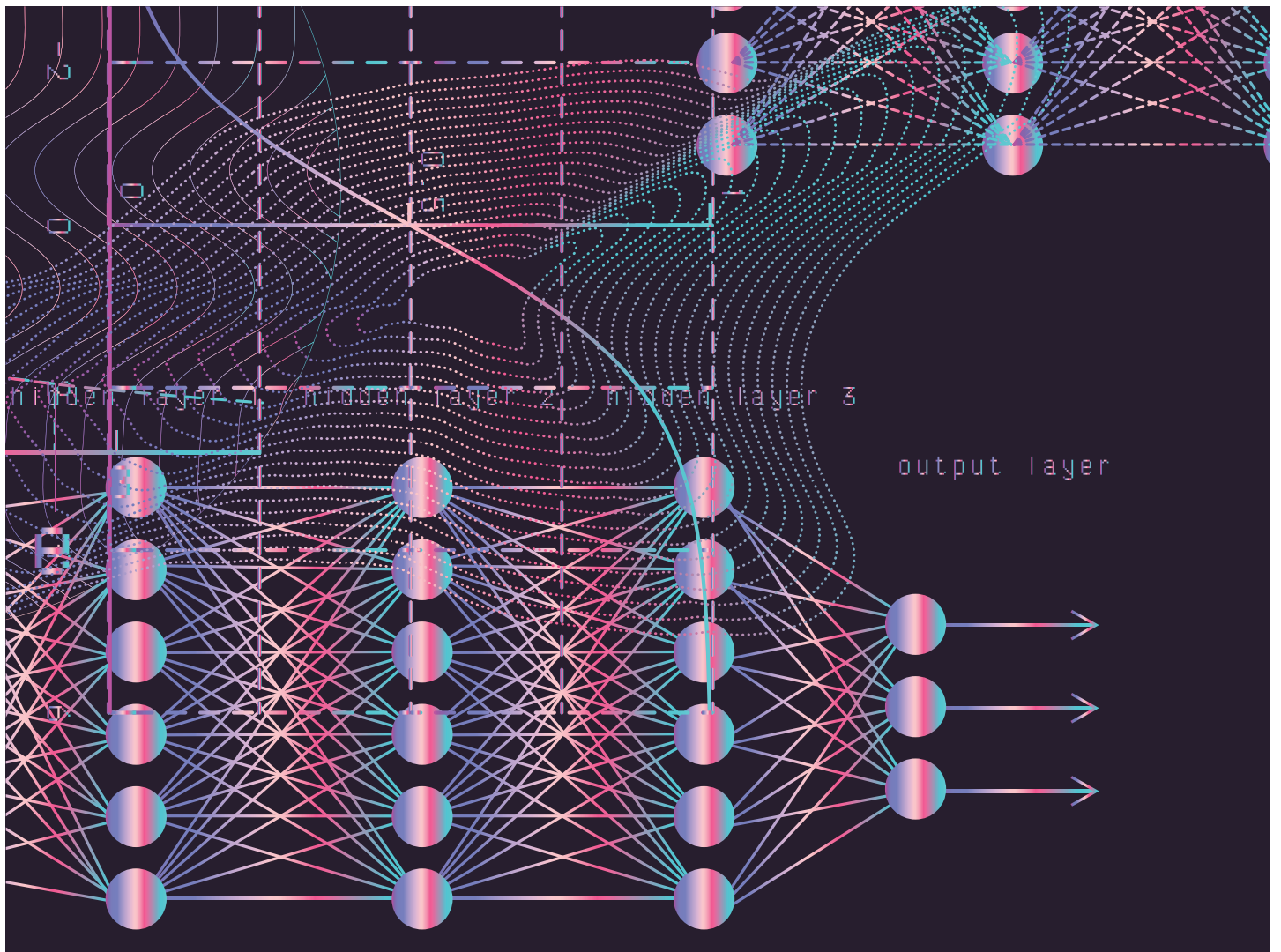As generative AI adoption reshapes the competitive landscape, businesses without a strong data strategy will find their ambitions limited.

Data strategies for AI leaders

Key takeaways

1 Executives' top ambition for generative AI adoption is driving increased efficiency or productivity (72%), far exceeding their interest in increasing revenue (30%) or reducing costs (24%).

2 Strong data capabilities will be essential underpinnings to these aspirations, but only 22% of businesses consider their data foundations "very ready" to support generative AI applications today.

3 The rise of AI exacerbates longstanding challenges in data management – data governance, security, and privacy (cited by 59%), data quality and timeliness (53%), and data integration (48%) – and may supply the urgency needed to finally address them.

O rganizations are starting the heavy lifting to get real business value from generative AI. As Arnab Chakraborty, chief responsible AI officer at Accenture, puts it, "2023 was the year when clients were amazed with generative AI and the possibilities. In 2024, we are starting to see scaled implementations of responsible generative AI programs."

Some generative AI efforts remain modest. As Neil Ward-Dutton, vice president for automation, analytics, and AI at IDC Europe, describes it, this is "a classic kind of automation: making teams or individuals more productive, getting rid of drudgery, and allowing people to deliver better results more quickly."

Most companies, though, have much greater ambitions for generative AI: they are looking to reshape how they operate and what they sell.

Great expectations for generative AI

The expectation that generative AI could fundamentally upend business models and product offerings is driven by the technology's power to unlock vast amounts of data that were previously inaccessible. "Eighty to 90% of the world's data is unstructured," says Baris Gultekin, head of AI at AI data cloud company Snowflake. "But what's exciting is that AI is opening the door for organizations to gain insights from this data that they simply couldn't before."

> "AI is opening the door for organizations to gain insights from unstructured data that they simply couldn't before."

Baris Gultekin, Head of AI, Snowflake

In a poll conducted by MIT Technology Review Insights, global executives were asked about the value they hoped to derive from generative AI. Many say they are prioritizing the technology's ability to increase efficiency and productivity (72%), increase market competitiveness (55%), and drive better products and services (47%). Few see the technology primarily as a driver of increased revenue (30%) or reduced costs (24%), which is suggestive of executives' loftier ambitions.

Respondents' top ambitions for generative AI seem to work hand in hand. More than half of companies say new routes toward market competitiveness are one of their top three goals, and the two likely paths they might take to achieve this are increased efficiency and better products or services.

Methodology

MIT Technology Review Insights conducted a poll on businesses' data foundations for generative AI and their ambitions and challenges when deploying and scaling generative AI applications in May 2024. The 275+ executive respondents represent a broad range of industries and work at organizations across the globe.

"There is no value without good business data."

Neil Ward-Dutton, Vice President for Automation, Analytics, and AI, IDC Europe

For companies rolling out generative AI, these are not necessarily distinct choices. Chakraborty sees a "thin line between efficiency and innovation" in current activity. "We are starting to notice companies applying generative AI agents for employees, and the use case is internal," he says, but the time saved on mundane tasks allows personnel to focus on customer service or more creative activities. Gultekin agrees. "We're seeing innovation with customers building internal generative AI products that unlock a lot of value," he says. "They're being built for productivity gains and efficiencies."

Chakraborty cites marketing campaigns as an example: "The whole supply chain of creative input is getting re-imagined using the power of generative AI. That is obviously going to create new levels of efficiency, but at the same time probably create innovation in the way you bring new product ideas into the market." Similarly, Gultekin reports that a global technology conglomerate and Snowflake customer has used AI to make "700,000 pages of research available to their team so that they can ask questions and then increase the pace of their own innovation."

The impact of generative AI on chatbots – in Gultekin's words, "the bread and butter of the recent AI cycle" – may be the best example. The rapid expansion in chatbot capabilities using AI borders between the improvement of an existing tool and creation of a new one. It is unsurprising, then, that 44% of respondents see improved customer satisfaction as a way that generative AI will bring value.
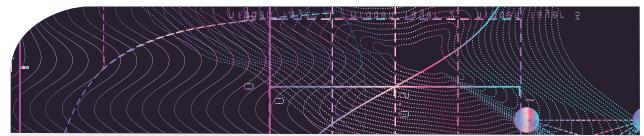
A closer look at our survey results reflects this overlap between productivity enhancement and product or service innovation. Nearly one-third of respondents (30%) included both increased productivity and innovation in the top three types of value they hope to achieve with generative AI. The first, in many cases, will serve as the main route to the other.

But efficiency gains are not the only path to product or service innovation. Some companies, Chakraborty says, are "making big bets" on wholesale innovation with generative AI. He cites pharmaceutical companies as an example. They, he says, are asking fundamental questions about the technology's power: "How can I use generative AI to create new treatment pathways or to reimagine my clinical trials process? Can I accelerate the drug discovery time frame from 10 years to five years to one?"

Data strategy underlies AI innovation

Behind the diversity of ways in which respondents hope to secure value from generative AI looms one commonality: the need for enormous quantities of the business's own data, accessibly stored and ready to use. Off-the-shelf AI tools will not differentiate businesses when their adoption will soon be universal. For any enterprise AI use case, says Ward-Dutton, "there is no value without good business data" of the company's own.

Executive aspirations for generative AI adoption

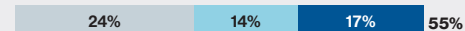What are the primary types of value your organization hope to achieve from its generative AI efforts?

☐ First choice ☐ Second choice ■ Third choice

Increased efficiency or productivity
| 36% | 23% | 13% | **72%** |

Increased market competitiveness
| 24% | 14% | 17% | **55%** |

Innovation in products or services
| 12% | 13% | 22% | **47%** |

Improved customer satisfaction
| 7% | 18% | 19% | **44%** |

Increased revenue
| 10% | 11% | 9% | **30%** |

Faster decision-making
| 6% | 10% | 13% | **29%** |

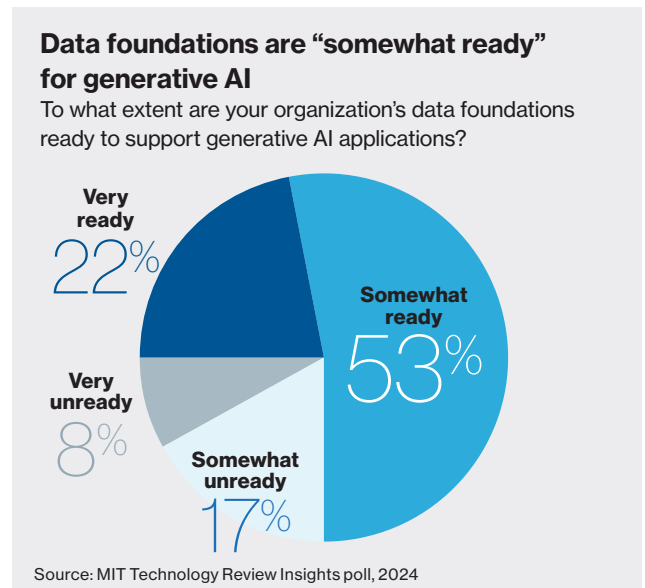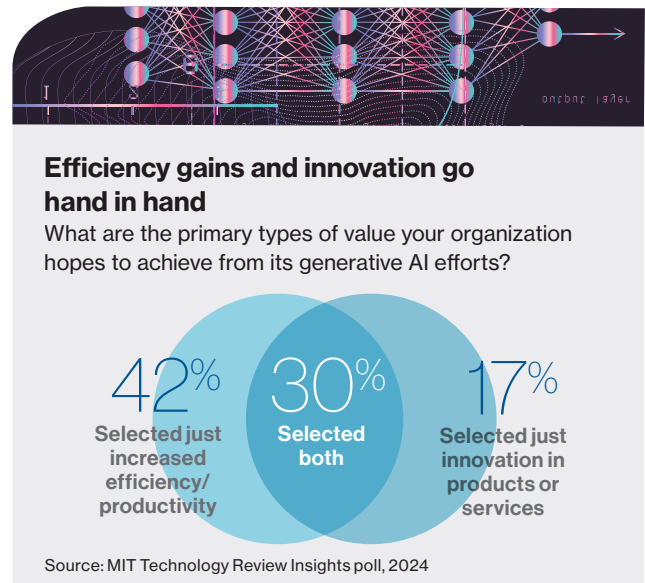**Reduced costs**
| 6% | 10% | 8% | **24%** |

Source: MIT Technology Review Insights poll, 2024

One of the great benefits of generative AI is that it can make use of unstructured data, which many companies have in great quantities, such as wikis, documents, and webpages. As Gultekin explains, "AI makes all of this data much more useful because now you can unlock a massive amount of it that wasn't easily accessible before." Moreover, humans can ask questions of their data using natural language. This has a significant impact on enterprises, helping them deliver insights from previously untapped sources of information.

This flood of newly accessible data, though, represents both an opportunity and a competitive requirement. Chakraborty describes the challenge: "To create significant differentiation that becomes a competitive asset, you need to start almost building your own large language models, which means you need to train on massive amounts of data harnessed from within your enterprise – from finance to HR to supply chain to marketing – and marry that with external data."

How ready, though, are most companies' data estates to support them in the race to generative AI value? Gultekin puts it plainly: "The data foundation is at the core of generative AI capabilities." Data foundations cover a broad collection of processes and assets involved in the gathering, aggregation, storage, and accessibility of organizational data.

Chakraborty identifies three specific data capabilities necessary to support the effective deployment of generative AI: the quality of the data; the ability to integrate multiple sources of data; and the timely democratization of data to relevant business users. These are persistent difficulties, he notes, saying "all have been issues that we have heard about for 10 years." These challenges will not easily be addressed this time, either, he warns: "We still will be 10 years from now, as

Efficiency gains and innovation go hand in hand
What are the primary types of value your organization hopes to achieve from its generative AI efforts?

42% Selected just increased efficiency/ productivity

30% Selected both

17% Selected just innovation in products or services

Source: MIT Technology Review Insights poll, 2024

Data foundations are "somewhat ready" for generative AI
To what extent are your organization's data foundations ready to support generative AI applications?

Very ready 22%

Very unready 8%

Somewhat unready 17%

Somewhat ready 53%

Source: MIT Technology Review Insights poll, 2024

well, because data is very dynamic and the speed at which it is getting created is magnifying every month."

At first glance, most poll respondents seem positive about the state of their company data foundations. The majority rate their business "very ready" (22%) or "somewhat

"To create significant differentiation that becomes a competitive asset, you need to start almost building your own large language models, which means you need to train on massive amounts of data harnessed from within your enterprise and marry that with external data."

Arnab Chakraborty, *Chief Responsible AI Officer, Accenture*

ready" (53%) to support generative AI applications, while about one-quarter acknowledge that their organization is either "somewhat unready" (17%) or "very unready" (8%).

This distribution is consistent with the experience of experts. Ward-Dutton, for example, says that, in his interactions with companies, "We consistently see a bell curve. The people at the front end have already made significant investments in things like data governance and data quality. There are also a bunch of people still using, basically, Excel for everything."

Chakraborty warns that being "somewhat ready" could conceal dangerous weaknesses. It might describe, for example, an organization that can integrate data sets from across the business, but that also knows that the underlying data is not fully clean. If so, he adds, "Somewhat ready is not good enough. On accuracy and the quality of the data, 50% ready is not good enough. It needs to be higher to create trust." He adds, "On the ability to democratize that data to business users" – to get users the data they need when they need it – "somewhat ready is also not good enough."

IDC's research, Ward-Dutton says, shows that "somewhat ready" is rarely "almost ready" in practice. The company's surveys have found that only 30 to 40% of businesses are confident in their ability to perform each of the following data controls in their AI work: strictly control sensitive data with certainty that none will be leaked during model training or use; manage use of third-party IP included in the models they are using and ensure that their own IP does not leak; and track and control how generative AI is interacting with their own internal data. His conclusion from such data is that "overall, readiness is pretty equivocal."

Nor is it all clear sailing for companies self-ranked "very ready" by respondents. Instead, investment of resources and time in their data foundations may just have made clear their next set of challenges.

The main challenges of deploying generative AI at scale are different for companies who say their data foundations are "very ready." Predictably, higher readiness correlates with fewer challenges related to access to scalable computing power, data silos and integration issues, and data governance. In contrast, "very ready" companies face more data quality issues, presumably because better access to the information has revealed
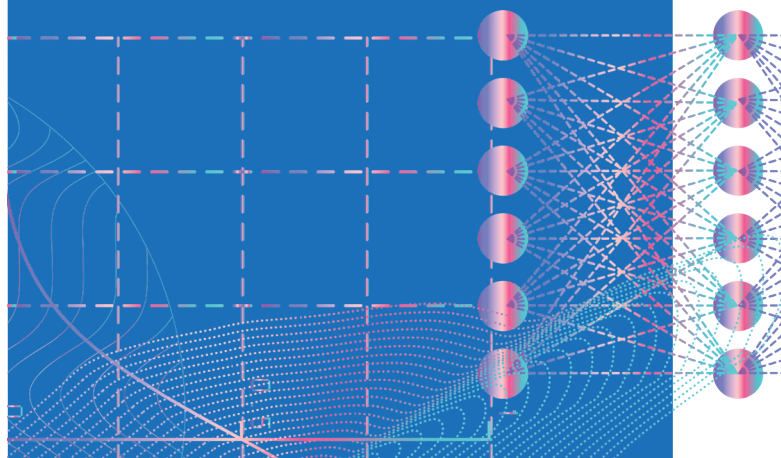
The hallucination challenge

Hallucinations—or the tendency of generative AI models to produce confident, but incorrect outputs—are a major challenge organizations face in deploying AI applications. After all, no company wants to deploy a chatbot that makes false promises about its return policies to its customers or an internal data tool that delivers unfounded insights to its analysts.

Baris Gultekin of Snowflake sees hallucinations as a major concern, and he has advocated for an insistence on higher accuracy rates in the next generation of AI-based tools. "For large enterprises, AI accuracy is imperative, and providing users with trustworthy AI will be the leading factor for businesses' AI success at scale," Gultekin notes. Ensuring the quality and cleanliness of the data used to train AI models is of course an essential first step to improving their outputs, he continues—"garbage in, garbage out" is a common refrain.

Companies are also addressing hallucinations through improved techniques for ensuring accurate AI responses. Retrieval augmented generation (RAG), for example, is a technique used to ensure that AI outputs are grounded in verifiable outside data. Guardrails can impose firm boundaries around the language or topics generative AI outputs can contain, and human and computer review of outputs can iteratively improve an AI application's results.

And sometimes it's best for AI to simply acknowledge that it doesn't know. Gultekin says, "We've been focused on building products that know when to abstain from answering a question. LLMs are blissfully ignorant about when they should not be answering, which is a dangerous slope for businesses' critical decision-making. We've been investing in making systems that know when not to answer."

its shortcomings. As Chakraborty says, "data-driven transformation is not something where you can say, 'I've done it for the last two to three years, and now I'm done.' It's a journey that will never be done."

That said, generative AI's benefits are becoming visible to those companies farther along the road. Gultekin reports that companies that have invested heavily in data foundations are "all now reaping the benefits, because they are able to bring AI onto that data." Ward-Dutton adds that this deployment of generative AI represents a business leap as well as a technological one. Companies that have invested in data governance and quality, he says, "typically have also got to the point where they're really engaging business people in how to manage data." Now, the business people see, "'Wow, this data actually has value for me.'"
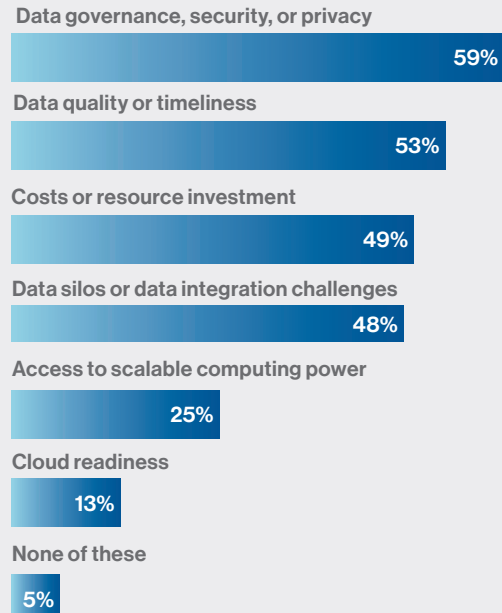
Roadblocks in the search for AI value

Nothing good comes easy. Nearly all poll respondents acknowledge challenges in deploying generative AI. Only 5% said that they did not face any of the challenges listed in the question.

A majority of respondents (59%) cite data governance, security, or privacy as a challenge. Nearly as many (53%) say the same of data quality or timeliness. Meanwhile, nearly half face difficulties with cost (49%) and with overcoming data integration challenges (48%). These are longstanding issues in data management, but the advent of AI complicates them further. It also, as Ward-Dutton explains, increases the urgency of finally grappling with them. "For a long time, many organizations have been able to limp along, or even do relatively well, without being very good at data-related technical, cultural, and organizational capabilities," he says. "We're seeing a lot having to catch up fast."

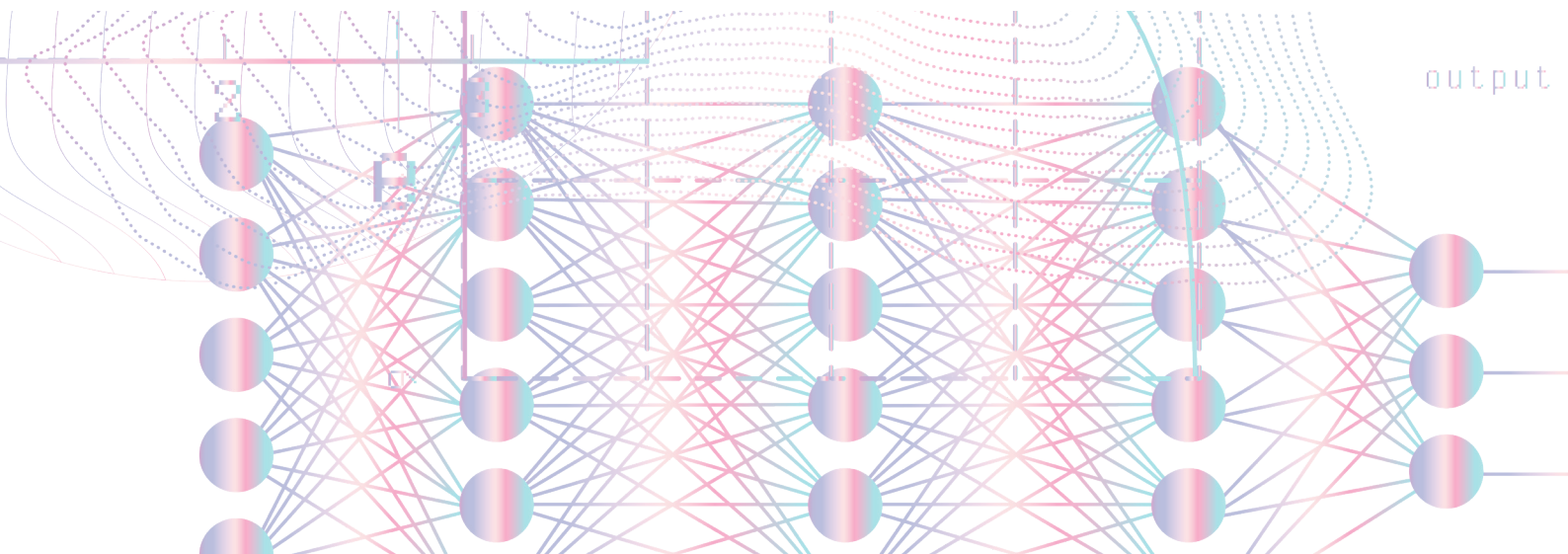### A wide range of AI deployment and scaling challenges

Which of the following challenges does your organization face when it comes to deploying AI at scale?

Data governance, security, or privacy
59%

Data quality or timeliness
53%

Costs or resource investment
49%

Data silos or data integration challenges
48%

Access to scalable computing power
25%

Cloud readiness
13%

None of these
5%

Source: MIT Technology Review Insights poll, 2024

It is unsurprising that data governance, security, or privacy is the most widespread challenge. Corporate data represent a company's "crown jewels," says Gultekin. Deep concern about it is inevitable. Ward-Dutton adds that, despite "quite good awareness at a high level" of the risks involved, "people are still flailing around a bit to find the right mix of responses."

Gultekin summarizes the many governance concerns surrounding generative AI: "At the foundational level, AI starts with securing and governing data. Companies want

to bring AI closer to their data, ensuring that data security and privacy are upheld. On top of that, customers want assurances that they are not liable for the data that the large language model was trained with, and they want to ensure that their data does not get used to improve the model for others without their permission. These are all table stakes and should be the industry standard."

Safe use of generative AI requires careful governance of its data sources. Chakraborty explains that the vast amount of public and company data used to train a large language model brings with it inherent risks. "You need a very surgical view in data governance," he says. From sourcing data to creating outputs, companies require very strong data governance, including clarity on who owns the data and who stewards it. Adding to the challenge, Gultekin explains, "is that the new data generative AI makes accessible may not itself be fully governed."

Data quality and timeliness is another predictable concern. As Chakraborty puts it, "if your data quality is not clean, you are going to get all kinds of garbage." Yet because of the opaque way in which the technology works, this may not be obvious. This is where data lineage becomes critical, ensuring transparency and traceability of data as it moves through systems. For those scaling up generative AI pilots, thereby giving the technology a greater role, such quality issues magnify, adds Gultekin.

While IT departments are typically told to fix low-quality data, the problem is caused where the information began "in the first place: that typically is the business," says Ward-Dutton. Employees who understand the value of their data "are much more likely to care about the quality of what they're providing," he says. "The technical or the capability piece and the organizational and cultural one are two sides of the same coin."

Data silos and data integration remain a challenge for nearly half of organizations. Gultekin notes that generative AI has heightened awareness of this problem as well. By making large amounts of data all "much more useful and much more accessible," he says, "it is shedding a lot of light on the data infrastructure that is needed." He argues that investment in a single, standardized data foundation across the organization will enable much more powerful generative AI uses. It will also reduce governance and security concerns: "When you keep your data in one place for one thing, another place for another thing, governing and securing that data becomes really difficult," he says.

Spending and resource decisions, including those needed to shore up data foundations, are of course a challenge when it comes to any technology investment. On the bright side, the cost of generative AI itself is decreasing substantially. In particular, Gultekin reports, in recent months, enterprises have begun creating smaller LLMs that remain extremely capable while being less expensive. "All these distilled models are still as good," he says. "They are also very cost-effective and very good from a latency perspective."

As organizations feel increasing urgency to deploy AI applications, they are realizing that their data is the key to how quickly and effectively they can unlock new value. Many will find that their generative AI ambitions are just castles in the air if they don't have the right data foundations to support them. A strong data strategy, however, can guide them to surmount their data challenges on the way to AI success.

"For a long time, many organizations have been able to limp along, or even do relatively well, without being very good at data-related technical, cultural, and organizational capabilities. We're seeing a lot having to catch up fast."

Neil Ward-Dutton, Vice President for Automation, Analytics, and AI, IDC Europe

"Data strategies for AI leaders" is an executive briefing paper by MIT Technology Review Insights. We would like to thank all participants as well as the sponsor, Snowflake. MIT Technology Review Insights has collected and reported on all findings contained in this paper independently, regardless of participation or sponsorship. Teresa Elsey was the editor of this report, and Nicola Crepaldi was the publisher.

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 US 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 entrepreneurs worldwide for surveys and in-depth interviews.

From the sponsor

Snowflake makes enterprise AI easy, efficient, and trusted. Thousands of companies around the globe, including hundreds of the world's largest, use Snowflake's AI Data Cloud to share data, build applications, and power their business with AI. The era of enterprise AI is here. Learn more at **snowflake.com** (NYSE: SNOW).

What is the AI Data Cloud?
The AI Data Cloud is a unified service used by over 10,000 companies to power their businesses with data, AI, and applications. Snowflake's AI Data Cloud consists of platform capabilities that support diverse data, AI, and applications workloads, as well as content — the datasets, models, and apps themselves — that are available to share and consume natively in the AI Data Cloud. The AI Data Cloud platform includes:
- Interoperable Storage: Snowflake-hosted and optimized storage (or extends to cloud data lakes)
- Elastic Compute: a unified engine supporting many languages and runtime needs
- Cortex AI: a managed AI layer including foundational LLMs, Chat API, and UI studio
- Cloud Services: infrastructure delivered as a service, including high availability and maintenance
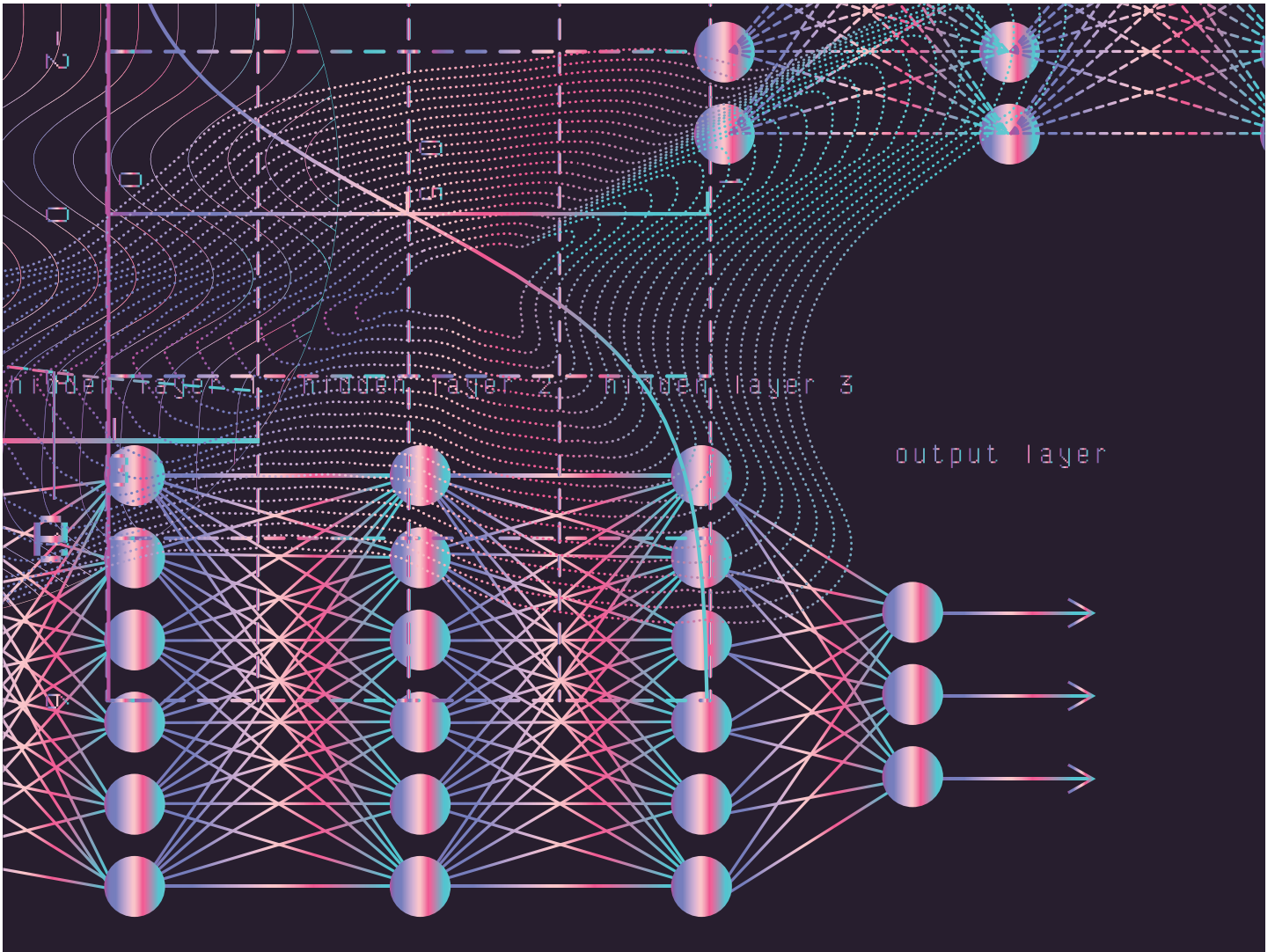- Snowgrid: a cross-region, cross-cloud network supporting data sharing and business continuity

❄️ snowflake®

MIT Technology Review Insights