



AI AGENTS

Scaling AI Requires New Processes, Not Just New Tools

By [Eric Jesse](#), [Zeeshan Shah](#), and [Rajeev Singh](#)

ARTICLE JANUARY 20, 2026 8 MIN READ

Two decades of automation technologies, from robotics to guided user interfaces to machine language, have enabled companies to achieve significant cost savings by reducing the amount of arduous human labor required. Freed from inefficient tasks, teams have been able to refocus on more value-added work that demands judgment, higher-level reasoning, or interpersonal relationships.

Now companies are turning their attention to the customer experience. The emergence of semi-autonomous and autonomous AI agents now promises companies a step change in productivity and value creation by reducing costs and increasing revenues. But seizing this opportunity requires a broader and bolder mindset around design and implementation. Instead of seeking incremental gains by automating discrete steps within a process, organizations should redesign entire processes end-to-end, with AI agents assuming different roles along the way. Some of these agents will execute purely transactional work—rule-based, high volume, low variance—on their own. Other agents will provide an unprecedented level of support to human teams, who will still manage complex, high-stakes relationships.

How It Works in Practice

One industrial goods company recently redesigned its quote-to-order process to improve efficiency and boost revenue by deploying AI agents end-to-end. By standardizing processes and linking discrete systems, it is reducing labor costs by between 30% and 40%. At the same time, improved quote turnaround time and gains from unmanaged requests for quote (RFQs) are generating tens of millions of dollars in additional revenue.

The new process, designed to accommodate local needs across a complex global footprint, relies on four AI agents (see Exhibit 1):

- **Assessment and Classification.** Automates front-end intake by evaluating and sorting inbound requests. It classifies emails, assists quoting, checks configurations, and suggests products.
- **Recording.** Streamlines order entry and processing across systems by booking orders, supporting RFQ changes, enabling direct shipping, and improving pricing and escalations.
- **Status.** Enhances visibility and customer communication by automating acknowledgments, enabling self-service, and boosting quote conversions.
- **Lead-Time Generation.** Supports both quoting and order fulfillment timelines by delivering accurate lead times using unified planning data.

EXHIBIT 1

A Redesigned, Seamless, End-to-End Quote Process, Augmented by Four Specialized AI Agents



Source: BCG experience.

Note: RFQ = request for quote.

The company has designed this process to have AI agents resolve around 70% of RFQs without human intervention. Those RFQs involve smaller transaction values and simpler products with established engineering specifications. Around 20% of RFQs would require some human intervention in collaboration with AI agents. The remaining 10% would comprise the most sophisticated or complex transactions and require intensive human intervention with the support of AI agents.

To develop and plan the implementation of this multi-agent approach, the company first assessed whether its systems and data were sufficient to provide the agents the necessary support context. It is implementing the new process in two releases over 15 to 18 months to allow for change management efforts and to ensure it is ready to adopt the structural changes. It is also setting up its operating model to sustain the AI agent platform, including the formation of agent-based solution teams with cross-function platform capabilities.

Four Key Decisions for an End-to-End Process Transformation

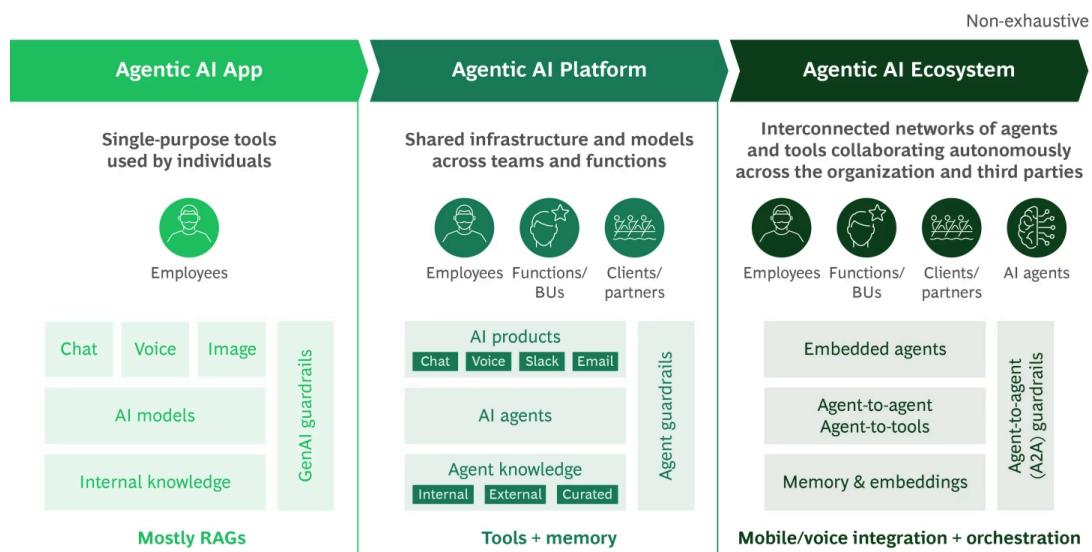
Leaders will need to make several business and technology decisions as they redesign processes end-to-end and embed AI agents.

Platform Versus Product. Leaders must decide whether to adopt a centralized infrastructure (platform) or use decentralized agents (products). A robust agentic platform, owned by platform teams, provides a shared infrastructure for memory, orchestration, tool registries, and governance. Such a platform enables AI agents to work seamlessly together across functions and business units. AI agent products, meanwhile, focus on delivering targeted capabilities and outcomes. This separation of the agents, which can be owned by a business unit with some support from a cross-functional IT team, ensures scalability, reusability, and value-driven execution at speed.

AI Agents Across or Within Business Units. Similar to previous waves of automation, companies began the deployment of single AI agents to solve a specific issue or automate a single step. Many organizations are now progressing to AI tools connected across a shared infrastructure orchestrated by an AI agent within defined guardrails. The final stage of this progression features multiple networks of agents collaborating in an ecosystem that can facilitate processes across the organization through agent-to-agent interactions. The organization needs to decide which deployment is the optimal fit. (See Exhibit 2.)

EXHIBIT 2

The Deployment of AI Agents Is Evolving from Apps into Ecosystems



Source: BCG analysis and experience.
Note: RAG = retrieval-augmented generation.

Serverless Versus Client-Server. Architects and product leads should weigh trade-offs in scalability, latency, integration, and operations to determine which approach best supports the organization's transformation objectives.

A serverless-native approach, such as Bedrock from Amazon Web Services, treats AI agents as on-demand services with no infrastructure to manage. These agents can span diverse systems with greater flexibility and can automatically scale to accommodate unpredictable spikes in workload without manual provisioning. The trade-off is that serverless agents may face latency and execution constraints.

The client-server approach sets up AI agents in a cloud vendor environment. This gives the company more flexibility and control over the governance of the AI agents, including workloads and performance. However, scaling and extending agents may require careful capacity planning or reliance on the vendor's infrastructure. Integrating beyond the native stack can add complexity.

Build Versus Buy. With an understanding of your requirements for the choices above, the final decision is whether to build your own agentic platform or to utilize and deploy a vendor platform such as Sem4AI or n8n. Exhibit 3 summarizes the key criteria for this decision.

EXHIBIT 3

The Build or Buy Decision Depends on Which Criteria Matter Most to Your Organization

Decision criteria	Build versus buy	Notes
Cost (upfront)	●	Prebuilt AI platform cost less to get started quickly
Cost (ongoing)	●	Well-architected in-house platforms may have half the total cost of ownership versus recurring costs per agent
Control and customization	●	Vendor offerings would need extensive customization for highly specialized or regulated processes such as financial services, health care, or defense
Ability to innovate	●	Buying allows rapid experimentation for early-stage use cases or organizations just starting with GenAI; this can reveal clear focus areas.
Strong in-house capabilities	●	Building is more viable and efficient if you already have strong engineering capabilities with AI, data platforms, MLOps, and engineering practices
Speed to deploy	●	Vendor platforms offering prebuilt AI agent capabilities can accelerate deployment
Speed to scale	●	Organizations have more flexibility and control on scaling with a build solution, assuming that process changes have been put in place to support it
Risk of obsolescence	●	Technology moves fast; the right level of resourcing and a sound roadmap are key, whether build or buy

● Advantage for build ● Even or independent ● Advantage for buy

Source: BCG experience.

Note: MLOps = machine learning operations.

The decision to build or buy will have a significant impact on the economics of the transformation. Leaders should weigh the short-term speed and convenience of buying against the long-term scalability and cost efficiency of building, especially as agent usage grows across the organization. Vendor platforms offering prebuilt AI agent capabilities can accelerate deployment but have higher annual run costs. Factoring in licenses and implementation services, those costs can reach up to \$1.5 million per use case or function. That's around three times an in-house platform's typical annual run costs, which are driven primarily by token usage from foundational models, cloud infrastructure, and internal engineering talent.

Regardless of the decision, leaders must also consider the risk of rapid obsolescence as AI technologies evolve. This means designing AI platforms as plug-and-replace systems so that the organization can swap out core components like LLMs, memory modules, orchestration layers, and tool registries as they become outdated. Leaders can consider tasking platform teams with scanning emerging components to ensure that they consider and adopt the latest paradigms—such as Model Context Protocol and agent-to-agent (A2A) protocols—in their releases.

How to Meet the Change Management Challenge

Through many AI transformations across sectors, BCG has established a guiding principle of 10/20/70 for resource allocation. That is, companies should devote 10% of their efforts to algorithms and 20% to technology and data; the remaining 70% of their efforts should focus on people and processes to make sure that the changes stick.

Within processes, the first step is to open up the thinking around all value-creating steps to find the optimal end-to-end process design. A well-designed agentic process should significantly reduce the number of checks because it eliminates the human uncertainty that prompts questions such as “Did I hear that correctly?” or “Did I forget something?” It can also make reviews definitive and final instead of iterative, this reducing review time from days to minutes.

The organizational impact will likely be far-reaching. We anticipate significantly fewer frontline employees and a corresponding reduction in the management layers. This will lead to a reworking of spans of control. Managers will have smaller teams focused on higher-level, higher-value tasks where humans still excel, augmented by AI tools they know how to use. The required skill sets will demand greater fluency in technology, with consequences for learning and development programs.

Two Prerequisites for an Agentic AI Transformation

Data readiness and the right team structure are prerequisites for making an end-to-end process transformation succeed. One common misconception is that an organization must wait until enterprise data is fully clean, structured, and integrated. In our experience, waiting for perfect data often leads to unnecessary delays. The latest models and AI agents, especially those using retrieval-augmented generation (RAG) and external tool APIs, can work effectively with semi-structured, decentralized, and even messy data. Adopting a “build with what’s good enough” mindset lets AI usage drive data maturity, not the other way around.

Whether the organization builds or buys its solution, it should create platform and product teams with clearly defined responsibilities across AI agents. The platform team owns the shared modular infrastructure, which includes LLM orchestration, memory services, tool registries, agent evaluation, governance, and observability. The optimal solution is to have common teams that scale across business units and enable multiple agents. But discrete platforms teams may be necessary to meet regulatory demands or other needs of a business unit.

The product teams focus on designing and iterating AI agents that solve domain-specific problems. They embed agents into processes such as quote-to-order. These teams should include AI product managers, user experience designers for human-agent interaction, and business process owners who can frame outcomes and not just features. You can have a single product team managing all your agents or set up multiple product teams, depending on number of processes having agentic support.

Deploying AI agents at scale is an ambitious and present opportunity. The technology is proven, the models are evolving rapidly, and the window for first-mover advantage is open. Now is the time for leaders to shift from piloting agents to redesigning the work, not just the tools.

Authors



Eric Jesse

Managing Director & Partner
Denver



Zeeshan Shah

Platinion Director
New York



Rajeev Singh

Platinion Managing Director
Dallas



ABOUT BOSTON CONSULTING GROUP

Boston Consulting Group partners with leaders in business and society to tackle their most important challenges and capture their greatest opportunities. BCG was the pioneer in business strategy when it was founded in 1963. Today, we work closely with clients to embrace a transformational approach aimed at benefiting all stakeholders—empowering organizations to grow, build sustainable competitive advantage, and drive positive societal impact.

Our diverse, global teams bring deep industry and functional expertise and a range of perspectives that question the status quo and spark change. BCG delivers solutions through leading-edge management consulting, technology and design, and corporate and digital ventures. We work in a uniquely collaborative model across the firm and throughout all levels of the client organization, fueled by the goal of helping our clients thrive and enabling them to make the world a better place.

© Boston Consulting Group 2026. All rights reserved.

For information or permission to reprint, please contact BCG at permissions@bcg.com. To find the latest BCG content and register to receive e-alerts on this topic or others, please visit bcg.com. Follow Boston Consulting Group on [Facebook](#) and [X \(formerly Twitter\)](#).