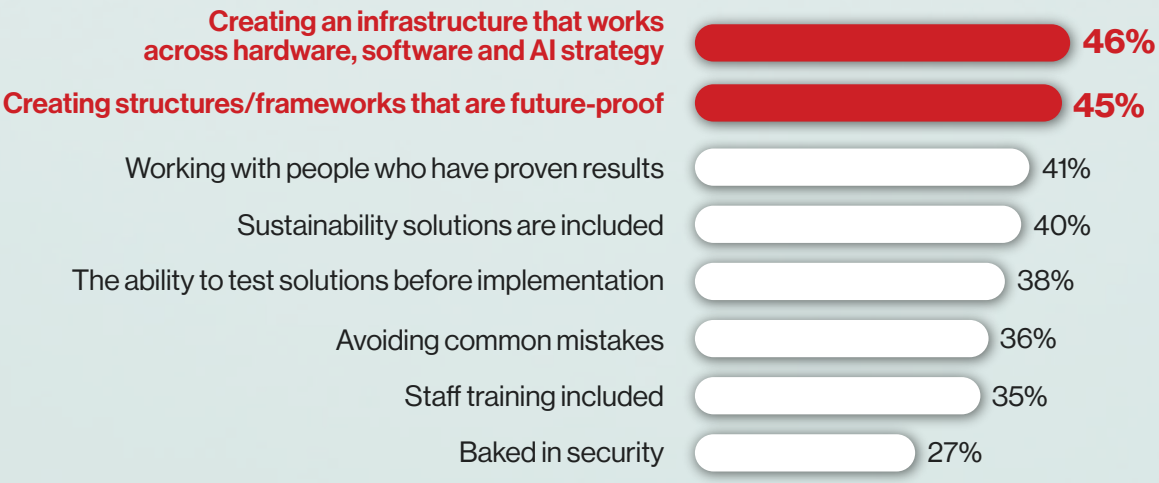


"As you shift AI up the value stack, it impacts your business model. You care more about your data and what you do with it. That translates into how you must position to solve very specific customer problems from what you know, and how solutions are packaged, offered and managed."

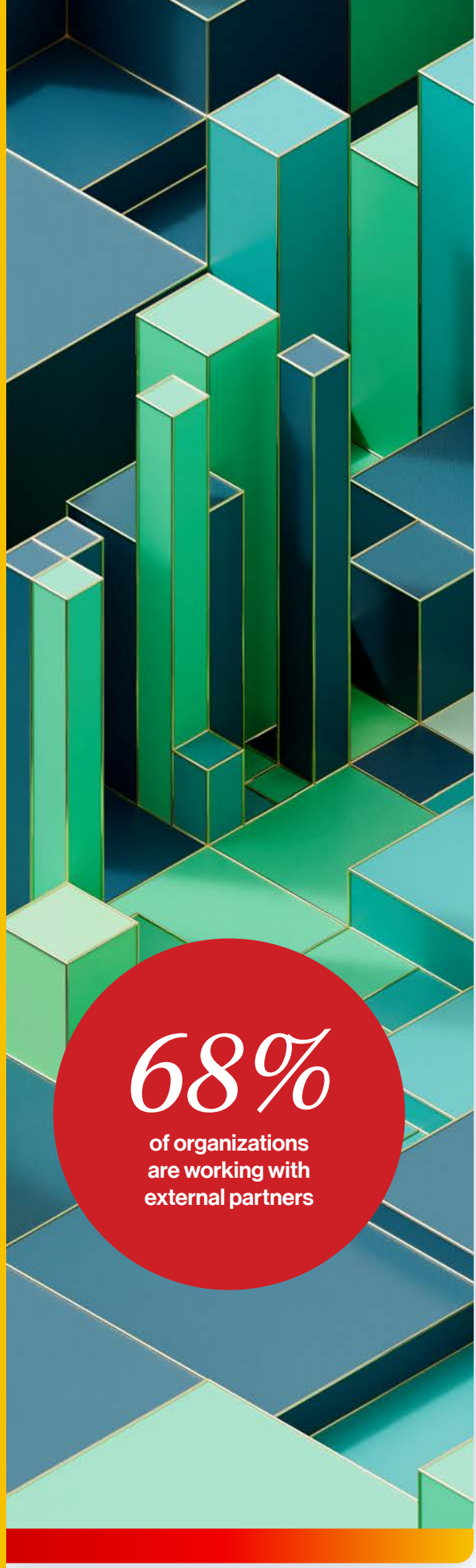
Simon Ninan
SVP of Business Strategy
Hitachi Vantara

When learning and skills acquisition need to scale, getting outside training can be beneficial. A third (35%) say the included training is a major advantage of working with a third party. IT leaders say IT staff is the main group that needs training to implement AI (31%). This is a present issue in Brazil where almost half (47%) require AI training for IT staff. However, a fifth (19%) say they also need help educating their leadership, and a quarter (25%) refer to getting help training users.

The Benefits of Working with a Partner



Most organizations have deep knowledge that is distinctly theirs. They can focus on **building** those areas where they are uniquely situated to make the most of AI. For the rest, there is an entire ecosystem of third parties who can provide products to **buy** or services to **borrow** to create scalable, sustainable and secure solutions, ultimately leading to success.



Buy, Build or Borrow Strategically

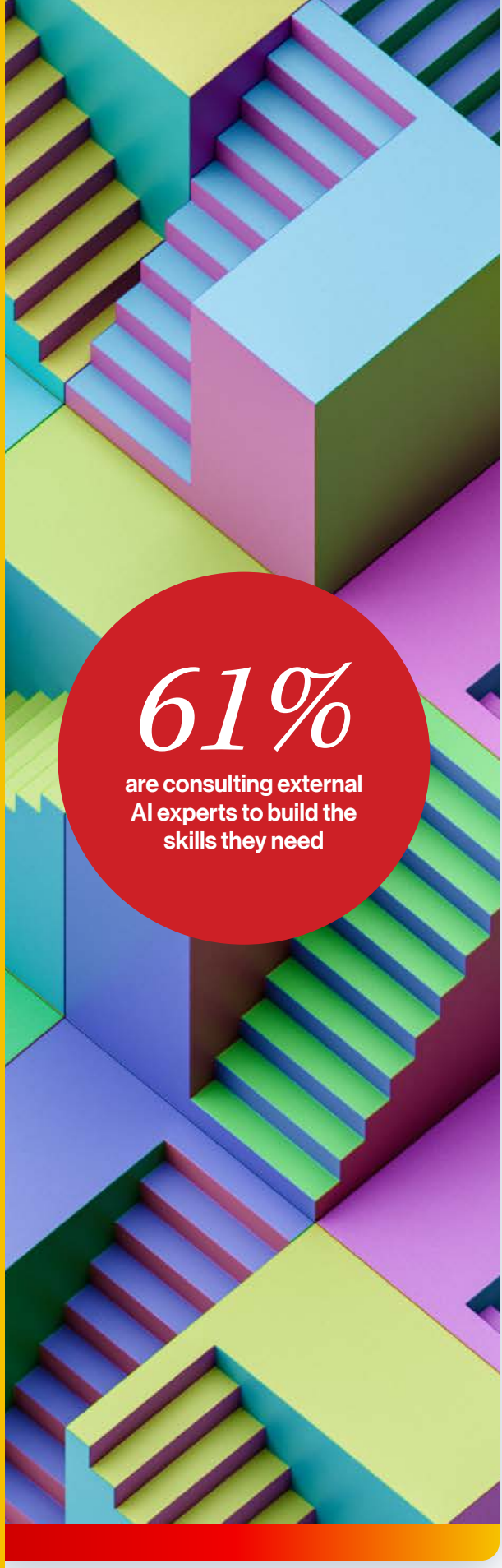
Most organizations are **buying** the help they need, by bringing in either skilled staff or technology tools. This is especially true of AI models themselves, where 56% have purchased an LLM/AI from another company, and 59% have rented a model. Regionally, the UK (81%) has the highest number of organizations renting models, and 65% say they are highly successful with those models. The use of publicly available LLMs is more common in Europe (75%) and Asia (71%), than in the Americas (62%).

Many organizations are also **borrowing** the skills and technologies they need. 68% are working with external partners, 46% are working with external experts to manage areas of AI where they need help and 37% are outsourcing to third parties.

Savvy IT leaders utilize outcome-based models to avoid paying for products that don't perform as expected and safeguard mission critical services with appropriate service level aagreements (SLAs). These agreements move away from buying products, and towards buying infrastructure as a service (IaaS), reducing risk.

"Companies want to work with partners that help them grow, help them be more efficient or reduce and mitigate risk. We are providing automation which translates into operational simplicity, so companies are more efficient. If companies get more insights out of the data, that will help them compete and grow."

Octavian Tanase
Chief Product Officer
Hitachi Vantara



61%

are consulting external AI experts to build the skills they need

Building solutions internally is an approach used by 39% of organizations who are training their staff to address their unmet AI needs. 60% have decided to build their own LLM/AI models, more often in professional services (75%) and IT sectors (71%). However, even those building their own solutions are getting help to do it. 63% are partnering with global systems integrators (GSIs) to build their models, and 61% are consulting external AI experts to build the skills they need.

"When core data is the critical piece to get the right AI, Hitachi Vantara ensures all your data is available to your AI wherever it is: block, file or object; and run cost-effectively on AI chips."

Dan McConnell
SVP Product Management and Enablement
Hitachi Vantara

Wherever IT leaders bring in third parties to help, just getting technical expertise isn't enough. Partners with deep industry knowledge of the business problems and outcomes are essential to understanding what the "right" accurate data means – **the data that really matters**.

06

The Hitachi Vantara POV: AI Demands Fresh Ground

Our research uncovers a glaring truth: IT leaders understand data quality is crucial for AI success, yet they fail to act on it. And while they battle urgent security threats, sustainability is left behind.

It's time to rethink AI's foundation. AI's brilliance demands more from its infrastructure. It must be robust and scalable. Prioritize high-quality data. Ensure tight security. Enforce total data governance. Integrate sustainability. Focus is needed on all these aspects, all the time. Neglect any, and you'll be left scrambling to catch up. Left in the dust. Or worse — facing the steep cost of undoing stakeholders' trust.

Hitachi Vantara's 10 Dimensions for Data Quality

Accurate
Does the data reflect reality?

Complete
Is any crucial data missing?

Consistent
Is the data uniform in formats, units and naming conventions?

Unique
Is there duplicate data?

Reliable
Is data based on typical scenarios?

Accessible
Is data accessible to models as needed?

Timely
Is the data current and relevant?

Traceable
Can data origins and processing be tracked?

Tagged
Is metadata available to add context and meaning?

Clean
Does our data contain personal, sensitive or proprietary information?

Prioritize Complete, High-Quality Data From the Start

AI models are only as good as the data they're trained on. Poor data quality means poor AI performance. If data quality is a widespread issue, restrict AI models to only use data that has been rigorously screened.

Good data quality rests on the foundations of a robust data infrastructure. Consider hybrid cloud environments to manage data while balancing cost, access and security requirements.

"IT leaders need virtualized access to their data at every site, whatever format, from edge-to-core-to-cloud, all the time. Hyperscalers want enterprises to shift all data to the cloud and most companies will shift some, but not their core, important, critical, sensitive data. The irony is this data is what will make an AI model be industry leading – this is only going to be possible through Hitachi Vantara."

Jason Hardy
VP & CTO - Artificial Intelligence
Hitachi Vantara

How good is good enough with data quality?

An organization's entire data system doesn't need to be perfect. To begin with, only the data used to build AI models and test the outputs needs to be fit for purpose. IT leaders can create a secure sandbox to test use cases with curated data they know and understand.

"Quality is specific to the use case that is being considered. No organization of a major size can clean all their data, it must be use-case specific. For example, if companies are going to implement a service chatbot, then you just need your product manuals. You don't need to worry about your financial data or your customer database."

Sasan Moaveni
Global Business Lead for AI &
High-Performance Data Platforms
Hitachi Vantara

More IT leaders attribute AI successes to using high-quality data than any other factor. It is crucial to guarantee that the data AI models use for training doesn't poison the outputs. Hitachi's dimensions of data quality can help IT leaders verify if they are using the correct data for their models.

Experiment Responsibly

As with any new technology, IT leaders need to experiment to realize the potential of generative and more traditional AI. Half of the IT leaders (48%) we spoke to said they were building the skills they needed to implement AI by experimenting.

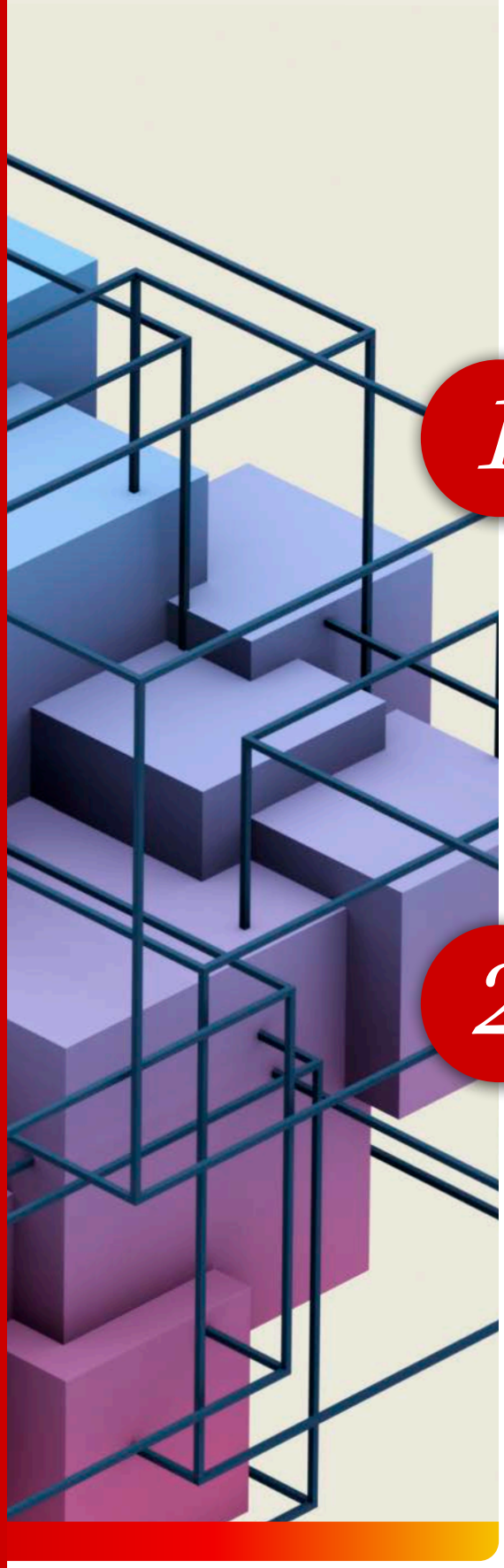
However, without considering the impact and mitigating the pitfalls of implementing AI, a strategy of experimentation risks failure before it has a chance to shine.

- Create sandboxes to isolate experimental AI until it can be understood.
- Monitor the data that AI uses to make sure it is high quality data.
- Match the accuracy of AI outputs to the quality of the data inputs. If the data is good, but the output is wrong, the model must be the problem.
- Use the data collected during experimentation to create better quality datasets for future use.

Implement Sustainable Solutions at Every Step in the Stack

Sustainability is something to consider at every level of the value chain:

- Choosing eco-friendly **data storage** is the first step. Most experts are choosing to outsource these solutions, letting third parties manage sustainability by maximizing economies of scale.
- Closer to home, **software and applications** can be optimized to run efficiently, though this can reduce processing power and data accessibility. This will be important for markets like Malaysia (43%) and Italy (41%), where processing power is still a top concern. Hybrid cloud solutions can give IT leaders the resources to monitor most of these applications from a single suite and already optimize many of the infrastructure tools for sustainable use.
- AI is a resource-intensive tool, and often, other applications are already built specifically to solve business problems more efficiently. Understanding when AI is the correct tool for the job requires **strategic decision-making** that can free up resources. IT leaders must integrate sustainable thinking into their infrastructure, applications, models, data practices and strategies from the start.



Stay Current With Security Best Practices

1

Create and implement effective policies

As staff are trained to follow technology policies, it becomes more important to ensure those policies are effective and up to date. As AI touches more areas of the organization, bringing in diverse resources to build robust policies and procedures becomes vital.

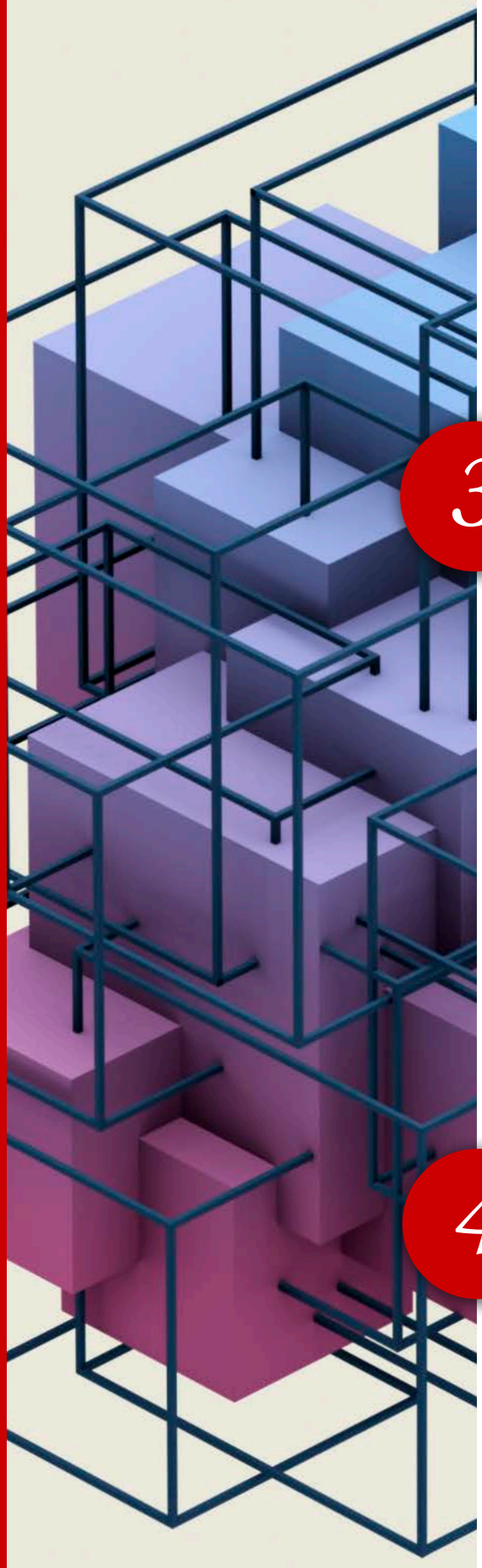
> **"We created an AI governance committee, with key resources, including obviously, information security, but also legal, compliance, risk management."**
>
> **CISO**
> Global financial services company
> USA

2

Data governance across all layers of risk

How will your organization recover when something goes wrong? Consider fallbacks and rollbacks. When hardware or software fails, what are the redundancy systems that will take over? When data gets corrupted, tainted or attacked, how can you roll back storage and AI models to mitigate the impact?

> **"If you're using internal documents, source code, images, or presentations, do you have the rights to that? Is there a simple way to trace to the data sources you trained your AI with it? Maybe it was some source code you don't have the right for or maybe it's tainted with a viral GPU license. That's now problematic. We're building a data-time-machine that can isolate the training data at particular model versions so that you can recover from issues like this."**
>
> **Octavian Tanase**
> Chief Product Officer
> Hitachi Vantara



3

Reduce complexity

57% of IT leaders are trying to protect public, private, on-premises and hybrid cloud environments simultaneously. They are handling more data in more different formats than ever before. The cloud has rapidly stretched the threat surface, creating more potential vulnerabilities.

> **"We have data centers that have private cloud environments, and then we have production workloads in four different public cloud environments. The Azure, AWS, GCP and Oracle Cloud Infrastructure, all over the place. Fun! I get to secure them all."**
>
> **CISO**
> Global financial services company
> USA

Reducing the complexity of data infrastructure through uniformly managed hybrid cloud environments allows for better vulnerability management. More than ever, vulnerability testing and management are necessary to reveal weaknesses in configurations or application design. Many of these tasks can be automated for greater simplicity.

4

Fight fire with fire

Although bad actors are leveraging AI and technology, these tools are also being considered for security purposes. A third (33%) of IT leaders see AI as a tool to identify risks. Already, there are solutions where generative AI understands prompts being used on other AI applications. It then vets those prompts, identifying whether they come from internal or external sources and provides guardrails and protection against unintended consequences.

Pick the Right Tool for the Right Job

AI, specifically generative AI, is the newest, latest and greatest technology. It has applications and use cases in previously unimaginable areas. That doesn't mean it's always the right tool for the job. If you want to add up a few numbers, a calculator will perform better — It's more reliable, well-tested for security, and uses much less computing power.

IT leaders must educate users on the best tools for their use cases and set expectations with CEOs, customers and users, many of whom believe AI is either magic or a silver bullet that can solve every problem. Although IT leaders are not immune to this temptation, they are more critical. 60% believe AI is only suitable for certain business situations.

"Don't solve problems that have already been solved. There's a temptation to use AI for everything, but GenAI is only one aspect of broader analytics. It's the shiny tool businesses are applying to everything, but it's is an addition — not a fix all. You don't need AI for 2+2."

Jason Hardy
VP & CTO - Artificial Intelligence
Hitachi Vantara

60%

of IT leaders believe that AI is only suitable for certain business situations

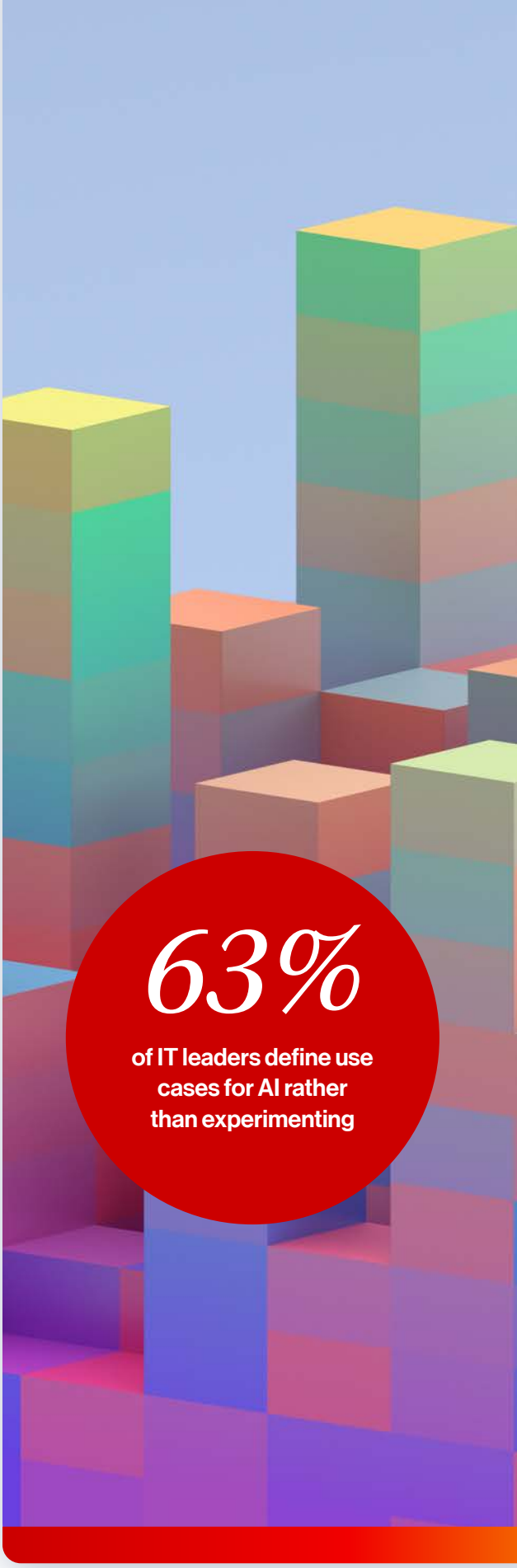
Set Good Goals

Finally, understanding what you want to achieve with AI is crucial to knowing if an implementation is successful. Solid governance and project management was the second most common marker (37%) for successfully implementing AI – this was top in Asia (45%) and Europe (37%).

"The threshold of trust is if the level of data quality is sufficient to help companies reach a critical point where they rely on AI systems. For companies like banks, this level is higher than in other sectors, but everyone needs a base level of quality."

Simon Ninan
SVP of Business Strategy
Hitachi Vantara

Currently, 63% of IT leaders say they define use cases for it rather than experimenting with AI on everything. This is most common in healthcare where 75% say they have defined use cases. Along with use cases, key performance indicators (KPIs) should be identified to describe success. Those KPIs are diverse. Customer and user satisfaction, conversion and adoption rates, among others, are critical to understanding if AI models have met the "Trust Threshold.[8]"



The End of the Beginning

AI has captured the world's imagination, but we're just scratching the surface. To lead the way, organizations must build scalable, sustainable and secure data infrastructure. Trustworthy data drives AI's transformative power. Only with unwavering attention to data quality can organizations unlock AI's full potential.

Learn More →

Research Methodology

The initial research for this report and analysis was conducted through an online survey with industry decision-makers and experts.

Between August 17 and September 13, 2024, 1,200 IT leaders were asked about how their companies are using AI and the state of their data infrastructure.

Leaders were either C-suite executives involved in data (CTO/CDO/CIO/CISO) or IT management.

All companies had a minimum annual global turnover of $100 million. We surveyed across 15 markets and from a wide variety of industries. The markets included are the USA, UK, India, China, France, Mexico, Brazil, Spain, Australia, Germany, Canada, Italy, Indonesia, Malaysia and Singapore.

The data is weighted to ensure results are comparable to 2023 data. IT leaders and C-suite executives were weighted 70:30 respectively.

The Experts We Spoke To

We would like to thank five experts from global companies around the world that we spoke to for their opinions on the state of data infrastructure and the impact of AI. Their comments and quotes are used with permission. In some cases, their identity has been anonymized at their request.

Jim Rutt, CIO, The Dana Foundation, USA

José Dantas, CTO, Voltz, Brazil

Srihari Udugani, VP of Technology Innovation and Operations, Borderless Access, India

CISO, Global financial services company, USA

CTO, Global telecoms company, Germany

Sources

[1] Modern Data Infrastructure Dynamics – Drowning in Data: A Guide to Surviving the Data-Driven Decade Ahead, Hitachi Vantara, 2023

[2] Calculated using 25GB per film and 65,000 films, Statista.com (NA), Statista.com (Europe), Criterionforum.org

[3] New Gen AI Research From Google Cloud and National Research Group Shows Strong ROI for Early Adopters, Aug., 2024

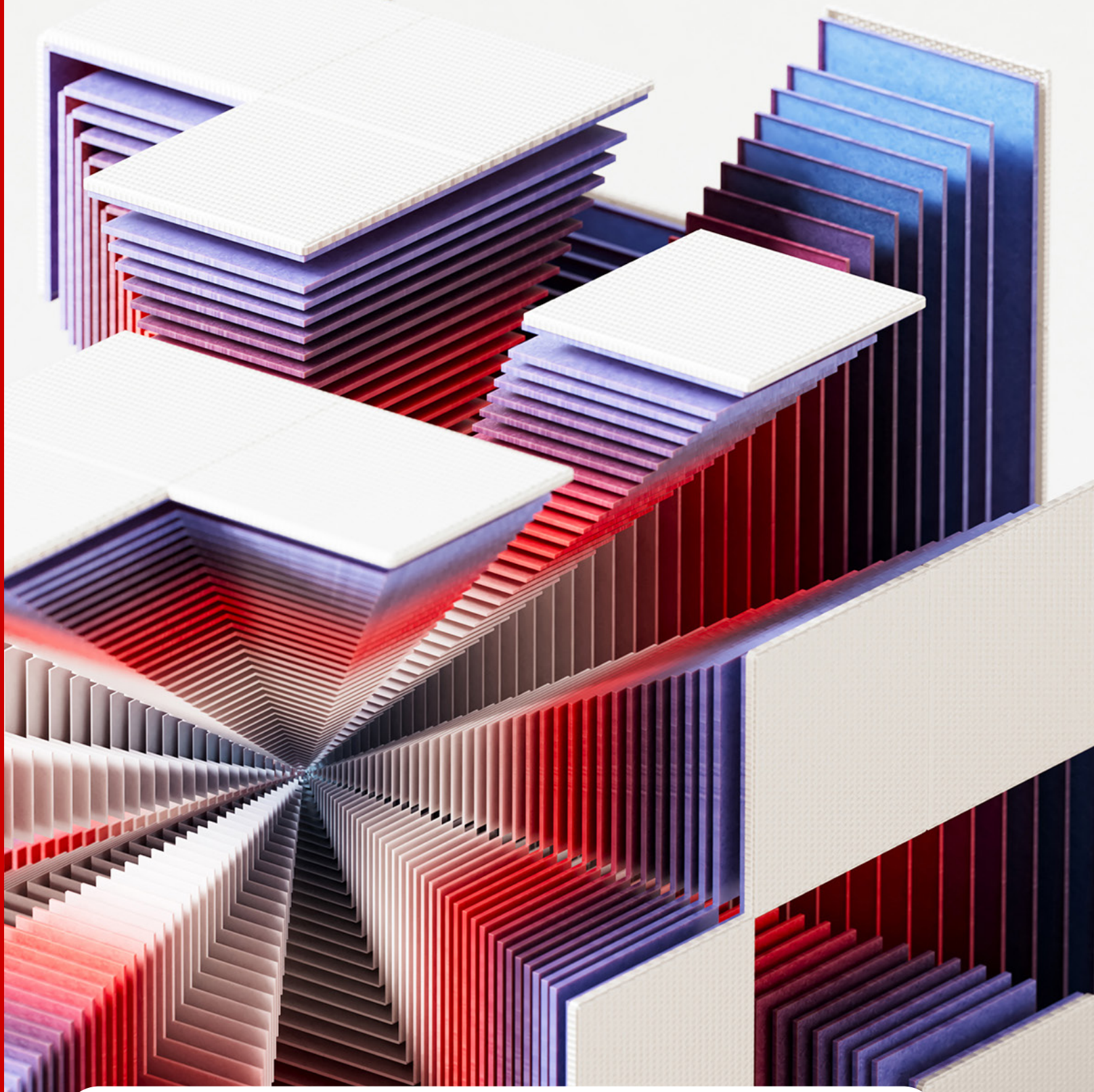
[4] Hungry for Energy, Amazon, Google and Microsoft Turn to Nuclear Power, New York Times, Oct. 2024

[5] Computational Power and AI, The AI Now Institute, Sept. 2023

[6] The Growing Energy Footprint of Artificial Intelligence, Alex de Vries, Joule, Oct. 2023

[7] SEC Adopts Climate-Related Disclosure Rules, Norton Rose Fulbright, April 2024

[8] Crossing the Trust Threshold, Upward, May 2023



Hitachi Vantara

About Hitachi Vantara

Hitachi Vantara is transforming the way data fuels innovation. A wholly owned subsidiary of Hitachi, Ltd., we're the data foundation the world's leading innovators rely on. Through data storage, infrastructure systems, cloud management and digital expertise, we build the foundation for sustainable business growth.

HV-GBS-RE-State-of-Data-Infrastructure-Global-Report-2024-20DEC24-B

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact