

Delegating Outcomes

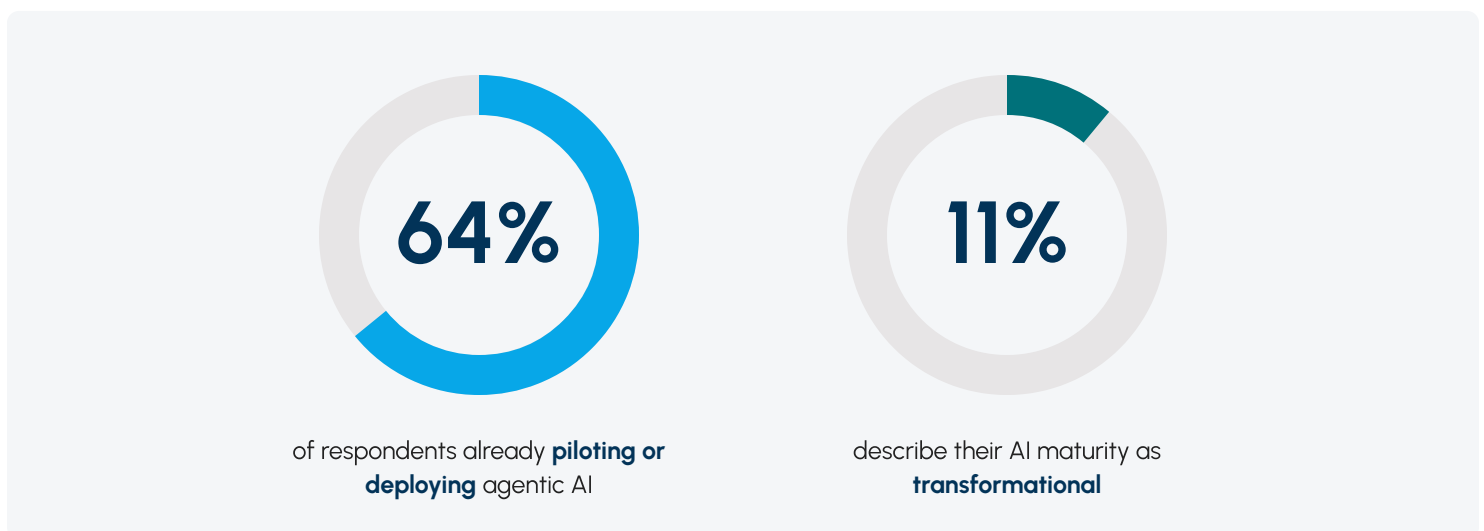
How a Universal Agent Changes the Way
Enterprises Work

The shift from assistance to delegation

For most of the past three years, enterprise use of generative AI has followed a familiar pattern: an employee asks a question, the model answers, and the employee does the rest. That pattern delivered real productivity gains, although it left the hardest part of most jobs, coordinating information and actions across many systems and people, firmly with the human. The next phase moves the unit of work from the prompt to the outcome, as employees assign an objective to an agent that plans the task, uses the necessary tools and data, and returns the finished work.

Futurum's IH 2026 AI Decision Maker Survey (n=838) suggests organizations are ready for that shift, with 64% of respondents already piloting or deploying agentic AI (see Figure 1).¹ However, only 11% describe their AI maturity as transformational, highlighting the problem most enterprises now face. For CxOs, IT leaders, knowledge workers and developers, the question has shifted to how to deploy agents in a way that scales, remains governed and pays its way.

Figure 1. Agentic AI Adoption and AI Maturity



Source: Futurum Research, IH 2026 AI Decision Maker Survey (n=838)

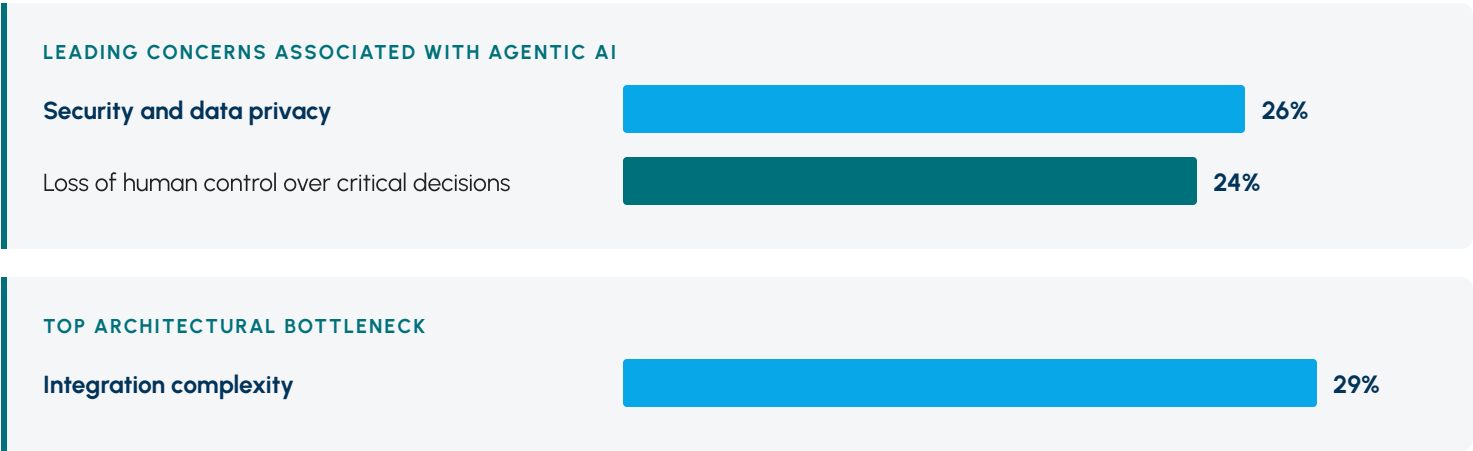
Why first-generation agent programs hit a ceiling

The first wave of enterprise agents generated considerable activity and productivity gains, although much of it was fragmented. Departments built single-purpose agents on different platforms, each with its own context, permissions, and model, and employees spent time re-briefing tools that could not see each other's work. Some organizations now run thousands of custom agents, which shows the appetite for automation and the management burden that comes with it.

1. Futurum Research, IH 2026 AI Decision Maker Survey, n=838 enterprise AI decision-makers.

As these programs grow, governance tends to slow them first. Security and data privacy are the leading concerns enterprise leaders associate with agentic AI in Futurum's survey, cited by 26%, followed closely by loss of human control over critical decisions at 24% (see Figure 2).¹ Integration complexity is the top architectural bottleneck at 29%, as agents confined to a single system cannot automate the cross-application work that justifies the investment. Costs climb as token volumes grow faster than per-token prices fall, and model churn adds rework, since the leading model changes every few months and workflows hard-wired to one model must be rebuilt each time.

Figure 2. Leading Agentic AI Concerns and Architectural Bottleneck



Source: Futurum Research, 1H 2026 AI Decision Maker Survey (n=838)

What a universal agent for work has to get right

We see five requirements that separate an agent platform capable of carrying real enterprise work from a collection of assistants.



01

Persistent context.

An agent that forgets the project between sessions, or between a laptop and a phone, pushes the coordination burden back onto the user. Long-running work needs memory that survives across devices, channels, and days.



02

Presence where work happens.

Employees will not adopt a tool that forces them out of their inbox, documents, and chat. The agent has to work within those applications, including those from other vendors, and via a command line for technical users.



03

Orchestration.

Complex objectives break down into parallel and sequential steps, so the agent needs to coordinate specialist sub-agents and, in some cases, hold a persistent role within a team.



04

Model choice.

Matching the model to the task raises accuracy on hard jobs and lowers cost on simple ones. Futurum data show that 71% of organizations use cloud APIs for inference and 51% run it on their own infrastructure, often both at once (see Figure 3),¹ so an agent tied to a single model family fits poorly, and keeping the choice open protects investment in context and skills as the leading model changes.

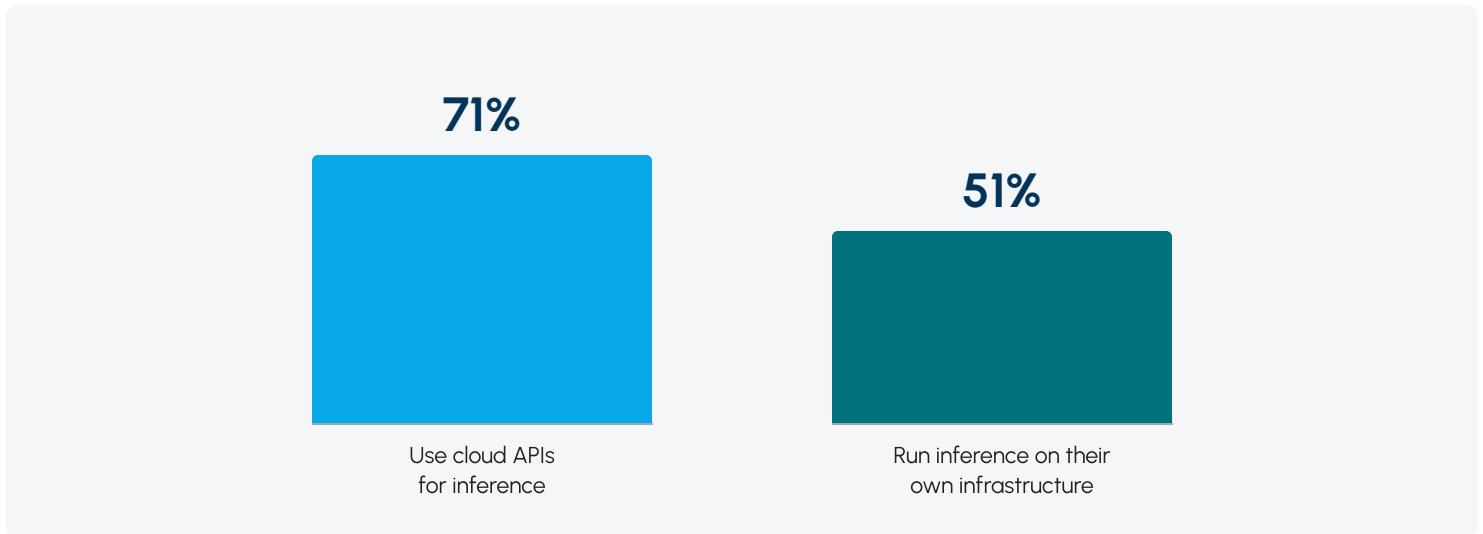


05

Governance and cost control by design.

Identity, permissions, auditing, and spend limits have to be built in from the start, because retrofitting them across hundreds of agents rarely works.

Figure 3. Where Organizations Run AI Inference



Source: Futurum Research, IH 2026 AI Decision Maker Survey (n=838); multiple responses allowed

What changes for each role

CxOs.

The strategic question shifts from tool adoption to operating model, as leaders decide which roles and processes to redesign around delegation and how to measure outcomes rather than usage.

IT decision makers.

One agent with a single identity model, policy layer, and audit trail is far easier to govern than dozens of disconnected assistants, and an agent that works across a mixed estate avoids a wholesale platform switch.

Knowledge workers.

Routine coordination, such as scheduling, assembling reports, or turning a request into a slide deck, can be handed off entirely, leaving more time for judgment, relationships, and strategic decision-making.

Developers.

A single agent and API spanning chat, autonomous work and code, together with shared registries of tools and skills, lets builders package departmental expertise once and reuse it across the company.

How Google approaches the universal agent

Google has been building toward this model for some time. When it launched Gemini Enterprise in October 2025, it positioned the platform as a single front door to AI at work, combining its models, a no-code agent builder, pre-built agents, connectors, and central governance.² That gave employees one place to start, although they still chose and assembled the agents themselves. The Gemini agent takes the next step by doing that assembly on the user's behalf, accepting an objective and returning finished work.³

Measured against the five requirements, the design holds up well. The agent maintains persistent memory in the cloud and operates within Google Workspace, Microsoft 365, Confluence, and Slack, so long-running jobs continue across devices and channels, including desktop, web, mobile, CLI, and browser. It also integrates with third-party development tools such as Git and Jira; enterprise platforms like Salesforce and ServiceNow; databases like BigQuery, Databricks, Postgres, and Snowflake; and files on the employee's desktop.

The Gemini agent can orchestrate a roster of sub-agents or takes a standing role as a coworker agent with its own identity, routes work across the Gemini family and Anthropic's Claude models, and connects to enterprise systems through tools, reusable skills and the Model Context Protocol (MCP). Governance is built in at the level of each agent, with attested identity, least-privilege access, an isolated sandbox, and policy enforced across all agent traffic, while spend caps and model routing keep costs in view.

Early customers report material gains. Commerzbank has cut manual document review from 20 hours to one, Nokia is utilizing an ecosystem of specialized agents to tackle the complexity of today's modern telecom networks – reducing issue resolution time by up to 80%, and DBS Bank runs chains of specialized agents, with humans kept in the loop, to prepare corporate credit memos.³

20 hrs → 1 hr

Commerzbank: manual document review

Up to 80%

Nokia: reduction in issue resolution time



ANALYST PERSPECTIVE

Google has put forward the most complete enterprise agent design we have seen from a hyperscaler, with identity, policy, cost control, memory, and model choice designed in from the start. Treating the model underneath the agent as a separate, swappable choice, and treating coworker agents as governed, non-human identities, maps directly to the five requirements for a universal agent. The next test is how far organizations extend it from bounded, human-supervised workflows into the long-running, cross-application work that delegation promises.

2. Google Cloud, Gemini Enterprise launch announcement, October 2025.

3. Google Cloud, Gemini agent announcement, Gemini at Work 2026; customer results as reported by Google.

Where to start

Organizations moving from assistance to delegation should consider four early steps.

- 1 Pick two or three cross-application workflows, such as credit memos, contract review, or executive reporting, where delegation has a measurable outcome.
- 2 Register business definitions and key data sets in a shared catalog before scaling, since grounding drives accuracy, and hallucination remains the top generative AI reservation for a quarter of buyers in Futurum research.⁴
- 3 Set identity, policy, and spend-cap guardrails centrally from the outset rather than retrofitting them.
- 4 Stand up a skills registry early so departmental expertise becomes reusable across the company.

The organizations that gain most from agentic AI will be those that treat it as a change in how work is assigned and governed, with the platform choice following from that.

4. Futurum Research, *IH 2026 Data Intelligence, Analytics, & Infrastructure Market Sizing & Five-Year Forecast*.

About This Paper

AUTHOR

Nick Patience

Vice President & Practice Lead, AI Platforms | The Futurum Group

PUBLISHER

Futurum Research

INQUIRIES

Contact us if you would like to discuss this report and The Futurum Group will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in context, displaying author's name, author's title, and "The Futurum Group." Non-press and non-analysts must receive prior written permission by The Futurum Group for any citations.

LICENSING

This document, including any supporting materials, is owned by The Futurum Group. This publication may not be reproduced, distributed, or shared in any form without the prior written permission of The Futurum Group.

DISCLOSURES

The Futurum Group provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.



ABOUT GOOGLE CLOUD

Google Cloud offers a powerful, optimized AI stack—including AI infrastructure, leading models like Gemini, data management capabilities, multicloud security solutions, developer tools and platform, as well as agents and applications—that enables organizations to transform their business for the Agentic Era. Customers in more than 200 countries and territories turn to Google Cloud as their trusted technology partner.



ABOUT THE FUTURUM GROUP

The Futurum Group is an independent research, analysis, and advisory firm, focused on digital innovation and market-disrupting technologies and trends. Every day our analysts, researchers, and advisors help business leaders from around the world anticipate tectonic shifts in their industries and leverage disruptive innovation to either gain or maintain a competitive advantage in their markets.

© 2026 The Futurum Group. All rights reserved.