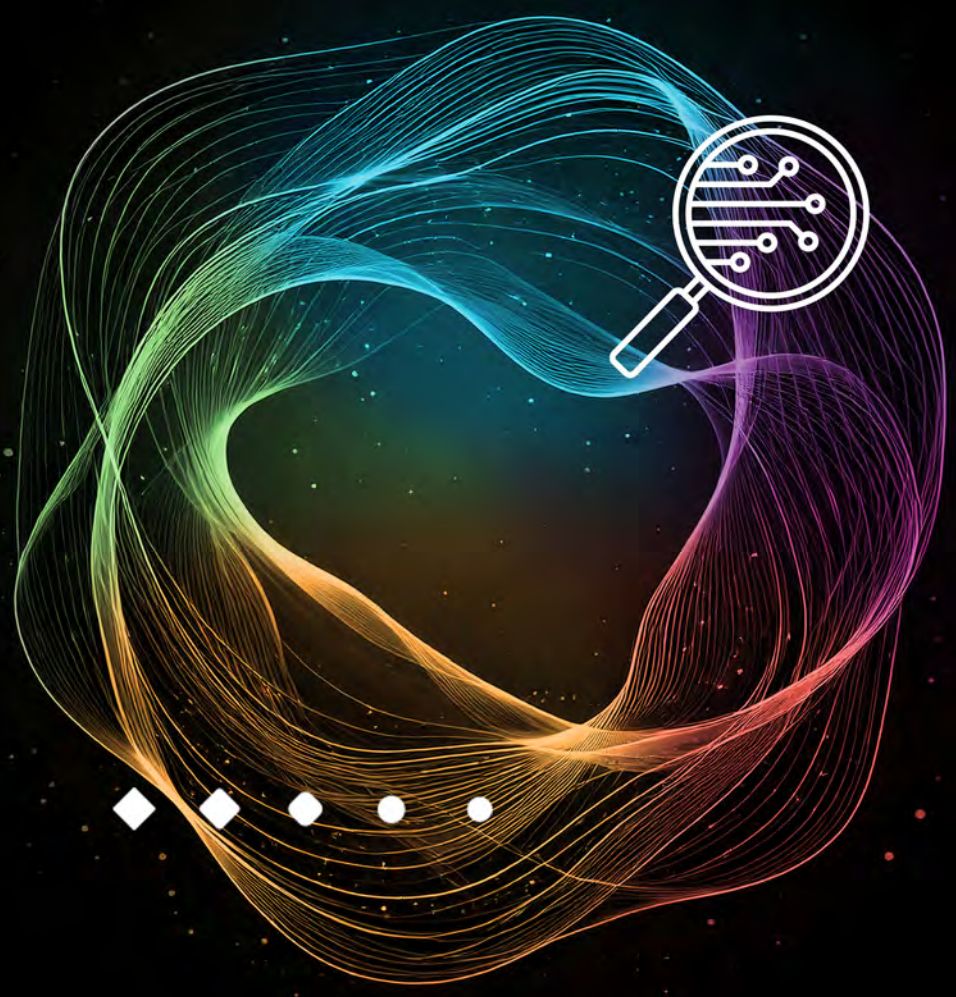
Deloitte.



The measured leap

Appraising AI agent impact with agent operations

December 2025

Content

Key takeaways

- AI agents can enable enterprises to shift human roles from execution to oversight, opening opportunities for improved productivity.
- New KPI frameworks and solutions will be needed to help appraise the performance and impact of AI agents while safeguarding against new and emergent risks.
- Business process decomposition—whereby existing processes are broken down into discrete tasks—can help identify not only where agents may provide value, but also the metrics and solutions needed to monitor, analyze and optimize their performance.
- An adaptable AI agent operations capability can help streamline AI agent implementation and observability across a range of use cases.

Introduction: Manufacturing a transformation 3
From potential to performance 4

Evolving roles: How digital workers change human work 5
From human in the loop to human *on* the loop 6
Automation with oversight 7

Reimagining KPIs: How to define the metrics of success 8
Decomposing tasks, reimagining KPIs 9
Smart KPIs for smarter agents 10
Designed for impact, primed for performance 12

Proving impact: How to monitor AI agent performance 13
Agent operations for effective observability 14
A reference architecture for AI agent operations 15

Making it real: AI agent operations in action 16

Transformation now: Improve business outcomes with AI agent observability 18

Get in touch & Endnotes 19

Introduction

Manufacturing a transformation

The software engineers call him Greg—a nod to his gregarious nature. No matter the conversation, he is friendly and inquisitive, eager to please and receptive to feedback.

Greg works as part of a team with Guru, a software incident classification specialist who seems to understand every feature (and known bug) of every application in the company's sprawling tech stack. There's also Dupin, the incident analyst whose analytical skills rival those of Edgar Allen Poe's fictional detective of the same name; and Vera, whose healthy skepticism and fixation on accuracy is much needed on her eager, fast-moving team.

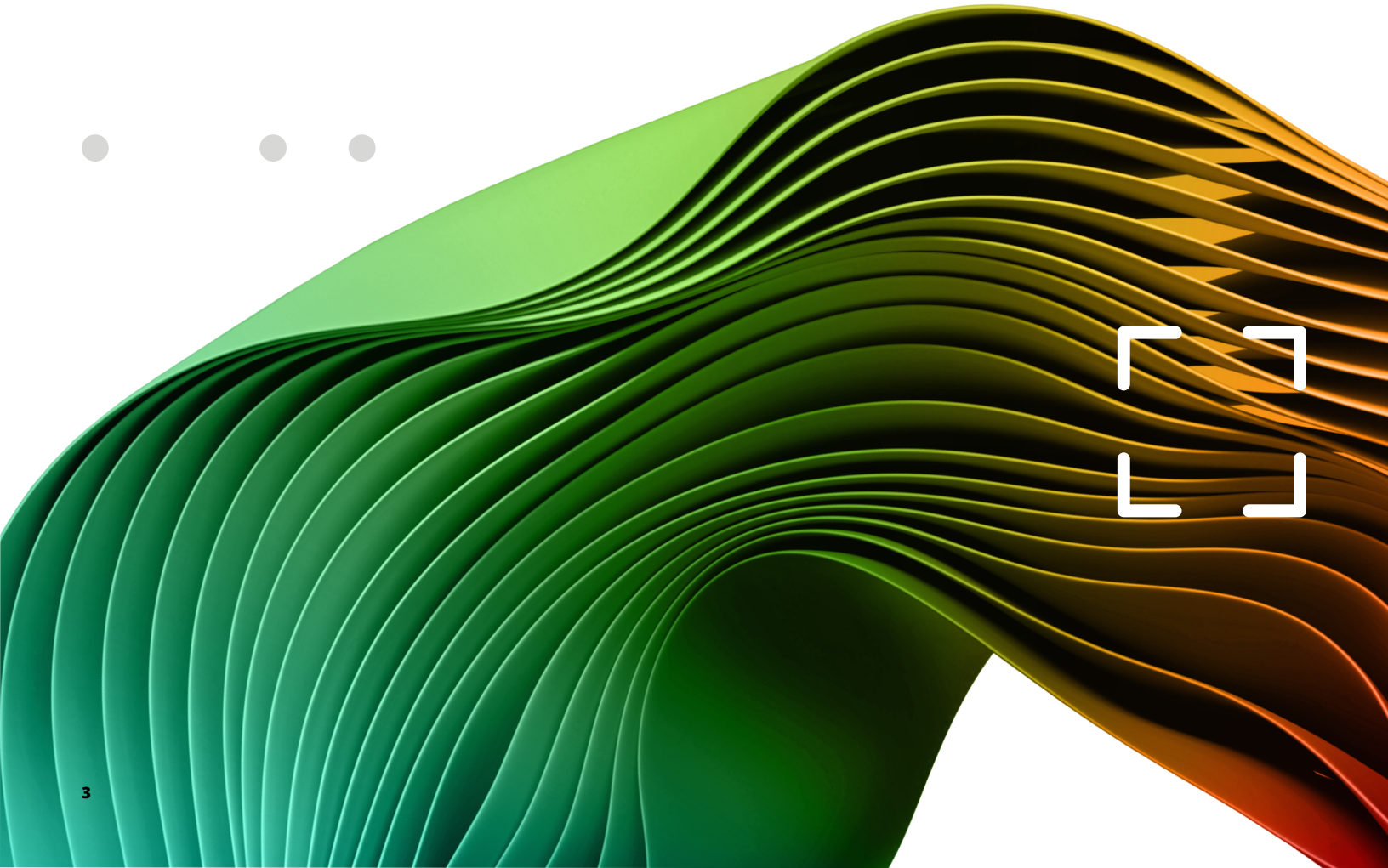
Though still in training, this application support team of four has accurately understood, diagnosed and resolved numerous software incidents—including several of the most common (and resource-sapping) issues faced by employees at the industrial equipment manufacturing company.

"If they keep learning and working this fast," one IT leader mused, "they might eventually help our customers too."

That's good news, considering what it has taken to bring Greg, Guru, Dupin and Vera on board and up to speed. The process has looked nothing like the company's usual hiring and training routine. That's because the new team members look nothing like the company's usual employees.

They're AI agents—digital workers with the ability to reason, plan, learn, remember and act based on user needs. Leaders at the company hope these four digital workers can alleviate the seemingly inexhaustible and growing backlog of level 1 and level 2 software support tickets—giving human workers time to address the more complex issues.

The team of agents has performed well in controlled testing. **The question is: How will they perform when it actually matters?**



From potential to performance

It's not just leaders at the manufacturing company* who see the potential of AI agents. Senior leaders across industries already recognize agentic AI as a breakthrough innovation poised to unlock the full potential of Generative AI (GenAI). Some leaders aren't waiting to add Gregs and Veras to their teams: 39% are already directing investments toward agentic AI as part of their new technology strategies.[1]

But as you know, *potential* doesn't pay the bills.

If you're going to make the case—and make the investment—you need *proof* that digital workers perform in ways that improve operational and business outcomes.

In this report we explore approaches, metrics and AI agent observability solutions that can help your organization monitor and enhance AI agent performance against business and operational goals. We present an adaptable framework for key performance indicators (KPIs) and a reference architecture for AI agent operations that can help accelerate impact and enhance transparency of the digital workforce.

By applying these recommendations, your organization can more rapidly progress on the path to agent-enabled workforce transformation—while protecting against unexpected costs and business risks.

AI agent ABCs

AI agents are reasoning engines that can understand context, plan workflows, connect to tools and data, and execute actions to achieve a defined goal. By combining these core abilities with knowledge of organizational policies, standard operating procedures and knowledge bases, agents can virtualize skills and serve roles as digital workers.

Multiagent AI systems employ a team of multiple, role-specific AI agents to understand requests, plan workflows, coordinate role-specific agents, streamline actions, collaborate with humans, and validate outputs.

AI agent observability is a technology-enabled, people-powered capability that helps organizations continuously monitor, analyze and optimize AI agent performance against desired business and operational objectives.

** Deloitte has helped companies across industries design and deploy AI agents in application support and other roles / processes.* ***The industrial equipment manufacturing company and named AI agents referenced in this report are fictionalized*** *and used for illustrative purposes only.*

Evolving roles

How digital workers change human work

Leaders at the manufacturing company understand that automation can be technically complex and expensive. Some in the C-suite question whether designing and implementing a new digital workforce for application support will yield positive return on investment (ROI). The chief risk officer also wants to know how hallucinations by the digital support team will be identified, traced to their source and corrected.

Meantime, the company's current team of application support specialists feels worried—and is voicing concerns to any leader who will listen. Surely Greg can't be as empathetic as a human specialist? Guru and Dupin may know *what* an application is supposed to do, but they can't really understand *why* people use it—so can they really be trusted to classify and resolve issues faster and more effectively than people can? And how will the agentic workforce move the goalposts—or, for that matter, change the whole *game*—for human support specialists?

From human in the loop to human *on* the loop

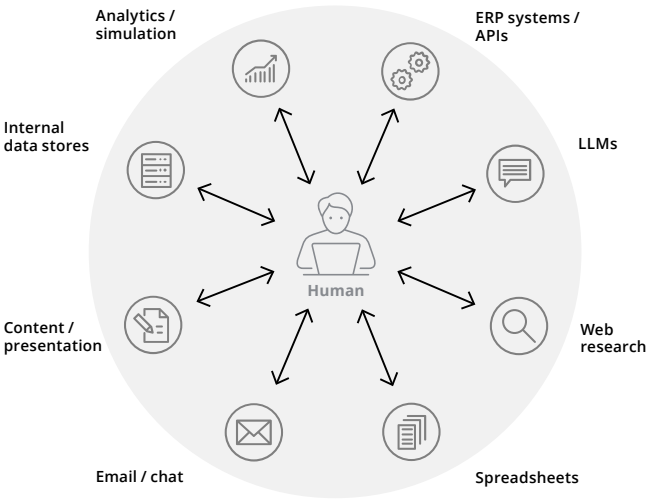
The integration of AI agents into the workforce is a complex process that demands careful consideration of the roles that humans and AI agents should play. This involves recognizing the strengths and limitations of all workers, determining how they can collaborate to achieve optimal results, identifying use case-specific processes and operating models that will provide the greatest benefits at the lowest cost, and continuously monitoring and improving performance.

To paraphrase the old proverb: Every transformational journey begins with a single step. Given the high stakes and many steps that lie ahead, it's important to develop a clear vision of where and how far you plan to go, chart the most efficient path to your destination, and continuously monitor your pace of progress.

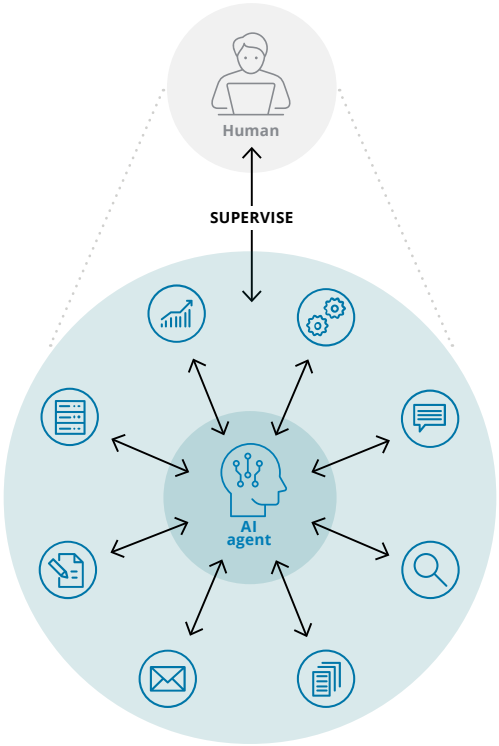
It's important, in other words, to take a measured approach to agentic transformation.

Many organizations have begun the agentic journey by deploying general-purpose large language models (LLMs) and prebuilt agents from enterprise software providers to help human workers complete tasks. While this approach remains heavily reliant on human input and interaction with AI, it has nevertheless produced measurable impact across a range of use cases. Our findings reveal that marketing leaders who adopt automation keep up with content creation demands more successfully and achieve better results across important metrics. In fact, leaders reported a 29% higher impact on revenue from their content marketing efforts and were 24% more likely to consistently meet their content goals compared to those not using automation tools.[2]

Multiagent AI systems represent a fundamentally different opportunity. By shifting human responsibilities away from the *execution of tasks* that are supported by AI agents and toward the *supervision of work* that is planned and executed by multiagent AI systems, whole new realms of enterprise productivity, speed, scalability and efficiency become possible. *(See illustration.)*



Currently, **a human drives** a set of digital tools (including LLM-powered AI)



Multiagent AI systems shift **human** responsibility away from execution of tasks and toward **supervision of tasks**

Automation with oversight

Automating complex interactions between multiple AI agents allows digital workers to perform in ways that standalone language models can't match. Through continuous human-aligned feedback, multiagent AI systems can be continuously provided new training that helps improve their output quality, amplify their efficiency and adapt swiftly to changing requirements.

But without proper transparency and monitoring, they can also go haywire.

One of the main benefits of digital workers—their scalability—is also a potential hazard: Whereas the damage done by a single human salesperson who misinterprets a client's order can damage one relationship, that danger pales in comparison to the risk of a single AI agent misinterpreting *every* client's order. Such failures in connected AI systems can quickly cascade across multiple interactions.

Behind every effective team—human or otherwise—is effective oversight. Effective oversight depends on clearly defined roles and expectations, coupled with clear-eyed monitoring of performance. In the next section of this paper, we explore how business process decomposition and a new KPI framework for AI agents can help turn these principles into practice.

Reimagining KPIs

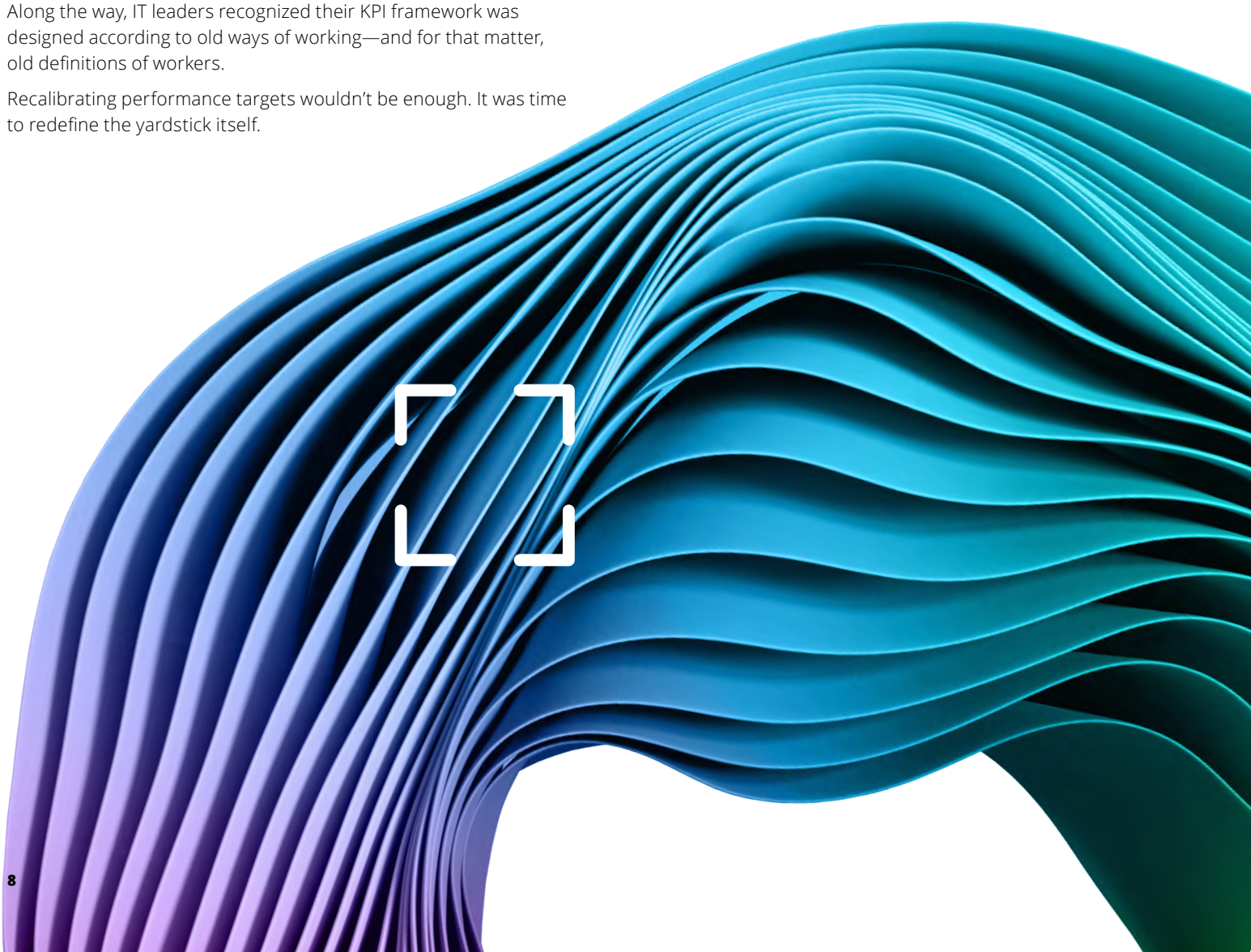
How to define the metrics of success

Greg, Guru, Dupin and Vera weren't designed just to improve the company's application support operating model, but rather to transform it. Before any coding began, IT leaders carefully analyzed existing data, tools and processes used by the application support team, along with the roles and responsibilities of each team member. They interviewed experienced support specialists to learn the tricks and shortcuts that have helped them circumvent system inefficiencies and fill knowledge gaps. Finally, they examined the application support function through the eyes of business users, noting where handoffs and verifications created pain points and inefficiencies.

Those insights proved crucial in defining the roles of each agent in the digital team and determining the language models, tools, data and memory that each agent would need to perform its role.

Along the way, IT leaders recognized their KPI framework was designed according to old ways of working—and for that matter, old definitions of workers.

Recalibrating performance targets wouldn't be enough. It was time to redefine the yardstick itself.



Decomposing tasks, reimagining KPIs

Across domains and functions of the enterprise, AI agents and multiagent AI systems open new possibilities to drive productivity and program delivery through business process automation. Use cases once considered too complicated for GenAI can now be enabled at scale—securely and efficiently.

It's not just about *improving* processes. Multiagent AI systems have the potential to impact every layer of enterprise architecture by *reinventing* the ways that work gets done.

Achieving that potential depends on identifying use cases where digital workers and teams have the greatest potential to improve outcomes, and then decomposing existing processes into discrete tasks. Business process decomposition enables your organization to identify not only the roles and activities needed in a workflow, but also the KPIs and observability solutions that will allow you to see, manage and optimize AI agent performance.

As with any transformation initiative, KPIs will serve as the North Star as you design and implement new solutions. By understanding processes at an atomic level, you're better able to orient them toward your North Star.

The primary business and operational KPIs against which you measure success—such as customer satisfaction and total cost—may remain essentially the same as you shift tasks and roles to digital workers.

But those KPIs will likely prove insufficient—especially during times of rapid change, when AI agents handle work at a scale and pace that human workers cannot match. Typically, multiagent AI systems will impact multiple organizations within the enterprise—and each of those organizations will need unique ways of understanding the impact.

Consider the work assigned to Greg, Dupin, Vera and Guru. As the team of digital workers expands its capacity to handle a wider range of software incidents, finance leaders need to understand the costs associated with their increased token usage—and should be alerted if costs suddenly spike due to agent runaway, emergent issues or unforeseen demand. IT leaders need to monitor trends such as latency and speed to resolution as agent workloads increase and diversify. Risk leaders will expect immediate alerts if the system or its users are accessing data that shouldn't be available to them. The list goes on.

Business process decomposition enables your organization to identify not only the roles and activities needed in a workflow, but also the KPIs and observability solutions that will allow you to see, manage and optimize AI agent performance.

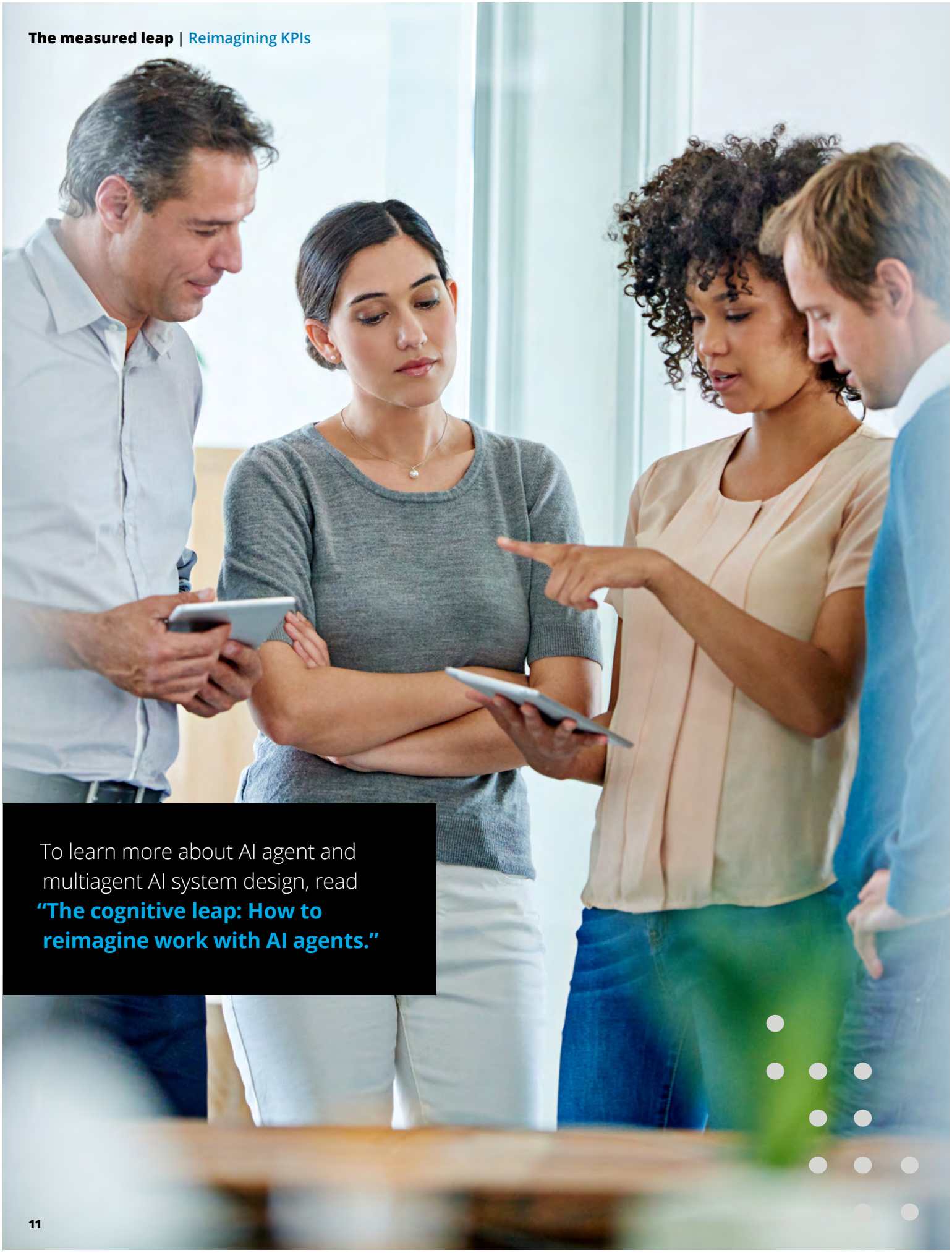
Smart KPIs for smarter agents

The complex ways that AI agents work individually and together—and the many ways they impact enterprise operations and results—demand a comprehensive KPI framework for assessing their performance.

We recommend a framework consisting of five primary categories of KPIs. *(See examples in graphic.)* Successful agentic transformation of business processes depends on assessing the operational, output and thought efficiency of each individual agent—and multiagent AI systems as a whole—against KPIs in *all five* categories.

While this approach may result in a larger number of KPIs than were measured as part of traditional operational processes executed by human workers, many will be applicable—and easily adapted—for *all* processes involving AI agents.

	COST	SPEED	PRODUCTIVITY	QUALITY	TRUST
PURPOSE	**Monitor and optimize the cost of operating agentic systems over time**	**Monitor and identify potential opportunities to improve latency of systems and components**	**Monitor and identify potential opportunities to improve system throughput**	**Monitor and identify potential opportunities to improve system response quality**	**Monitor and measure human user feedback trends**
EXAMPLE KPIs	**Cost** *How much does it cost to run and maintain the agent workforce?* **Token usage** *How much computational power is consumed per task?*	**Retrieval latency** *How much time does it take to retrieve context?* **Generation latency** *How much time does it take to generate a response?* **Tool call latency** *How much time does it take for tools to respond?*	**Success rate** *What percentage of tasks are completed successfully?* **Average handling time** *How is the time to resolution changing as AI agents handle tasks?* **Productivity gain** *How much time it took to complete the process?*	**Tool selection efficiency** *How many interaction turns does it take for an agent to select and use the correct tool?* **Correct tool utilization** *Is the agent selecting the most appropriate tool for the task?* **Plan efficiency** *How efficient is the agent in planning a task?*	**User feedback scores** *How do users perceive the agent's understanding of their feelings?* **Usage metrics** *How is the usage of the agent or multiagent system changing over time?*



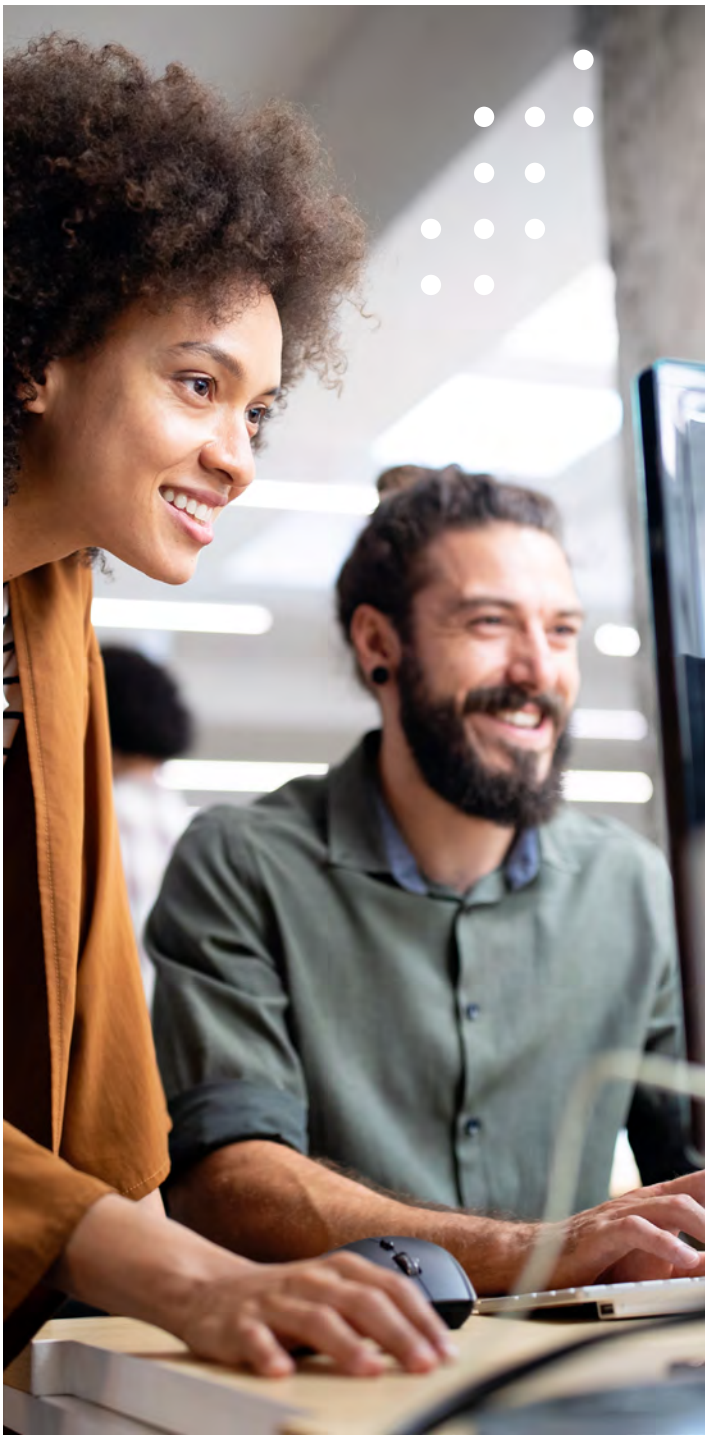
To learn more about AI agent and multiagent AI system design, read **"The cognitive leap: How to reimagine work with AI agents."**

Designed for impact, primed for performance

AI agents perform better when they have what they need to succeed. That starts with clearly defining their roles through business process decomposition and then designing, training and equipping them to do their jobs. This will help ensure AI agents are *ready to perform* the work assigned to them.

When it comes to ensuring AI agents *actually perform* their work and meet their KPIs, **consider the following:**

- **Data quality and context:** AI agents are dependent on the data they can access and the context they are given. These inputs should be of high quality to enable effective performance.
- **Continuous improvement:** The potential for continuous improvement and learning should be built into AI agent KPIs and solutions, ensuring their adaptation and evolution over time is a feature, not a bug.
- **Humans on the loop:** As AI agents evolve and improve their performance, the human workforce will continue to be an essential supervisor and collaborator in their processes. Human-aligned evaluation helps ensure higher quality and more efficient output by applying enduring human capabilities[3] such as imagination, adaptive thinking, emotional intelligence and sense-making.
- **Change management:** AI agent integration may produce unexpected challenges and emotional responses among human counterparts. In this evolving context, change management through employee support, training and incentives will be key to help ensure these issues don't undermine the performance of digital or human workers.
- **Ethical considerations:** Burgeoning ethics, transparency and trustworthiness concerns as applied to agentic AI will require monitoring and operational actions to achieve desired outcomes.



Proving impact

How to monitor AI agent performance

It's been just a few months since Greg began working with real business users, and he's already generating buzz around the company. Recently, when a sales leader found himself unable to access a key application during a client meeting, Greg was not only quick with a solution but also empathetic in ways that defused his panic. Others have started calling the chatbot "Go-to Greg"—he seems to have an answer for every software-related problem.

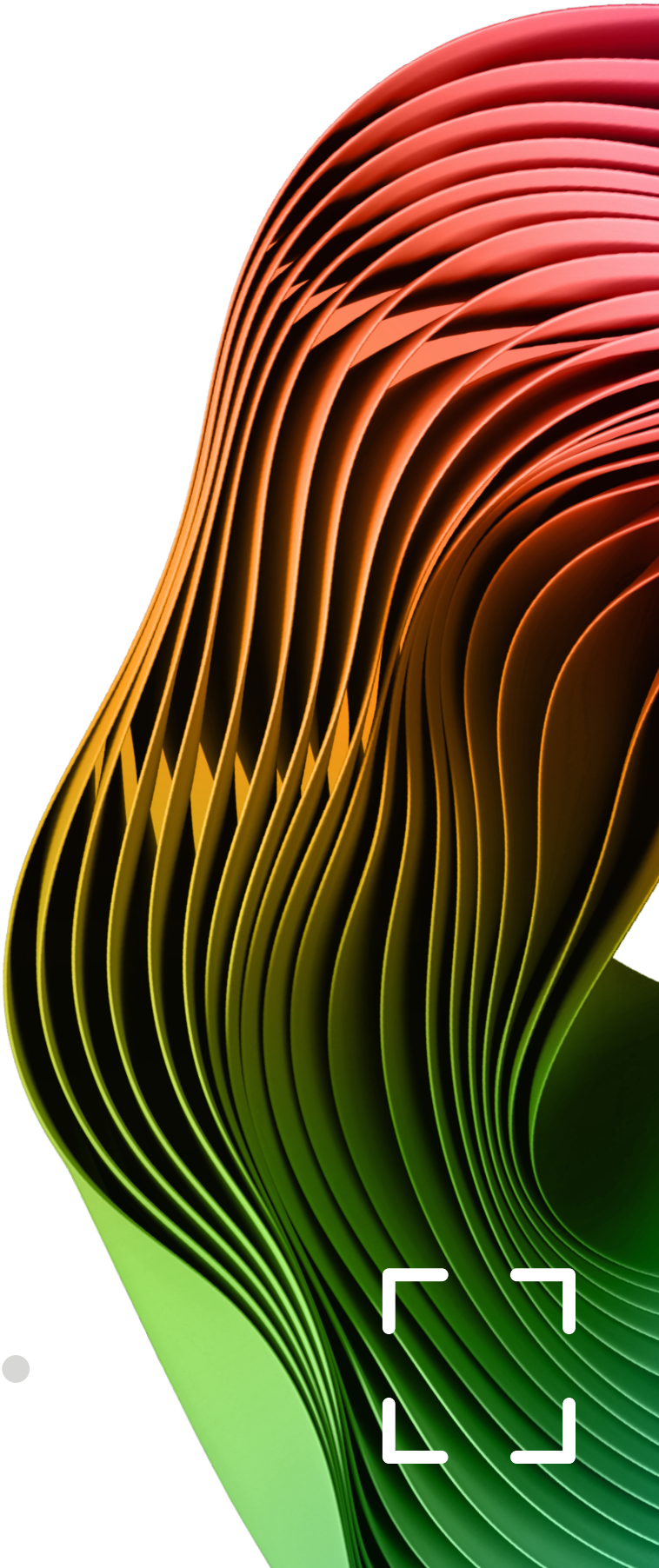
These bits of feedback indicate Greg may be providing value. But anecdotal stories are no substitute for proof—especially when it comes to ascertaining whether Greg is using resources efficiently, following proper protocols and interacting well with human users and other agents.

To prove Greg's performance and value against operational and business KPIs, his organization needs the ability to …

1. **Collect** comprehensive, ever-fresh data about everything Greg does
2. **Manage** all that data in ways that are persistent, secure and trustworthy
3. **Understand** why and how Greg acts in real time
4. **Monitor** and optimize Greg's efficiency, decisions and actions over time

These requirements become even more important when assessing the performance of Greg's behind-the-scenes team of AI agents. By design, business users only interact with Greg. When Greg is slow to respond because Guru is undergoing maintenance, the user may assume Greg isn't able to work fast. When a software fix proposed by Dupin and checked by Vera proves insufficient, the user may assume Greg is prone to mistakes.

AI agent observability is key to addressing these needs. It's about putting the right solutions in place to ensure every input, decision and action of every AI agent can be seen, understood and improved.



Agent operations for effective observability

In "The cognitive leap: How to reimagine work with AI agents," we explored how a reference architecture for multiagent AI systems can allow enterprises to more rapidly scale, expand and reuse AI agents and multiagent frameworks across a range of use cases—while also streamlining governance, monitoring, operation and improvement of agentic outputs. We recommended a reference architecture comprising four loosely coupled but independent layers: *interactions*, *workflows*, *agents* and *agent operations*.

In essence, agent operations serves as the performance and risk management function for digital workers and teams.

Through the capability of AI agent observability, agent operations can provide other functions of the enterprise with alerts and insights regarding the operational performance and business impact of AI agents and multiagent AI systems.

It also plays a critical role in enforcing governance guardrails and regulatory compliance—helping to ensure that AI agents don't silently propagate bad output, expose sensitive data or mislead users.

Because some risks are more immediate than others, **consider structuring agent operations in ways that enable both a "hot" and a "warm" path for insight processing and delivery**.

- **Hot path:** Operational signals are captured immediately as the agent produces them, enabling **observability of risks** such as toxicity, data leakage or regulatory noncompliance. By establishing clear governance guardrails and alert thresholds, critical issues can be surfaced as they happen.

- **Warm path:** Agent actions and interactions across sessions are aggregated and analyzed in ways that enable **observability of patterns** such as reasoning loops, response drift or repeated unnecessary turns to choose tools. While "warm path" insights require extra processing through analytics solutions—and thus extra time to produce—they are a key element of agent performance management, risk mitigation and ongoing optimization.

Given the diversity of KPIs and use cases involved in agentic transformation—and the speed at which risks can emerge—agent operations should be implemented, not as a set of *solutions*, but as an ecosystem of connected *capabilities* with its own reference architecture.

Let's explore what that looks like.

A reference architecture for AI agent operations

Telemetry layer

Purpose:
Capture the fine-grained, real-time interactions and performance of AI agents and systems.

Raw session data is captured and output in the form of ...

- **Logs** *Timestamped system-level records*
- **Traces** *Complete journey of a request or transaction*
- **Spans** *Single unit of work within a trace*
- **Events** *Records of significant occurrences within a system*

Operational insights layer ("Hot path")

Purpose: Capture signals produced by AI agents and systems in near real time, allowing rapid alerts and response to security risks, compliance issues or failures.

Example insights include ...

- **Retrieval, generation & tool call latency**
- **Token usage**
- **Errors & rejected operations**
- **Cost**

Analytical insights layer ("Warm path")

Purpose: Ingest, normalize, enrich and analyze telemetry data to reveal the efficiency and effectiveness of AI agents and systems over time.

Example insights include ...

- **Customer satisfaction scores**
- **User acceptance rate**
- **Response / consistency drift**
- **Average handling time**

Both insights layers include storage solutions to support the persisting agent state for both operational continuity and retrospective analysis.

Monitoring layer

Purpose:
Provides insights regarding the live status and performance of AI agents and systems against business and operational KPIs.

Supports observability through ...

- **Alerts & notifications**
- **Dashboards**
- **Ad-hoc reports**

Making it real

AI agent operations in action

One Monday morning, the top sales representative at the industrial equipment manufacter opens a support ticket through the company's IT support portal and begins chatting with **Greg**, the gregarious interface agent. By asking a few questions, Greg quickly learns the rep is receiving error notifications as he tries to forecast a new client's spend in the coming fiscal year. Greg recognizes the issue involves the company's customer relationship management platform, and activates the appropriate software incident workflow.

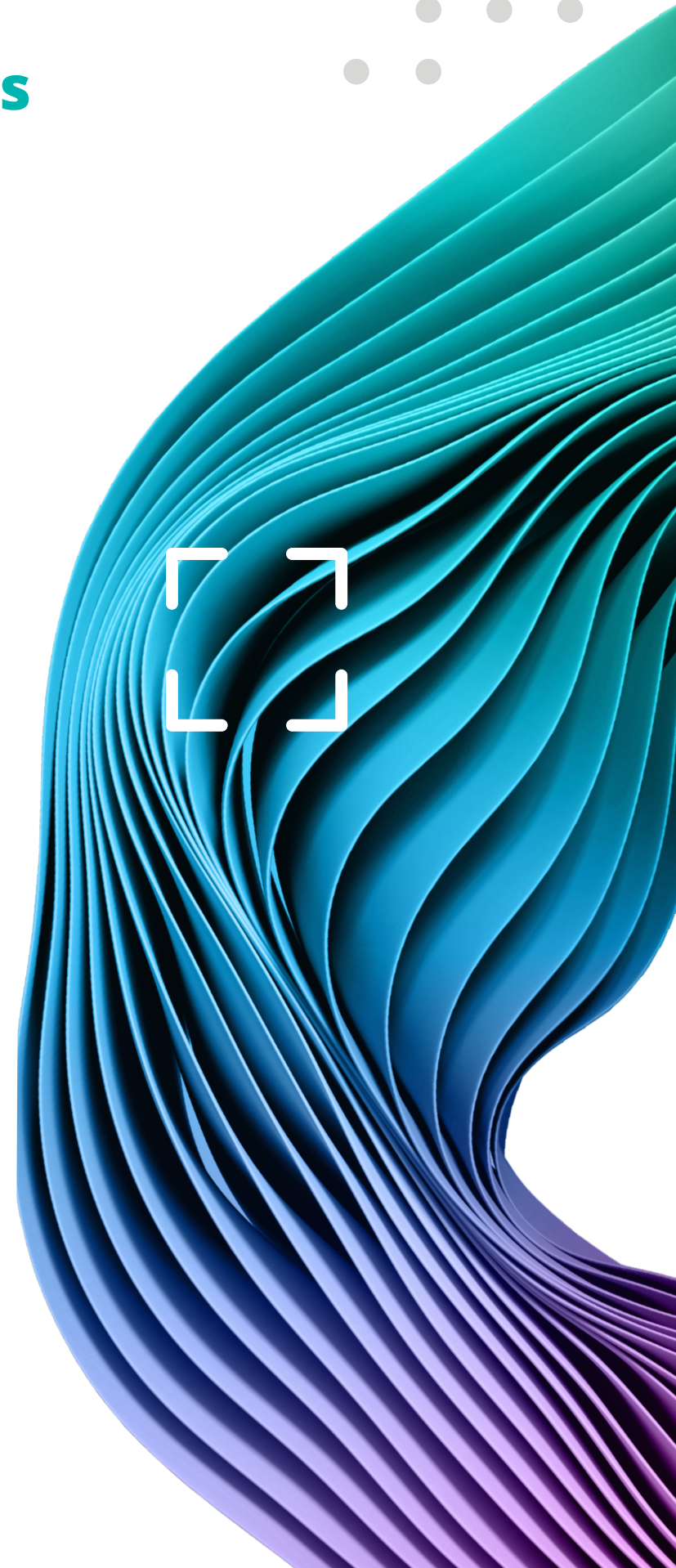
Guru, the incident classification agent, identifies this as a data issue and engages **Dupin**, the software incident analysis agent. Dupin quickly scans the company's knowledge resources, where he learns the rep's data permissions are most likely at fault.

Dupin proposes a change to the rep's permissions to **Vera**, the incident resolution agent. However, Vera notices Dupin's proposal would allow the rep to access *too much* data—including personally identifiable information (PII) about the client.

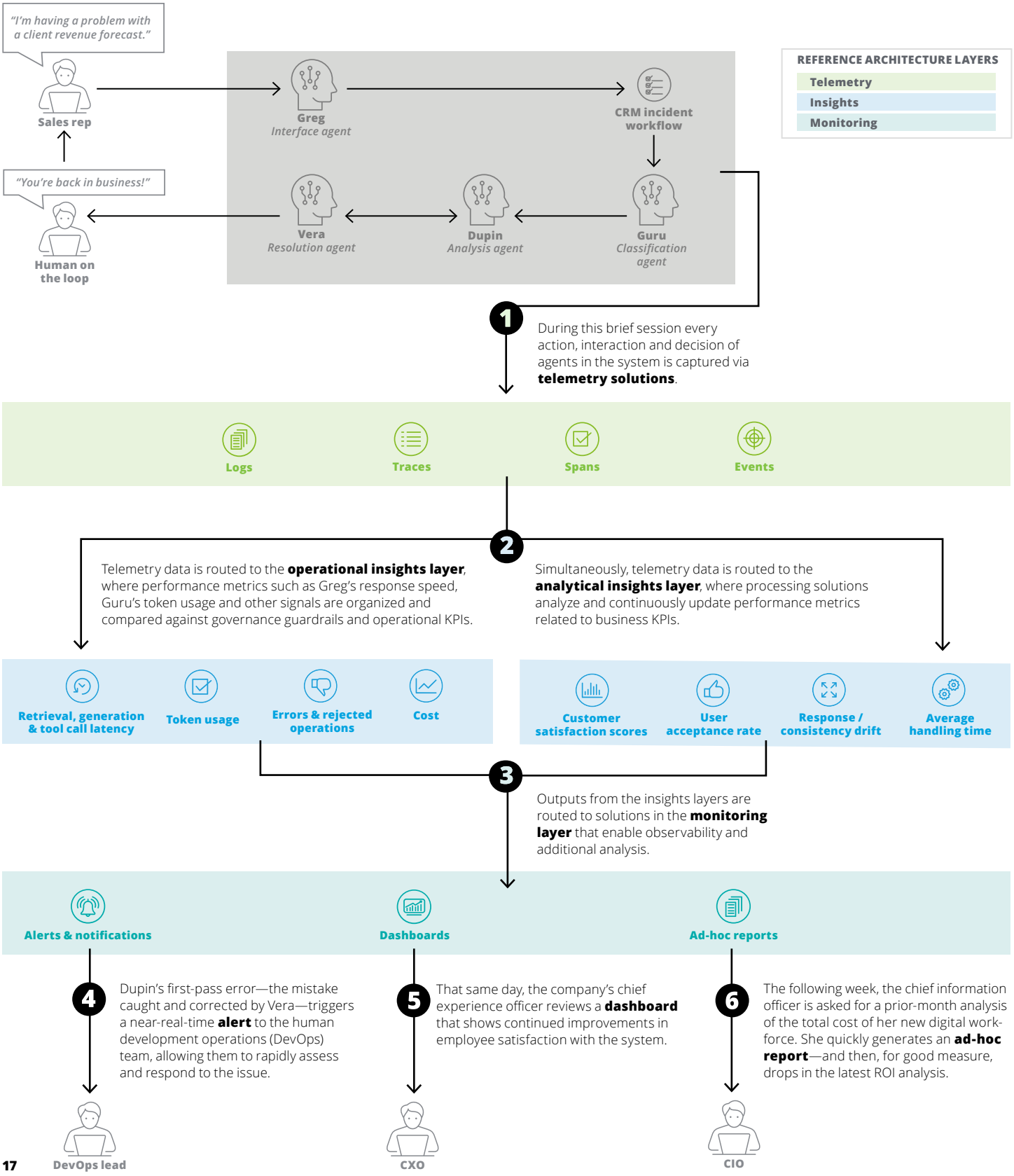
Vera provides feedback to Dupin, who narrows his recommendation to only allowing access to the client's past purchase data. Dupin resubmits the solution to Vera, who confirms it solves the problem and conforms to data governance standards.

Vera passes the solution to a **"human on the loop,"** who confirms and implements it for the sales rep. The rep is then invited to rate his satisfaction with the experience.

In the illustration on the next page, we detail how an agent operations capability can help ensure these new digital workers are meeting user, operational and business expectations.



AI agent operations in action



Transformation now

Improve business outcomes with AI agent observability

The transition from human-powered to human-led agent-enabled operations demands a thoughtful approach to agent architecture and process design, performance measurement and continuous oversight. As our fictional manufacturing company recognized, the effective design and deployment of digital workers is not simply about automating tasks through the use of AI agents.

It's about reimagining how work gets done, how value is created … and how digital worker performance is made observable and actionable.

Organizations that succeed will be those that define clear roles for both human and digital workers, decompose business processes to identify where agents can drive the greatest impact, and establish robust KPI frameworks and observability solutions that enable transparency, accountability and ongoing improvement.

Only then can the Gregs, Gurus, Dupins and Veras of *your* digital workforce succeed.

As you consider your own path forward, ask yourself:

- Where can digital workers most effectively augment or transform my operations?
- How will I measure and monitor their performance to ensure business value and manage risk?
- What processes do we need to decompose to identify the most valuable opportunities for agentic automation?
- Which KPIs—operational, output, thought efficiency—are needed to support our organization's goals?
- How will we maintain transparency and traceability of agent decisions and actions across workflows?
- What mechanisms will we use to continuously improve agent performance and adapt to changing requirements?
- How will we monitor and address the potential emotional and cultural impacts of AI agents on our human workforce?
- What safeguards can we put in place to detect and mitigate risks such as agent hallucinations or runaway costs?
- How will we integrate human-on-the-loop oversight to ensure quality, ethical standards and compliance?

Get in touch

Prakul Sharma
Principal, AI & Data
Deloitte Consulting LLP
praksharma@deloitte.com

Parth Patwari
Principal, AI & Data
Deloitte Consulting LLP
ppatwari@deloitte.com

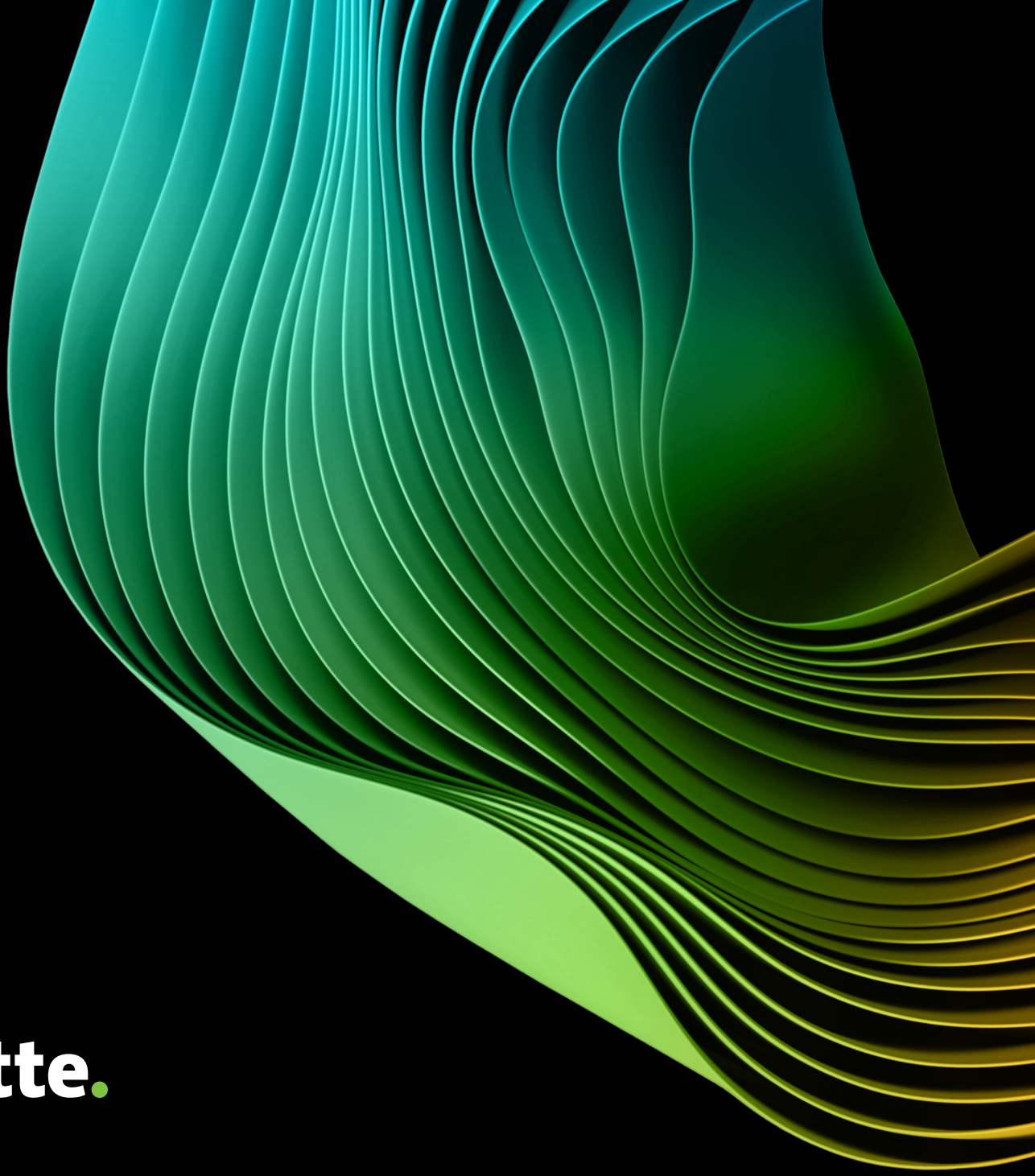
Brijraj Limbad
Senior Manager, AI & Data
Deloitte Consulting LLP
blimbad@deloitte.com

Contributors to this report:
Jim Rowan, Danielle Boutwell, Giannis Doulamis, Pradeep Gorai, Siva Muthu, Sanghamitra Pati, Laura Shact, Greg Vert

To learn more, visit our website

Endnotes

1. Tim Smith, Garima Dhasmana, Diana Kearns-Manolatos, Gregory Dost, Parth Patwari, and Iram Parveen, "AI is capturing the digital dollar. What's left for the rest of the tech estate?", Deloitte Insights, October 23, 2025, accessed November 24, 2025.
2. Deloitte Digital, "Marketing content automation: Harness AI for your marketing content supply chain", January 28, 2025, accessed November 24, 2025.
3. John Hagel III, John Seely Brown and Maggie Wooll, "Skills change, but capabilities endure," Deloitte Insights, August 30, 2019, accessed September 17, 2025.



Deloitte.

About Deloitte

As used in this document, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see www.deloitte.com/us/about for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting.

Deloitte provides industry-leading audit and assurance, tax and related services, consulting, financial advisory, and risk advisory services to nearly 90% of the Fortune Global 500® and thousands of private companies. Our people deliver measurable and lasting results that help reinforce public trust in capital markets, enable clients to transform and thrive, and lead the way toward a stronger economy, a more equitable society, and a sustainable world. Building on its 175-plus year history, Deloitte spans more than 150 countries and territories. Learn how Deloitte's approximately 457,000 people worldwide make an impact that matters at www.deloitte.com.

This publication contains general information only and Deloitte is not, 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 business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor. Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

Copyright © 2025 Deloitte Development LLC. All rights reserved.