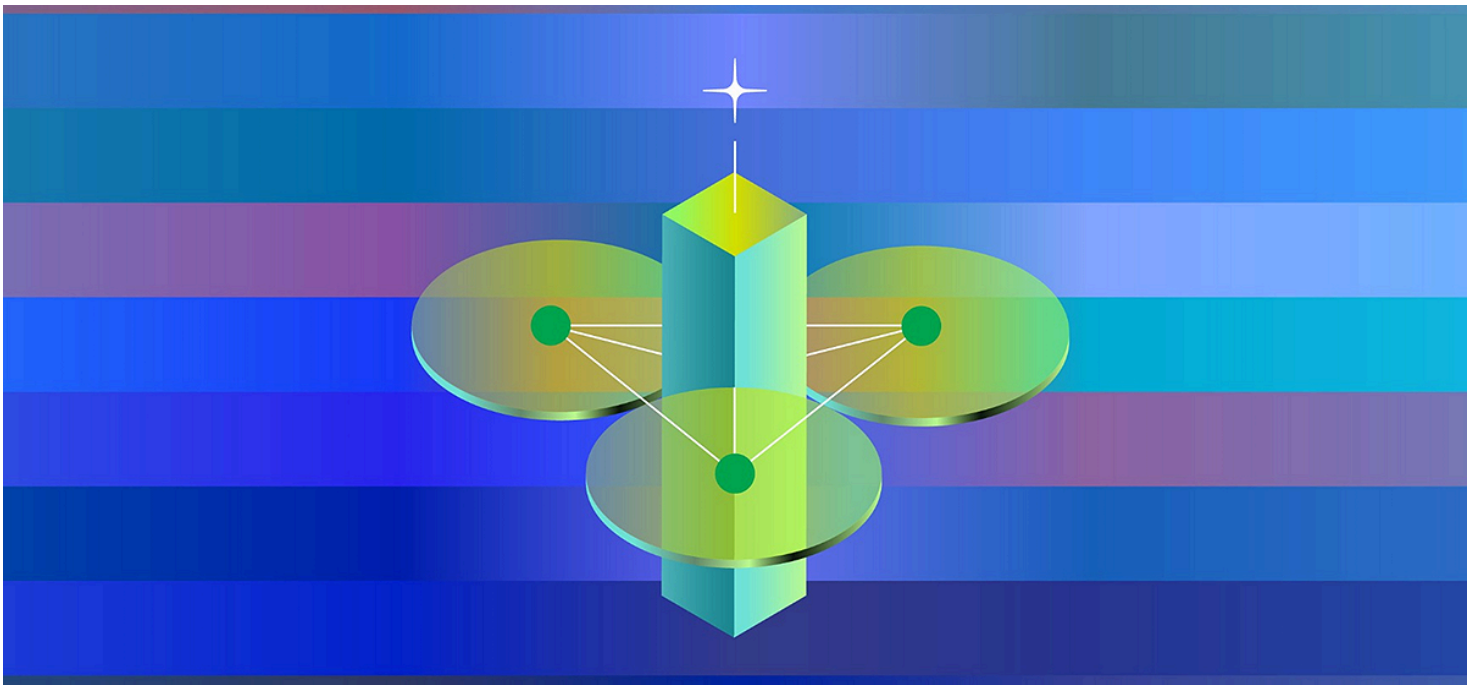
Deloitte. Insights



The next tech infrastructure advantage is intelligence orchestration

As AI spreads across cloud, edge, applications, and devices, organizations should consider a common framework for coordinating how it acts

*By **Chris Thomas**, **Parth Patwari**, **Ram Ravi**, **Oniel Cross**, **Diana Kearns-Manolatos**, **Iram Parveen***

Article • 12-min read • 21 August 2026 • Deloitte Center for Integrated Research

Artificial intelligence is changing what enterprise technology infrastructure needs to be able to do. Simply supporting applications, data, and transactions might no longer be enough, as tech infrastructure increasingly has to coordinate models, agents, workloads, and decisions across the enterprise.

As AI becomes more agentic and the landscape expands to include dozens of frontier, open-source, and derivative models, enterprises may need to govern model selection, routing, and life-cycle management as an ongoing operational discipline rather than a one-time technical choice.[1]

In this technology landscape, organizations could benefit from an enterprise control plane: a governed, intelligent layer that mediates decisions across systems, determining who can act, what data can be accessed, how workloads execute, and how AI agents operate within defined cost, latency, and regulatory boundaries.[2]

The benefit of having this coordinating layer is becoming more urgent as the next five years are likely to be shaped by **two major uncertainties**: whether platforms become more concentrated or federated, and whether interfaces remain human-mediated or agent-mediated. Each uncertainty may introduce entirely different organizing logic. Rather than optimizing for a single expected end state, enterprises will likely need adaptable infrastructure and governance.

Based on more than 30 expert interviews and foresight analysis, our research identifies three capabilities that could help C-suite leaders orchestrate intelligence across distributed systems:

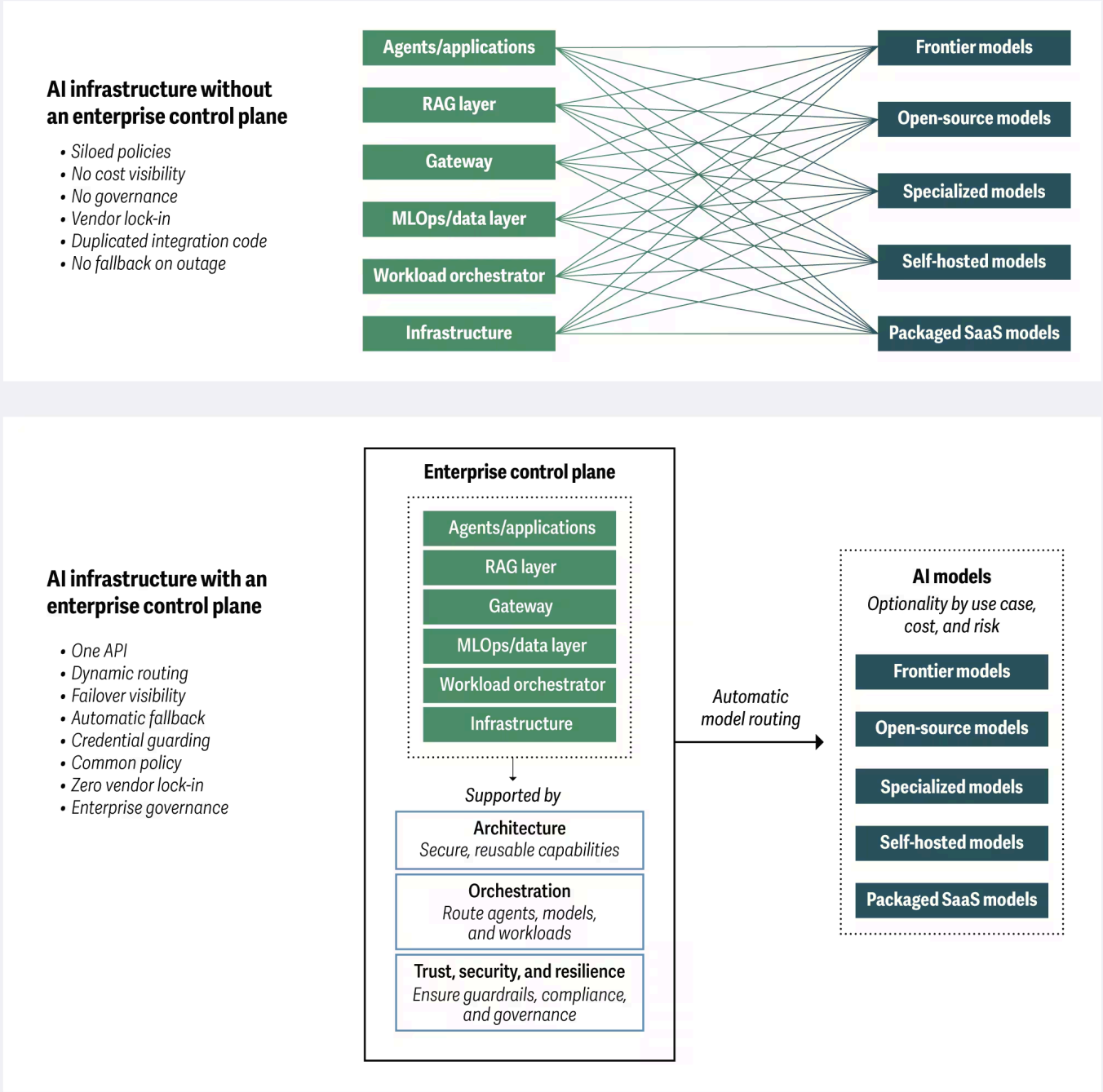
- **Architecture** that exposes enterprise capabilities securely and flexibly
- **Orchestration** that coordinates models, agents, workloads, and decisions enterprisewide
- **Cybersecurity resilience** capabilities that establish identity, authority, integrity, and control across human and machine interactions

These capabilities can form the foundation of a control plane (figure 1). A control plane resembles the service-oriented architecture but is more event-driven: instead of simply exposing reusable services, it should coordinate workflows as conditions evolve.

One challenge for leaders over the next 12 to 24 months might be to begin architecting the controls that can support an AI-native enterprise.

Figure 1

From fragmented AI infrastructure to a governed enterprise control plane



Note: RAG layer = retrieval-augmented generation layer; SaaS = software as a service; API = application programming interface.
Source: Deloitte interviews conducted March to May 2026.

Architecture for an agentic enterprise

Architecture is the foundation that makes a control plane possible. It determines whether enterprise capabilities can be exposed securely and flexibly across channels, agents, platforms, cloud, edge, and on-premises environments or whether they will remain trapped inside systems that are hard to access, govern, and orchestrate.

Many organizations are still modernizing toward a point-in-time target state. But an agentic enterprise needs a horizon architecture[3] that provides a two- to five-year vision tied to business priorities and maturity gaps.

There's no market consensus on what that target state looks like. Some expect hyperscale platforms to consolidate control, while others suggest that a defining constraint—sovereignty, latency, or resilience, for example—should be the organizing principle for the system. As one banking leader told us in an interview, "I believe we are heading toward a world where infrastructure becomes almost invisible to the end user. It's going to be a handful of hyperscalers plus a few AI platform providers. They're going to dominate the control plane."[4]

While the precise future remains uncertain, the architectural requirements are becoming clearer. As customer discovery and enterprise interaction shift toward AI-mediated interfaces,[5] organizations should consider rearchitecting for a resilient, agent-first future. Enterprise capabilities should be accessible through secure service layers; computation should be placed across cloud and edge; and identity and orchestration should be built in, not bolted on.

The following architectural principles can help build the foundation that the enterprise control plane may require.

Start with capabilities, not channels

Enterprises have traditionally designed architecture around the places they expect to interact with users, such as websites, mobile apps, call centers, and other channels. But as agents increasingly become the primary consumers of enterprise services, architecture might need to shift from channel-centric design to capability-centric design, organized around functions such as search, pricing, or identity verification.

This builds on the "headless architecture"[6] principle, in which capabilities are decoupled from any fixed presentation so they can be used across interfaces and autonomous agents can act on them directly. When an agent, rather than a human, becomes the primary consumer of an application programming interface, the definition of good service design changes.

Capabilities should be exposed through secure, auditable, and standardized service layers. They should also be designed for failure, allowing critical functions to continue with downgraded capabilities if necessary.

Choose the architectural paradigm before the platform

As different architectural patterns for AI emerge, leaders should understand the design choices available to them and the trade-offs each may create. Our foresight research points to multiple plausible futures for technology infrastructure, ranging from a concentrated integrated platform ecosystem to an embodied edge. Each implies different choices for architectural orchestration.

Deloitte's client experience suggests that leaders should first choose the architecture paradigm that best supports the business strategy, and only then select platforms. Some broad patterns may help leaders define the best choice:[7]

- **Embedded**: Agentic capabilities are woven into each layer, often through software as a service
- **Dedicated orchestration**: A purpose-built agentic tier that centralizes runtime and governance
- **Hybrid or federated**: A central control plane governs identity, audit, cost, and policy while execution remains distributed across domains

The specific choice will often depend on the organization's strategy, maturity, and constraints. But the guiding principle remains the same: paradigm first, platform second.

Architect around outcomes and economics

Architecture should also move beyond the applications and systems through which work currently moves and reflect the business outcomes AI is expected to produce. An outcome-first architecture can help leaders determine how much of the process requires deterministic, rules-based execution, and where a more flexible, AI-driven approach is needed.[8]

One hyperscaler's AI leader told us that "The real questions for AI system design are: What is the task? How do you define it? Who designs it? And is the workforce ready for that change?"

Cost-aware architecture shouldn't be a runtime-only decision. It should pivot as financial decisions change, with parameters that can

be adjusted. A centralized product management team with integrated workflows can help guide architectural decisions toward a more cost-aware architecture based on the business case.

Balance composability with sovereignty and control

Enterprises should decide how much of the stack should be modular and composable, and where nonnegotiable requirements like sovereignty, latency, or control should anchor the architecture.

The answer to these questions will likely vary. A tier-one US bank executive described a full-stack model where hyperscalers provide computing, managed services, AI models, and developer platforms. A defense technology executive shared a different picture: Sovereign AI may require a separate environment because data may need to remain pinned to a specific country.[9]

Neither approach is universally correct. The architectural choice should follow the organization's operating requirements and risk profile. As one pharma company's director of digital architecture notes: "Multi-cloud only makes sense where there is a strong operational or resilience rationale. Every movement of data creates additional cost and complexity, and engineering overhead."[10]

Control boundaries should be embedded from the start, not retrofitted. At the same time, leaders should design for planned impermanence: replaceable components, explicit re-evaluation, and an architecture capable of change as the environment evolves.[11]

Orchestrating intelligence across the enterprise

Orchestration is the decision layer that determines who and what can act, in which environments, and under what constraints. It translates governed capabilities into coordinated action.

Agents need boundaries and rules governing access, authority, and autonomy, but many current efforts focus instead on connecting individual platforms and protocols. The greater source of operating leverage may be the ability to orchestrate intelligence across the enterprise.

A central component of that capability is model orchestration: an intelligent router that determines, in real time, which model should handle a task, where inferencing should occur, and what cost, latency, and regulatory constraints should govern the decision. As contexts shift, the control plane will need to coordinate a distributed network of agents while accounting for their dependencies and intent.

Agentic systems rely on context engineering, in which the context available to an agent shapes what it can decide and do. As open-source AI[12] and thousands of autonomous or semi-autonomous agents are integrated into enterprise systems, humans won't be able to anticipate every interaction or encode every possible path in advance. The chief of technology operations at an oil exploration company says, "Future systems will behave like cognitive entities, reasoning all the data, areas of planning, workflows—more like organizational agents rather than static tools."[13]

Enterprises will likely need a combination of rules-based and behavioral controls across software, platforms, and infrastructure. These controls allow the control plane to coordinate intelligence from the interface, through the model layer, to the environment where work is executed.

Coordinate software and agent behavior

Laptop, mobile, robot, and drone interfaces all run on software. Physical AI systems may look different from traditional apps, but they require the same kind of end-to-end orchestration. The control plane needs a consistent way to determine what an agent is permitted to do across those environments. Enterprises can begin by tiering agent autonomy according to levels of authority and connecting those tiers to specific controls, monitoring, and rollback paths.[14] This creates a governable system where agents operate within explicit boundaries that can change according to the task, context, and level of risk.

Enable platforms to coordinate dynamically

Many agent orchestration platforms today are hard-coded, with functions linked through rules, workflows, and controls baked into system architecture. That approach becomes a liability as agents act across tools, files, and open networks[15] in ways engineers can't fully anticipate.

Context engineering[16] can help govern those interactions by shaping what agents know and what they can do. Emerging approaches, such as distributed dynamic dependency injection,[17] also point toward systems that allow agents to discover and coordinate by capability rather than through fixed instruction—closer to how human organizations adapt than how traditional software is wired.

Model orchestration is another important platform capability. As the model mix changes, an intelligent routing layer can normalize interactions across different models and determine which one is best suited to a particular task. Model context protocol servers and similar standards can support interoperability without requiring enterprises to rewrite systems for every provider or model-tool combination. Practically speaking, this means one model could hand off a request to another model mid-workflow, without requiring bespoke connectors for every possible handoff.

Route workloads across infrastructure

Orchestration should also extend to the infrastructure layer. As AI systems become more interdependent, enterprises need a digital gateway that can route workloads dynamically based on cost, latency, security, and data sensitivity.

This is the strategic nervous system of the enterprise control plane, coordinating AI agents, models, policies, identity, governance, and workloads across the tech stack. The difference is a shift from raw computing to the ability to orchestrate intelligence securely and at scale.

That shift means moving from static hosting decisions to policy-based workload routing. Infrastructure should decide, in real time, where inference is best executed. A latency-sensitive workload in a manufacturing or hospital environment, for example, may need to run at the edge even when cloud execution is less expensive. The control plane applies the enterprise's priorities and constraints to decision-making, and the result is a coordinated intelligence system in which software, platforms, and infrastructure can adapt together. Over time, this may extend beyond a single enterprise to an ecosystem in which multiple organizations need shared standards for trust, cost, compliance, and payments.

Building trust, security, and resilience into the control plane

Cybersecurity is the trust layer that makes the control plane safe to operate. It should verify the integrity of interactions, so architecture and orchestration can act with confidence rather than relying on assumed trust.

Many enterprises are still modernizing infrastructure based on yesterday's security model. Risk management often remains a separate function,[18] even as agentic systems create new attack surfaces through prompt injection, data poisoning, hallucinations, and adversarial attacks.[19]

The control plane therefore needs security and risk controls that operate continuously rather than only at the point of deployment. It should be able to identify both human and machine identities, enforce runtime guardrails, validate outputs and provenance, isolate compromised paths, and preserve auditable records of who acted, on whose authority, and under what conditions.

Without that foundation, organizations risk embedding weak trust boundaries and brittle controls into the systems they're building for the future. The following priorities can help leaders establish the trust, security, and resilience the enterprise control plane requires.

Extend identity and authority beyond human users

An agentic enterprise should have a next-generation identity fabric that extends identity and access management to agents, devices, workloads, and services.[20] That identity model should be anchored in cryptographic identity, hardware roots of trust, and explicit authority controls, so every actor is traceable, but none is implicitly trusted. Agents should not receive default entitlements or automatically inherit the full identity context of a human user. Access should be explicitly delegated, limited to the task, and bounded by time, data, authority, and risk.[21]

Authentication and authorization should also remain distinct. The control plane should attach verification to the moment of action and preserve the reasoning behind each request, making intent part of the verifiable record. One hyperscaler's AI and data leader notes, "Right now, agents or bots are outside the authority authentication, and that's where organizations are trying their best to put the guardrails in place."[22]

Blockchain is one option for a sovereign identity control plane; centralized policy with distributed enforcement is another. A services registry or gateway can resolve identity, authority, and revocation consistently across a distributed control plane without forcing every action through a centralized choke point.

Modernize the cryptographic foundation

As quantum computing advances, some of today's cryptography (specifically, asymmetric cryptography) will likely become easier to break, creating a significant risk. Adversaries could store encrypted data now and try to decode it later when quantum capabilities arrive. At the same time, always-on agentic systems are creating more vulnerable connections that are also at risk. For example, recent security testing[23] showed that an advanced AI model was able

to break a weakened version of a widely used encryption method, highlighting how AI could accelerate the speed and sophistication of online attacks.

This is beyond a technical issue for the security team. US technology standards bodies recommend that organizations begin moving toward quantum-resistant cryptography now to protect information technology systems from future attacks.[24] A recent US executive order set a deadline for federal agencies and contractors to adopt quantum-resistant cryptography by 2030 to 2031.[25]

These updates to quantum-resistant technology should be treated as part of the build-now foundation for digital trust.[26] Organizations should map where asymmetric encryption is currently used, focusing first on the most critical data and transactions, and upgrade their system in phases. The transition will likely require board oversight and cross-functional coordination, rather than implementing it as a narrow security team initiative.[27]

Establish AI-specific runtime controls

A perimeter-only security model[28] is likely not sufficient for systems that act continuously across models, tools, networks, and edge locations. Enterprises need an adaptive security architecture that can respond dynamically and faster than an attacker.

This is the next evolution of zero trust (a security approach that assumes no trust by default and requires verification before access is granted). Runtime controls should separate traffic by use case, isolate higher-risk paths, and design the environment so that one compromised edge can be pulled without disrupting core web activity.

As enterprises route sensitive queries through multiple model providers, each handoff increases exposure. A model gateway can strengthen runtime control by enforcing consistent guardrails and logging every model interaction. One US bank executive says, “It’s moving from ‘Let’s build a better wall’ to ‘Let’s detect, decide, and respond faster than the attacker.’”

Govern AI output integrity and protect human judgment

Trust should extend beyond access and model performance to output integrity and the decisions that follow from it. The control plane should be able to verify provenance, continuously validate outputs, and detect behavioral anomalies across models and agents.[29] This can help the control plane identify inaccurate, manipulated, or out-of-policy results before they influence a human decision[30] or customer interaction. These controls can reduce the risk of automation bias, keep human judgment central where it matters most, and prevent bad outputs from being scaled across the enterprise.

From infrastructure to intelligence: Moves that matter

The challenge now goes beyond simply modernizing infrastructure and extends to building the architecture, orchestration, and trust foundations required to coordinate intelligence across a distributed system. Market advantage will likely come from a governed, intelligent system that decides which model runs which task and under what constraints, and that enforces security at every step with modernized permissions and cryptography. Investing now in a control plane that can adapt as technologies, interfaces, business needs, and risks evolve can help enterprises preserve the flexibility to respond to whatever comes next and avoid locking themselves into today’s platforms. The priority for the next 12 to 24 months should not be completing the transition to an agentic enterprise, but rather architecting a stronger foundation for competitive advantage.

Continue the conversation

Meet the industry leaders



Chris Thomas
Principal, Hybrid cloud infrastructure offering leader | Deloitte Consulting LLP



Parth Patwari
US AI and engineering leader | Deloitte Consulting LLP



Ram Ravi
Managing director, Strategy | Deloitte Consulting LLP



Oniel Cross
Principal | Government and public sector hybrid cloud infrastructure offering leader | Deloitte Consulting LLP



Diana Kearns-Manolatos
Senior manager, Subject matter specialist | Deloitte Services LP

BY

Chris Thomas
Global

Parth Patwari
United States

Ram Ravi
United States

Oniel Cross
United States

Diana Kearns-Manolatos
Global

Iram Parveen
India

ENDNOTES

1. Harsha Kotikela, "Agentic AI at scale can break your infrastructure before it transforms your business," *Forbes*, July 1, 2026.
2. Andy Packham, "Control planes: A strategic lens for enterprise platform decisions," HCL Tech, April 21, 2026.
3. Ram Ravi and Jagjeet Gill, "Horizon architecture: Enterprise architecture strategy," Deloitte, July 1, 2026.
4. Deloitte interview conducted between March and May 2026.
5. Siva Muthu, Anantha Ramadas, Jayanta Ghosh, and Harshad Deshpande, "Preparing digital channels for agentic AI evolution," Deloitte, May 5, 2026.
6. Beenu Ji, "Salesforce headless 360: The API-first future of CRM and AI agents," Salesforce, May 10, 2026.
7. Ram Ravi et al., "Architecting the agentic enterprise," Deloitte, 2026.

8. Deloitte interview conducted between March and May 2026.

9. Ibid.

10. Ibid.

11. Ravi et al., "Architecting the agentic enterprise."

12. Katherine Noyes, "**Open-source AI agents: What happens when personal bots go to work?**" Deloitte and *The Wall Street Journal*, May 30, 2026.

13. Deloitte interview conducted between March and May 2026.

14. Chirag Agrawal, "**AI governance strategy: A 12-month plan to build responsible, trusted AI autonomy**," *CDO Magazine*, May 14, 2026.

15. Noyes, "**Open-source AI agents**."

16. Pedro Lopez, "**What is AI agent context management?**" AirByte, March 10, 2026.

17. MCP Mesh, "**Dependency injection**," accessed Aug. 14, 2026.

18. Emily Mossburg et al., "**The Global Future of Cyber Survey, 5th edition**," Deloitte, May 2026.

19. Kieran Norton, Tim Li, Tim Davis, Emily Mossburg, Diana Kearns-Manolatos, and Saurabh Bansode, "**How can tech leaders manage emerging generative AI risks today while keeping the future in mind?**" *Deloitte Insights*, Feb. 20, 2025.

20. Abhishek A. Hemrajani, "**What's new in IAM: Security, governance, and runtime defense**," Google Cloud, May 7, 2026.

21. Shuva Jyoti Kar, "**The temporal identity paradox: Architecting zero-trust for long-running data agents on ephemeral compute**," Medium, June 27, 2026.

22. Deloitte interview conducted between March and May 2026.

23. Dustin Volz, "**Anthropic AI model finds flaws in tough-to-crack encryption algorithms**," *The New York Times*, July 28, 2026.

24. Deloitte interview conducted between March and May 2026.

25. Richard L. Wells, "**Quantum computing stocks drop 35% as federal mandate reshapes investment case**," MSN, June 17, 2026.

26. Deloitte interview conducted between March and May 2026.

27. Isobel Markham, "**Quantum computing threat elevates cryptography to board-level risk oversight**," Deloitte and *The Wall Street Journal*, June 10, 2026.

28. Microsoft, "**Microsoft digital defense report 2025**," October 2025.

29. James Delbridge, Val Srinivas, and Shivalik Srivastav, "**How AI-native banking products could reshape institutional banking**," *Deloitte Insights*, May 20, 2026.

30. Shaina Raza, Ranjan Sapkota, Manoj Karkee, and Christos Emmanouilidis, "**TRiSM for agentic AI: A review of trust, risk, and security management in LLM-based agentic multi-agent systems**," *AI Open* 7, February 2026.

ACKNOWLEDGMENTS

We would like to thank Deloitte's subject matter experts—**Adnan Amjad**, **Arpan Tiwari**, **Colin Soutar**, **Dhyan Raj**, **Duncan Stewart**, **Faruk Muratovic**, **Gopal Srinivasan**, **Jason Chmiel**, **Prakul Sharma**, **Scott Buccholz**, **Tim Davis**, and **Vinit Shah**—for taking the time to speak with us and for sharing their valuable insights, which helped shape this research.

We are also grateful to the marketing and PR team—**Andrew Ashenfelter**, **Anushka Bose**, **Christian Parsons**, **Cindy Chang**, **Jen Reid**, **Kaneez Fizza**, **Matt Merwrill**, **Nicole Bostock**, **Rachel Freya Rosenberg**, **Rebecca Lalez**, **Saurabh Rijhwani**, **Tafline Laylin**, and **Winslow Sowards**—for their guidance and leadership in extending the impact of these insights.

Finally, we appreciate the support of our research partners at 10EQS for collating input from market leaders and providing data analysis and synthesis that were instrumental to this work.

Editorial (including production and copyediting): **Corrie Commisso**, **Elisabeth Sullivan**, **Shyamili M**, **Pubali Dey**, and **Anu Augustine**

Design: **Molly Piersol** and **Sonya Vasilieff**

Cover image by: **Jim Slatton** and **Sonya Vasilieff**

Knowledge services: **Vanapalli Viswa Teja**

COPYRIGHT

Copyright © 2026 Deloitte Development LLC. All rights reserved. Member of Deloitte Touche Tohmatsu Limited

About Deloitte Insights

Deloitte Insights publishes original articles, reports and periodicals that provide insights for businesses, the public sector and NGOs. Our goal is to draw upon research and experience from throughout our professional services organization, and that of coauthors in academia and business, to advance the conversation on a broad spectrum of topics of interest to executives and government leaders.

Deloitte Insights is an imprint of Deloitte Development LLC.

About this publication

This publication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or its and their affiliates are, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your finances or your business. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser. None of Deloitte Touche Tohmatsu Limited, its member firms, or its and their respective affiliates shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one of more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulation of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms.