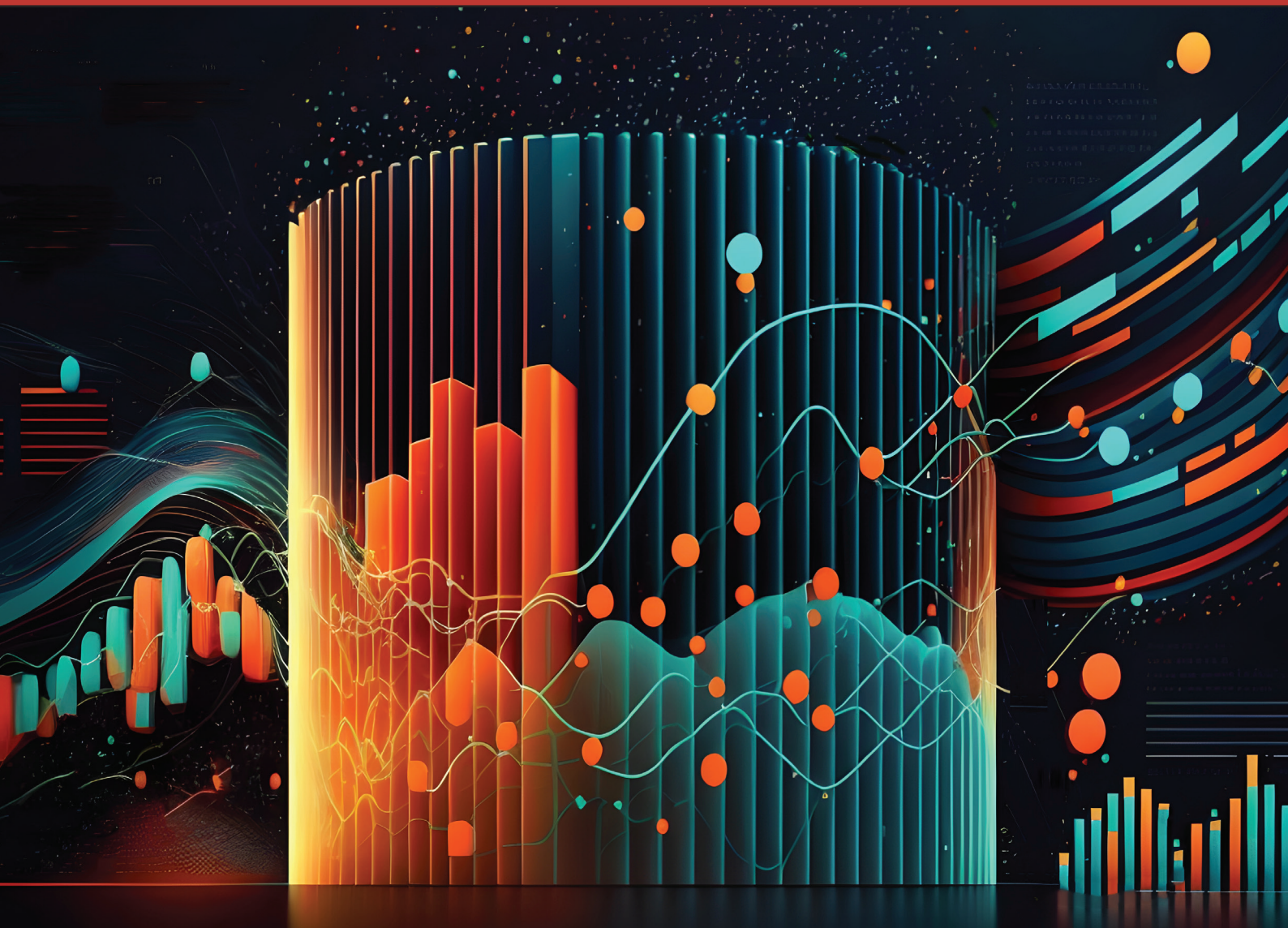# Laying the foundation for data- and AI-led growth

Preface

"Laying the foundation for data- and AI-led growth" is an MIT Technology Review Insights report sponsored by Databricks. To produce this report, MIT Technology Review Insights conducted a global survey of senior data and technology executives. The report also draws on in-depth interviews with CIOs, CTOs, and CDOs of large private and public sector organizations, conducted between June and August 2023.

Denis McCauley was the author of the report, Teresa Elsey was the editor, and Nicola Crepaldi was the publisher. The research is editorially independent, and the views expressed are those of MIT Technology Review Insights.

We would like to thank the following executives for their additional time and insights:

Jack Berkowitz, Chief Data Officer, ADP

Sanjay Bhakta, Chief Product and Technology Officer, Condé Nast

Murali Brahmadesam, Chief Technology Officer and Head of Engineering, Razorpay

Jon Francis, Chief Data and Analytics Officer, General Motors

Deb Hall Lefevre, Executive Vice President and Chief Technology Officer, Starbucks

Yemi Oshinnaiye, Chief Information Officer, U.S. Transportation Security Administration

Jeffrey Reid, Chief Data Officer, Regeneron Genetics Center

John Roese, Global Chief Technology Officer, Dell Technologies

Naveen Zutshi, Chief Information Officer, Databricks

About the survey

The survey that forms the basis of this report was conducted by MIT Technology Review Insights between June and August 2023. Responses were collected from 600 senior technology executives.

The survey respondents all work in large enterprises or public sector organizations, headquartered in 12 countries in North America, Europe, Asia-Pacific, and the Middle East. The majority (75%) are C-level executives, with the balance consisting of SVPs, VPs, and heads of IT, AI, data, engineering, and similar roles. They represent eight different industry sectors, and all respondents work at organizations earning $500 million or more in annual revenue.

Job titles included
• Chief information officer
• Chief technology officer
• Chief data/analytics officer
• Chief data scientist
• Chief enterprise/data architect
• SVP/VP/head of IT, AI, data, engineering, or similar

Eight industries are represented
• Retail & CPG
• Media & entertainment
• Telecommunications
• Health care & life sciences
• Financial services
• Energy
• Manufacturing
• Government & public sector

Respondents come from twelve countries, representing four regions

NORTH AMERICA
• Canada
• United States

EUROPE
• France
• Germany
• Netherlands
• United Kingdom

ASIA-PACIFIC
• Australia
• India
• Japan
• Singapore
• South Korea

MIDDLE EAST
• Israel

# CONTENTS

01

Executive summary

Enterprise adoption of AI is ready to shift into higher gear. The capabilities of generative AI have captured management attention across the organization, and technology executives are moving quickly to deploy or experiment with it. Many organizations intend to increase their spending on the wider family of AI capabilities and the data infrastructure that supports them by double digits during the next year. And notwithstanding concerns about unfavorable economic conditions, executives see opportunities to leverage data and AI to deliver more growth to their organizations, to both the top and bottom lines.

Based on a global survey of 600 technology leaders and a series of in-depth interviews, this report finds that organizations are sharply focused on retooling for a data and AI-driven future. Everything from data architecture to AI-enabled automation is on the table, as technology executives strive to find new efficiencies and new sources of growth. At the same time, the pressure to democratize the power of data and AI creates renewed urgency to bolster data governance and security.

Following are the study's key findings:

• **CIOs are doubling down on their investments in data and AI.** Faced with increasing audience expectations, new competitive pressures, a challenging economic backdrop, and an unprecedented speed of innovation, technology leaders need their data and AI assets to deliver more growth to the business than ever before. They are investing to secure this future: every organization surveyed will boost its spending on modernizing data infrastructure and adopting AI during the next year, and for nearly half (46%), the increase will exceed 25%.
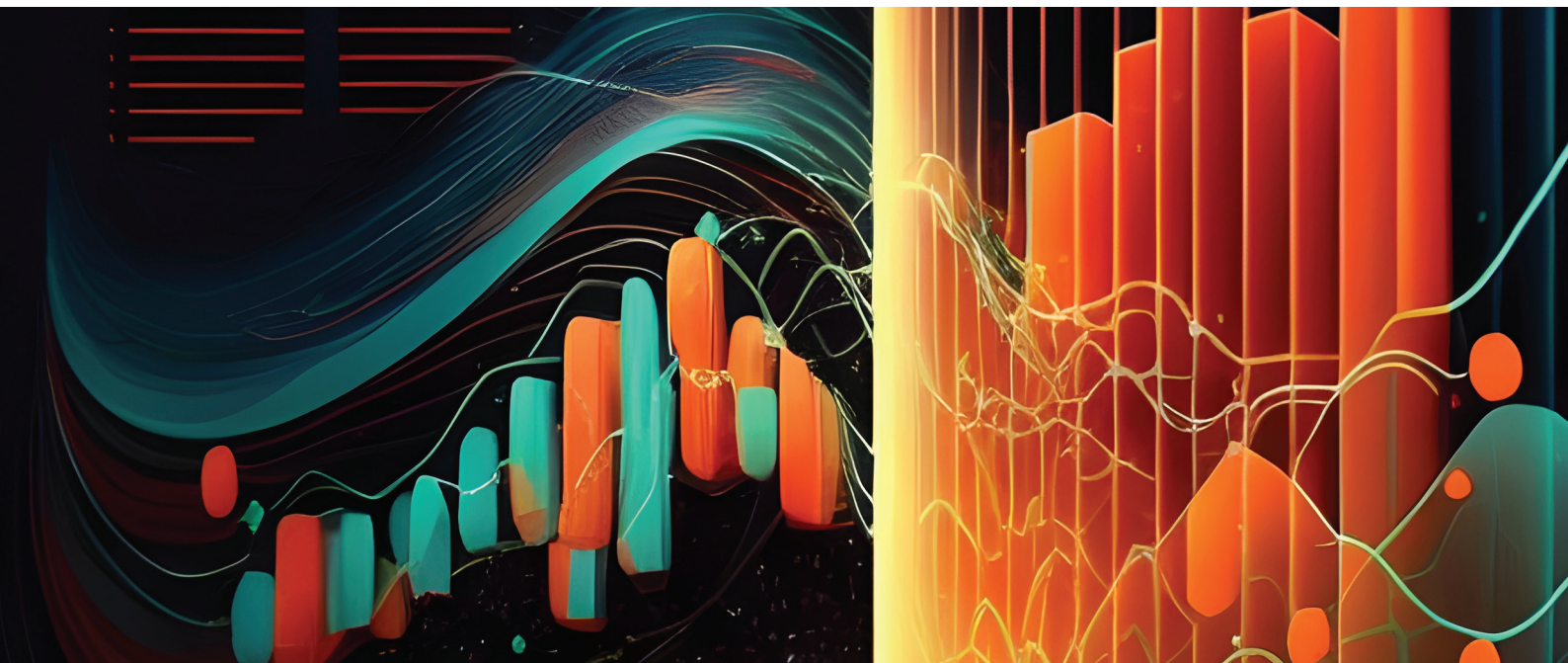
• **Consolidation of data and AI systems is a priority.** The proliferation of data and AI systems is particularly extensive in the survey's largest organizations (those with annual revenue of more than $10 billion). Among these, 81% operate 10 or more of these systems, and 28% use more than 20. The executives we interviewed aim to pare down their multiple systems, connecting data from across the enterprise in unified platforms to break down silos and enable AI initiatives to scale.

• **Democratization of AI raises the stakes for governance.** As business units and their staff clamor to use generative AI, executives seek assurance that governance frameworks for the technology can provide not only the needed data accuracy and integrity but also adequate data privacy and security. That's probably why 60% of respondents say a single governance model for data and AI is "very important."

• **Executives expect AI adoption to be transformative in the short term.** Eighty-one percent of survey respondents expect AI to boost efficiency in their industry by at least 25% in the next two years. One-third say the gain will be at least 50%.

• **As generative AI spreads, flexible approaches are favored.** Eighty-eight percent of organizations are using generative AI, with one-quarter (26%) investing in and adopting it and another 62% experimenting with it. The majority (58%) are taking a hybrid approach to developing these capabilities, using vendors' large language models (LLMs) for some use cases and building their own models when IP ownership, privacy, security, and accuracy requirements are tighter.

• **Lakehouse has become the data architecture of choice for the era of generative AI.** Nearly three-quarters of surveyed organizations have adopted a lakehouse architecture, and almost all of the rest expect to do so in the next three years. Survey respondents say they need their data architecture to support streaming data workloads for real-time analytics (a capability deemed "very important" by 72%), easy integration of emerging technologies (66%), and sharing of live data across platforms (64%). Ninety-nine percent of lakehouse adopters say the architecture is helping them achieve their data and AI goals, and 74% say the help is "significant."

• **Investment in people will unlock more value from data and AI.** In our survey, talent and skills gaps overshadow organizations' other data and AI challenges. When asked where their company's data strategy needs to improve, the largest share of respondents (39%) say investing in talent. The number-one difficulty they face with their data and AI platforms, with 40% citing this as a top concern, is training and upskilling staff to use them.

A subsequent report will examine these survey results in detail, accompanied by insights from additional executive interviews across six sectors: financial services, health care and life sciences, retail and consumer packaged goods, manufacturing, media and entertainment, and government.

Faced with increasing audience expectations, new competitive pressures, a challenging economic backdrop, and an unprecedented speed of innovation, technology leaders need their data and AI assets to deliver more growth to the business than ever before.

02

New vistas for AI-led growth

Amid uncertainty in the macroeconomic environment, chief executives are demanding that their technology investments, including those in data and AI, work harder and deliver more value to the organization than ever before.

The emergence of generative AI provides technology executives with an additional set of tools to achieve that. At the same time, for all its widely publicized promise, generative AI adds complexity to this challenge. As more parts of the enterprise clamor to experiment with it, CIOs must ensure that their data infrastructure is robust enough to cope with the enormous data processing demands and governance challenges posed by these advances.

The technology leaders participating in this study see this challenge as an opportunity for AI to deliver considerable growth to their organizations, in both their top and bottom lines. While 70% of survey respondents say it's very important for AI projects to help reduce costs, the same percentage say it's very important that these projects enable new revenue generation (although there is variation across industries and geographies). Executives clearly do not consider this a time to batten down the hatches.

According to Deb Hall Lefevre, Starbucks executive vice president and chief technology officer, AI will be powering experiences across the business, "from bean to cup." She envisions, for example, her company's researchers, and even its in-store baristas, using generative AI to help discover and develop new

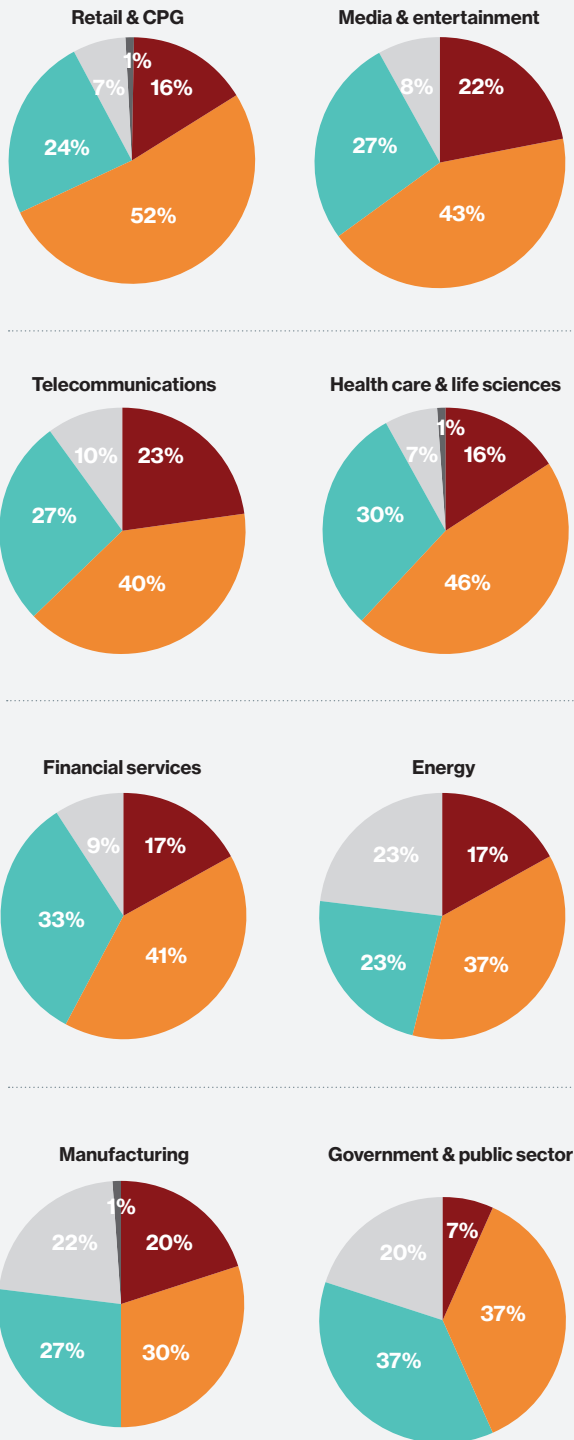"With advances in AI and machine learning and the investments we've made, we're better able now than a few years ago to create efficiencies. We'll do so in the back office with HR chatbots, on the factory floor with predictive maintenance, and in IT operations by scaling and productionizing software development."

Jon Francis, Chief Data and Analytics Officer, General Motors

Figure 1: Rate of adoption

How would you rate the current pace of AI adoption in your industry or field?

Legend: ■ Very fast ■ Fast ■ Moderate □ Slow ■ Very slow

Retail & CPG
- Very fast: 16%
- Fast: 52%
- Moderate: 24%
- Slow: 7%
- Very slow: 1%

Media & entertainment
- Very fast: 22%
- Fast: 43%
- Moderate: 27%
- Slow: 8%

Telecommunications
- Very fast: 23%
- Fast: 40%
- Moderate: 27%
- Slow: 10%

Health care & life sciences
- Very fast: 16%
- Fast: 46%
- Moderate: 30%
- Slow: 7%
- Very slow: 1%

Financial services
- Very fast: 17%
- Fast: 41%
- Moderate: 33%
- Slow: 9%

Energy
- Very fast: 17%
- Fast: 37%
- Moderate: 23%
- Slow: 23%

Manufacturing
- Very fast: 20%
- Fast: 30%
- Moderate: 27%
- Slow: 22%
- Very slow: 1%

Government & public sector
- Very fast: 7%
- Fast: 37%
- Moderate: 37%
- Slow: 20%

Source: MIT Technology Review Insights survey, 2023

---

personalized beverages and to find ways of taking friction out of the drive-through experience. "The scale at which we're doing such things is the differentiator," says Hall Lefevre. "When you can drive innovation at scale quickly, you can really build a competitive advantage."

By and large, technology leaders are not being asked to do more with less. Gartner forecasts worldwide IT spending to grow by 4.3% in 2023 and 8.8% in 2024.[1] Most of that growth will be concentrated in the software category that includes spending on data and AI. Indeed, over the next year, all the surveyed executives expect their organizations' spending on modernizing data infrastructure and adopting AI to increase, and almost half of them (46%) say their spending will rise by more than 25%.

That may be bullish, but most survey respondents perceive similar actions by their peer organizations – 60% of survey respondents say AI adoption in their industry is "fast" or "very fast" (see Figure 1). Though the dynamics vary somewhat by industry (a topic we will address further in a follow-up industry report), the trend is that AI adoption in the enterprise is moving ahead at pace and data modernization to support it remains a priority.

AI-driven efficiency gains promise business growth

The surveyed executives are decidedly upbeat about the efficiency gain they see AI bringing to their industries over the next two years: 81% expect a gain greater than 25% during that period, and 33% believe it could exceed 50% (see Figure 2).

The executives interviewed agree that much of their AI-led growth may come from gains in efficiency. "With advances in AI and machine learning and the investments we've made, we're better able now than a few years ago to create efficiencies," says Jon Francis, chief data and analytics officer at automaker General Motors (GM). "We'll do so in the back office with HR chatbots, on the factory floor with predictive maintenance, and in IT operations by scaling and productionizing software development."

Some organizations aim to use AI-enabled automation to find new efficiencies. "If we can automate our core processes with the help of self-learning algorithms, we'll be able to move much faster and do more with the same amount of people," says Yemi Oshinnaiye, chief information officer with the US Transportation Security

Administration (TSA). "Ultimately for us it will mean automation at scale and at speed."

Accelerated AI requires stricter prioritization

While organizations are eager to adopt AI, technology leaders today are likely to be more careful about how they invest and manage their portfolio of AI projects. "In the current climate, we have to be careful about conserving our resources," says John Roese, global chief technology officer at Dell Technologies. "But we cannot stop adopting AI and advanced data technologies. We have to be much more structured and programmatic about our investments. This is a time where you have to choose which projects you will prioritize because you don't have infinite resources."

Among other things, this means a laser focus on AI project prioritization. This is partly dictated by economic uncertainty. "It has led us to prioritize projects much more ruthlessly than before," says Naveen Zutshi, chief information officer at Databricks. "We've reduced the number of projects we'll pursue, but we're working on

> "But we cannot stop adopting AI and advanced data technologies. We have to be much more structured and programmatic about our investments."
>
> John Roese, Global Chief Technology Officer, Dell Technologies

Condé Nast: AI for social media optimization

Big web platforms are making life difficult for content publishers. That is certainly true for news providers, as Facebook and Google are slashing the volume of news links they include on their sites and declining to pay providers for them. But publishers of lifestyle and specialist magazines, such as Condé Nast, are also finding it difficult to gain space for their content on these platforms. "The tech giants increasingly want to keep audiences on their own platforms," says Sanjay Bhakta, the company's chief product and technology officer. "As a result, we no longer get the volumes of traffic to our sites that we used to get. Our challenge now is to attract people to our online properties and keep them there. That's where we monetize our content."

AI is integral to meeting that challenge, says Bhakta. His team, for example, has built a model to optimize the content that Condé Nast places on social media. Called

Falcon, the model analyzes the content people consume on social media, and then, comparing that with the company's own content, predicts which content is most likely to drive social media users to its websites. Ingesting the large volumes of data the model needs and training it has taken a great deal of time, Bhakta says, but it's been worth it for the impact the model has had on the number of eyeballs migrating to the company's sites.

"Falcon has been phenomenal in helping our social media team figure out what we should post," says Bhakta. "We've tested the model against a human making those decisions, and we've invariably found that what the machine recommends always has a higher impact. We use it aggressively now because it's consistently shown it can drive traffic back to our properties. It's a very important marketing tool for us."

more longer-range ones that will drive significant impact to the company — needle-movers. We've reduced all of the other ankle-biters that IT organizations often spend time on."

AI's rising popularity across business units and departments — given a big injection by generative AI — is another reason to exercise greater discipline in project prioritization. According to Hall Lefevre, that requires more rigor in applying existing metrics and frameworks rather than creating new ones: "There's a tsunami of use cases and proofs of concept coming at us from every part of our business, at a pace that is probably greater than past emerging tech trends," she says. "From a framework perspective, we're asking the same questions — about the expected business impact, the complexity, and the cost to test if it's something that we can truly scale. The key point is that whatever the prioritization frameworks you use, it's never been more important to make sure that you're focused on the right things."

Figure 2: Efficiency gains expected from AI
How much efficiency gain will AI bring to your industry in the next two years?

Up to 25%	20%
26–50%	48%
51–75%	31%
76–100%	2%

Source: MIT Technology Review Insights survey, 2023

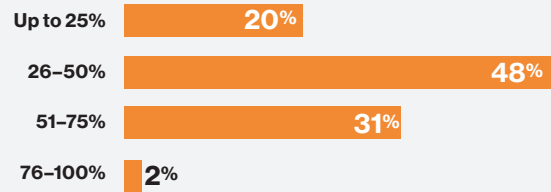# Dell Technologies: Raising the bar for AI use cases

In the accelerated digitalization of enterprises that accompanied the covid-19 lockdowns, technology companies were given greater rein to invest in building new AI and related capabilities than they might have previously enjoyed. That period is over, says John Roese, global chief technology officer at Dell Technologies. "As an industry, we could be a bit more aggressive with projects then, without fully considering the short- and long-term costs and benefits," he says. "That was acceptable because it was all about serving a rapidly expanding market. The result was a lot of activity that wasn't necessarily quantifiable, even as it started to scale and consume resources." The industry has to be much more careful now, he says.

Dell continues to encourage employees to experiment with data and AI projects and to move workloads to the cloud. But Roese requires more rigor to be applied to use cases. His message is twofold. First, he says, "When that experiment looks viable in proof of concept and you decide

to scale it, you must be able to quantify both the near- and long-term costs associated with the activity. If you cannot answer that question, you should not proceed to scale it."

Second, any data and AI project proposal must include a detailed assessment of the business impact it will drive, whether that is customer value creation, cost reduction, or something else. "Failing to do so will probably involve reinterpreting the ROI after starting to scale," says Roese. "It will be based on what you did versus what you aspired to do. And in an AI and data project, you should be aspiring to transform your business." That, he adds, requires clarity about the metric that will be applied to gauge success: "There has to be some permanence around measurement."

03

Modernize and simplify

Data infrastructure and AI systems are becoming increasingly intertwined. The main reason for that is the enormous demands that AI models place on the collection, processing, storage, and analysis of data. Razorpay, an India-headquartered financial services company, began using AI in its operations about five years ago, says Murali Brahmadesam, the company's chief technology officer and head of engineering, and since then the volume of payments it processes has grown enormously. "Our business has grown 10x every two years, as has the scale of our data and processing needs," he says.

"Data infrastructure that worked five years ago doesn't work now," says Brahmadesam. "For example, when smaller we could serve data from an operational database. But when we reached this kind of scale, we had to build a performance-oriented architecture to be able to store data more efficiently in a cost-effective manner while also serving customer needs."

This type of growth, along with advances in AI capabilities, including those demonstrated by generative AI, increase the urgency to modernize legacy data architectures. Another requirement is a simplification of that architecture, consolidating disparate data and AI platforms and unifying governance to the extent possible across those that remain.

Modernize

Generative AI and the LLMs that support it will multiply the workload demands on data systems. It also makes the tasks such systems must handle increasingly complex. "The implications of generative AI for data architecture are enormous," says Roese. "I see three major challenges: feeding huge volumes of mostly unstructured data into models to train them; storing data long-term in ways that are conducive to AI consumption; and putting adequate security around models."

"Data infrastructure that worked five years ago doesn't work now."

Murali Brahmadesam, Chief Technology Officer and Head of Engineering, Razorpay
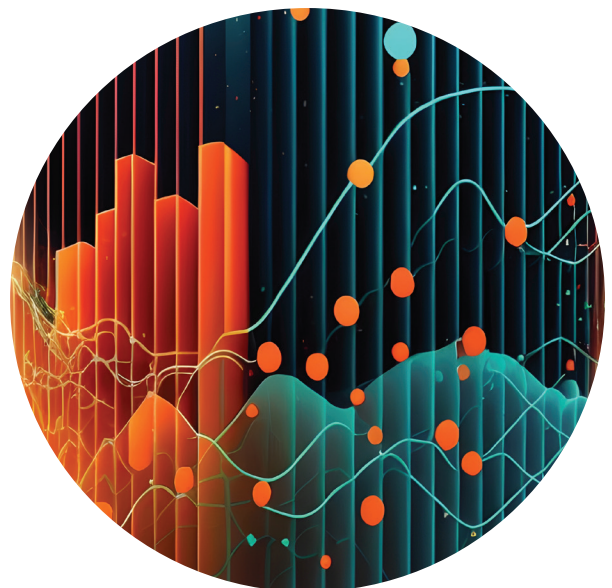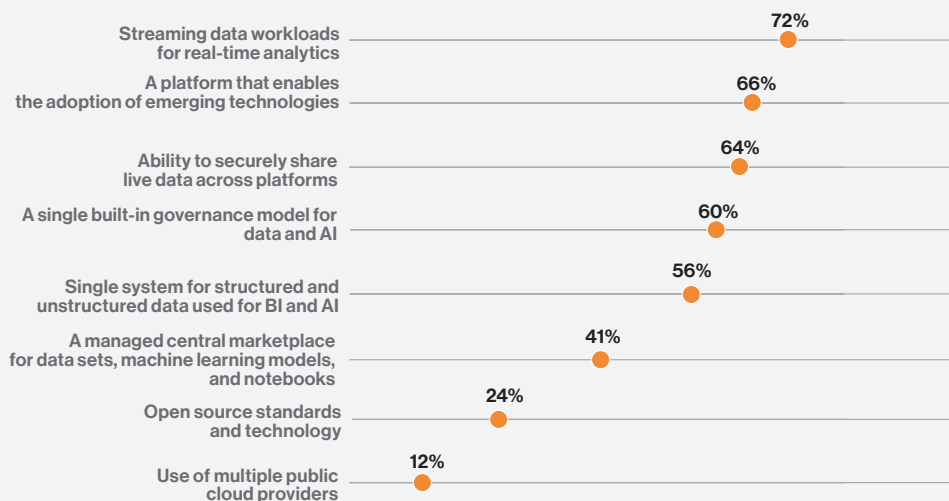
Figure 3: Infrastructure attributes necessary for modernization
In the next two years, how important are the following to achieving your organization's overall technology goals? (percentage responding "very important")

Streaming data workloads for real-time analytics — 72%
A platform that enables the adoption of emerging technologies — 66%
Ability to securely share live data across platforms — 64%
A single built-in governance model for data and AI — 60%
Single system for structured and unstructured data used for BI and AI — 56%
A managed central marketplace for data sets, machine learning models, and notebooks — 41%
Open source standards and technology — 24%
Use of multiple public cloud providers — 12%

Source: MIT Technology Review Insights survey, 2023

The great acceleration, our 2023 report about CIO's perspectives on generative AI, emphasized that organizations supporting LLMs need a data infrastructure that is flexible, scalable, and efficient.[2] The technology leaders surveyed for this study — the majority of whose organizations are already using generative AI — make clear their requirements for their data infrastructure (see Figure 3). Their priorities are streaming data workloads for real-time analytics (cited by 72%), a fundamental capability for many AI applications; the ability to adopt emerging technologies (66%); the ability to securely share live data across platforms (64%); and a single governance model for data and AI (60%).

In modernizing their data infrastructure to deliver these capabilities, three-quarters of the surveyed technology leaders have adopted a lakehouse architecture. Of the one-quarter who have not, nearly nine in ten expect to do so in the next three years (see Figure 4). A lakehouse combines the features of two previously dominant architecture approaches, the data warehouse and the data lake. "The lakehouse architecture we've chosen helps us scale responsibly, with a good balance of cost versus performance," says Brahmadesam.

According to Bhakta, lakehouse architecture is at the heart of everything Condé Nast does with data: "It's enabled us to bring all our data together in one place, ensure that it's clean and that we can stand behind it." With a lakehouse, says Bhakta, his team has created a single data platform with bronze, silver, and gold layers

of data. "Gold is ready for consumption," he explains. "Silver is also ready, but you can do your own visualization on top of it. Bronze is raw data on which you can run your own sophisticated analytics."

TSA's modernized data architecture has given its CIO a clearer view into the data it collects and stores. According to Oshinnaiye, "I can now organize my law enforcement data, I can organize my airline checkpoint data, I can organize my rail data, and I can organize my inspection data. And I can at the same time make correlations and glean understandings from all of that data, separate and together."

Simplify

Data silos are the bane of today's data and technology executives — a legacy of the disparate approaches different parts of organizations took to storing (and protecting) the data they collected or generated, and the different platforms IT acquired or built to manage them. Multiple data, analytics, and AI systems that run internally and in vendor clouds have added to the complexity. The result is a myriad of platforms, a vast amount of data duplication, and, often, separate governance models.

While most organizations in the survey employ fewer than 10 data and AI systems, proliferation of systems is most extensive in the largest ones (see Figure 5). Among organizations with annual revenue of more than $10 billion, 81% have 10 or more of these systems, and 28% employ more than 20.
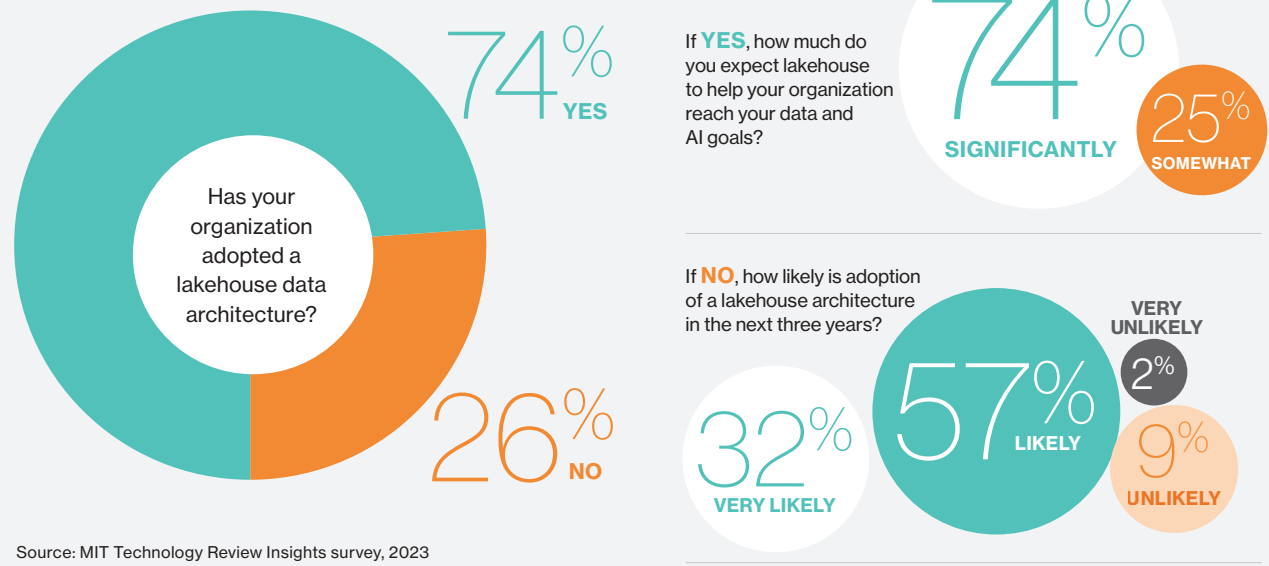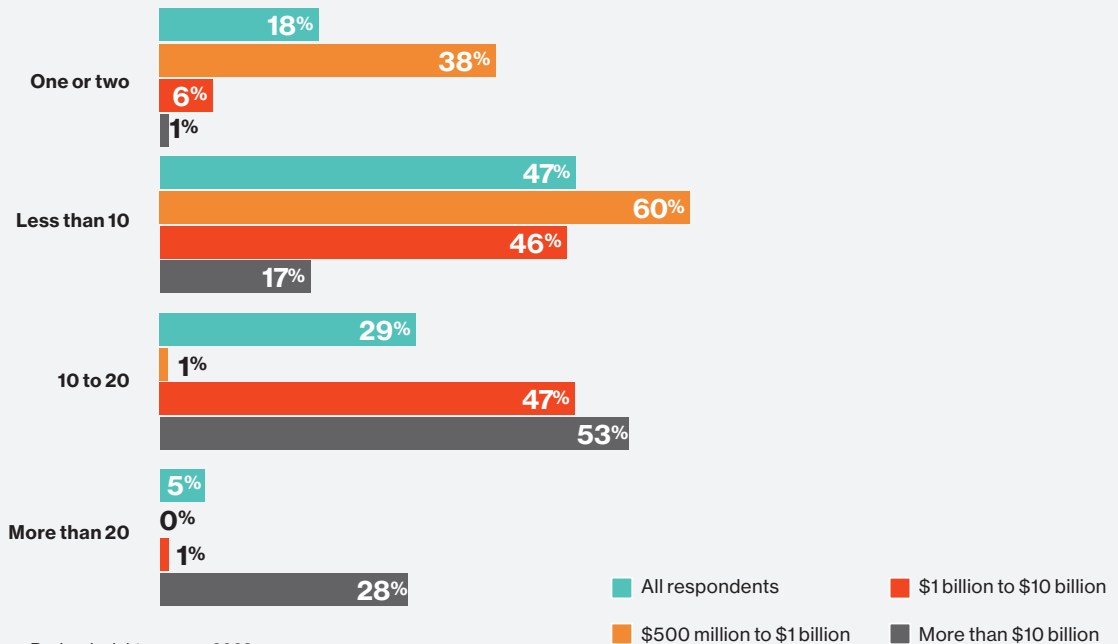
Figure 4: Adoption of lakehouse data architecture

74% YES

26% NO

Has your organization adopted a lakehouse data architecture?

If **YES**, how much do you expect lakehouse to help your organization reach your data and AI goals?

74% SIGNIFICANTLY

25% SOMEWHAT

If **NO**, how likely is adoption of a lakehouse architecture in the next three years?

32% VERY LIKELY

57% LIKELY

2% VERY UNLIKELY

9% UNLIKELY

Source: MIT Technology Review Insights survey, 2023

Figure 5: Proliferation of data and AI systems

How many different data, AI, and ML systems does your organization use? (by company revenue)

One or two
- 18%
- 38%
- 6%
- 1%

Less than 10
- 47%
- 60%
- 46%
- 17%

10 to 20
- 29%
- 1%
- 47%
- 53%

More than 20
- 5%
- 0%
- 1%
- 28%

- All respondents
- $500 million to $1 billion
- $1 billion to $10 billion
- More than $10 billion

Source: MIT Technology Review Insights survey, 2023

"Lakehouse has enabled us to bring all our data together in one place, ensure that it's clean and that we can stand behind it."

Sanjay Bhakta, Chief Product and Technology Officer, Condé Nast

Our interviewees aim to consolidate the number of platforms they use and seamlessly connect data across the enterprise. Zutshi says this type of simplification comes with many benefits. "You don't have to have multiple policy and governance engines, multiple warehouses and separate lake houses; you can have one lakehouse platform with a unified policy and governance engine," he says. "You're not constantly shuttling data back and forth for each persona's use; instead different personas use one copy of data to perform their tasks."

The interviewees' organizations are taking different approaches to simplification. According to Hall Lefevre, Starbucks is centralizing data across its organization by building a set of cloud-centric, domain-specific data hubs. "This is the key piece of the puzzle for us in reducing redundancy, and it provides a strong foundation for our AI use cases," she says.

For GM's data and analytics team, reusable technologies are a focus of its efforts to simplify the infrastructure and avoid duplication. In previous businesses where he has worked, Francis witnessed a proliferation of purpose-built platforms supporting specific applications. "We need to start thinking about more reusable frameworks and technologies when deploying AI, to avoid replicating technology stacks or building bespoke technologies with different licenses," he says. "My data team is working with IT to build more of those reusable frameworks."

Organizations also need space to innovate, and sometimes this is more easily done away from mainstream data systems. While centralization and complete interoperability of data are unambiguously desirable, striking a balance between full integration and freedom to innovate can be a challenge, says Jeffrey Reid, chief data officer of the Regeneron Genetics Center. "We try to achieve this balance by having some functions that are about data management — bringing data together — and other functions where there's a lot of greenfield exploration. Once success has been reached with that exploration, we help bring the data into the fold," he says.

"We need to start thinking about more reusable frameworks and technologies when deploying AI, to avoid replicating technology stacks or building bespoke technologies with different licenses."

Jon Francis, Chief Data and Analytics Officer, General Motors
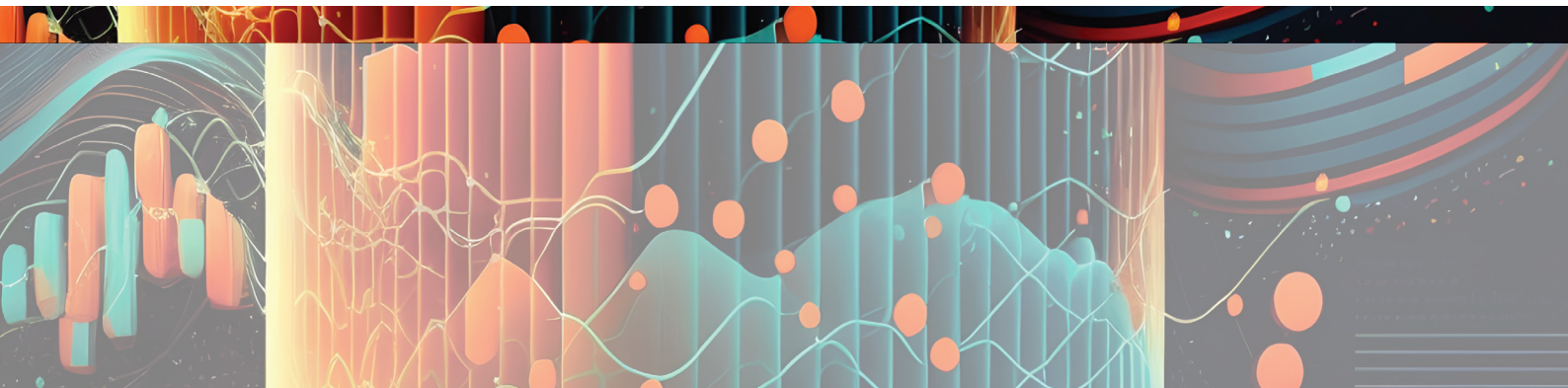
04

The generative AI accelerator

As our report **The great acceleration** makes clear, generative AI has the potential to be a game-changer.[3] ChatGPT and similar chatbots have elevated the possibilities of AI to an everyday conversation topic among the general public and technologists alike.

"You can ask English-like questions that are converted into SQL, and you can convert the responses into insights, rather than waiting on IT or SQL analysts to give you an answer," says Zutshi. "You will see a huge productivity boom as you see those tasks being done by businesses themselves. There's enthusiasm around this right now, and as a CIO you want to leverage it."

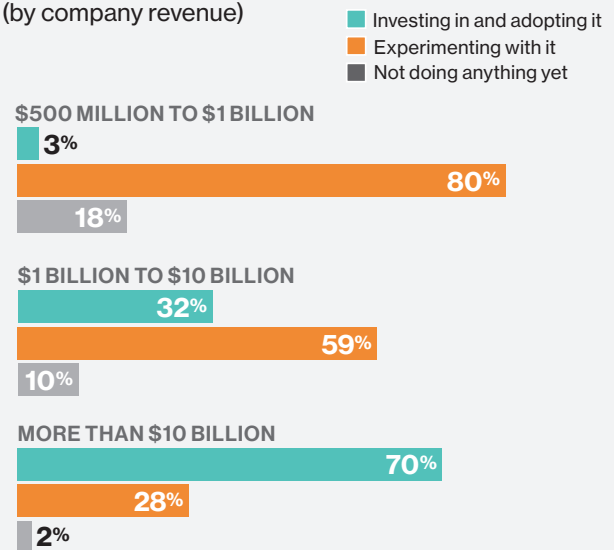## Mounting adoption brings challenges into focus

One-quarter of the surveyed organizations (26%) are currently investing in generative AI systems (see Figure 6). The majority (62%) are experimenting with the technology. The biggest organizations in the survey (those with revenue exceeding $10 billion) appear to have few doubts about its utility: 70% are already investing in and adopting generative AI, compared with just 3% of the smallest ones (those with revenue between $500 million and $1 billion).

The generative AI wave is a global phenomenon — not one just limited to Silicon Valley. Companies across the globe are investing in and experimenting with the technology, seeking to gain competitive advantage. Survey respondents from Singapore and Japan are moving the fastest to adopt generative AI, and despite looming regulatory pressure in the EU, European countries are also leaning in to the technology (see Figure 7).

Personalization, supply chain optimization, and quality control are the generative AI use cases cited most often overall as likely to deliver value (see Figure 8). The relative importance of use cases varies widely from industry to industry, however. For example, real-time data analysis is the top-ranked use case among government organizations and is also of prime importance to financial services and energy providers. Executives from manufacturing, retail, media and entertainment, and telecoms expect generative AI to deliver considerable value in automation and efficiency. And financial services and energy providers look especially favorably on the risk management benefits such models can confer.

Figure 6: Generative AI adoption grows
How is your organization currently using generative AI?
(by company revenue)

- Investing in and adopting it
- Experimenting with it
- Not doing anything yet

$500 MILLION TO $1 BILLION
- 3%
- 80%
- 18%

$1 BILLION TO $10 BILLION
- 32%
- 59%
- 10%

MORE THAN $10 BILLION
- 70%
- 28%
- 2%

Source: MIT Technology Review Insights survey, 2023

Most of the technology leaders interviewed are developing generative AI applications now or are likely to be doing so in the future. Among those yet to take the plunge, there remain unanswered questions about LLMs. "LLMs are definitely in our future," says Oshinnaiye. "It's very important from a customer engagement perspective that we are able to leverage things that are going to become normal." He points out, though, that existing LLMs come with limitations, such as their ignorance of current events: "ChatGPT, for example, may be over a year behind. It's getting better, but you won't always have current information. So how do we take our data sources and our current information and leverage them in an LLM?"

Zutshi adds that lock-in to any particular model is inherently limiting: "You don't want to get wedded to one model, because the models will keep iterating. You should be able to switch models over time, to take advantage of improving efficacy." Purpose-built models and infrastructure may become important options in this space, allowing organizations to clean and vet the data used, feed in their own proprietary data, and take advantage of different models' strengths. "What Databricks did," Zutshi says, "is build an abstraction layer on top of the models, a gateway that gives you the choice. You can pick the right model for the right job."

Generative AI also poses new governance and risk challenges. "In the preceding era of AI, most projects were mainly domain or process-specific," says Roese. "They

"You can ask English-like questions that are converted into SQL, and you can convert the responses into insights, rather than waiting on IT or SQL analysts to give you an answer. You will see a huge productivity boom as you see those tasks being done by businesses themselves."

Naveen Zutshi, Chief Information Officer, Databricks

Figure 7: Regional differences in generative AI adoption
How is your organization currently using generative AI? (by country)

● Investing in and adopting it
● Experimenting with it
● Not doing anything yet

Country	Experimenting with it	Investing in and adopting it	Not doing anything yet
U.S. & Canada	60%	26%	14%
France	70%	28%	2%
Germany	66%	24%	10%
Netherlands	63%	30%	7%
United Kingdom	59%	22%	19%
Israel	63%	20%	17%
India	68%	16%	16%
Japan	60%	35%	5%
Korea	58%	28%	15%
Singapore	53%	37%	20%
Australia	68%	30%	2%

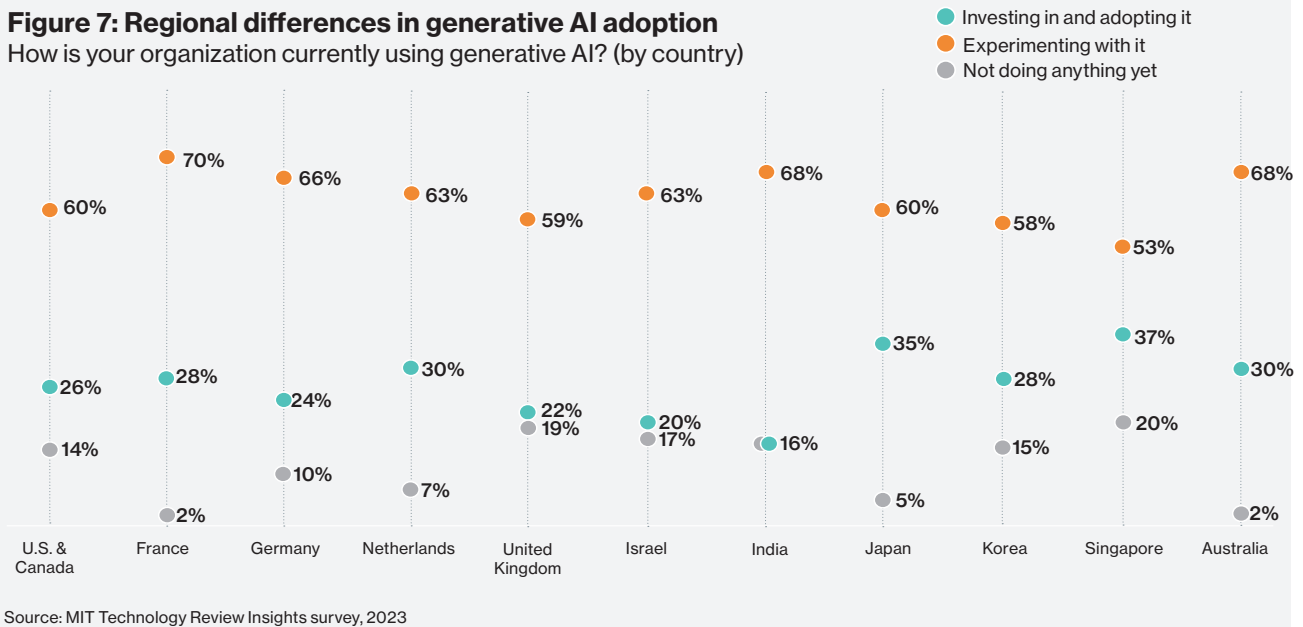Source: MIT Technology Review Insights survey, 2023

Figure 8: Top generative AI use cases by industry
In the next two years, what are the generative AI use cases that will bring the most value to your organization?

Survey respondents were asked to identify their top three choices.

	ALL RESPONDENTS	FINANCIAL SERVICES	GOVERNMENT & PUBLIC SECTOR	HEALTH CARE & LIFE SCIENCES	RETAIL & CPG	MANUFACTURING	MEDIA & ENTERTAINMENT	ENERGY	TELECOM
Personalization and customer experience	1	1	3 tie	1	2		1	3 tie	
Supply chain optimization	2			2	1	1		1	
Quality control	3			3		2		2	1
Real-time data analysis and insights		2	1					3 tie	2
Automation and efficiency			2		3	3	2		3
Innovating products and services						3			
Predictive maintenance			3 tie						
Risk management		3						3 tie	

Source: MIT Technology Review Insights survey, 2023

> "At the end of the day, it's about the data quality of service to feed our machine learning models."
>
> Jack Berkowitz, Chief Data Officer, ADP
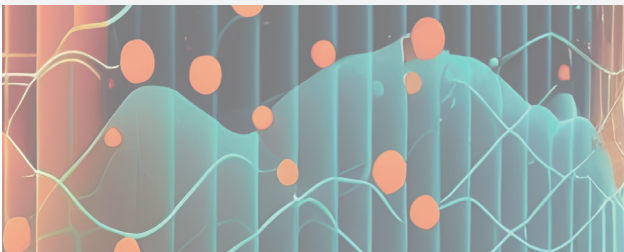
weren't generalized, large-scale AI projects like we're starting to see now. As we move into this new era, the risk of getting it wrong is amplified by orders of magnitude."

Hall Lefevre believes a holistic governance model is the key. "We think data and AI governance as a whole need to be synched consistently, because the effectiveness of AI is highly dependent upon data quality and how the data is governed," she says. "Educating and aligning the entire organization on the governance models is also indispensable. With AI use exploding across our enterprise, everyone must have a complete understanding of our governance policies."

Reid notes that governance needs to strike a balance between control and enablement. "We need to have the data set up so that when somebody has an idea on the way to work, they don't have to wait for a week before they can test it. So we try to enable agile experimentation while safeguarding security and privacy."

ADP: No stranger to the challenges of generative AI

When ADP started developing internal LLM capabilities in 2019, chief data officer Jack Berkowitz and his team quickly realized that they needed to re-evaluate their data architecture needs. The problem was multiple new data silos sprouting from different AI projects. Teams were building independent data pipelines and using APIs to connect with one another. The result, recalls Berkowitz, was a "confused mess where we were replicating data over and over. Managing data freshness, quality, and governance began to get out of hand."

ADP's solution, implemented during the succeeding two years, was to bring together the structured and semi-structured data the projects were using. "We thought that if we can bring those together, we'll be able to build repeatable processes that will help us to develop machine learning applications," says Berkowitz. "So that's what we did. We now have the bulk of our structured and semi-structured data in a data lakehouse. There are upwards of 10 petabytes of data in those workloads, refreshing within 15 seconds from our transactional systems in the U.S. and Canada, all adhering to the respective national data protection and privacy rules."

According to Berkowitz, among the most important benefits this architecture solution has conferred is reusability of the classifiers and techniques his team began developing a few years before. "Project teams no longer need to build their own," he says. But the biggest benefit has been reliability about data accuracy and integrity. "At the end of the day," says Berkowitz, "it's about the data quality of service to feed our machine learning models."

LLMs consume huge volumes of data in training and can expose sensitive data if the proper controls are not in place. There are also issues around demonstrating the provenance of the data LLMs use, according to Jack Berkowitz, chief data officer at payroll and HR systems provider ADP. "We've been working with LLMs since 2019, but the latest versions make ensuring data quality of service more complex," he says. And if it's not clear what data a model was trained on, he adds, it's difficult to ensure that advice being provided to clients is precise.

Yet another generative AI challenge is putting adequate security around completed models. Indeed, data security and privacy are the respondents' main concerns about the use of generative AI, together cited by 50%. "A trained model is incredibly valuable," says Roese. "How do we make sure that if there's a ransomware attack, no one can get at it? And when we put it into production, is it operating on a platform that can't be hacked?"

Zutshi notes, however, that AI's impact to security will be double-edged. "I also feel security is one of the areas that will get disrupted the most with AI in a positive way," he says, "because you have an enormous amount of data that can be used for detecting anomalous behavior and acting on it much more quickly. I believe you will see a lot more AI-based automation in the security posture of every company."

Build or buy?

Organizations must decide whether to build their own LLMs internally or to make use of proprietary models provided by tech vendors. Surveyed executives are mostly taking a hybrid approach: 58% say they are open to both building and buying. Thirteen percent say they are opting solely to build their own LLMs, based on open-source models, while 29% are exclusively using external models (see Figure 9).

"LLMs are definitely in our future. It's very important from a customer engagement perspective that we are able to leverage things that are going to become normal."

Yemi Oshinnaiye, Chief Information Officer, U.S. Transportation Security Administration

The survey sample's smallest organizations are twice as likely as the largest to employ vendor models as their sole approach to building generative AI. This suggests that a perception of high cost when building internally is keeping many organizations from taking this option. Cost concerns should diminish, however, as inexpensive open-source options emerge for training LLMs.
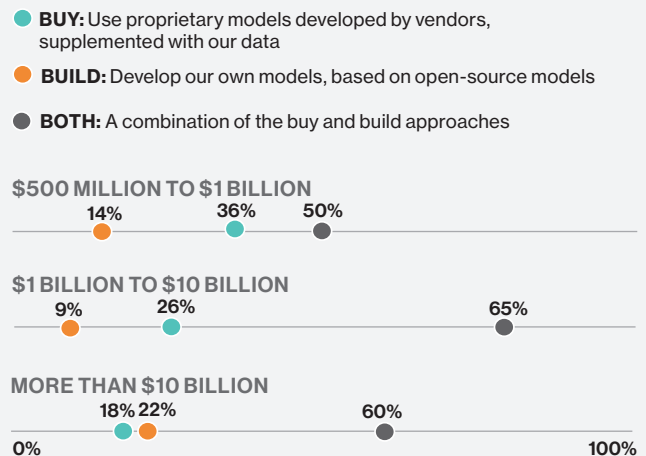
For many organizations, using vendor LLMs comes with risks, including IP ownership, security, privacy, and accuracy. A hybrid approach thus makes sense, enabling case-by-case decisions about which option to employ. "Publicly trained LLMs are perfectly good for use cases such as developer productivity," says Brahmadesam. "But when it comes to KYC [Know Your Customer regulations], we have to build models that are unique to our needs, as the documents the model will analyze are not publicly available. In KYC, accuracy is critical."

The hybrid approach also makes sense for Starbucks. Security is paramount in whichever option is chosen, says Hall Lefevre. "We want to make sure whatever data we use to populate and tune a model is very secure and in our own kind of cluster. If we do choose to look at open-source vendors, we want to make sure their products and tools can be run in a secure way in our cluster so that we can maintain IP."

In the biotech industry, building internally is often a must, says Reid. "Scientific discovery requires this because we need to represent biological concepts more compactly than they're represented in the tokenization scheme that's used for public models such as ChatGPT." He nevertheless wants the possibility of using a foundation model for less-demanding use cases. "When there's a foundation model that already does something more brilliantly than we ever could," he says, "why wouldn't we start with that?"

Figure 9: Approaches to developing generative AI
Which approach is your organization taking or likely to take when adopting generative AI/ LLMs?
(by company revenue)

● **BUY:** Use proprietary models developed by vendors, supplemented with our data

● **BUILD:** Develop our own models, based on open-source models

● **BOTH:** A combination of the buy and build approaches

$500 MILLION TO $1 BILLION
14%   36%   50%

$1 BILLION TO $10 BILLION
9%   26%   65%

MORE THAN $10 BILLION
18% 22%   60%
0%   100%

Source: MIT Technology Review Insights survey, 2023

"If we do choose to look at open-source vendors, we want to make sure their products and tools can be run in a secure way in our cluster so that we can maintain IP."

Deb Hall Lefevre, Executive Vice President and Chief Technology Officer, Starbucks
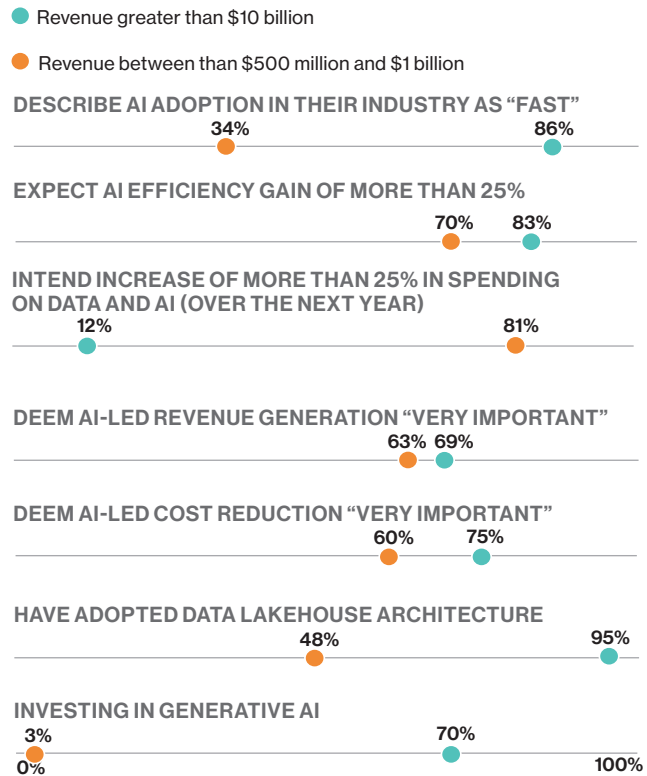
The AI scale advantages of big enterprises

Across the survey, stark differences are evident between the sample's largest and smallest enterprises in many aspects of their data and AI capabilities. Some differences may simply be attributable to the financial resources organizations have at their disposal to invest. For example, 70% of the largest firms have invested in generative AI compared with just 3% of the smallest ones. And while eight in 10 respondents from the largest organizations expect to boost their spending on data and AI during the next year by more than 25%, far fewer of the smallest ones (12%) intend to do so (see Figure 10).
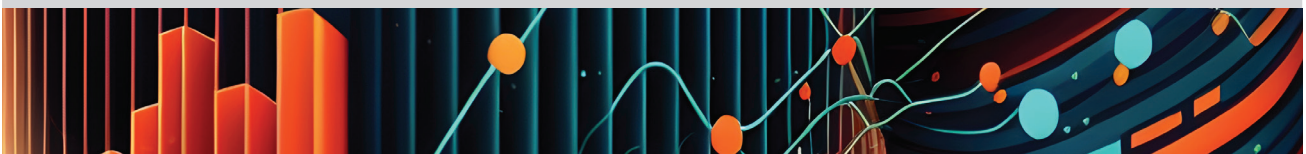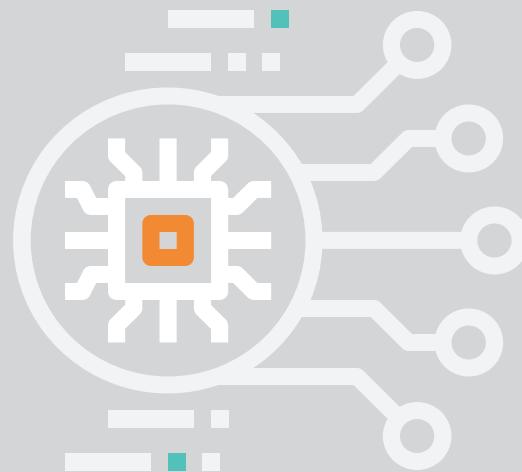
These disparities may also reflect a difference in outlook. Technology executives in the biggest organizations, for example, are clearly more optimistic than their peers in the smallest about the speed of AI adoption in their industries: 86% of the former describe the pace as "fast," compared with just 34% of the latter. This may help explain why fewer respondents from the smallest enterprises anticipate a large AI-led efficiency boost to their industry in the next two years. But lower expectations of industrywide adoption could also influence investment intentions, if executives do not see their competitors pulling ahead in building data and AI capabilities.

These differences suggest that smaller firms risk falling irretrievably behind their larger peers in the race to master AI. Limited resources may dictate caution on the part of some, but open-source options are available to access generative AI models and other cloud-based AI resources at relatively low cost. Smaller enterprises have as much to gain, in relative terms, from improvements in customer experience, operating efficiency and product innovation as their larger brethren. They cannot afford to let the growth opportunities that AI offers pass them by.

Figure 10: The AI gap: survey responses by the largest and smallest enterprises

● Revenue greater than $10 billion

● Revenue between than $500 million and $1 billion

DESCRIBE AI ADOPTION IN THEIR INDUSTRY AS "FAST"
34%   86%

EXPECT AI EFFICIENCY GAIN OF MORE THAN 25%
70%   83%

INTEND INCREASE OF MORE THAN 25% IN SPENDING ON DATA AND AI (OVER THE NEXT YEAR)
12%   81%

DEEM AI-LED REVENUE GENERATION "VERY IMPORTANT"
63% 69%

DEEM AI-LED COST REDUCTION "VERY IMPORTANT"
60%   75%

HAVE ADOPTED DATA LAKEHOUSE ARCHITECTURE
48%   95%

INVESTING IN GENERATIVE AI
3%   70%
0%   100%

Source: MIT Technology Review Insights survey, 2023

05

Rise of the citizen AI developer
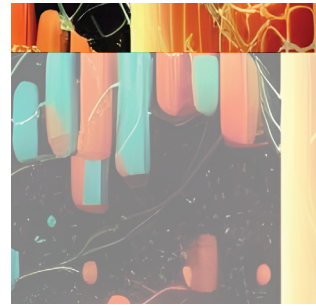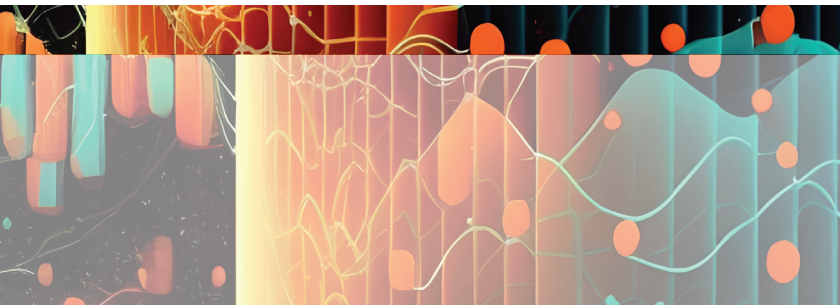
The term *democratization* entered the enterprise technology lexicon when cloud and software-as-a-service (SaaS) took root, enabling employees in any business unit easy access to work applications. It soon became clear that employees' easy access to data across business units is also critical. Since then, technology executives across industries have sought to make the democratization of data, as well as of insights derived from analytics, a reality.

To facilitate this, 64% of survey respondents want their data architecture to enable secure sharing of live data across platforms, considering it "very important" to achieving their technology goals. Another 41% deem it very important to have a managed central marketplace for data sets, ML models, and notebooks (see Figure 3). Advances in data-sharing software, such as APIs, and

"Generative AI is the first big step in AI democratization. I'm very optimistic about this within our organization."

Naveen Zutshi, Chief Information Officer, Databricks

in architecture, such as data lakehouses, have proven to be catalysts of data democratization. The high-level objective of Condé Nast's single data platform with its bronze, silver, and gold layers was, according to Bhakta, "to democratize data across our organization."

When ADP started bringing the different data repositories in its organization together on a unified platform, Berkowitz was told he would go against the grain of all their owners. "The exact opposite happened," says Berkowitz. "We now have 60 or 70 teams around the company on the platform, independently working, sharing data, and collaborating."

Democratizing AI

AI is a new frontier of democratization within organizations today. "Generative AI is the first big step in AI democratization," says Zutshi. "I'm very optimistic about this within our organization." Francis concurs. "Six months ago, I would have said democratization of AI is probably a decade away," he says. "But with generative AI, it's happening in front of us." His team at GM is exploring how to use generative AI as a "copilot" to help employees in other business units learn to develop code and build applications. "Obviously, someone senior will need to review the code," he says, "but having that capability will help us scale AI and enable wider innovation with its use."

The simple text and voice interfaces of generative AI models will eventually replace the specialist interfaces that software developers use today, says Reid. "A more casual interface will make it easier for everyone, even those with no coding skills, to query data," he says. And generative AI's value goes far beyond analytics, Reid says: "It provides an opportunity to help you remember something, or to validate something that you already know, in order to support decision-making. We're seeing that enable everyone across the board."

With AI democratization, of course, comes risks. Those may be familiar to legal, compliance, and security teams, but they loom especially large now with the increasing adoption of generative AI. "You can empower anyone at any level of the organization today to go build an AI application," says Francis. "But are we fully thinking through all the legal, privacy, and security and implications, or the commercial ones, such as what it means from a brand perspective?"

Executive decisions about safeguarding the organization – and consumers – from harm caused by AI will likely be influenced by lawmakers and regulators. Senior management would be wise to review their existing policies and ensure they're ready for the generative AI era before government steps in.

The copilot is on board

Using technology to empower people – to innovate and create value – is the main object of democratization. But people also augment technology, making technology and people decisions closely intertwined.

Technology executives surveyed worry that their organization are not sufficiently focusing on the human dimension of data and AI. When asked how their company's data strategy needs to improve, investment in talent and the workforce (cited by 39% of respondents) is the top response, ahead of such critical areas as governance and data processing speed. An even larger share (72%) say it will be "very important" to encourage innovation that will help attract and retain talent.

Figure 11: Data and AI platform challenges
What are the top pain points of your organization's current data and AI platforms?

Need to train or upskill workforce to use platforms	**40%**
Inadequate governance frameworks	**26%**
Siloed legacy systems	**25%**
Inadequate security frameworks	**25%**
Too many disparate systems	**18%**

Source: MIT Technology Review Insights survey, 2023

Asked what concerns them most about their current data and AI platforms, executives name people bottlenecks ahead of technology weaknesses (see Figure 11). Respondents cite the need to train and upskill employees (40%) as the foremost "pain point" they are experiencing with these platforms. Substantially less concerning are technology-specific issues such as inadequate governance (cited by 26%), silos (25%), and inadequate security frameworks (25%).

People challenges help make the generative AI copilot role all the more attractive to technology leaders. As organizations struggle to fill talent gaps in data science, Francis believes the copilot role could be a solution: "Wherever we can leverage technology to help people across the business to build applications will, I think, be very additive."

Reid firmly believes that everyone in the organization needs to learn how to use AI. "Something you learn as a scientist at a very young age is to constantly be aware of what's going on in the field. You need always to be curious. Everyone should constantly be challenging themselves to learn and train on new things, including AI. After all, this is how science has progressed."

"Six months ago, I would have said democratization of AI is probably a decade away, but with generative AI, it's happening in front of us."

Jon Francis, *Chief Data and Analytics Officer, General Motors*

Databricks: Operationalizing generative AI across the business

To put generative AI to use across the enterprise, CIOs need to find ways to build innovation into the business, developing strategic use cases and finding ways to operationalize LLMs that make these tools true copilots to their employees. Naveen Zutshi, CIO at Databricks, argues for cultivating as many use cases as feasible: "My approach to generative AI is you take a portfolio approach to ideas. Let's say you have a hundred ideas; maybe 10 or 20 of those will actually pan out. You want to have a fail-fast philosophy and you want to drive a lot of ideas within the company."

When a technology like generative AI has so uniquely captured the popular imagination, many employees are eager to investigate ways it can solve problems in their own work. "Previously you'd be pushing to the business, 'Hey, I have this idea, can we do this?'" Zutshi says. "But right now, the business is asking you, 'What are we doing about generative AI?' There's a pull from business, and you want to leverage that enthusiasm."

Databricks took advantage of this popular momentum by developing a generative AI hackathon that, rather than being limited to technical teams, invited in stakeholders from sales, HR, legal, and finance. In addition to enabling the team to crowdsource ideas directly relevant to real business problems, this wider participation meant that key nontechnical product elements like business value validation and risk management were built into the tools developed from the beginning. And more importantly, this created ready-made generative AI champions for change management in the business.

As a result of the first hackathon, the participants produced fifteen working generative AI–based software tools, for use cases ranging from meeting summarization, contract clause retrievals, and résumé and job parsing to documentation generation and troubleshooting. Three of these will be fully put into production across the business, and these hackathons are planned to continue every six months. Zutshi is optimistic about the power of these approaches to both help employees grow in their technical skills and to create a culture that enables scaling AI within the business – "to drive the sense of education across the enterprise."

06

Conclusion
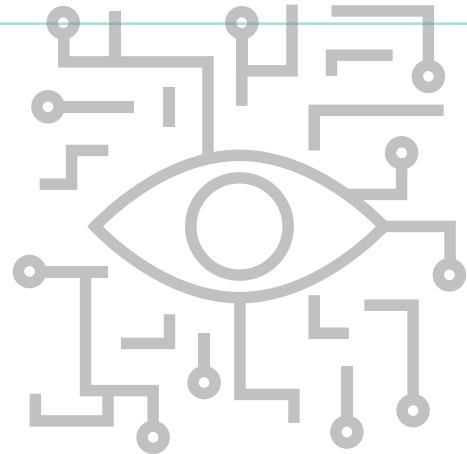
f it triggers a much wider diffusion of AI throughout economies, generative AI will be a technology inflection point. Experts predict that it will, unleashing a new wave of productivity and potentially adding trillions of dollars of new value across industries.[4] While realization of that impact is to come, generative AI is influencing organizations' data and technology infrastructure and the investment decisions technology executives are making to modernize.

The stakes involved in these decisions are rising due to the infrastructure demands these advances in AI capabilities pose. As Roese says, "the risk of getting it wrong is amplified by orders of magnitude." The upside of getting it right, by contrast, will be enormous even if only some of the predictions for generative AI's impact on value generation prove accurate.

Getting it right creates several imperatives for the data and other infrastructure that support organizations' AI capabilities. Three emerge in clear profile from this study's findings. Modern, future-ready data infrastructure should ensure the following:

Simplification amid complexity. Consolidating the multiple AI and ML systems that have sprung up since organizations began deploying these capabilities, along with the data platforms that support them, such as a lakehouse, will create greater congruence across models and the data sources that feed them. And simplification of user interfaces will promote AI democratization.

Empowerment with control. AI democratization, along with that of data — which organizations have been striving to embed for years — is needed to unleash the kinds of innovation that create new value. But all forms of democracy need guardrails lest chaos ensue, and this is no exception. The ease of using generative AI in particular bolsters the case for strong and unified governance.

Sharing with trust. CIOs emphasize the importance of teams being able to share live data securely across platforms. These capabilities are the bedrock of data ecosystems that tie together analytics and AI models within enterprises and that are coming to extend more widely, across an organization's suppliers and partners. In a subsequent report that assesses the implications of this research for different industries, we will also explore the emergence of such industry data ecosystems.

The upside of getting it right will be enormous even if only some of the predictions for generative AI's impact on value generation prove accurate.

About MIT Technology Review Insights

MIT Technology Review Insights is the custom publishing division of *MIT Technology Review,* the world's longest-running technology magazine, backed by the world's foremost technology institution—producing live events and research on the leading technology and business challenges of the day. Insights conducts qualitative and quantitative research and analysis in the U.S. and abroad and publishes a wide variety of content, including articles, reports, infographics, videos, and podcasts. And through its growing MIT Technology Review Global Insights Panel, Insights has unparalleled access to senior-level executives, innovators, and thought leaders worldwide for surveys and in-depth interviews.

About Databricks

Databricks is the Data and AI company. More than 10,000 organizations worldwide — including Comcast, Condé Nast, and over 50% of the Fortune 500 — rely on the Databricks Lakehouse Platform to unify their data, analytics and AI. Databricks is headquartered in San Francisco, with offices around the globe. Founded by the original creators of Apache Spark™, Delta Lake and MLflow, Databricks is on a mission to help data teams solve the world's toughest problems. To learn more, follow Databricks on **X** (formerly known as Twitter), **LinkedIn**, and **Facebook**.

databricks

Endnotes

1. "Gartner Forecasts Worldwide IT Spending to Grow 4.3% in 2023," Gartner, July 19, 2023, https://www.gartner.com/en/newsroom/press-releases/2023-07-19-gartner-forecasts-worldwide-it-spending-to-grow-4-percent-in-2023.
2. "The great acceleration: CIO perspectives on generative AI," MIT Technology Review Insights, July 2023, https://www.databricks.com/resources/ebook/mit-cio-generative-ai-report.
3. "The great acceleration: CIO perspectives on generative AI," MIT Technology Review Insights, July 2023, https://www.databricks.com/resources/ebook/mit-cio-generative-ai-report.
4. "The economic potential of generative AI," McKinsey & Company, June 14, 2023, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier.

Illustrations

All illustrations assembled from Adobe Stock and The Noun Project icons.

MIT Technology Review Insights

www.technologyreview.com

insights@technologyreview.com