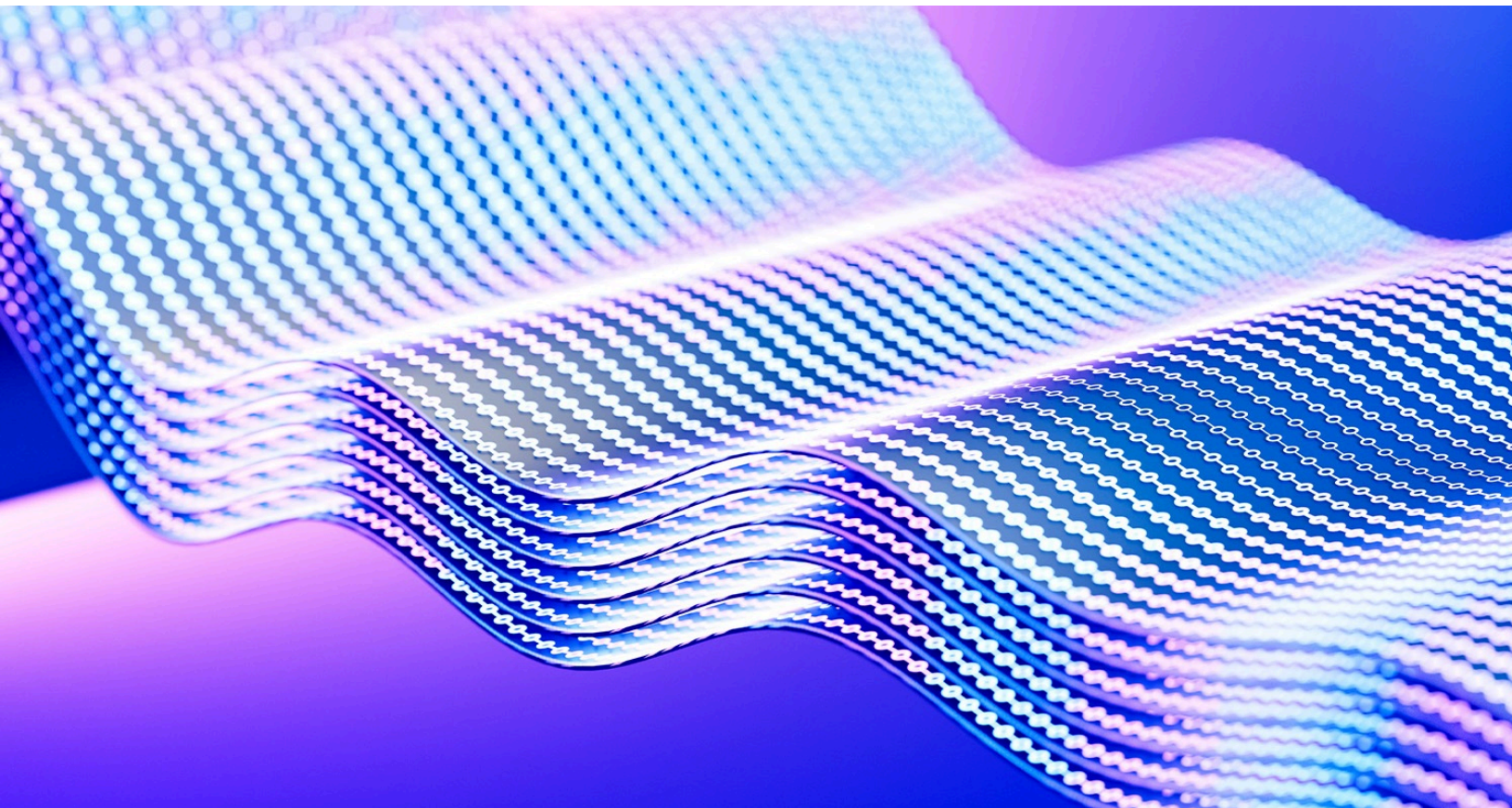


McKinsey Technology

Building the foundations for agentic AI at scale

Agentic AI scales on strong data. To capture value, tech leaders can agentify high-impact workflows, modernize data architectures, enforce data quality, and evolve operating models.

This article is a collaborative effort by Asin Tavakoli, Brian Goodman, Henning Soller, and Kayvaun Rowshankish, with Akshat Kumar, Carlos Barreto, Satyajit Parekh, and Tancredi Bernard Litta Modignani, representing views from McKinsey Technology and QuantumBlack, AI by McKinsey.



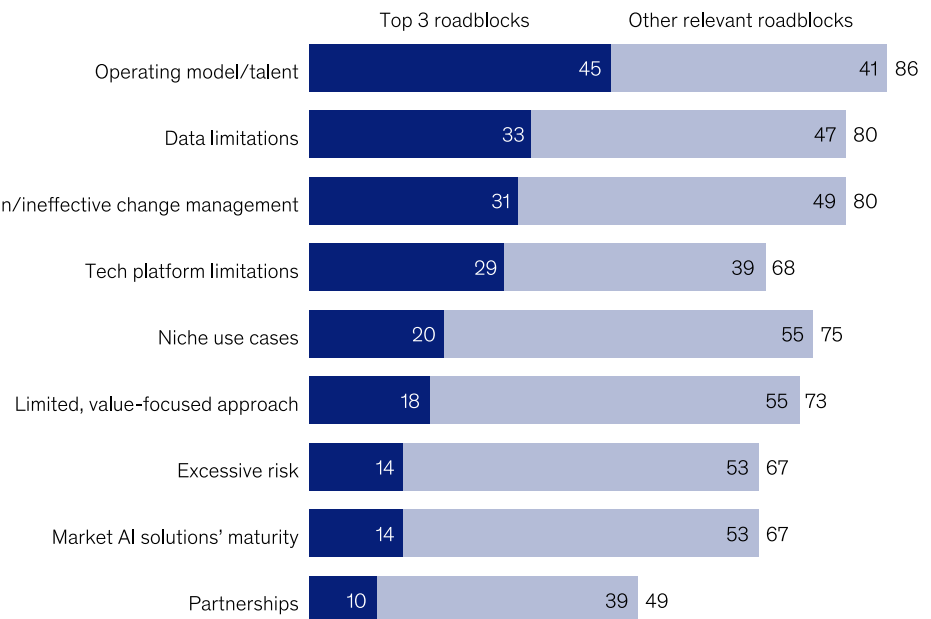
A house is only as strong as its foundation. That's what companies are quickly coming to understand about agentic AI as well. Nearly two-thirds of enterprises worldwide have experimented with agents, but fewer than 10 percent have scaled them to deliver tangible value.¹ Shaky data is often to blame; eight in ten companies cite data limitations as a roadblock to scaling agentic AI (Exhibit 1). Addressing this issue is a core element of building a solid capability foundation—and that's what distinguishes companies that create value from AI from those that don't, according to our [Rewired](#) research.

While companies have often muscled through issues of fragmented and siloed data, those issues are impossible to manage at scale. Inconsistent governance has just increased the challenge of preserving data context while enforcing access control, lineage, and auditability.

Exhibit 1

Data limitations are a primary constraint that companies face in scaling AI.

Biggest roadblocks to scaling AI,¹ % of respondents



¹Question: What are the biggest roadblocks you have observed in scaling impact from AI/gen AI? Percentages do not sum to 100%, as this question was multiselection.
Source: McKinsey survey of senior executives at global telecom operators, December 2025 (n = 49); findings apply to telecom industry but are illustrative of other industries

McKinsey & Company

¹"[Seizing the agentic AI advantage](#)," McKinsey, June 13, 2025.

Data is the backbone of agentic AI

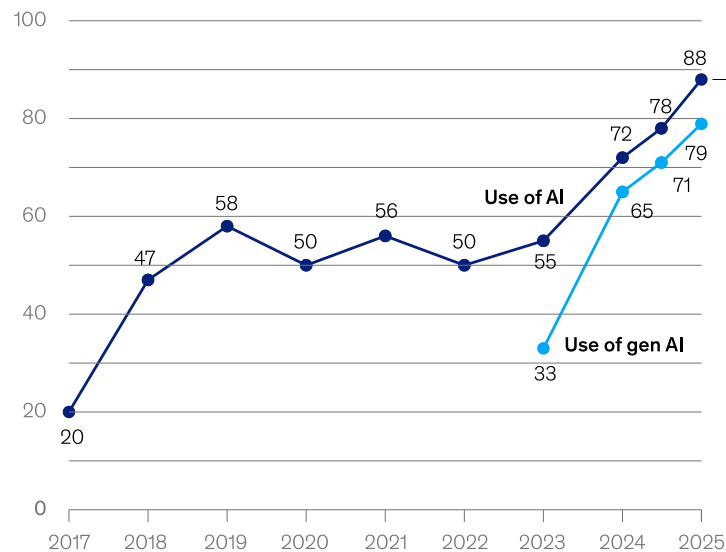
Many companies have embedded AI into their operations (Exhibit 2). Now, they are taking the next step to [implement agentic AI](#), striving to automate complex business workflows. But to function reliably at scale, agentic AI needs a steady flow of high-quality data.

Exhibit 2

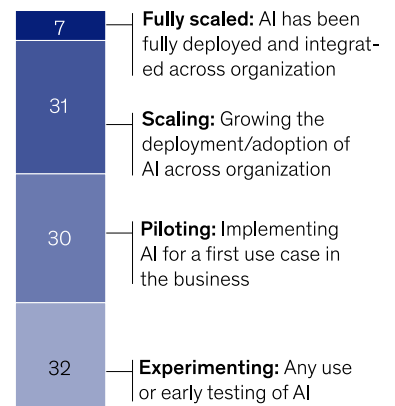
AI is becoming a critical enterprise accelerator, with the majority of companies having deployed it.

Use of AI by respondents' organizations, % of respondents

Organizations that use AI in at least 1 business function¹



Phase of AI use among organizations using AI in 2025



¹In 2017, the definition of AI use was using AI in a core part of the organization's business or at scale. In 2018–19, the definition was embedding at least 1 AI capability in business processes or products. In 2020, the definition was the organization has adopted AI in at least 1 function, and in 2025, the definition was regular use of AI in at least 1 function.

Source: McKinsey Global Surveys on the state of AI, 2017–2025

McKinsey & Company

Success with agentic AI depends on a data architecture that can support increasing levels of autonomy, coordination, and real-time decision-making. This often looks like modular, interoperable frameworks that give agents reliable access to the data they need to operate safely (see sidebar, “Seven data architecture principles that enable scale”). While gen AI has

already shown the need for data access control, lineage, and traceability, agentic platforms place greater operational pressure on these foundations. Because agentic AI coordinates multiple models and data sources continuously, often without human intervention, it requires tighter, more automated governance to ensure reliability and control at scale.

Seven data architecture principles that enable scale

1. **Treat data ingestion like a product.** Make it easy and consistent for all data—batch, real-time, structured, or unstructured—to enter the company once and be usable by everyone.
2. **Share meaning, not just data.** Ensure data comes with clear, common definitions so analytics, AI models, and agents all understand it the same way.
3. **Use one data foundation for analytics and AI.** Build data once and use it everywhere—reports, machine learning, and gen AI—rather than running separate pipelines and platforms.
4. **Build trust into the platform by default.** Security, access controls, privacy, and AI governance should be automatic, not added later or managed manually.
5. **Expose capabilities through stable interfaces.** Provide clear APIs and model access points so teams can reliably build applications and AI solutions without rework.
6. **Make behavior visible and measurable.** Continuously track data quality, model performance, speed, and cost so issues are caught early and systems improve over time.
7. **Provide a controlled way to run AI agents and applications.** Coordinate AI agents and applications through a shared execution layer that enforces enterprise rules and guardrails.

Two agentic archetypes are emerging: single-agent workflows, where one agent uses multiple tools and data sources sequentially; and multi-agent workflows, where specialized agents collaborate through shared knowledge graphs and fine-grained data access. Both require consistent, interoperable data, without which agents could break down. Single agents could make inconsistent decisions from fragmented data, while multi-agent systems could lose coordination and propagate errors.

How to prepare data for agentic AI

To enable a scaled transformation into an agentic organization, companies can start by building foundational data capabilities. This requires not just a technology reboot, but also an organizational one. That's because a company's data strategy and operating model is just as important as its underlying data quality and architecture. Success depends on taking four coordinated steps that link strategy, technology, and people:

- *Step 1: Identify high-impact workflows to “agentify.”* Organizations can identify a small number of high-value, end-to-end workflows where increased autonomy could unlock impact. Building on [established approaches to building data products](#), leaders can prioritize agentic use cases based on value potential, feasibility, and strategic fit before scaling more broadly.
- *Step 2: Modernize each layer of the data architecture for agents.* Rather than rebuilding everything from scratch, leaders can modernize existing platforms to support interoperability and governance across systems. While some may be tempted to lean on advancements in AI to shortcut data architecture best practices, the strongest organizations build modular, evolutionary architectures with components that can be replaced as new technologies emerge.
- *Step 3: Ensure that data quality is in place.* Organizations must move from periodic data cleanup to continuous, real-time quality management. They can do this while ensuring that both structured and unstructured data, as well as agent-generated outputs, meet consistent standards for accuracy, lineage, and governance.
- *Step 4: Build an operating and governance model for agentic AI.* Scaling agentic AI requires [rethinking how work gets done](#). Human roles are shifting from execution to supervision and orchestration of agent-driven workflows. In a [hybrid human–agent work environment](#), clear governance is essential to allow agents to operate transparently and safely at scale.

Below we dive more deeply into each of the four steps.

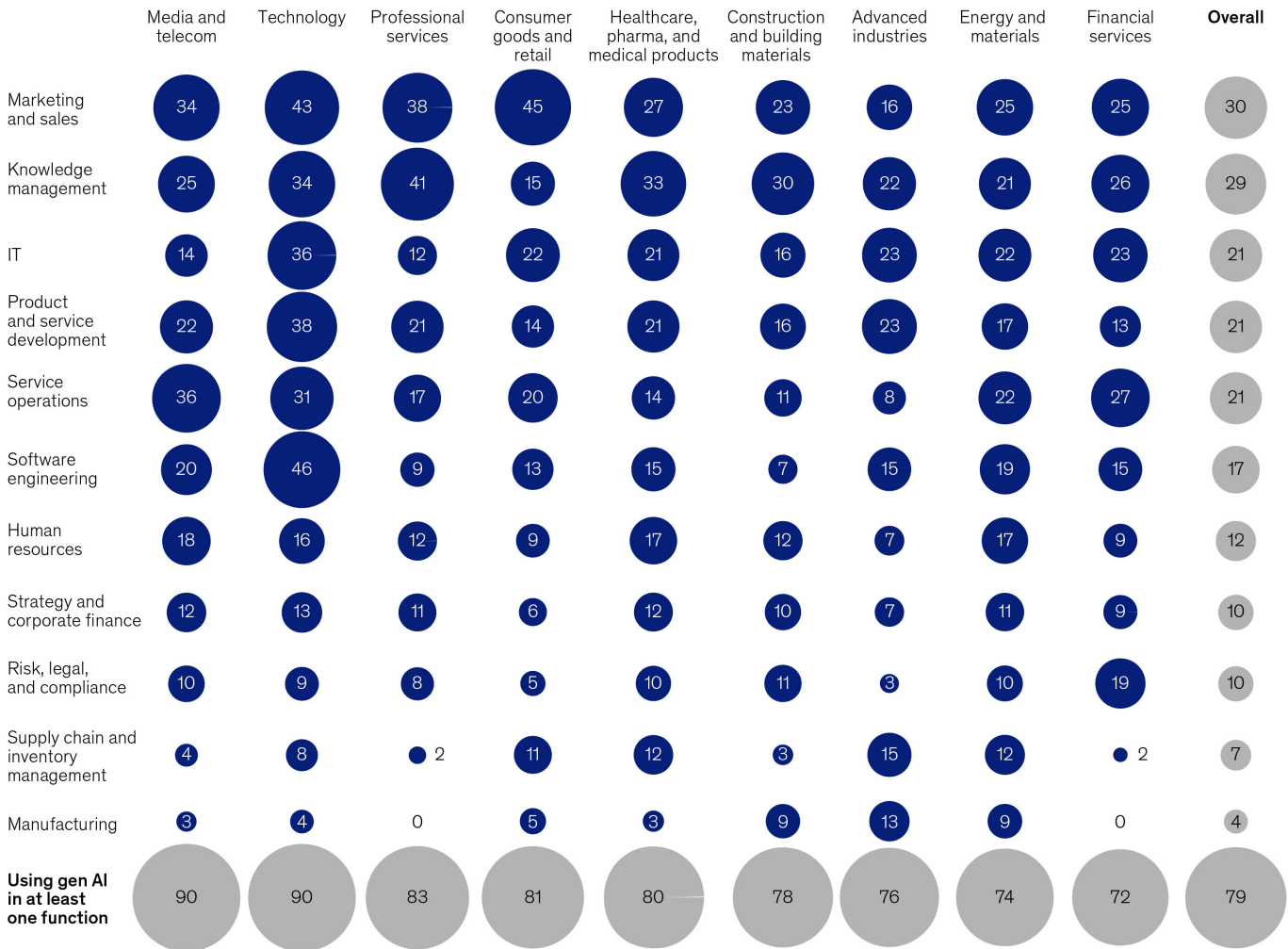
Identify high-impact workflows to ‘agentify’

For most organizations, the path to generating value from agentic AI starts not with redesigning everything at once, but with deliberately rewiring a few critical workflows in high-impact domains (Exhibit 3). Leading companies identify domains such as knowledge management or marketing, analyzing existing data to pinpoint where the greatest value lies and where increased autonomy could materially change business outcomes. This requires first mapping end-to-end workflows and then identifying steps where agentic capabilities could add value and the corresponding data needed for those agents to do their tasks. This helps companies prioritize use cases based on their value potential and feasibility.

Exhibit 3

AI adoption is led by a few functions, including marketing, knowledge management, and IT.

Business functions in which respondents' organizations are regularly using gen AI, by industry,¹
% of respondents



¹Question: In which of the following business functions is your organization regularly using gen AI—that is, machine learning algorithms such as ChatGPT that can create new content, including audio, code, text, and images?

Source: McKinsey Global Survey, June 25–July 29, 2025 (n = 1,993); participants came from 105 nations representing the full range of regions, industries, company sizes, functional specialties, and tenures; the data were weighted by the contribution of each respondent's nation to global GDP

McKinsey & Company

Targeted pilots with clear metrics can validate early impact. At the same time, teams should identify the data that can be reused across tasks and workflows. This focus on reuse is critical to developing scaled models.

Modernize each layer of the data architecture for agents

Preparing the data architecture for agentic AI requires strengthening and adjusting the layers of the data stack. Rather than rebuilding systems from scratch, companies can modernize each layer to improve visibility and governance across workflows.

To illustrate what a modernized data architecture looks like in practice, consider a typical omnichannel retail journey. Product data and purchase histories sat in silos, so context was broken as customers moved across channels, leading to inconsistent recommendations and service experiences. An agent-ready architecture connects systems and data to [support the entire customer commerce journey](#). It evolves toward data interoperability by combining traditional machine learning, gen AI, and agentic techniques.

In our omnichannel example, the *data source layer* is where customer data such as views, wish lists, purchase history, and support interactions is ingested into the enterprise. Unstructured data is continuously ingested, transformed, and recombined as it flows into models, which means governance must travel with it. Data quality checks, security controls, and lineage tracking need to be automated and embedded directly into the pipelines, not handled as one-time reviews. Data teams also build preprocessing pipelines that clean, enrich, and label data, adding business context and quality checks that enable agents to interpret information and act reliably.

The *data platform layer* connects data from different systems and makes it usable for applications and AI models by orchestrating access, synchronization, and real-time interaction across systems. In our omnichannel example, this layer ensures that customer preferences, interaction history, and transaction status remain accessible when needed as different agents engage across channels.

Vector stores and embedding services are essential for working with unstructured data, so they need to be part of the data platform. These services make documents, images, and other unstructured content searchable based on meaning rather than keywords, while keeping those representations up to date as content evolves.

Adding agent-specific interoperability standards² to this layer could automate integration and access processes, allowing for structured context sharing, direct agent-to-agent coordination, and secure transactional exchanges. Multiple agents working together could, for example, retrieve and update data across inventory, fulfillment, customer relationship management, and payments systems in real time, coordinating actions without losing continuity.

In this layer, agents can maintain dedicated working memory or scoped context, while access to shared data and contextual information is governed dynamically based on use cases and permissions. As agent autonomy increases, identity-management controls become essential to preserve data quality, reconciliation, and auditability.

²Some examples of these interoperability standards include model context protocol (MCP), which structures how agents access and share context, agent-to-agent (A2A) communication frameworks, which enable direct coordination between agents, and agent payments protocols (AP2), which facilitate trusted transactional interactions.

The semantic layer turns data into knowledge. It sits between raw data and AI applications and codifies the business meaning of data into a machine-readable form that humans can understand. Rather than treating data as disconnected tables or files, the semantic layer defines what things are, how they relate, and what rules govern them.

In practice, this layer is most often implemented through ontologies and knowledge graphs. Ontologies define how the attributes and relationships between data add up to business reality. Knowledge graphs operationalize this vocabulary by linking real-world data across systems into a connected network of entities. Without this shared semantic foundation, agents may act on incomplete or conflicting interpretations of the same data, increasing error rates and operational risk as scale grows.

Data products turn curated data into reusable, business-ready assets. They package data with clear ownership, quality standards, semantics, and interfaces for consumption. Through a product mindset, companies treat data as a performance asset that can be reused across multiple use cases and domains. Reusable data products allow agents to draw on trustworthy predictive and generative insights at scale, while observability records how agents use data, creating the traceability needed for oversight and enabling feedback loops that improve upstream data and models.

Data consumption sits atop the entire tech stack. This capability delivers data and intelligence into workflows and applications. It includes analytics and reporting tools, data APIs, retrieval interfaces, and model outputs embedded directly into business processes. Agentic orchestration and retrieval services are embedded in this layer. AI systems dynamically assemble context—often from unstructured data—rather than relying on predefined queries. Orchestration enables AI systems to decide what to retrieve, how to refine it, and when to iterate, while retrieval services ensure this access is secure, efficient, and governed at scale.

Modern architectures analyze model outputs to strengthen the data itself. Gen AI applications can generate labels, usage patterns, and context that improve quality and support future models—and these outputs, just like any other, need to be captured for training.

Governance and access controls provide core components that control how agents interact with data, tools, and models in a controlled way. A medallion architecture progressively curates and enriches data from raw to agent-ready form, while preserving lineage and auditability. Data access is then governed through APIs, queries, and permissions. As models retrieve and use unstructured data dynamically, an AI gateway is needed to control access and usage. The gateway governs model access to unstructured data, enforces usage policies, and records how data is retrieved and used in prompts and responses.

Ensure that data quality is in place

Curated, high-quality data becomes a strategic differentiator in the agentic AI era. Foundation models can be costly to deploy, given the costs of large-scale inference, fine-tuning, infrastructure, and governance. Organizations with well-structured internal data sets can reduce technology investment costs by fine-tuning smaller, domain-specific models on their own data. These models are not just more cost- and resource-efficient, but more resilient and compliant.

Making unstructured data usable requires improving its quality through tagging, classification, vector embeddings, and graph-based structuring. This allows agents to reliably understand entities, relationships, and context. Unstructured data must be held to the same standards as structured data.

Companies must also evolve how they manage structured data. Instead of periodic cleanups, organizations can engage in continuous, real-time data quality monitoring. This process is supported by AI-enabled automated validation, anomaly detection, and enrichment pipelines that prevent issues from propagating across workflows. Metadata management provides lineage and business context so that agents can trace and justify decisions.

Finally, as agents generate new data, organizations must apply the same quality, lineage, and reconciliation standards to their outputs. This includes data retrieved or written through agent-invoked tools and APIs, which should operate through governed, reconcilable interfaces rather than bypassing enterprise quality control. Shared fit-for-purpose definitions embedded into automated quality checks can ensure that agents act on reliable information at scale.

Build an operating and governance model for agentic AI

As agentic systems scale, governance becomes the primary mechanism for control. Clear and explicit policies are needed to define what agents can do, what data they can access, and when human approval is required, with access checks evaluated automatically for each agent based on their role and scope. Importantly, agents should not introduce new data quality or governance rules; they should follow the same standards as other systems, applied automatically as autonomy increases.

Well-developed agents can help with this process, with guardrail agents operating within a well-defined control function to continuously monitor agent activity to [ensure transparent and compliant behavior](#). For example, creative compliance agents can review images and multimedia outputs for brand misrepresentation or policy violations and trigger corrective actions.

IT and governance functions must also manage the agentic life cycle. This requires issuing credentials, tracking activity logs, monitoring performance, and enforcing policy compliance through automated checks. Agent activity is automatically captured through built-in telemetry, ensuring that actions, data access, and decisions are consistently logged and traceable.

Clear accountability for agent behavior—spanning business outcomes, risk management, and policy compliance—is critical to scale. In practice, business domains own day-to-day governance of agent-enabled workflows, including domain models and ontologies. Meanwhile, central data and AI teams maintain shared platforms, guardrails, and oversight. This federated model balances domain autonomy with enterprise-wide accountability

Find more content like this on the
McKinsey Insights App



Scan • Download • Personalize



In the agentic age, technology leaders are finding that their data foundations increasingly define competitive positioning. Yet despite the promise of using data to generate real value from agentic AI, many organizations still struggle to make data accessible and governable for digital agents. The time has come to drive data transformations that pave the way for an agentic future.

Asin Tavakoli is a partner in McKinsey's Düsseldorf office; **Brian Goodman** is a partner in the New York office, where **Kayvaun Rowshankish** is a senior partner and **Akshat Kumar** is an associate partner; **Henning Soller** is a partner in the Frankfurt office; **Carlos Barreto** is a distinguished data engineer in the São Paulo office; **Satyajit Parekh** is a senior knowledge expert in the Boston office; and **Tancredi Bernard Litta Modignani** is a knowledge expert in the Lisbon office.

The authors wish to thank Artur Freitas Gonçalves and Francesco Scorca for their contributions to this article.

This article was edited by Kristi Essick, an executive editor in the Bay Area office.

Core elements of this article are based on the "No data architecture, no AI advantage" chapter from [Rewired: How Leading Companies Win with Technology and AI](#) (Wiley, April 14, 2026), by Eric Lamarre, Kate Smaje, and Rob Levin, with Alex Singla and Alexander Sukharevsky.

Copyright © 2026 McKinsey & Company. All rights reserved.