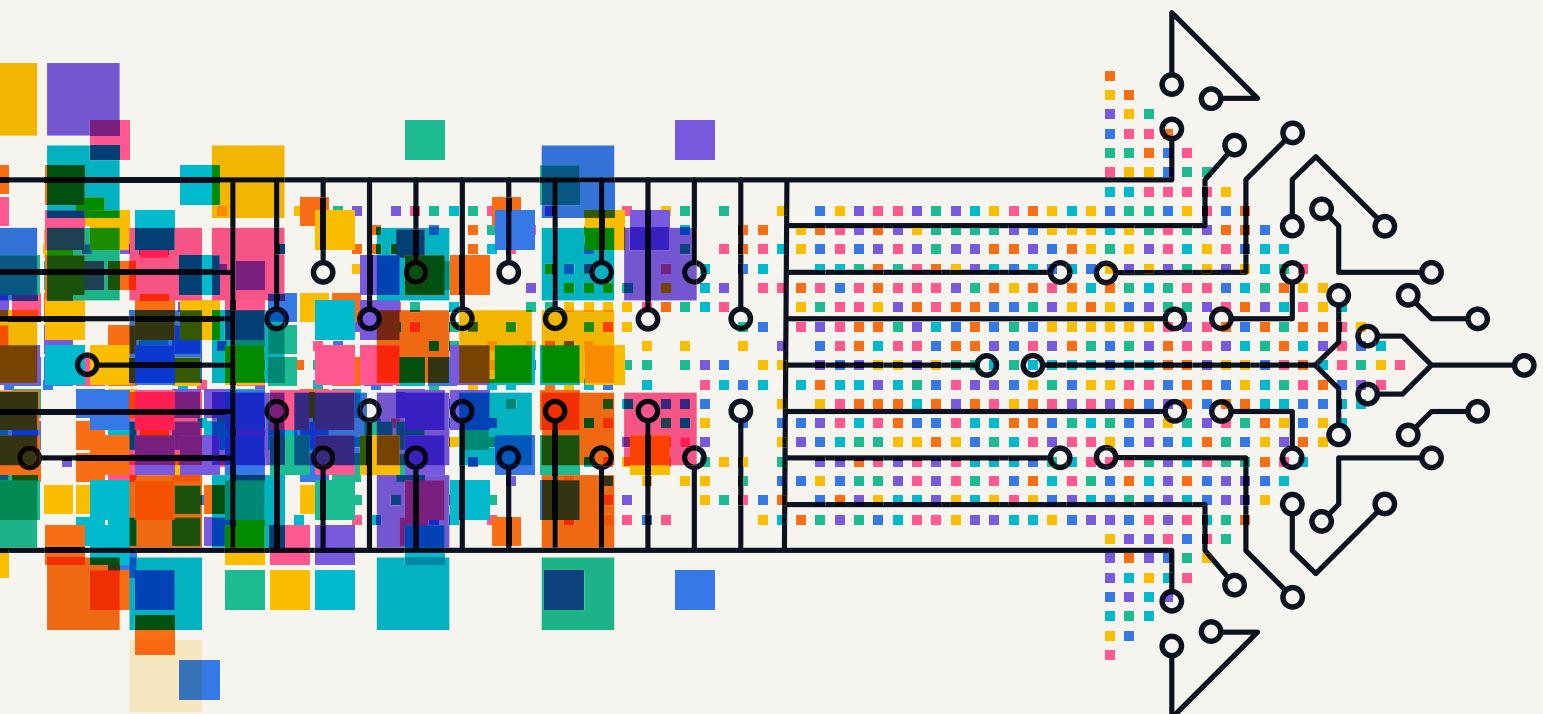


Redefining data engineering in the age of AI



Preface

“Redefining data engineering in the age of AI” is an MIT Technology Review Insights report sponsored by Snowflake. This report, based on survey research and executive interviews, seeks to understand how the role of data engineering teams is evolving as AI becomes central to enterprises and their success. Denis McCauley was the author of the report, Virginia Wilson 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 time and insights:

Chris Child, vice president of product, data engineering, Snowflake

Ritu Jyoti, chief executive officer, stealth AI startup (formerly general manager and group vice president of AI, automation, and data and analytics, IDC)

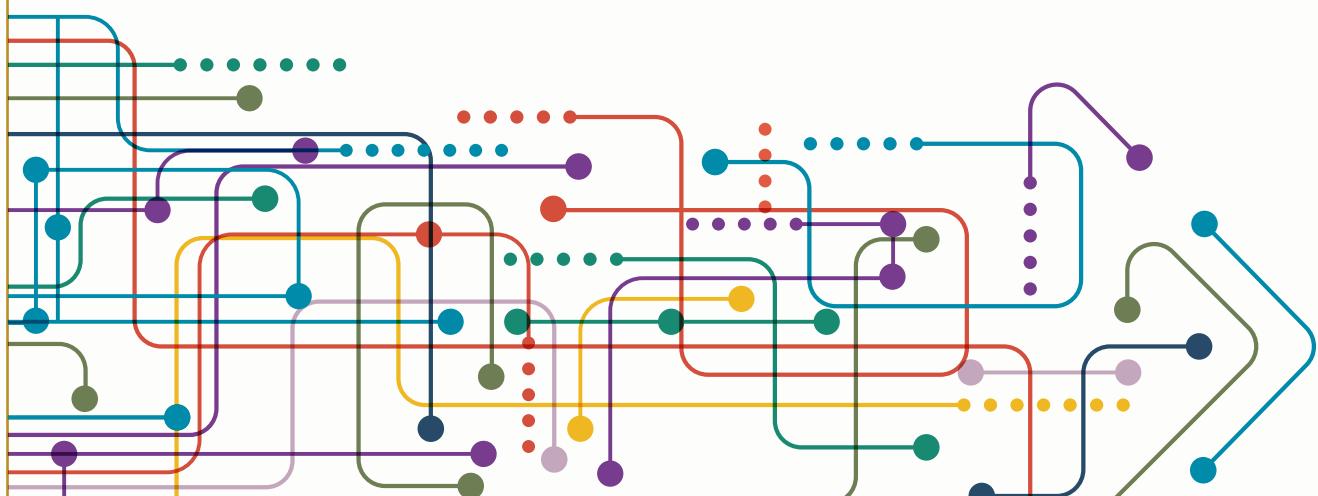
Dave Masino, senior director, data and intelligence, Slalom

George Westerman, senior lecturer and principal research scientist, MIT Sloan School of Management

Methodology

In June 2025, MIT Technology Review Insights, in collaboration with Snowflake, conducted a survey of 400 chief information officers, chief technology officers, chief data and analytics officers, and other senior data and technology executives. The survey respondents work in organizations that span seven industries, each headquartered in one of 10 countries. All organizations earn \$500 million or more in annual revenue.

In addition to the quantitative research from the survey, a series of in-depth interviews with senior technology executives and other experts offer firsthand insights into the evolution of the data engineering team’s role in the AI era.

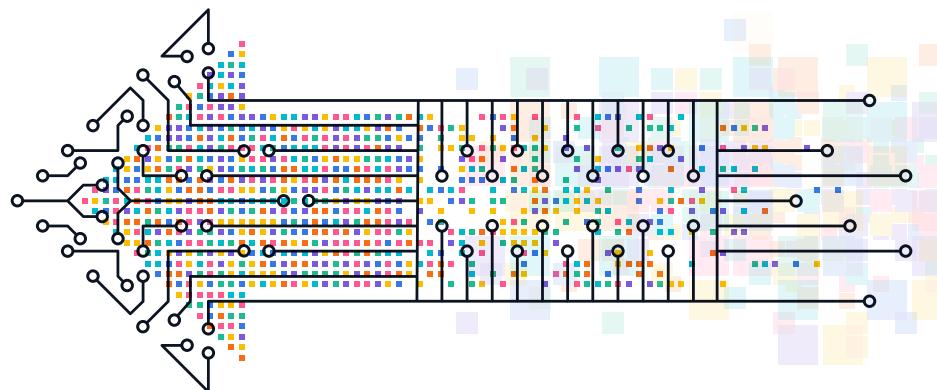


CONTENTS

01	Executive summary	4
02	Role change: The AI impact	5
	Growing workloads and a shifting focus	7
	Learning from software engineers	9
03	The agentic challenge	11
	Agents of change	12
	Risk equation	13
04	AI-powered data engineering	14
	More tools, greater complexity	15
	Focus on finance and manufacturing	16
05	Building influence	17
	Engineer and architect	18
06	Conclusion: Embracing the change	19



01 Executive summary



As organizations weave AI into more of their operations, senior executives are realizing data engineers hold a central role in bringing these initiatives to life. After all, AI only delivers when you have large amounts of reliable and well-managed, high-quality data. Indeed, this report finds that data engineers play a pivotal role in their organizations as enablers of AI. And in so doing, they are integral to the overall success of the business.

According to the results of a survey of 400 senior data and technology executives, conducted by MIT Technology Review Insights, data engineers have

become influential in areas that extend well beyond their traditional remit as pipeline managers. The technology is also changing how data engineers work, with the balance of their time shifting from core data management tasks toward AI-specific activities.

As their influence grows, so do the challenges data engineers face. A major one is dealing with greater complexity, as more advanced AI models elevate the importance of managing unstructured data and real-time pipelines. Another challenge is managing expanding workloads; data engineers are being asked to do more today than ever before, and that's not likely to change.

Key findings from the report include the following:

- **Data engineers are integral to the business.**

This is the view of 72% of the surveyed technology leaders – and 86% of those in the survey's biggest organizations, where AI maturity is greatest. It is a view held especially strongly among executives in financial services and manufacturing companies.

- **AI is changing everything data engineers do.**

The share of time data engineers spend each day on AI projects has nearly doubled in the past two years, from an average of 19% in 2023 to 37% in 2025, according to our survey. Respondents expect this figure to continue rising to an average of 61% in two years' time. This is also contributing to bigger data engineer workloads; most respondents (77%) see these growing increasingly heavy.

- **AI boosts data engineers' productivity.**

Nearly three quarters of the business leaders surveyed (74%) say AI has led to an increase in the output of data engineering teams in the past two

years. And for 77%, it has brought an improvement in the quality of their teams' work.

- **AI agents and tools increase efficiency, but also complexity.**

Over half of the surveyed organizations (54%) expect to begin deploying agentic AI within the next year (20% have already done so). For data engineering teams, agentic AI and other AI-powered tools should further enhance efficiency by automating tasks in pipeline optimization, data integration, orchestration, and other areas. But respondents also worry that new AI tools will increase integration complexity and add to tech stack fragmentation.

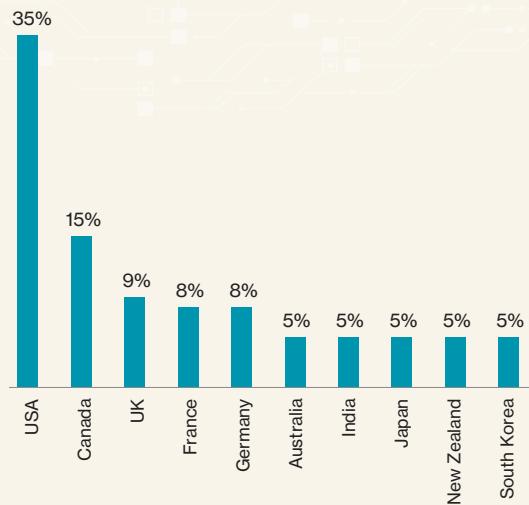
- **Data engineers' influence extends beyond pipeline management.**

Two-thirds of respondents (66%) say their data engineers are influential in decisions about investment in data tools and on vendor selection. Over half say the same about overall data strategy (51%), AI use-case feasibility (53%), and business units' use of AI models (56%).

Role change: The AI impact

Data engineers have worked behind the scenes for years but are now capturing the attention of their organizations' senior leaders. This is largely thanks to AI. Of the senior technology leaders in our survey, 72% deem data engineers to be integral to the overall success of their business (see Figure 2a). This figure rises to 86% of respondents in the survey's largest organizations – those with annual revenues of more than \$10 billion (see Figure 2b). Sector-wise, executives in financial services and manufacturing are particularly convinced of the importance of data engineers.

Figure 1a: Survey respondents by country (where their organization is headquartered)
(% of respondents)



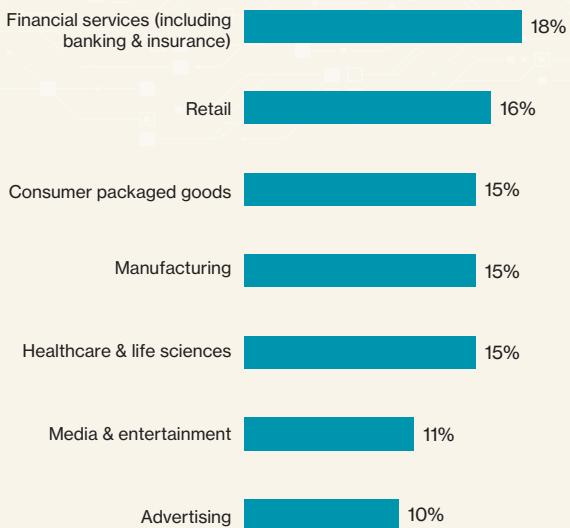
Source: MIT Technology Review Insights survey, 2025

According to Ritu Jyoti, formerly a senior executive at International Data Corporation (IDC) and now chief executive officer of a stealth AI startup, the explanation for data engineers' elevated profile is straightforward: "AI is increasingly a foundation of business success, and there is no AI without data."

Chris Child, vice president of product, data engineering at Snowflake, makes a similar point: "To realize AI's full potential, a strong data foundation isn't optional – it's mission critical. And we're seeing the workforce evolve to follow suit, with data engineers being viewed less

Figure 1b: Survey respondents by primary industry

(% of respondents)



Source: MIT Technology Review Insights survey, 2025

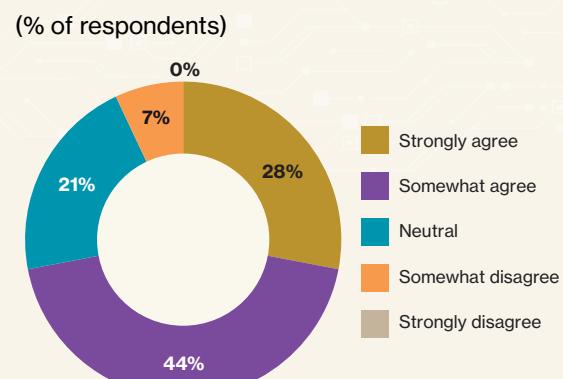
“To realize AI’s full potential, a strong data foundation isn’t optional – it’s mission critical. If your C-suite still considers data engineering as a support role, you’re already five years behind – and probably training your future competitors.”

Chris Child, Vice President of Product, Data Engineering, Snowflake

as a backend plumbing function, but instead as an integral, strategic business partner responsible for enterprises’ most valuable asset: their data. If your C-suite still considers data engineering as a support role, you’re already five years behind – and probably training your future competitors.”

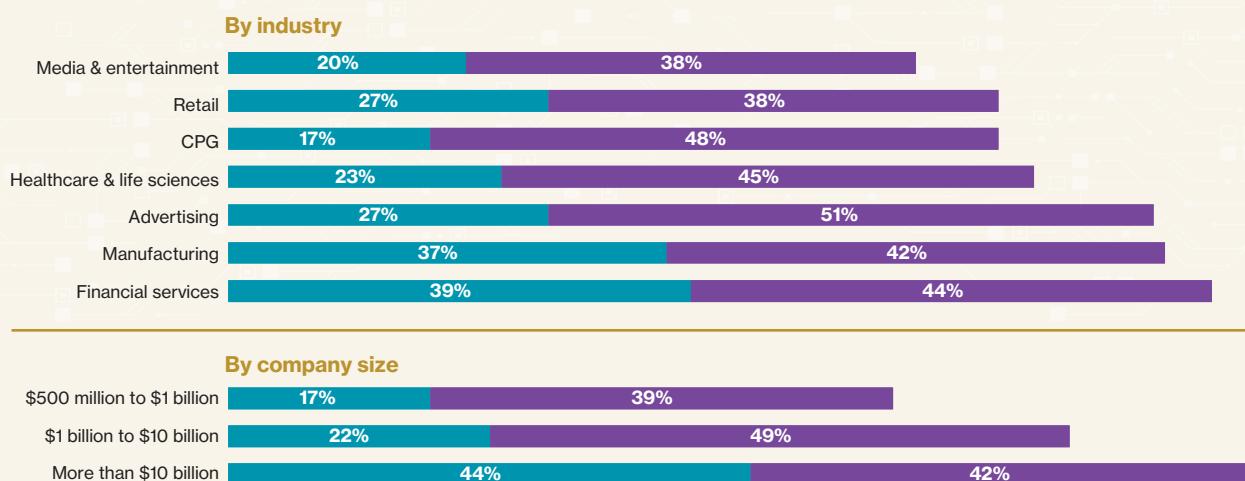
For the survey’s biggest companies, which placed even greater emphasis on the importance of data engineers, this is likely due to their more advanced AI maturity. Considerably more of the biggest companies have already deployed generative, multi-modal, and semi-autonomous agentic AI, as well as AI-powered data tools. “Some of those early AI adopters – the organizations that are really moving the needle on AI – are looking at their data engineers as strategic enablers in business transformation,” says Jyoti.

Figure 2a: Nearly 3 in 4 data and technology executives agree data engineers are integral to the success of their business



Source: MIT Technology Review Insights survey, 2025

Figure 2b: Nearly 9 in 10 of the largest surveyed organizations say data engineers are integral to the success of their business



Source: MIT Technology Review Insights survey, 2025

■ Strongly agree ■ Somewhat agree

Growing workloads and a shifting focus

A large majority of surveyed executives (81%) say the job description for data engineers in their organization has changed radically due to AI. This can be seen in a shift in their day-to-day activities. Two years ago, data engineers spent an average of 19% of their workday on AI projects (such as building and monitoring model pipelines or data cleansing), say respondents, and 81% on other activities (data warehousing, quality checking, and infrastructure management, for example). In 2025, the analogous figures are 37% of time spent on AI and 63% on other activities (see Figure 3).

"The role has evolved significantly," says Jyoti. "In the earlier days of AI and machine learning, data engineers focused on building data pipelines. They were very siloed, focusing on one task at a time. Now they're focused on designing and architecting reusable data platforms, governance, and oversight. They're playing a pivotal role as key enablers of AI."

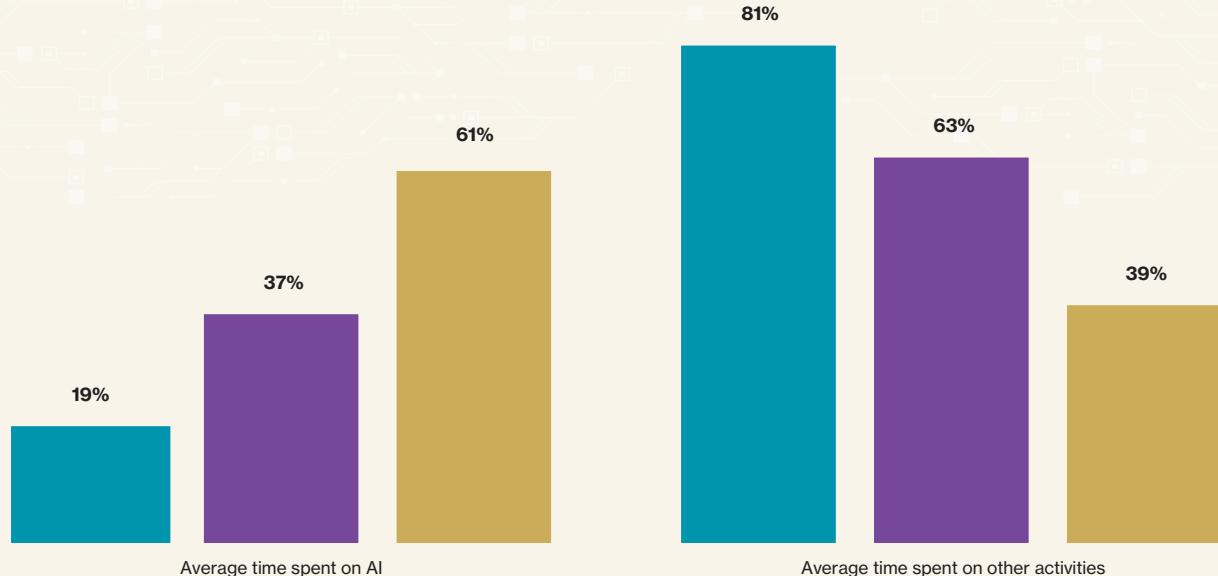
The respondents expect the change to accelerate: two years from now, they believe the ratio of time spent on AI to that on other activities will almost be the reverse of today – 61% to 39% in favor of AI (see Figure 3). As priorities change, so too does the volume of work.

"Some of those early AI adopters – the organizations that are really moving the needle on AI – are looking at their data engineers as strategic enablers in business transformation."

**Ritu Jyoti, Chief Executive Officer,
Stealth AI Startup**

Most respondents (77%) say their data engineers' workload is becoming increasingly heavy (see Figure 4a). "They're struggling to keep up," says Jyoti. "Data volumes and velocity have exploded. Data is now multimodal and is being ingested in real time." Interestingly, considerably more chief AI officers observe this than their peers (see Figure 4b).

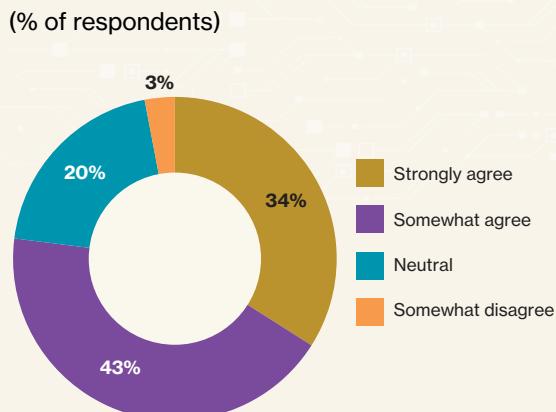
Figure 3: Data engineers are spending increasing amounts of time on AI projects versus other activities (% of respondents)



“Models need so much more data and in multiple formats. Where it used to be making sense of structured data, which was relatively straightforward, now it’s: ‘What do we do with all this unstructured data? How do we tag it? How do we organize it? How do we store it?’ That’s a bigger challenge.”

George Westerman, Senior Lecturer and Principal Research Scientist, MIT Sloan School of Management

Figure 4a: 3 in 4 respondents agree data engineers’ workloads are growing



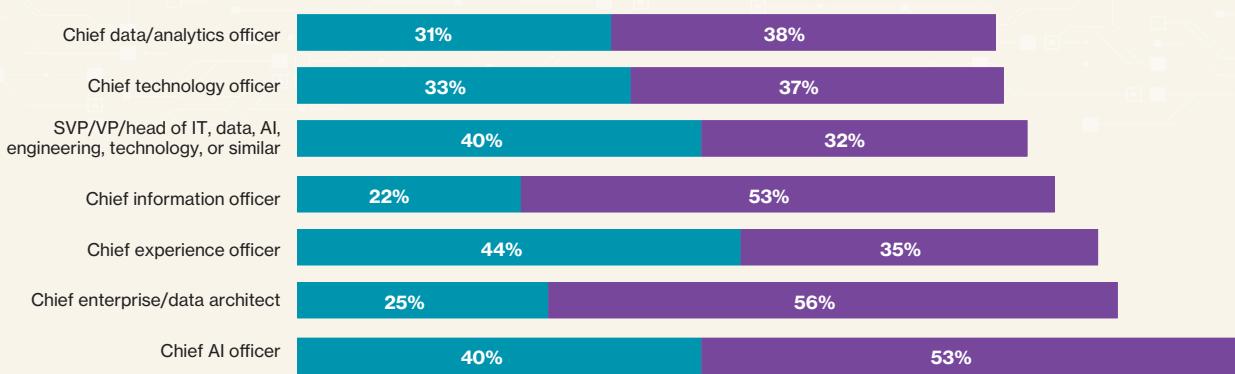
Source: MIT Technology Review Insights survey, 2025

“Models need so much more data and in multiple formats,” says George Westerman, senior lecturer and principal research scientist at the MIT Sloan School of Management. “Where it used to be making sense of structured data, which was relatively straightforward, now it’s: ‘What do we do with all this unstructured data? How do we tag it? How do we organize it? How do we store it?’ That’s a bigger challenge.”

Dave Masino, senior director of data and intelligence at technology consulting company Slalom, sees the growing workload as an opportunity. “AI has increased the amount of work data engineers are doing, but you inherently want your team to be busy,” he says. “I see the advantage we’re getting from AI-enabled acceleration eclipsing the amount of additional workload.”

Figure 4b: CAIOs are the most likely to agree that data engineers’ workloads are becoming increasingly heavy

(By executive role)



Source: MIT Technology Review Insights survey, 2025

Strongly agree Somewhat agree

Learning from software engineers

Data engineers' success in mastering AI can in large part be credited to their ability to adapt software engineering practices to their own needs, says Chris Child, vice president of product and data engineering at Snowflake. "To my mind this is the biggest shift in data engineering in the past five to seven years," he says, "and it's still ongoing. It's helped data engineers understand how to build and manage complex software projects, including AI projects, at scale."

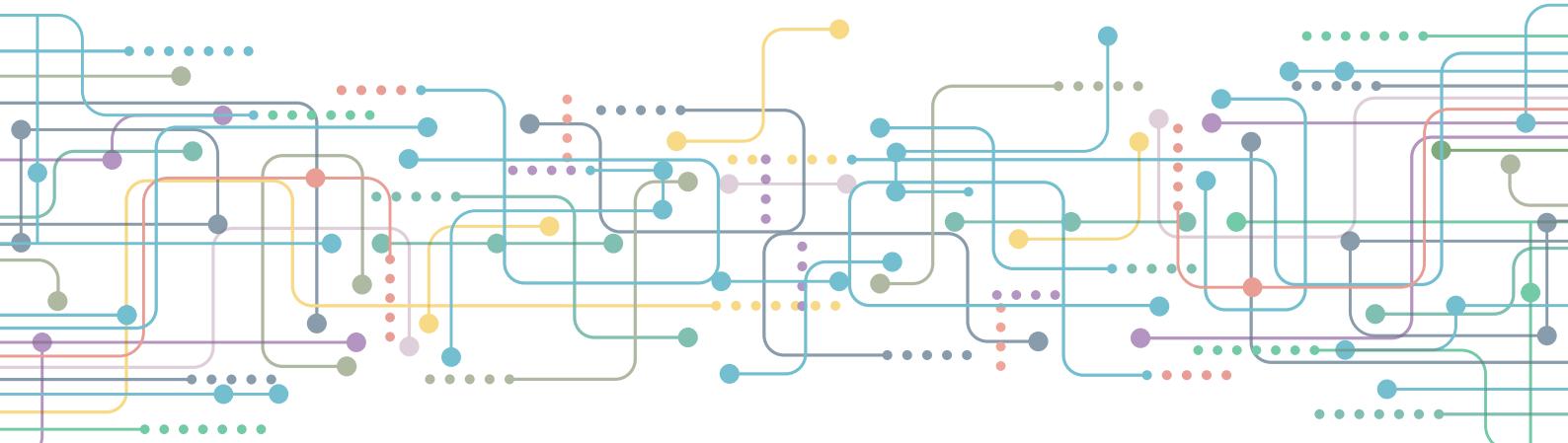
Child provides examples: "It's how to treat infrastructure as code, using tools like Terraform to set up your data engineering infrastructure.

It's making sure all your data engineering code is checked into version control deployed through continuous integration and continuous deployment." This is helping with managing AI projects, Child says, because many AI tools in use today have been trained on software engineering problems.

"It's been a massive positive," says Child. "Data engineers are now thinking in a software kind of way. They're now better able to move quickly and to change things without fear of breaking something," he says. "It's been a change in philosophy and team culture as much as a change in tooling."

"AI has increased the amount of work data engineers are doing, but you inherently want your team to be busy. I see the advantage we're getting from AI-enabled acceleration eclipsing the amount of additional workload."

Dave Masino, Senior Director of Data and Intelligence, Slalom



Partner perspective: Snowflake

Entering a golden age for data engineering

The rise of AI has made data engineers more important than ever. With growing workloads and responsibilities, they are now strategic partners who directly shape business outcomes. At Snowflake, we see this as more than a simple shift in responsibilities. It's the dawn of a new, inspiring era for data engineering as a whole.

Adapting to meet complex and growing demands is nothing new for data engineers; these days, they're handling more unstructured data and managing real-time pipelines for AI. And this is where Snowflake's platform is designed to help them thrive. Our fully managed, serverless architecture handles the operational overhead, allowing data engineers the freedom to focus on strategic, high-value tasks instead of worrying about infrastructure provisioning, tuning, and scaling. We don't just provide the building blocks; we make it easy to create a resilient data foundation that is ready for the future.

Delivering business value through trustworthy data without getting bogged down in mundane maintenance requires tools that simplify processes, and these are built directly into the Snowflake platform. Starting with ingestion, Snowflake Openflow makes it easy to handle multimodal data with high throughput and low latency – which we've made even more cost-effective. With Snowpark, data engineers can use familiar languages like Python to build, manage, and deploy data pipelines and machine learning models, bringing the code to the data instead of the other way around. For real-time needs, Dynamic Tables offer a declarative framework that streamlines the creation of both batch and streaming pipelines – meaning a single, consistent framework can work for both types of data flows without the traditional complexity.

Given the shift in focus toward AI projects, as the survey revealed, we see the value in having a built-in suite of AI-powered capabilities, like Snowflake Cortex AI. By integrating data and AI on a single, secure platform, Snowflake can help data engineers move beyond basic data management and become the essential partners their businesses need to succeed with AI today.

This, of course, also underscores the critical importance of data governance. New AI tools can often increase the complexity of integration and add to tech stack fragmentation, making robust governance more difficult to maintain. Snowflake is built to address this head-on. Our platform's unified architecture consolidates structured, unstructured, and semistructured data in a single, secure environment, minimizing the fragmentation that leads to gaps in governance. Providing features that help with secure data access, lineage tracing, and data quality – essential for maintaining the integrity of the data feeding AI models – Snowflake can help data engineers trust their data.

The role of the data engineer is expanding. They're architects, foundation-setters, orchestrators, and so much more. Simply put: They are the operational lifeblood of any data-driven organization.

The agentic challenge

The relentless progression of AI means that data engineers must keep pace and grasp the data implications of the field's advancing capabilities. "Data engineers will need to keep increasing their efficiency," says Child. "For one thing, the amount of data that AI models require is growing exponentially, as is the number of AI projects." But it's more than volume, he says. "Data engineers are managing more complexity, such as unstructured data and real-time pipelines. They're also managing the ever-increasing expectations of business stakeholders."

Data engineers in the surveyed organizations are clearly striving to meet those expectations, judging by how their workday has shifted since 2023 (see Figure 3). That change has coincided with a surge in enterprise adoption of generative AI. Most of the surveyed businesses have by now deployed generative AI, and most of the rest will begin doing so in the next 12 months (see Figures 5 and 6).

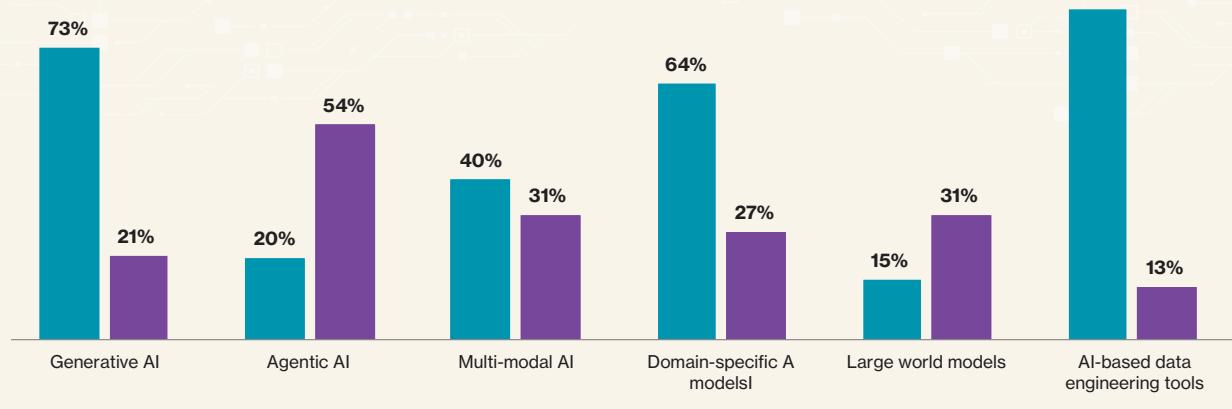
"Data engineers are managing more complexity, such as unstructured data and real-time pipelines. They're also managing the ever-increasing expectations of business stakeholders."

Chris Child, Vice President of Product, Data Engineering, Snowflake

Adoption of multi-modal AI, which incorporates a wide variety of data types, is also growing. Four in ten of the surveyed businesses have already begun deploying multi-modal AI, while 31% plan to start doing so within the next year (see Figure 5). Multi-modal AI illustrates the imperative for data engineers to learn how to manage and optimize flows of unstructured and semi-structured data. The diversity of data types – images, video, audio, sensor data, and other formats – is not new. But the volumes and varieties consumed by large language models (LLMs) and the speed of data ingestion are much greater than in pre-generative AI days.

Figure 5: More than 8 in 10 technology leaders say their organization has already deployed AI-based data engineering tools

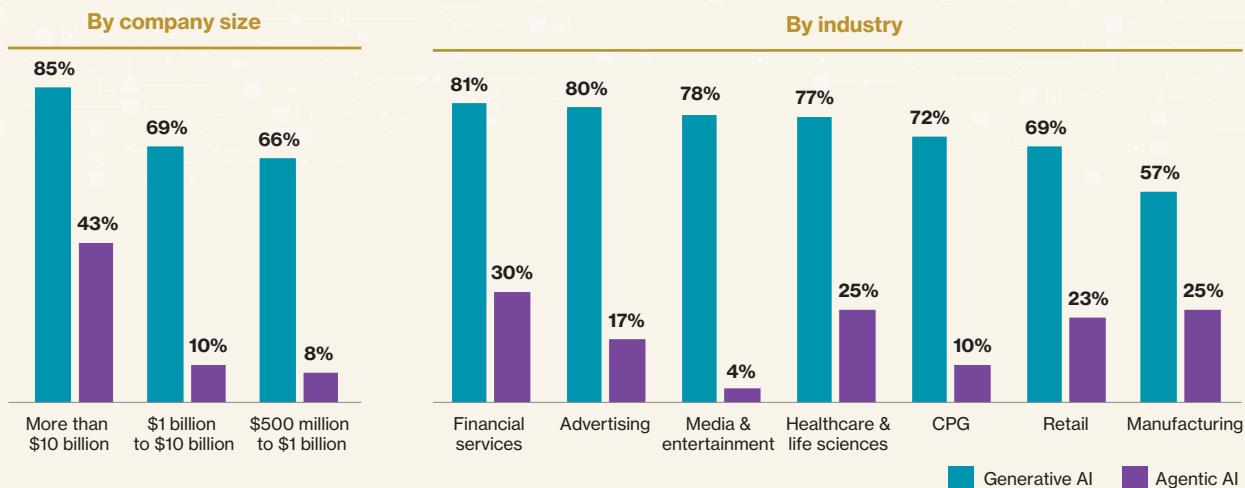
(% of respondents)



Source: MIT Technology Review Insights survey, 2025

■ Have begun deploying ■ Will begin deploying within 12 months

Figure 6: The majority of companies are already deploying generative AI. The largest are most advanced with agentic AI



Source: MIT Technology Review Insights survey, 2025

Agentic AI with its autonomous capabilities poses the next big challenge to data engineers. Its adoption thus far is limited. Just one-fifth (20%) of the surveyed organizations have begun working with agentic models. Over half (54%) say they will start doing so in the next 12 months (see Figure 5).

Agents of change

For enterprises, the business benefits agentic AI offers are not vastly different from generative AI: gains in operational efficiency, employee productivity, and automation-driven cost economies, for example. The difference with agentic AI lies in its ability not only to analyze and inform decision-making as generative AI does, but to be goal-driven, adaptive, and autonomously make decisions and act on them. “Agentic AI relies on foundation models or LLMs as does generative AI, but it provides more advanced reasoning, context, interpretation, and self-correction,” says Jyoti. “Agentic AI will give us systems that not only research, analyze, and plan, but that act on plans in a dynamic and agile way.”

For data engineering teams, agentic AI offers a raft of efficiency-enhancing benefits. The top two, according to surveyed respondents, are better pipeline debugging and optimization (cited by 42%) and better data integration (38%). The latter refers to the consolidation of data from different systems and formats to create single, consistent datasets. Improved orchestration across teams (34%) and stronger data governance and compliance (33%) are other frequently mentioned agentic AI advances for data engineering (see Figure 7).

“Agentic AI relies on foundation models or LLMs as does generative AI, but it provides more advanced reasoning, context, interpretation, and self-correction. Agentic AI will give us systems that not only research, analyze, and plan, but that act on plans in a dynamic and agile way.”

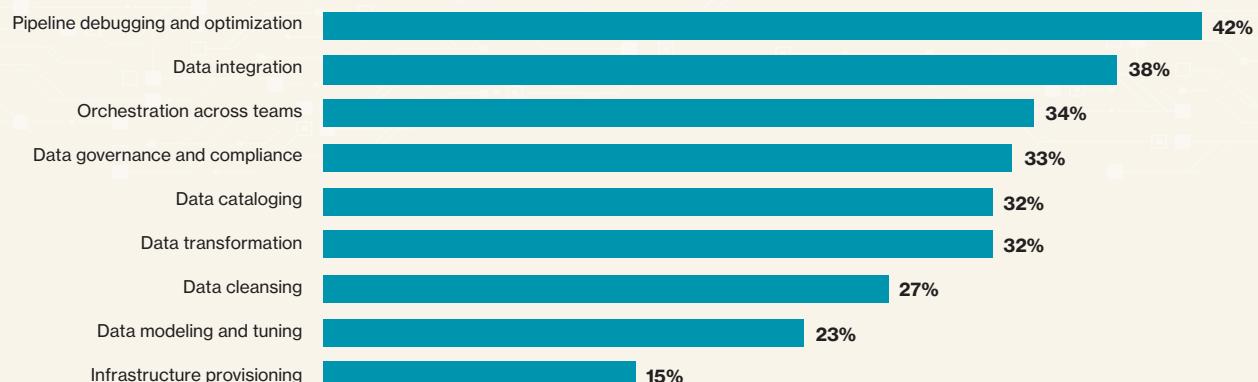
Ritu Jyoti, Chief Executive Officer, Stealth AI Startup

Among the survey’s biggest businesses, improved orchestration is the most attractive potential benefit (cited by 52%). For the smallest – those with annual revenues of between \$500 million and \$1 billion – it is data integration (47%).

Child predicts higher value benefits coming from agentic AI. “We’ll start to see more agentic data engineering where having [AI] agents do a larger chunk of their operational work allows data engineers

Figure 7: Pipeline debugging and optimization and data integration are the top two benefits of agentic AI for data engineering teams

(% of respondents)



Source: MIT Technology Review Insights survey, 2025

and teams to think about the bigger picture," he says. "They'll ask, 'What are our overarching goals? What budget do I give to which of these agents to process data? How do we think about our overall data estate rather than just individual pipelines?' Then you'll start to see a larger shift in the role of the data engineers."

"I'm looking forward to offloading all the repetitive engineering work to [AI] agents and having my team focus on the things that are interesting, which are architecture, systems thinking, and solving real business problems," says the head of data and analytics at a large retail organization.

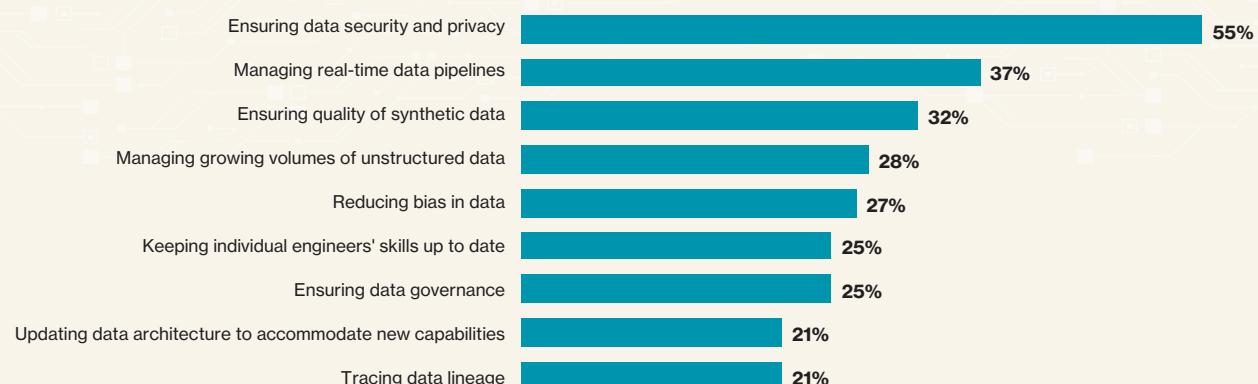
Risk equation

With advances in technologies like agentic AI also come challenges. For survey respondents, by far the most critical challenge is in ensuring data security and privacy, as cited by 55% (see Figure 8).

The defining characteristics of agentic AI, as autonomous systems that take actions themselves, magnifies the potential fallout that could occur as a result of a data breach. "The best case scenario is that a breach results in some embarrassment. The worst case is that your business is forced to shut down, says Masino.

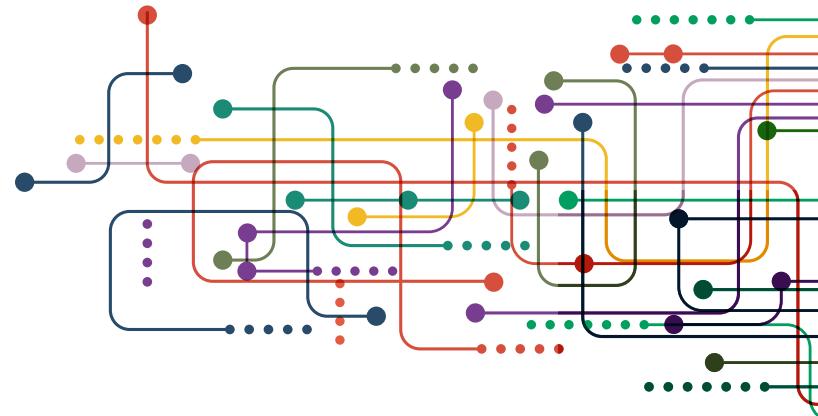
Figure 8: Ensuring data security and privacy is rated the greatest challenge for data engineering teams as AI capabilities advance

(% of respondents)



Source: MIT Technology Review Insights survey, 2025

AI-powered data engineering



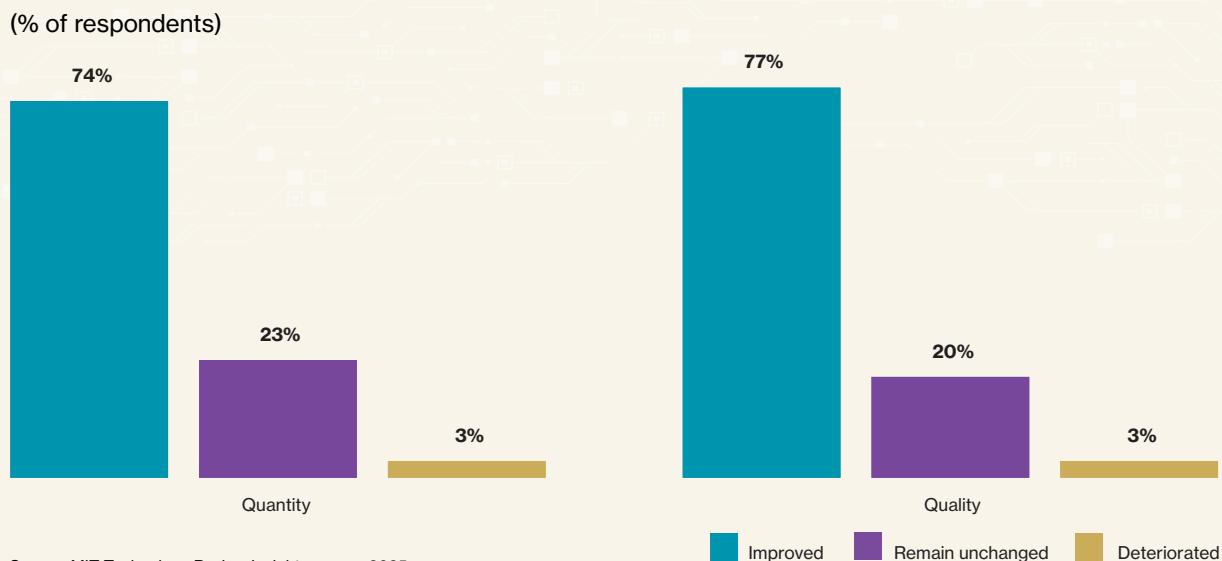
While data engineers are working to make AI more effective, AI is returning the favor. The technology has proven a boon to data engineer productivity, in terms of output quantity and quality.

Nearly three-quarters of respondents (74%) report AI-led improvement in their data engineering teams' productivity in the past two years in terms of quantity. This could include the number of outputs such as projects delivered and new code generated. Even more (77%) say the same about the quality of their data engineering teams' work (data freshness, for example) (see Figure 9). One-fifth of the surveyed executives say the improvement in output quantity has been "significant," and 29% say the same about the improvement in quality.

These improvements are no surprise to Masino. "With advancements in generative AI over the past two to three years and, most importantly, its integration into software development tooling, data engineers now have a very powerful accelerator at their disposal," he says.

The survey respondents' expectations of pipeline optimization, data integration, orchestration, and governance gains from agentic AI (see Figure 7) indicate a strong belief that AI can deliver yet more productivity gains to data engineering teams. This is further supported by the finding that 83% of respondents say their companies have begun using AI-based data engineering tools (see Figure 5).

Figure 9: AI has improved the quantity and quality of data engineering teams' output in the last 2 years, according to 3 in 4 respondents



“With advancements in generative AI over the past two to three years and, most importantly, its integration into software development tooling, data engineers now have a very powerful accelerator at their disposal.”

Dave Masino, Senior Director of Data and Intelligence, Slalom

Many such organizations are likely to be using AI code-generation tools of the type that appeared on the market earlier this decade. More recently, tools have emerged that use AI to automate or assist data engineers with data cleansing, integration, pipeline monitoring, metadata management, workflow orchestration, feature engineering, governance and compliance, and other tasks. “With the advanced tools we have now, it’s a lot easier to do things that once were very difficult,” says the head of data and analytics at a large retail organization.

“Such tools are making data engineers much more efficient,” says Westerman.

Masino agrees, believing that AI-enabled code editors and command line tools have had the most significant impact thus far. “My favorite tools today are Cursor, Claude Code, and Codex CLI,” says Masino. “What began as code completion and chat interfaces in development environments are now code agents that can act with autonomy.”

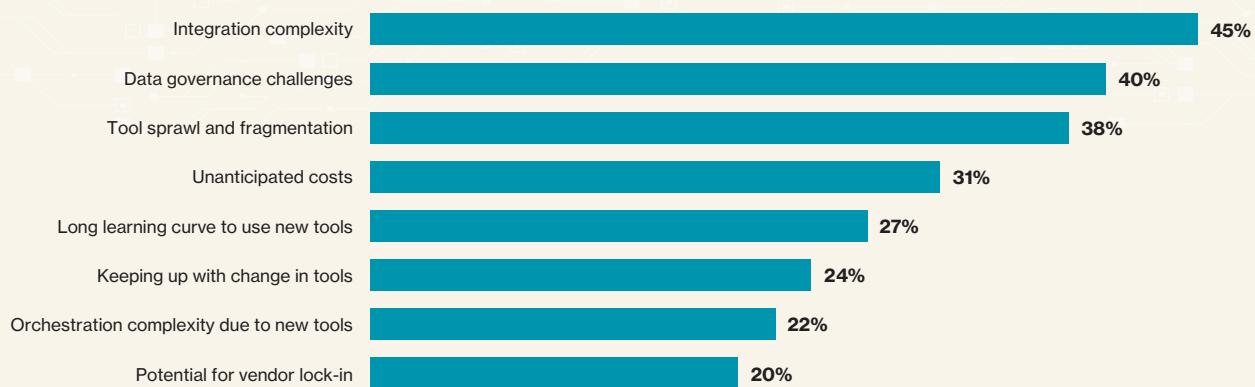
More tools, greater complexity

AI tools are not a panacea for data engineers, however. Many such tools run on LLMs, and some of the challenges AI tools present are not dissimilar to those in AI use cases elsewhere in the enterprise.

When asked to list the main challenges for data engineers posed by AI tools, our survey respondents put integration complexity and data governance at the top (cited by 45% and 40%, respectively) (see Figure 10). Integration challenges arise from, for example, the need to pull together data from different, often legacy, tooling systems used by a team. Models underpinning AI engineering tools rely on clean, accurate, and high-quality data as much as other models do, thus requiring robust governance processes. For instance, AI data tools must be able to trace the lineage of data being ingested, to monitor for data drift, and to manage access control securely. The biggest companies in the survey rate data governance their greatest challenge.

Figure 10: Integration complexity, data governance, and tool sprawl are the highest-rated challenges of new data tools, including those powered by AI

(% of respondents)



Source: MIT Technology Review Insights survey, 2025

Focus on finance and manufacturing

When dissecting the survey results by industry, financial services and manufacturing companies stand out from others in several aspects of the data engineering challenge. Take overall performance. More data engineering teams in these industries, as well as in media and entertainment, have registered productivity improvements in terms of both quantity and quality of output thanks to AI.

It may be no coincidence that data engineers in manufacturing and financial services spend more time working on AI projects today than their peers in other industries (40% and 39%, respectively). And the same is expected to be true two years from now (65% and 64%, respectively) (see Figure 11).

Cost and integration concerns are closely connected to another challenge high on the respondents' list: tool sprawl and fragmentation (cited by 38%) (see Figure 10). Aside from direct costs (AI tools often do not come cheap), respondents are also likely worried about indirect costs, like the time lost to data engineering teams from having to maintain a growing array of often disconnected tools.

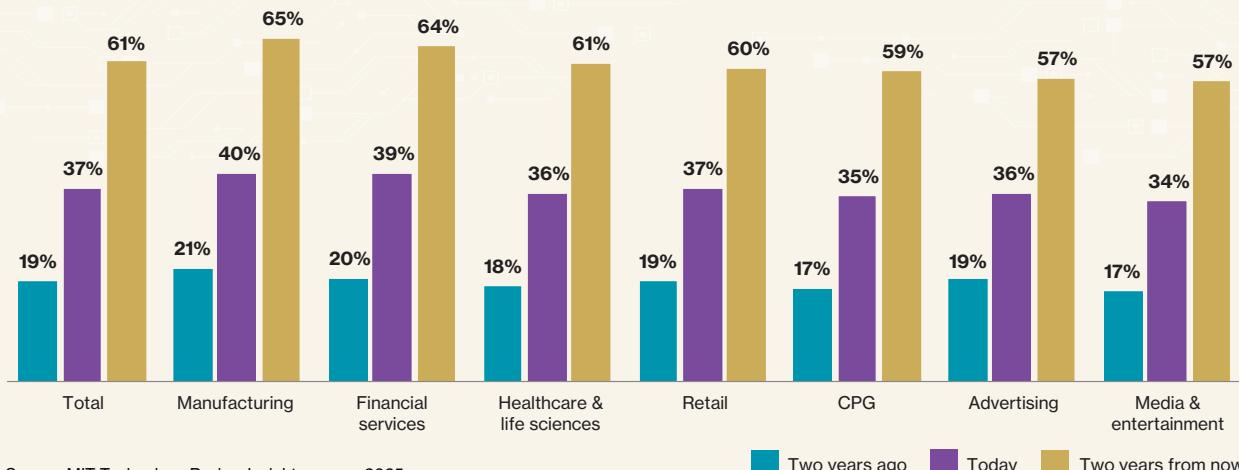
Their comparatively greater time spent with AI may help to explain a slightly greater prominence that data engineers in financial services and manufacturing enjoy in their organizations than their peers elsewhere. When asked about decision-making in a range of areas, executives from those industries are more likely to describe their data engineers as influential. This applies to areas such as investment in data tools, vendor selection, AI use-case feasibility, and overall data strategy.

Perhaps a more telling manifestation of their influence is that more data engineers from those two industries, along with those from advertising, are deemed by respondents to be integral to the success of their business (Figure 2b).

To avoid the hazards of fragmentation, teams need to rigorously review and consolidate their existing stack of tools before investing in newer, AI-enabled ones. "Tooling is continuing to evolve," says Masino. "There are myriad different choices, and it's best to avoid hyper-analyzing all of them. They all have rough edges here and there, and my advice is pick one, start using it, form your opinion on it, and continue from there."

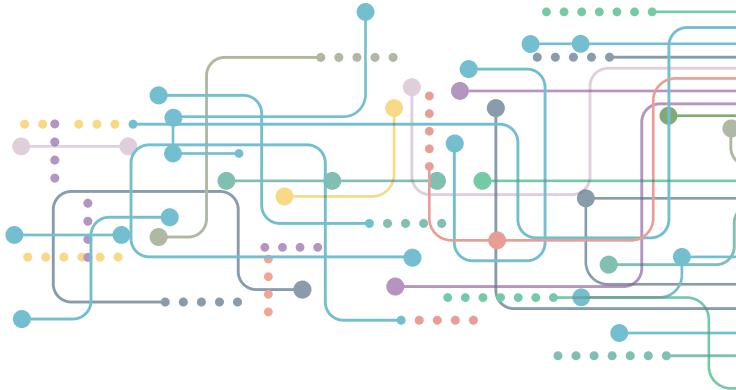
Figure 11: Data engineers in manufacturing and financial services spend more time, on average, on AI projects than other industries

(By industry; % of respondents)



Source: MIT Technology Review Insights survey, 2025

05 Building influence



Data engineers have long been described as the plumbers of the data world – technicians who maintain the pipelines critical to the smooth running of systems. In their traditional role as masters of extract, transform, load (ETL) work, data engineers have also been likened to bricklayers, says Westerman. “They may not seem as important as the people who design the building or who come to use it, but it can’t be built without them,” he says.

Limited to this role, data engineers would not be expected to exert significant influence with AI, data science, or other technology teams, to say nothing of the wider business. “If you’re basically doing ETL work, you don’t need a voice outside of your team, you just do what you’re asked to do,” says Westerman. “If you’re doing the architect job, on the other hand, then a voice becomes much more necessary.” In most organizations, the data architect has typically played the role of designer of data systems and infrastructure and definer of data standards.

“If you’re basically doing extract, transform, load work, you don’t need a voice outside of your team, you just do what you’re asked to do. If you’re doing the architect job, on the other hand, then a voice becomes much more necessary.”

George Westerman, Senior Lecturer and Principal Research Scientist, MIT Sloan School of Management

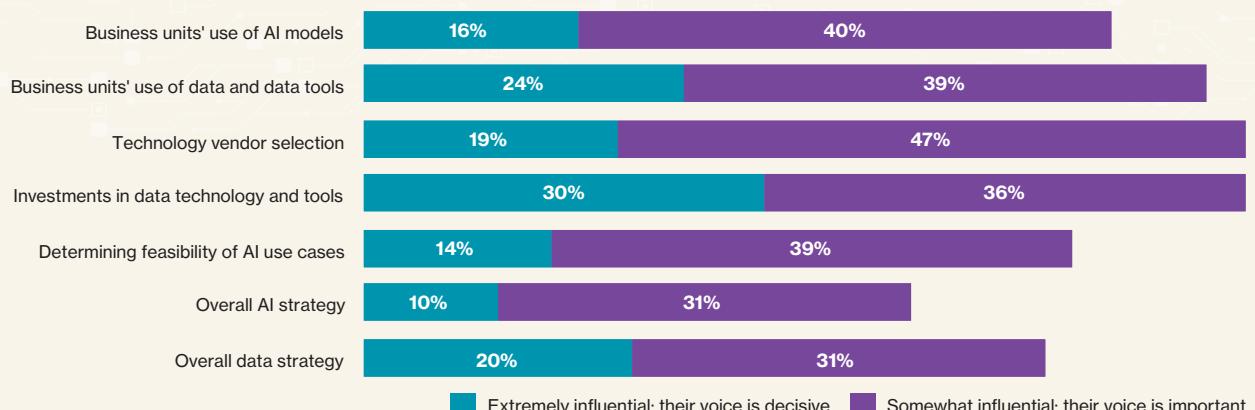
Our research points to some convergence of the data engineer and architect roles taking place in many organizations. In the survey, that is evidenced in the influence data engineers are seen to exert in areas that have historically been considered the domain of data architects.

For example, 51% of respondents say their data engineers influence their organization’s overall data strategy to a greater or lesser extent. More say data engineers’ influence extends to investments in data technology and tools (66%), technology vendor selection (66%), and business units’ use of data and data tools (63%) (see Figure 12).

In most organizations, data engineers’ influence also extends to the use of AI. Over half say this is the case with decisions made on AI use-case feasibility (53%) and on business units’ use of AI models (56%). Four in ten respondents (41%) say their data engineers’ voice is heard on questions regarding overall AI strategy (see Figure 12).

Figure 12: Data engineers are seen to exert decision-making influence in areas historically seen as the domain of data architects

(% of respondents)



Source: MIT Technology Review Insights survey, 2025

There is something of a disconnect on these points, however, between senior technology leaders. In particular, chief information officers (CIOs) have a weaker perception of data engineer influence than chief data officers (CDOs) and chief AI officers (CAIOs). The gap between CIO and CDO/CAIO perceptions is especially wide when it comes to data engineers' influence over matters of AI use. This suggests that data engineers' value to the business is not fully appreciated across the senior leadership team; it certainly isn't by all CIOs. Supporting this view, far fewer CIOs (55%) than CDOs (80%) or CAIOs (82%) deem data engineers to be integral to the business.

One plausible explanation for this divergence is the degree of visibility into data engineers' day-to-day activities. CDOs and CAIOs simply have more of it than CIOs. "It's hard to place value on the things you can't see," says Masino. "CDOs and CAIOs typically work more closely with data engineering teams than CIOs and know how difficult it can be to get value out of data."

The divergence in C-suite views can also be explained by the nature of the executives' respective roles, according to Child. "Part of the CDO's and CAIO's job is to think about the future, while the CIO is thinking a lot more about the infrastructure. The latter is likely thinking about data engineers as part of the machine that's functioning but not about how it's going to shift. That said, AI is pushing leaders to reconsider these processes. In a few years' time, the share of CIOs agreeing that data engineers are integral to the business is likely to be significantly higher."

Engineer and architect

Nevertheless, these findings provide a clear indication of an expanding data engineer role. "Some data engineers are starting to think about business problems that need to be solved," says Child. "They're asking, 'Where should we be investing resources? What should these different agents focus on?'"

"Over time, the data engineer role will shift from writing code for all pipelines toward managing the infrastructure that these are running in, orchestrating across a lot of these, and setting the rules and tests to make sure the right data is coming in," says Child.

Jyoti maintains that the data engineer and data architect roles will converge, and she sees that occurring now at a handful of organizations. "Data engineers need to become fluent in AI. That will help them grow into the roles of system or enterprise architects."

Jyoti sees enterprise architects playing an important role in driving the company's transformation. "They can proactively see who the players are in the industry and what the customers are demanding," she says. "They'll seek ways to innovate and to help prioritize the company's AI use cases. They'll provide input on where the company should invest to ensure those use cases deliver a competitive advantage. That will definitely raise their profile among the senior business leaders."

Conclusion: Embracing the change

Some data engineers worry that AI will automate their jobs away. While AI may indeed automate away some of what they currently do, data engineering jobs are likely safe for the foreseeable future. After all, there will always be a need for problem-solving around data. AI is also giving data engineers a valuable opportunity to grow.

Among other benefits, by taking on core data engineering tasks, AI-powered automation will lighten heavy data engineering workloads to allow a shift in focus from execution to strategy. “As AI agents take on more work, data engineers will spend less time on ETL and more on the data and AI strategizing that will help take the business forward,” says Child.

For data engineers, three calls to action emerge from the research:

Become an expert in AI. Data engineers need to understand how existing and emerging AI models, particularly LLMs, work and how they ingest, process,

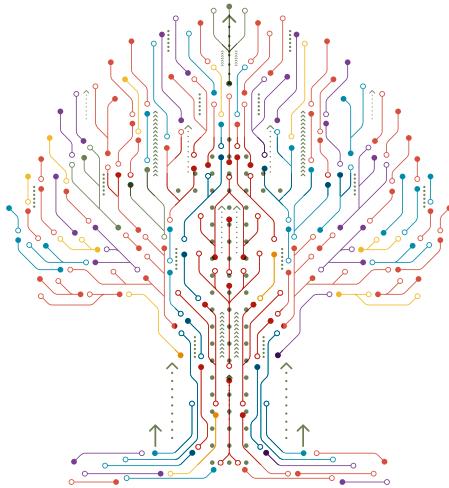
and verify data. Organizations should help their data engineers find and enroll in courses on, for example, machine learning frameworks, deep learning, feature engineering, and model evaluation. Heads of AI could also organize internal training in these areas.

Understand the business. Data engineers need a strong understanding of key business objectives to be able to deliver value. They need to have conversations within different business units to better understand what they are trying to do and what data they need. Technology leaders can facilitate this by pushing the creation of cross-functional teams in which data engineers are temporarily embedded with product owners and other stakeholders in AI use cases.

Build strong communication and presentation skills. As data engineers expand their influence, they need to be able to communicate in language that business units can understand. Data engineers who can strike a good balance between technical (including AI) acumen and the ability to collaborate, listen, and persuade will be most valuable to the business.

“As AI agents take on more work, data engineers will spend less time on extract, transform, load work and more on the data and AI strategizing that will help take the business forward.”

Chris Child, Vice President of Product, Data Engineering, Snowflake



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. This content was researched, designed, and written entirely by human writers, editors, analysts, and illustrators. This includes the writing of surveys and collection of data for surveys. AI tools that may have been used were limited to secondary production processes that passed thorough human review.

About Snowflake

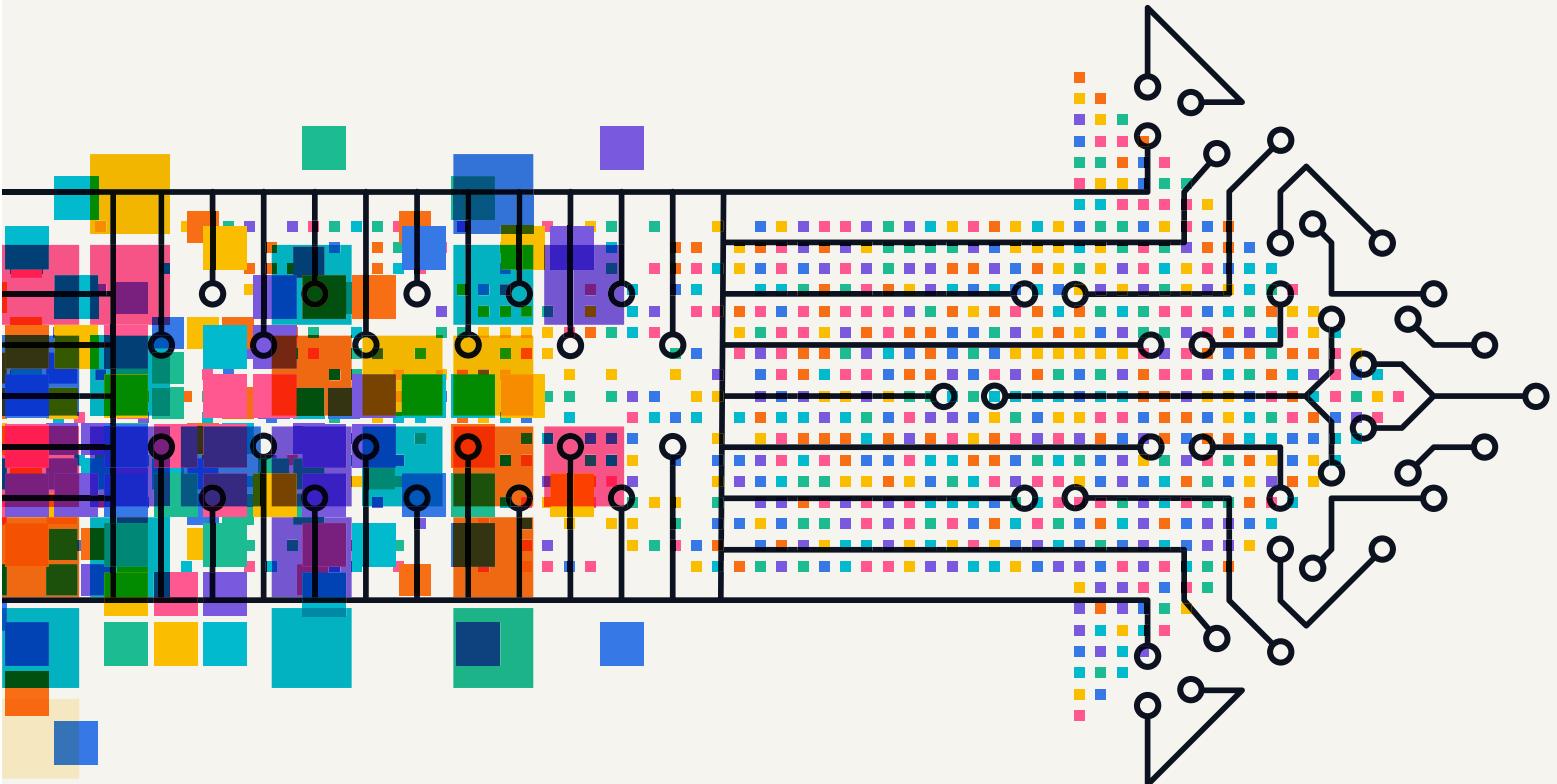
Snowflake is the platform for the AI era, making it easy for enterprises to innovate faster and get more value from data. More than 12,000 customers around the globe, including hundreds of the world's largest companies, use Snowflake's AI Data Cloud to build, use, and share data, applications, and AI. With Snowflake, data and AI are transformative for everyone. Learn more at [snowflake.com](https://www.snowflake.com) (NYSE: SNOW).



Illustrations

Illustrations assembled by Tim Huxford with elements from Shutterstock and Adobe Stock.

While every effort has been taken to verify the accuracy of this information, MIT Technology Review Insights cannot accept any responsibility or liability for reliance by any person in this report or any of the information, opinions, or conclusions set out in this report.



MIT Technology Review Insights

www.technologyreview.com

insights@technologyreview.com