The agentic scale trap

Why enterprise AI pilots succeed and why they fail in production

The better the question. The better the answer. The better the world works.

Shape the future with confidence

Contents

01 Introduction

02 The pilot-to-production paradox

03 AI adoption has outpaced readiness

04 The agentic inflection point

05 Pilot-to-production failure patterns

06 Industry perspectives

07 From ‘bolt-on’ to ‘built-in’ AI

08 A proposed enterprise AI readiness framework

09 Implications for CEOs, CIOs, CDOs and CAIOs

10 Conclusion: Intelligence is not the constraint

1. Introduction

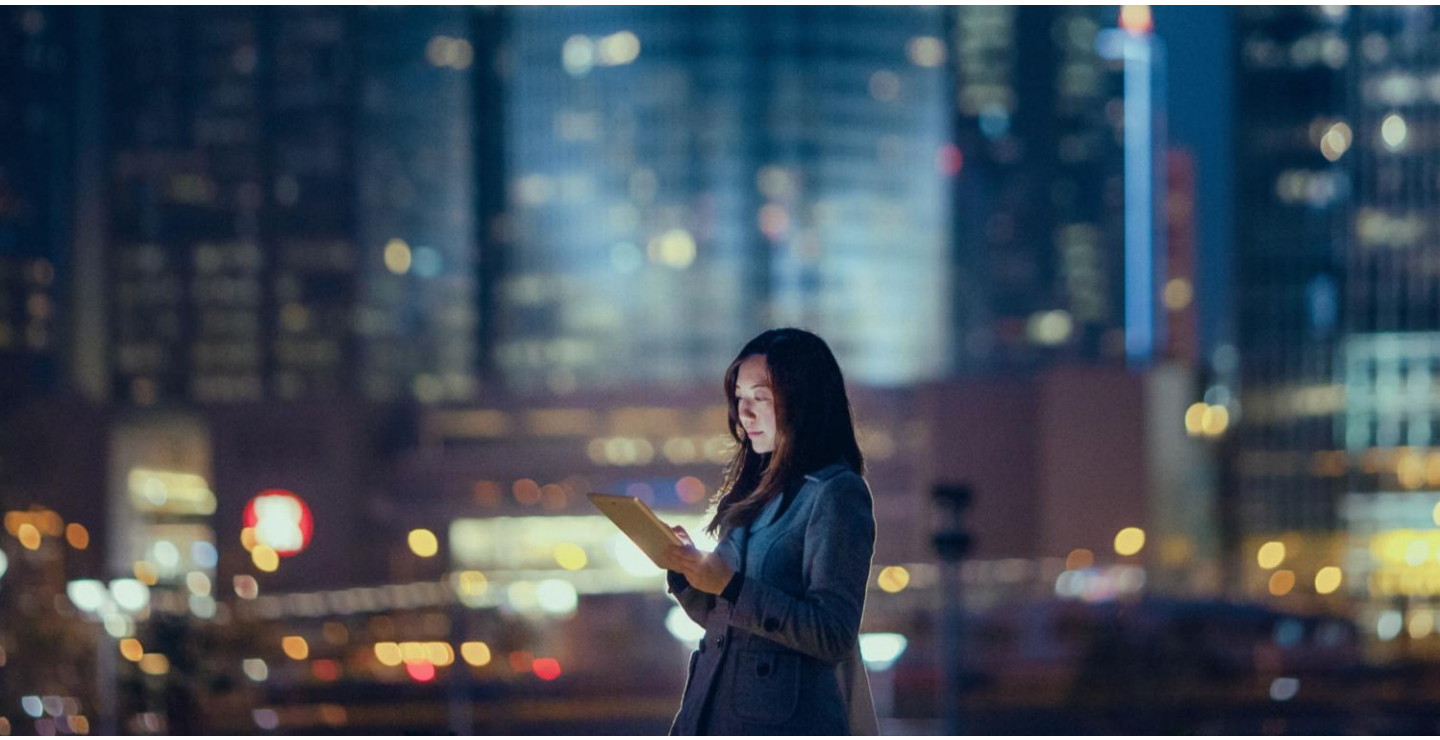
Artificial intelligence (AI) has moved from promise to proliferation. Across industries, organizations are piloting generative AI (GenAI) copilots, workflow assistants, decision-support tools and, increasingly, agentic systems that can autonomously plan and act. In many cases, these pilots appear successful, with models performing well in demonstrations and users reporting productivity gains.

When these same systems are pushed into production, many organizations find progress stalls. Integrating AI models into core platforms and workflows, exposing them to real data, and deploying them within regulatory and accountability constraints reveal underlying challenges that see initiatives paused, scaled back or abandoned altogether.

Many AI pilots, though intelligent, are not designed for the realities of production environments. Most aren't equipped for real-world conditions that are often messy, disjointed and unpredictable. While pilots suppress complexity, production environments surface and amplify it. As AI systems become more autonomous, the cost of that mismatch grows exponentially.

In the era of agentic AI, the focus for leaders shifts from AI's ability to generate answers, to its ability to operate safely, predictably, and accountably inside the enterprise. Confident AI transformation demands new architectures, new control mechanisms, new operating models and new definitions of ownership.

Enterprises must view AI as an end-to-end transformation opportunity rather than a technology upgrade. Success requires moving from 'bolt-on' AI — isolated pilots layered onto existing processes — to 'built-in' AI, where intelligence, governance and execution are designed together from the start.



2. The pilot-to-production paradox

2.1 Why pilots succeed and production fails

AI pilots are intentionally designed to succeed. They operate in constrained environments with:

- Curated and relatively clean data
- Simplified workflows
- Limited system integrations
- Narrow user populations
- Continuous human oversight
- Minimal regulatory exposure

These conditions are necessary to test feasibility, but they create a false sense of readiness. Performance observed in a pilot says more about the environment than the enterprise.

Production reverses these conditions:

- Data becomes heterogeneous, delayed, and imperfect
- Workflows span multiple systems and teams
- Decisions carry financial, regulatory, and reputational consequences
- Ownership becomes diffuse
- Failures must be explainable, auditable, and reversible

This is the point where many AI systems break down because enterprise processes, controls and ownership models have not evolved alongside them.

2.2 The shift to autonomous AI and its enterprise implications

As AI evolves from assistive to agentic, we see this gap become more pronounced. Traditional analytics and even early GenAI primarily supported human decision-making. Agentic systems, by contrast, can:

- Plan across multiple steps
- Invoke tools and application programming interfaces (APIs)
- Modify records
- Trigger workflows
- Interact with other agents

Autonomous AI systems pose more difficult questions for enterprise leaders:

- What is the agent authorized to do?
- What must it never do?
- When must humans intervene?
- Who is accountable for outcomes?

Most enterprises have not yet answered these questions at scale.

3. AI adoption has outpaced readiness

AI adoption is advancing more quickly than the operating models required to scale it. Many organizations now use AI across business functions and are experimenting with AI agents, but for most, enterprise level financial impact remains limited[1] [2].

While access to models is increasing, workflow redesign and governance oversight remain the bigger barriers to scaled value realization because they determine how effectively AI is embedded into day-to-day operations[3]. Evidence suggests that most AI projects fail because of misaligned objectives, weak data, inadequate infrastructure and poor integration into business workflows[4].

This is why many companies remain stuck in the "last mile" of AI transformation, where technical capability collides with ownership, incentives, accountability, process debt and architectural complexity[5]. In contrast, organizations with high AI maturity are materially more likely to sustain AI initiatives over several years, centralize governance and infrastructure, appoint dedicated AI leaders, and measure value across financial, customer and risk dimensions[6].

Agentic AI further intensifies the issue. EY research shows that 76% of organizations are already using or planning to use agentic AI, yet only 56% are aware of the associated risks. This highlights a critical gap as systems move from generating outputs to making decisions and executing actions autonomously, introducing risks such as unintended outcomes and unpredictable behavior[3].

Together with the governance expectations embedded in the EU AI Act, these signals point to a clear conclusion: scaled AI success depends heavily on an organization's readiness to govern autonomy in production[7].

4. The agentic inflection point

Agentic AI represents a qualitative shift. These systems behave more like digital workers than tools: reasoning, learning, coordinating, and acting across systems. Their deployment creates significant opportunities for organizations to reduce cycle times and manual coordination, improve process efficiency and transform operating models. But they also introduce distinct challenges including permission creep, non-transparent decision logic, and systemic, cascading failures.

These risks are increased when enterprises deploy agentic AI in environments that lack clear decision rights, runtime controls, behavioural observability and reversibility mechanisms. Without these platform-level controls, autonomous systems can become increasingly difficult to govern at scale. The same agent that performs impressively in a pilot can create material risk in production.

5. Pilot-to-production failure patterns

The gap between pilot and production is not new but failure within agentic systems presents a distinct set of problems[3].

5.1 Architecture and engineering — weak engineering foundations break autonomy at scale

Enterprise AI does not scale on prompt wrappers alone. Teams need a resilient digital core with clear orchestration logic, test discipline and security controls built in from day one[8]. Common challenges include:

- **Overly complex multi-agent design**: When too many agents are chained without clear state boundaries, a small upstream error can cascade through the workflow and trigger large production failures[8].
- **Weak engineering discipline**: If continuous integration, regression testing and fallback paths are missing, agents can stall, loop or produce unstable behavior under real work conditions[8,9].
- **Security and privacy as afterthoughts**: Controls for prompt injection, data leakage and access management must be embedded at design time. Retrofitting them later increases risk and cost[3].

The practical goal is a simple, modular, testable, and secure architecture that remains reliable as requirements, integrations and business needs evolve.

5.2 Data and grounding — poor data grounding erodes trust and accuracy

Model quality in production is constrained by quality of data and contextual reference. When data is incomplete, stale or weakly connected, outputs become confidently wrong[10]. Common causes include:

- **Low-quality or incomplete data**: Sophisticated reasoning on messy enterprise data amplifies existing errors instead of correcting them[10].
- **Weak retrieval pipelines**: Basic vector search often misses business context and relevance, which introduces noise and increases hallucination risk[10].
- **Data silos and stale context**: Agents restricted to isolated systems act on outdated assumptions, especially in fast-moving operations[10].
- **Poor memory strategy**: Without a clear design for short-term versus persistent memory, agents lose continuity and force users to repeat context[10].

Reliable grounding requires continuous data synchronization, strong retrieval design and explicit memory policies[10].

5.3 Cost and economics — pilot economics rarely hold at production scale

Pilot costs are often misleading because they exclude operational complexity. At scale, usage patterns and support demands change the cost profile materially[1]. Taking pilots into production can result in:

- **Token inefficiency**: Multi-step agent loops repeatedly pass large context windows, which can multiply model spend quickly[1].
- **Incomplete cost models**: Vendor model invoices are only one part of cost, runtime infrastructure, evaluation, monitoring, and tuning often account for most spend[1].
- **Unclear Return on Investment (ROI) and scope creep**: When business outcomes are not explicit, programs can expand in scope without proportional value realization[1].

Sustainable scaling requires clear unit economics, well-defined ROI gates, and disciplined prioritization of high-value use cases[1].

5.4 Governance and security — control gaps create operational and regulatory risk

As autonomy increases, governance must move from policy documents to runtime controls. The key question is how organizations define and enforce the boundaries of agent behavior in production[3,11]. Risks include:

- **Missing guardrails and verification**: Without continuous evaluation and policy checks, unsafe actions can pass into live workflows[11].
- **Limited human oversight**: Critical decisions still need explicit handoff points, exception paths and accountable ownership[3].
- **Auditability and compliance gaps**: If decisions and tool actions are not traceable end-to-end, incident response and regulatory review become high-risk events[9].

In production, governance needs to be embedded into day-to-day operations rather than treated as a one-time approval step[3].

5.5 Scalability and integration — integration bottlenecks limit enterprise-wide value

AI value depends on how well agentic systems connect to existing enterprise platforms. Fragile integration patterns constrain both reliability and scale. Two recurring barriers are:

- **Fragile enterprise integrations**: Custom point-to-point connectors break easily when upstream systems change, causing repeated service interruptions.
- **Infrastructure limits versus expanding scope**: Architectures designed for a single workflow often fail under concurrency, multi-tenant demand or global rollout.

Durable scale requires standardized APIs, loose coupling, and infrastructure sized for real production variability.

5.6 Monitoring and observability — limited runtime visibility slows recovery and learning

Non-deterministic systems need deeper operational telemetry than traditional software. Teams need visibility into both outcomes and the processes, tools and data that produced them[12]. Key observability challenges include:

- **Limited tracing**: Without end-to-end execution traces, teams cannot pinpoint where a failure started in multi-step agent flows[12].
- **Lack of auditable evaluation**: Without stable performance baselines and drift signals, organizations cannot reliably judge whether behavior remains within defined trust boundaries[12].
- **Difficult debugging**: Prompt changes and model variance can alter logic paths, so reproducibility tooling is essential for incident diagnosis[12].

Strong observability closes the loop between deployment, governance and continuous improvement[12].

Why agentic systems break at scale in production

1 Can we build it reliably?	2 Can we ground it accurately?	3 Can we trust and control it?	4 Can we operate it profitably?	5 Can we run it at enterprise scale?
Architecture and engineering	**Data, context and grounding**	**Governance, risk and trust**	**Economics and business value**	**Production operations**
Poor system design creates fragility at scale	Agents are only as good as their context	Lack of control prevents enterprise adoption	Many agents fail financially before they fail technically	The biggest failures emerge after deployment
▪ Over-engineered multi-agent architectures ▪ Weak agentic engineering foundations ▪ Poor resilience and failure handling ▪ Security, privacy and compliance added too late	▪ Low-quality or incomplete enterprise data ▪ Weak Retrieval-Augmented Generation (RAG) and retrieval pipelines ▪ Fragmented knowledge silos ▪ Outdated or stale context ▪ Poor memory and context management	▪ Missing guardrails and policy enforcement ▪ Limited human-in-the-loop oversight ▪ Compliance and regulatory gaps	▪ Token and inference cost inefficiencies ▪ Poor production-scale cost modeling ▪ Unclear ROI and business value ▪ Lack of value realization metrics	▪ Infrastructure constraints at scale ▪ Limited observability and tracing ▪ Poor auditability and debugging ▪ Inability to monitor agent behavior in production

6. Industry perspectives

6.1 Financial services: More autonomy calls for more accountability

In financial services, pilots often prove that AI can accelerate compliance triage, third-party review, underwriting support or decision augmentation. This is because the scope is narrow, historical datasets are curated, and humans remain close to the decision.

Production introduces a different operating reality. Agents must work across fragmented customer, risk and transaction systems; reconcile conflicting data definitions; navigate layered approval workflows; and support decisions that are regulated, contestable and auditable. This is particularly critical in areas such as credit scoring, insurance and essential services, which the EU AI Act identifies as high risk because of their impact on individual rights and outcomes. This complexity and scale can cause accuracy failures but also highlight dangerous gaps around accountability. When agents trigger multiple downstream actions — including across compliance, servicing and operations — ownership is diffused.

Agentic autonomy amplifies these failures because the cost of acting on incomplete context is materially higher than the cost of generating an imperfect recommendation. Success depends on creating the conditions for defensible judgment, traceability, reversibility and clear boundaries for action in highly controlled environments, alongside better predictions.

6.2 Healthcare: Fragmented systems and context gap weaken agent effectiveness

In healthcare, pilots can demonstrate the value of AI in situations where data is preselected, pathways are simple and human judgement remains firmly in the loop. For example, triage support, prior-authorization workflows or clinical-service orchestration.

Production exposes the sector's inherent variability, which tends to be suppressed in pilots. Records are fragmented across facilities, coding practices are inconsistent and access to information is dependent on roles. Organizations operate differently, and legal and ethical obligations will always influence patient pathways. In this complicated environment, the interplay between data, workflow and escalation is often missed by AI models. An agent may reason plausibly but still fail because it does not have critical information about context, accountability or clinicians' roles.

Agentic autonomy amplifies these risks because errors are not merely informational; they can affect service prioritization, patient communication or treatment-adjacent processes. In this sector, production success often stems from how well the system fits into institutional practice, not just from its technical performance.

6.3 Manufacturing and supply chain: Operational variations impact agentic scale

In manufacturing and supply chain, pilots often succeed where telemetry is clean, workflows are bounded, and local teams can supervise outputs closely. Predictive maintenance, scheduling refinement and exception handling can all look effective when tested on modernized assets or controlled production lines.

When systems scale, enterprise realities can tell a different story. Legacy execution systems, as well as uneven sensor quality and edge constraints can undermine model effectiveness. Operating variations differ across locations and many processes, including planning and logistics, are deeply interdependent.

In real-world operating environments, model performance can degrade. For example, an autonomous recommendation that works in one facility may create issues in another — misbalancing inventory, conflicting schedules or even raising safety risks. When agents can initiate or influence physical operations, these failures are intensified and difficult to reverse.

6.4 Retail and customer operations: Lack of cross-functional coherence limits agentic value

In retail and customer operations, agentic systems can improve service, order orchestration, resolution handling and merchandising decisions because pilots often begin in tightly defined workflows with limited product categories, simple routing logic and close supervision. Production reveals the broader operational dependency chain: inventory, pricing, promotions, logistics, returns, customer communication and policy enforcement must all align in real time.

Cross-functional dependencies can expose limitations that may not appear during pilots. For example, a system may streamline one step — such as service response or recommendation quality — while creating friction elsewhere through inventory mismatches, policy exceptions, or fulfillment errors. Agentic autonomy amplifies this because actions taken in one layer can trigger customer-facing consequences across others. The EY.ai Value Blueprint example for an EY life sciences client illustrates this principle clearly: value did not come from a standalone AI point solution, but from orchestration across more than 100 enterprise resource planning (ERP) systems and end-to-end process redesign. Ultimately, durable production value depends on how well AI is embedded into the broader enterprise system, not just the individual use case.

From a cross-industry perspective, **in enterprise software environments built on legacy systems,** such as low-code and workflow environments, pilots frequently demonstrate that AI can generate configuration files, flows or deployment logic from requirements. These pilot successes typically occur in controlled settings where the configuration domain is narrow, schemas are known, and the number of integration dependencies is limited.

Production exposes a more complex reality. Legacy platforms operate through opinionated schemas, proprietary abstractions, undocumented constraints and environment-specific dependencies that general-purpose AI tools struggle to target with precision. Organizations often encounter challenges with both artifact quality and the delivery predictability deemed critical for production environments. Teams find themselves needing to repeatedly refine outputs to satisfy platform semantics, governance controls, naming conventions, security policies and downstream deployment requirements.

When AI agents are deployed, this mismatch is amplified. A system may appear capable of generating deployable logic but still fail when required to act inside tightly bounded platform rules. These failures often arise from the gap between AI's general reasoning capability and the specificity of production-grade platform semantics.

7. From ‘bolt on’ to ‘built in’ AI

EY is supporting companies across multiple industries to redesign the enterprise through AI – so that the technology becomes intrinsic to how value is created and governed, rather than an incremental addition to existing operations. Our experience suggests that, while ‘bolt-on’ AI produces isolated pilots, ‘built-in’ AI reshapes workflows, roles, control boundaries and architecture with AI at the core. This shift is what separates experimentation from scaled transformation.

7.1 The EY AI-native enterprise architecture

In EY, AI-native enterprises operate across seven interconnected layers:

1	**Customer** Differentiated experiences, products and business models	Hyper-personalized experiences	Innovative products and services	Ecosystems and new business models	
2	**Workforce** Human-AI collaboration models leveraging role-based interfaces and operating systems	Role-based interfaces	Human-AI collaboration	AI-augmented productivity	Skills, enablement and change
3	**Processes** Methodologies and EY proprietary accelerators to reimagine outcomes and processes enabled by AI	Outcome-driven design	Reimagined end-to-end processes	EY proprietary accelerators	Continuous improvement
4	**Trust** Controls, guardrails and security in systems and data, as outlined in EY Responsible AI framework	Responsible AI principles	Security and privacy by design	Risk, compliance and governance	AI guardrails and safety
5	**Intelligence** Encoded enterprise knowledge and advanced AI and analytics that enhance decision-making and insights	Enterprise knowledge engine	Advanced AI and analytics	Insights and decision intelligence	Semantic context layer
6	**Agentic platform** Leverage the EY.ai agentic enterprise tech stack, powered by EY Fabric, to upgrade technology infrastructure, to support AI at scale	Agent orchestration and runtime	Model and tool management	APIs, services and integrations	Scalable, secure infrastructure
7	**Systems of record** Expose sources of truth through reliable connectors for agent and human interaction	Authoritative systems	Reliable connectors and APIs	Data consistency and integrity	Agent and human interaction

Engineered as a coordinated system

- Aligned to business strategy and outcomes
- Seamless human-AI collaboration
- Built on trust, security and responsible AI
- Intelligence embedded across the enterprise
- Scalable, modular and future-ready
- Continuous feedback and improvement

When the seven layers work together, AI drives measurable, sustainable enterprise value.

Cross-cutting enablers (foundation for all layers)

- Data governance and quality
- Identity and access management
- Integration and interoperability
- DevSecOps and MLOps excellence
- Observability, monitoring and audit
- Change management and adoption

7.2 Confidence by design

EY's experience supporting clients suggests that agentic AI scales when confidence is designed into the architecture from the outset — through governed data, bounded autonomy, explicit accountability, observability and reversibility embedded into the operating model itself.

For example, in one of its use cases, EY supported a financial services platform to embed confidence by design. The company wanted to rebuild its approach to reporting suspicious activity through an AI-augmented delivery model across five geographies and more than 10 integrated microservices. The model used enterprise AI coding tools to accelerate some repetitive development tasks, but their performance degraded sharply when the delivery challenge shifted from local code generation to distributed architectural reasoning. EY teams found that the challenge stemmed from gaps in contextual continuity, architectural coherence, governance resilience and refinement discipline that the broader enterprise had not yet fully operationalized. Put simply, the enterprise's lack of readiness for AI was the source of failure.

8. A proposed enterprise AI readiness framework

To close the pilot-to-production gap, enterprises need a readiness model built for autonomous systems. Five pillars matter most.

8.1 Architecture principles

'Built-in' AI starts with architectural discipline: designing the enterprise so agents can coordinate across systems without inheriting an uncontrolled blast radius.

Building open, secure-by-design architecture that supports orchestration, interoperability and scalable execution across models, tools and systems is a foundational requirement. The goal is to provide the technical conditions that allow autonomous systems to operate safely, predictably and accountably across the enterprise. This includes reliable systems of record, clear separation between data, execution and control layers, and runtime environments that can support policy-aware orchestration across business functions.

In production, architectural readiness depends on concrete mechanisms: identity-aware tool access, environment-specific permissioning, deterministic workflow boundaries around probabilistic systems, isolation of high-risk actions and resilient connectors into core platforms. Architecture must be able to distinguish experimentation from operational deployment through promotion gates, rollback paths and environment segmentation.

8.2 Operating model design

Enterprises that define how humans and agents work together are best positioned to realize sustainable value from autonomous AI. That means making authority boundaries clear, setting escalation paths, redesigning roles and clarifying business ownership of outcomes.

Ownership models must be explicit — defining who sponsors the use case, who governs controls, who monitors behavior in production and who is accountable when outcomes diverge from intent. This requires named decision rights, structured handoff points between human and agent, escalation rules for exceptions and lifecycle stewardship that persists after deployment.

In our experience, organizations that treat AI systems like managed operating capabilities outperform others. They have designated business owners responsible for value realization, technical owners responsible for reliability, risk owners responsible for control effectiveness and empowered frontline users ready to intervene when trust signals deteriorate.

8.3 Governance, risk and control

Governance must move from committee oversight to runtime enforcement. This is increasingly mandated through regulatory changes but is also best practice to drive successful, confident AI implementation. In EY, our "compliance as code" approach to responsible AI and control mechanisms keep autonomous systems aligned, traced and reversible.

For example, when deploying enterprise AI coding tools at scale, organizations that include per-task cost and productivity measurement as part of runtime governance can quickly course correct when results indicate issues, including by shifting routine work to flat-rate tools and reserving metered tools for high-value tasks.

8.4 AI engineering and observability

Production reliability extends beyond delivery pipelines. It requires AI-native observability, continuous evaluation, rapid rollback capabilities, and comprehensive incident reconstruction. In practical terms, enterprises also need visibility into how a system reasoned, which tools it invoked, which sources informed its output, what policies were triggered and where behavior deviated from expected baselines.

Achieving this requires engineering for operational transparency from the outset:

- Structured traces for multi-step execution
- Provenance capture for retrieved or referenced data
- Audit logs for tool use and action approval
- Evaluation pipelines that test groundedness and task completion quality
- Alerting for abnormal behavior patterns rather than only uptime failures

Feedback loops are also critical to the design, helping organizations learn from production incidents and improve future performance through prompt refinements, tool constraints, model selection or workflow redesign.

8.5 Metrics and value accountability

Pilot metrics often focus on model performance or user sentiment. Production metrics must go further, connecting task-level execution to enterprise outcomes. Enterprises need to measure both value creation and control integrity simultaneously.

A robust measurement model should therefore track:

- Baseline versus actual effort
- Rework introduced by AI outputs
- Decision latency and human override frequency
- Policy exceptions
- Adoption depth and cost-to-value ratios by workflow

Governance metrics should measure whether risky actions were escalated appropriately, whether explanations were available when needed, whether trusted data sources were used and whether autonomous behavior stayed within defined boundaries.

Without this dual view of value creation and control integrity, leaders cannot distinguish apparent adoption from durable operational value, nor can they calibrate where autonomy should expand or remain bounded.

The five pillars to close the pilot-to-production gap and scale autonomous AI with confidence are summarized below:

Vision: trusted, accountable, and value-driving autonomous AI — built-in, governed in and measured in.

1 Architecture principles

Design the enterprise for safe, scalable autonomy

Build open, secure-by-design architecture that supports orchestration, interoperability and scalable execution across models, tools and systems, enabling autonomous systems to operate safely, predictably and accountably.

Focus areas

- Reliable systems of record
- Clear separation of data, execution and control layers
- Runtime environments for policy-aware orchestration across business functions

Key enablers

- Identity-aware tool access
- Environment-specific permissioning
- Deterministic workflow boundaries around probabilistic systems
- Isolation of high-risk actions
- Resilient connectors to core platforms
- Promotion gates, rollback paths and environment segmentation

2 Operating model design

Define how humans and agents work together

Make authority boundaries clear, set escalation paths, redesign roles and clarify business ownership of outcomes.

Focus areas

- Clear decision rights and accountability
- Structured handoffs between human and agent
- Escalation rules for exceptions
- Lifecycle stewardship after deployment

Key enablers

- Business owners for value realization
- Technical owners for reliability
- Risk owners for control effectiveness
- Empowered frontline users to intervene when trust signals decline
- Treat AI systems as managed operating capabilities

3 Governance risk and control

Enforce governance at runtime

Move from committee oversight to runtime enforcement. Regulatory mandate and best practice for confident AI.

Focus areas

- Compliance as code and policy enforcement
- Traceability, auditability and reversibility
- Risk-based controls embedded in execution
- Continuous monitoring of control effectiveness

Key enablers

- Compliance as code (policies embedded in workflows)
- Real-time logging, traceability and immutable audit trails
- Per-task cost and productivity measurement
- Adaptive controls and guardrails
- Reversible actions and rollback capability

Example: use per-task cost and productivity measurement (e.g., AI coding tools) to course correct and allocate tools by value.

4 AI engineering and observability

Make behavior visible, explainable and improvable

Engineer for operational transparency, resilience and continuous improvement.

Focus areas

- End-to-end observability of behavior
- Evaluation of quality, safety and intent
- Rollback and incident reconstruction
- Feedback loops for continuous learning

Key enablers

- Structured traces for multi-step executions
- Provenance capture for data and tool usage
- Audit logs for tool use and action approvals
- Evaluation pipelines (groundedness, task completion, quality)
- Alerting for abnormal behavior patterns
- Feedback loops to refine prompts, tools, models and workflows

5 Metrics and value accountability

Measure value creation and control integrity

Correct task-level execution to enterprise outcomes. Measure value and control integrity simultaneously.

Focus areas

- Value realization across workflows
- Control integrity and risk management
- Cost-to-value and efficiency
- Calibrate where autonomy should expand or remain bounded

Key enablers

- Baseline vs. actual effort
- Rework introduced by AI outputs
- Decision latency
- Human override frequency
- Policy exceptions
- Adoption depth
- Cost-to-value ratios by workflow
- Governance metrics: escalations, explanations, trusted data use and boundary adherence

Foundational enablers (across all pillars)

Trusted data	Security and privacy	Change and skills	Ecosystem and partnerships	Culture and ethics
High quality, current, accessible and governed enterprise data	Built-in security, privacy by design and data protection	Workforce readiness, AI literacy and role evolution	Leverage partners, tools and platforms responsibly	Responsible AI culture with ethics at the core

Outcome: close the pilot-to-production gap and scale autonomous AI that delivers business value safely, predictably and sustainably.

9. Implications for CEOs, CIOs, CDOs and CAIOs

For CEOs, the issue is strategic. The focus is shifting from AI's ability to automate tasks to redesigning the enterprise and enhance its ability to govern autonomous work at scale. The competitive divide will form between organizations that industrialize trustworthy autonomy and those that remain stuck in pilot portfolios. EY.ai Value Blueprints are right to frame this as the difference between renovating the past and rebuilding for the future.

For **CIOs and CTOs,** the implication is architectural. Agentic systems require secure, interoperable and observable runtime environments alongside access to foundation models. The challenge is to build a control plane that allows agents to operate with bounded autonomy, governed tool access, resilient orchestration and auditable behavior. Without robust architecture and observability, scale may not be possible at all.

For Chief Data Officers (CDOs), the priority is to see that data remains current, governed and accessible across enterprise workflows. Agents need context that is current, explainable, permissioned and continuously usable across workflows. That requires semantic consistency, lineage, integration across silos and access models that support autonomous execution without compromising control. As discussed in the data architectures increasingly need to be designed around trusted, connected and consumable data foundations that can support AI and agentic workflows at scale.

For Chief AI Officers (CAIOs), the role is transforming to that of an orchestrator rather than holding siloed ownership. The CAIO must connect architecture, governance, data, evaluation, workforce readiness and business value into one operating system for AI, with clear separation of responsibilities between central platforms and domain delivery teams. Evidence suggests that organizations with dedicated AI leadership and centralized capabilities are materially better positioned to sustain AI in production.

10. Conclusion: Intelligence is not the constraint

Enterprise AI success increasingly depends on organizational readiness to deploy and govern AI at scale. As systems become more autonomous, organizations that fail to redesign around context, control, workflow and accountability will experience more production failures. Long-term value will come from converting autonomous capabilities into governed, auditable and resilient enterprise performance.

The agentic scale trap is real, but it is avoidable. The path forward is to match AI ambition with enterprise redesign. Organizations that shift from 'bolt-on' AI to 'built-in' AI can create tangible long-term value.

References

1. **EY, "How Tech companies can break out of the AI ROI trap."** Adoption is advancing faster than organizations' ability to translate it into consistent enterprise value. EY research indicates that 16% of companies report no ROI from GenAI initiatives and fewer than half achieve substantial returns, with value often emerging more slowly and unevenly than expected. Source: https://www.ey.com/en_gl/insights/tech-sector/how-can-you-break-out-of-the-ai-roi-trap

2. **EY, "The EY Work Reimagined study: How can a rebalance of power help re-energize your workforce?"** Reaping value from AI investments requires mastering the tensions between talent and tech or risking an AI productivity drop of 40%. "Nearly nine out of 10 employees now use AI at work"; "only 28% of organizations are positioned to turn AI deployment into high-value outcomes" Source: https://www.ey.com/en_us/insights/workforce/work-reimagined-survey

3. **EY, "AI adoption outpaces governance as risk awareness among the C-suite remains low."** The constraint is no longer access to AI models, but the ability to redesign workflows and embed consistent governance. EY research shows that while organizations are investing heavily in AI, many remain focused on pilots, and value realization is limited by gaps in execution, governance and end-to-end process transformation. Source: https://www.ey.com/en_gl/newsroom/2025/06/ey-survey-ai-adoption-outpaces-governance-as-risk-awareness-among-the-c-suite-remains-low

4. **RAND Corporation, "Avoiding the anti-patterns of AI: the root causes of failure for artificial intelligence projects and how they can succeed" (2024), by James Ryseff, Brandon De Bruhl and Sydne J. Newberry.** The report notes that more than 80% of AI projects fail, primarily because of misaligned objectives, weak data, inadequate infrastructure and poor integration into business workflows. Source: https://www.rand.org/content/dam/rand/pubs/research_reports/RRA2600/RRA2680-1/RAND_RRA2680-1.pdf

5. **Harvard Business Review; "The last mile problem slowing AI transformation" (2026), by Karim R. Lakhani, Jared Spataro, and Jen Stave.** Many companies are stuck in the 'last mile' of AI transformation, where technical capability collides with ownership, incentives, accountability, process debt, and architectural complexity. Source: https://hbr.org/2026/03/the-last-mile-problem-slowing-ai-transformation

6. **Gartner, "Gartner survey finds 45% of organizations with high AI maturity keep AI projects operational for at least three years" (2025).** Gartner found that 45% of leaders in high-maturity organizations said AI initiatives remain operational for at least three years, compared with 20% in low-maturity organizations. Source: https://www.gartner.com/en/newsroom/press-releases/2025-06-30-gartner-survey-finds-forty-five-percent-of-organizations-with-high-artificial-intelligence-maturity-keep-artificial-intelligence-projects-operational-for-at-least-three-years

7. **European Union (EUR-Lex), "Rules for trustworthy artificial intelligence in the EU" (2025 update to Regulation (EU) 2024/1689 summary).** The EU AI Act emphasizes transparency, oversight, human intervention and documentation for higher-risk AI uses. Source: https://eur-lex.europa.eu/EN/legal-content/summary/rules-for-trustworthy-artificial-intelligence-in-the-eu.html

8. **EY, "Architecting an agentic workforce at the EY organization" (2025).** The paper outlines the runtime environments, protocols, multi-agent workflows and layered control mechanisms needed to scale agentic systems safely and effectively in production environments. Source: https://www.ey.com/content/dam/ey-unified-site/ey-com/en-gl/technical/documents/ey-gl-architecting-an-agentic-workforce-at-the-ey-organization-01-26.pdf

References

9. **EY, "ModelOps frameworks bridge AI governance and value" (2025).** ModelOps extends beyond DevOps and MLOps to address the unique governance challenges of AI systems. Source: https://www.ey.com/en_us/insights/ai/modelops-frameworks-bridge-ai-governance-and-value

10. **EY, "AI business results depend on data quality."** Unlocking the true power of artificial intelligence (AI) depends on high-quality, well-governed data as the foundation for smart outcomes. Source: https://www.ey.com/en_us/alliances/ai-business-outputs-intelligent-data

11. **EY, "AI use case management"** Driving value and success in the age of agentic intelligence and multi-AI systems. Source: https://www.ey.com/en_us/cro-risk/ai-use-case-management

12. **Microsoft Learn, "Observability for generative AI and agentic AI systems" (2026).** Agentic and GenAI systems require observability beyond traditional logs, metrics and traces, including identity context, retrieval provenance, tool invocations, policy decisions and behavioral baselines to support monitoring, governance and incident response. Source: https://learn.microsoft.com/en-us/security/zero-trust/sfi/observability-ai-systems

Authors

Pablo Cebro

Partner, EY Platforms and Emerging Technologies Leader, Client Technology

EY GDS (CS) Argentina S.R.L.

Biju Varghese

Partner, NITRO Leader – Client Technology

EY Global Delivery Services (India) LLP

Tyler Buffie

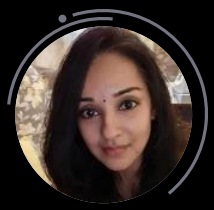
Principal, Applied Innovation Strategy Lead, EY Growth and Innovation

Ernst & Young U.S. LLP

Sumanta Mukherjee

Associate Director, NITRO Technology Research Lead – Client Technology

EY Global Delivery Services (India) LLP



Asha George

Associate Director, NITRO Innovation Lead – Client Technology

EY Global Delivery Services (India) LLP

EY | Building a better working world

EY is building a better working world by creating new value for clients, people, society and the planet, while building trust in capital markets.

Enabled by data, AI and advanced technology, EY teams help clients shape the future with confidence and develop answers for the most pressing issues of today and tomorrow.

EY teams work across a full spectrum of services in assurance, consulting, tax, strategy and transactions. Fueled by sector insights, a globally connected, multidisciplinary network and diverse ecosystem partners, EY teams can provide services in more than 150 countries and territories.

All in to shape the future with confidence.

EYG no. SCORE 113921-26-GBL

ED None