

THE EXECUTIVE IMPERATIVE

Moving AI from pilots into production

Why most enterprises stall and what it actually takes to scale AI

The time for pilots is coming to an end. Production is the new standard.

Enterprise AI investment has entered a new phase. The era of experimentation is no longer enough. Boards and executives are now asking a harder question: **Where is the return?**

Across industries, organizations report progress in pilots but struggle to move AI into reliable production. Models appear promising in controlled environments but break down when embedded into real workflows. Outputs become inconsistent. Trust erodes. Momentum slows.

Central conclusion

AI production does not fail because of model limitations or team capability. It fails because organizations attempt to scale AI without addressing leadership decisions and operational readiness.

Part I
Executive mandate



Part II
Operational readiness

Part I is written for executives. It explains why leadership involvement is required and identifies the three responsibilities that unlock production.

Part II is written for teams responsible for execution. It outlines the six operational pillars required to run AI reliably in production, with special emphasis on context orchestration as the primary failure point.

[Explore the interactive Executive Imperative microsite →](#)

THE PROBLEM

The AI execution gap

Enterprise AI adoption has accelerated rapidly. Pilots are common. Demos are impressive. Early signals often look encouraging. Yet production deployment remains rare.

In most large organizations, this creates a dangerous illusion: the pilot looks like proof of readiness. In reality, the pilot is proof that a small team can stabilize variables by hand. It does not prove the business can run the same workflow reliably across real data, real systems, and real edge cases.

Boards and executive teams are now asking for outcomes, not activity. The question is no longer "Are we experimenting with AI?" It is "Where is the measurable impact, and can we trust it?" The gap between those two questions is where most programs stall.

Reality check: MIT Sloan Management Review reports that 95% of enterprise AI initiatives fail to deliver measurable ROI even as investment continues to rise.

The pilot-to-production gap

Pilots succeed

Manually stabilized
Curated data
Human oversight



Production fails

Fragmented systems
Inconsistent data
No human buffer

This gap is often misinterpreted as a tooling issue or a skills gap. In reality, it is an **execution gap** rooted in how organizations operate.

Pilots succeed because teams manually stabilize the environment. Data is curated.

Assumptions are explicit. Humans review outputs before action is taken. These conditions create the appearance of readiness.

Production removes those buffers.

When AI is exposed to fragmented systems, inconsistent data, undocumented workflows, and competing definitions, it inherits the instability of the environment. **AI does not compensate for this complexity. It amplifies it.**

The result is predictable: Pilots persist. Production stalls.

This is why executive involvement matters. Teams can keep running pilots and tuning prompts indefinitely, but without leadership action to stabilize the environment, production will continue to behave like an experiment. Part I lays out the three responsibilities that determine whether AI becomes a reliable operating capability or remains stuck in perpetual proof of concept.

Why AI pilots stall

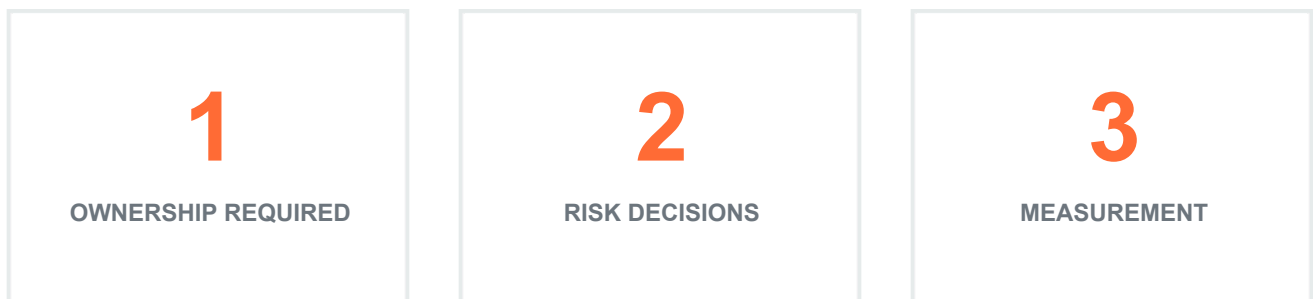
AI pilots are designed to learn. Production systems are designed to operate.

What changes in production is not ambition or talent. It is exposure to the real enterprise environment: inconsistent fields, competing "systems of record," undocumented exceptions, and fragile integrations. In pilots, teams quietly compensate. In production, that compensation turns into permanent, expensive scaffolding.

Once outputs vary without a clear explanation, trust breaks. Operators start double checking results, adding manual review steps, and creating workarounds. From the outside, this looks like slow adoption. In reality, it is the rational response to an unstable system.

In pilots, success is subjective. In production, success must be measurable.

Most AI initiatives stall because they avoid three realities of production:



First: Production requires ownership

Someone must be accountable for outcomes, not just experimentation.

AI systems cannot operate in production without clear ownership. In early experimentation, multiple teams can contribute ideas and share responsibility because the goal is learning, not reliability. Production is different. Once AI is embedded into core workflows, decisions about data sources, system behavior, failure handling, and escalation paths must be made consistently and enforced decisively. When ownership is unclear, these decisions are delayed or avoided. Issues surface repeatedly but remain unresolved. Standards are debated but not applied. As a result, AI systems remain trapped in pilot mode, stable only when manually supervised. Production AI requires a single accountable owner with the authority to make tradeoffs, resolve conflicts, and stand behind outcomes. Without that ownership, reliability cannot be achieved.

Second: Production requires risk decisions

Acceptable error must be defined explicitly.

AI does not behave deterministically. It produces probabilistic outputs that vary based on inputs, context, and conditions. In pilots, this variability is tolerated because humans review results and intervene when necessary. In production, undefined risk becomes a structural blocker.

Teams cannot design automation, validation, or escalation mechanisms without knowing what level of error is acceptable and where it is not. When risk tolerance is left implicit, teams default to caution. Human review expands. Automation is constrained. Systems remain technically capable but operationally unusable at scale. Decisions about acceptable error are not technical decisions. They are business decisions that define trust, exposure, and accountability. Until those decisions are made explicitly, AI cannot move into reliable production.

Third: Production requires measurement

Outputs must be evaluated against clear baselines and targets.

Production systems must be measured. AI is no exception. In experimentation, success is often described through demonstrations, qualitative feedback, or isolated examples. These signals are insufficient once AI is expected to operate continuously and autonomously. Without defined baselines, performance targets, and monitoring thresholds, organizations cannot determine whether AI is improving outcomes, introducing risk, or simply shifting effort elsewhere. Measurement is how trust is established and maintained. It enables leaders to approve broader deployment, detect degradation, and intervene when performance drifts. Without measurement discipline, AI remains advisory by default. It may appear useful, but it cannot be relied upon. Production requires evidence, not promise.

When these conditions are absent, teams do what they can. They iterate. They refine prompts. They test new models. But they remain constrained by decisions that sit above them.

This is not a team failure. It is an organizational one.

Why AI production fails without executive action

Enterprise AI is not failing because teams lack ideas, tools, or technical capability. It is failing because organizations are attempting to operationalize AI without changing the conditions under which their teams operate.

Across industries, the pattern is consistent. AI pilots show promise in controlled environments, yet stall or break down when exposed to production realities. Models behave inconsistently. Outputs cannot be explained or defended. Trust erodes. Momentum slows. Executives grow frustrated with the gap between investment and results.

This gap is often misdiagnosed as an execution issue. It is not.

The evidence is clear that experimentation is not the problem. Most enterprises are experimenting aggressively and spending heavily on AI, yet only a small fraction have achieved meaningful scale or measurable return. The issue is **readiness**.

40%

Agentic AI projects expected to be delayed or canceled due to governance and integration challenges

Source: Gartner

44%

Organizations report having a formal AI policy in place

Source: Vanta

\$12.9M

Average annual cost of poor data quality per organization

Source: Gartner

These are not technology constraints. They are leadership constraints. Governance, risk tolerance, and measurement discipline cannot be delegated to pilots and task forces.

Why production environments are different

AI behaves predictably only when the environment around it is predictable. Pilots succeed because teams manually stabilize the system. Data is curated. Exceptions are handled by humans. Outputs are reviewed before action is taken. These conditions do not exist in production.

Production environments are fragmented, governed by multiple systems of record, inconsistent definitions, undocumented workflows, and competing priorities. **AI does not compensate for this complexity. It exposes it.**

Teams cannot resolve fragmented data ownership. They cannot define acceptable enterprise risk. They cannot impose measurement discipline across functions. **These are leadership responsibilities.**

AI failure is not a technology problem. It is a systems management problem. And systems management is an executive function.

The three executive responsibilities that unlock AI production

AI moves into production only when executives resolve three issues that teams cannot solve on their own.

1. Establish clear ownership for AI outcomes

AI production fails when ownership is fragmented. Most AI initiatives span data teams, IT, engineering, operations, and business units. Each group contributes, but accountability is diffuse. Decisions stall. Tradeoffs are avoided. Integration is delayed.

Executives must assign ownership for AI outcomes, not experimentation. That ownership must include authority to enforce standards, resolve conflicts, determine readiness, and shut down initiatives that cannot meet operational requirements.

Without this clarity, AI remains a collection of pilots.

2. Define and Accept an Explicit Risk Posture

AI introduces probabilistic behavior into environments designed for determinism. In pilots, risk is absorbed by humans. In production, risk must be designed for.

Executives must define where variability is acceptable and where it is not. This includes decisions about human oversight, validation requirements, escalation paths, and fallback mechanisms.

When risk tolerance is implicit or inconsistent, teams default to caution. Automation stalls. Production is delayed indefinitely.

3. Enforce Measurement Discipline

Pilots survive on promise. Production survives on proof.

Executives must require AI to be measured like any other operational investment. Baselines, targets, and ongoing performance tracking are not optional.

Measurement is not reporting. It is the mechanism by which AI earns the right to operate autonomously. Without it, AI remains advisory at best.

What these responsibilities unlock

When leadership resolves ownership, risk, and measurement, the constraint shifts. The question is no longer whether AI should move into production. It becomes whether teams have the operational capabilities required to execute reliably.

That is the focus of Part II.

[Visit the companion microsite for a more interactive experience →](#)

The six pillars required to run AI in production

Once executive responsibilities are fulfilled, execution becomes possible. It is also where most organizations fail.

The most common operational failure is not model selection or prompt quality. It is the absence of context orchestration.

Without context orchestration, every downstream pillar degrades. Pilots stabilize. Production never arrives.

Most teams believe they are failing later in the AI lifecycle.

In practice, they fail at the very beginning.

Teams often focus their effort on prompts, models, and validation layers because those are the most visible components of AI systems. But these efforts are compensating behaviors. They exist because the underlying context supplied to AI systems is incomplete, inconsistent, or unstable.

When context is not orchestrated:

- Prompts must embed business logic that should live upstream
- Models receive conflicting or stale signals
- Validation becomes reactive instead of preventative
- Human review expands instead of shrinking

This creates a self-reinforcing cycle. Each downstream fix stabilizes pilots temporarily while making production less achievable.

1. Context orchestration

The ability to reliably assemble the right information, constraints, and signals at the moment of execution. This is the primary failure point.

2. Prompt management

Prompts refine behavior when context is stable. Without context, prompts compensate and become brittle.

3. Model orchestration

Routing across models adds value only after inputs are stabilized.

4. Hallucination management

Reactive validation catches symptoms. Preventative context reduces hallucinations at the source.

5. Hybrid integration

AI creates value only when embedded where work happens. Inconsistent outputs undermine adoption.

6. KPI and ROI measurement

Measurement assumes stability. Without consistent context, metrics capture noise rather than signal.

Pillar 1: Context orchestration

The primary failure point

Context orchestration is the ability to reliably assemble the right information, constraints, and signals at the moment of execution.

Most teams fail here first. When context is inconsistent, everything downstream degrades. Prompts become fragile because they are trying to compensate for missing or contradictory inputs. Models appear unreliable because they are forced to guess. Hallucination controls trigger constantly. Integrations get flooded with exceptions. Measurement becomes noisy because there is no stable baseline. In pilots, a small team manually stabilizes context every time. In production, that stabilization must be enforced by the system with explicit rules for required fields, conflict resolution, edge cases, and formatting.

In pilots, context is curated manually. In production, context must be assembled dynamically across fragmented systems.

When context is unstable, AI is forced to infer meaning from incomplete or conflicting inputs. Outputs become inconsistent. Trust erodes.

This is where most AI initiatives break.

What context orchestration includes in practice

Context orchestration is not a single system or dataset. It is the operational capability to assemble and govern the information AI requires at runtime.

- Authoritative data sources and ownership definitions
- Consistent field definitions and business rules
- Timeliness and freshness guarantees
- Explicit constraints and exclusion logic
- Documented assumptions that survive handoffs

In pilots, humans assemble this context manually. In production, it must be assembled automatically, consistently, and audibly.

The minimum viable context for production

Organizations do not need perfect data to reach production. They need stable data.

- One agreed source of truth per decision domain
- Clear ownership for resolving conflicts
- Defined failure modes and fallback behavior

Without these foundations, no amount of model tuning will produce reliable outcomes. Failures in context orchestration trigger predictable downstream effects across all other pillars:

Pillar 2: Prompt management

When context is missing or unstable, prompts are forced to compensate. Business logic and assumptions are embedded into prompts, creating brittle systems that are difficult to govern or explain.

In production, prompts must refine behavior, not replace missing context.

Pillar 3: Model orchestration

Teams often blame models for instability. Without reliable context, routing across models only accelerates inconsistency. Model orchestration adds value only after inputs are stabilized.

Pillar 4: Hallucination detection and remediation

Hallucinations increase when context is incomplete. Reactive validation catches symptoms after the fact.

Preventative context stabilization reduces hallucinations at the source.

Pillar 5: Hybrid integration into existing workflows

AI creates value only when embedded where work already happens. Inconsistent outputs undermine trust.

Adoption stalls. Automation is bypassed. **Reliable context enables adoption.**

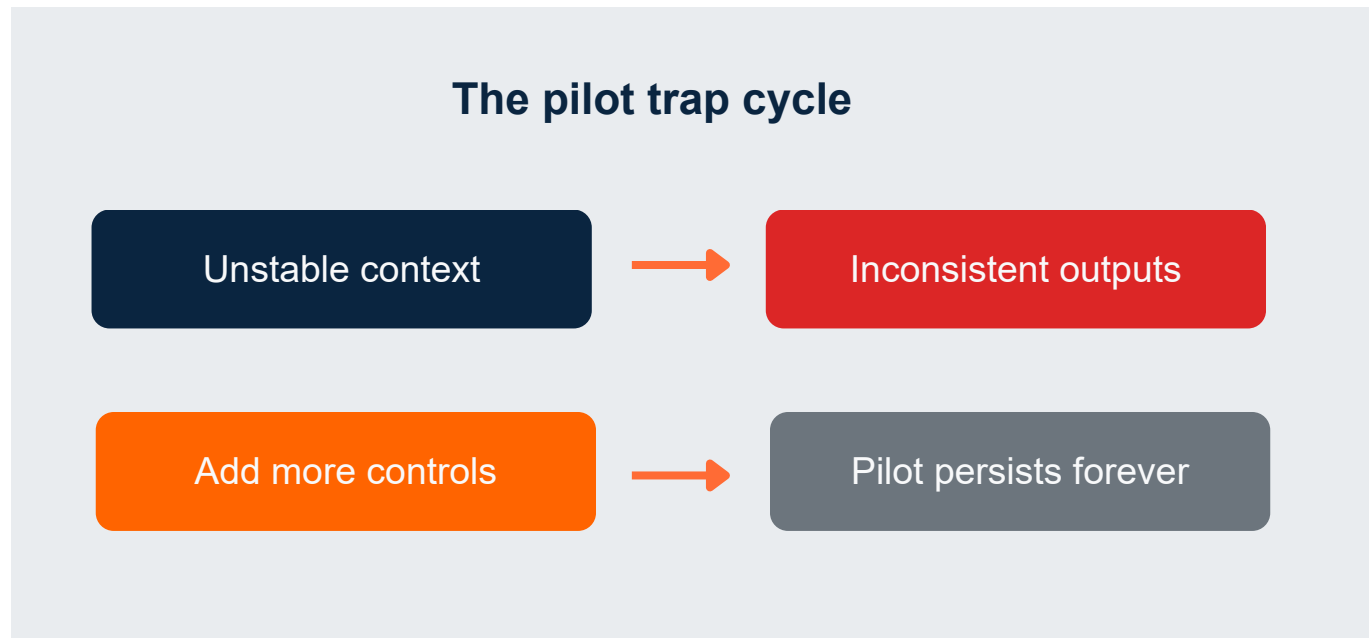
Pillar 6: KPI and ROI measurement

Measurement assumes stability. Without consistent context, performance metrics capture noise rather than signal. Executive confidence erodes. Pilots persist. **Stable context makes measurement meaningful.**

Why pilots never escape this trap

Pilots tolerate instability because humans compensate for it. Production cannot.

Organizations add downstream controls to stabilize pilots—validation layers, human review, escalation workflows—making systems safer but less scalable.



Production AI requires starting with context, not reacting to its absence.

Closing the loop between leadership and execution

Executives create the conditions for production. Teams build the systems that make production sustainable.

The path forward

Organizations that align both—executive commitment and operational readiness—move beyond experimentation and begin realizing durable returns from AI.

What success requires

From leadership:

- Clear ownership assigned to individuals with authority to enforce standards
- Explicit risk decisions that define where AI can operate autonomously
- Measurement discipline that treats AI like any other operational investment

From teams:

- Context orchestration as the foundation—not an afterthought
- Operational infrastructure across all six pillars
- Integration into existing workflows where work already happens

The organizations that succeed will not be those with the most sophisticated models. They will be those that understand AI is an operational discipline, not a technology experiment.

Next steps

Moving from pilots to production requires both strategic commitment and operational capability. Organizations ready to make this transition should assess:

- Whether executive ownership, risk tolerance, and measurement expectations are explicit
- Whether context orchestration infrastructure exists to support production workloads
- Whether operational teams have the authority and resources to execute

[Access additional tools and resources →](#)

The era of AI pilots is ending. Production is the new standard.

OPENRISE®

Learn more at openprisetech.com

[Visit The Executive Imperative Microsite →](#)

Research sources and evidence base

This whitepaper and the accompanying AI Pilots → Production microsite are informed by a synthesis of publicly available analyst research, academic studies, industry surveys, and company disclosures. Statistics and examples are cited for directional insight and should not be interpreted as guarantees of outcome.

Enterprise AI adoption, scale, and ROI

Only a small minority of organizations have achieved AI at scale

McKinsey's State of AI research consistently finds that a limited percentage of companies report AI deployed at scale and delivering material business impact.

Source: McKinsey & Company, The State of AI (multiple annual editions)

Enterprise AI investment continues to outpace realized returns

Industry and analyst reporting indicates tens of billions of dollars in annual enterprise AI spend, with most organizations still struggling to demonstrate sustained ROI.

Sources: McKinsey; Gartner; industry analyst synthesis

Early AI ROI signals do not reliably translate into durable outcomes

Research shows many organizations report early ROI indicators within months, but far fewer sustain or scale those gains into production systems.

Source: Rackspace Technology, AI Readiness and Adoption Report, 2025

Governance, risk, and production failure

A substantial portion of AI projects are expected to be delayed or canceled

Analyst projections highlight governance gaps, data access issues, and integration complexity as leading causes of AI project failure.

Source: Gartner, AI and Agentic Systems research and press releases (2024–2025)

Many organizations lack formal AI governance frameworks

Surveys indicate that fewer than half of organizations report having a formal, company-wide AI policy governing validation, risk, and usage.

Source: Vanta, State of AI Governance, 2025

Lack of defined success metrics is a common blocker to AI production

Analyst research consistently identifies unclear KPIs and success criteria as a barrier to scaling AI initiatives beyond pilots.

Source: 451 Research, enterprise AI adoption analysis (qualitative findings)

Data volume, data quality, and operational readiness

Global data creation continues to accelerate rapidly

Forecasts project global data volumes reaching approximately 175 zettabytes by the mid-2020s and exceeding 600 zettabytes by the end of the decade.

Sources: IDC Global DataSphere; Statista data forecasts

Organizations spend significant time preparing and validating data

Research shows a majority of organizations devote a substantial share of employee time to data preparation, cleansing, and validation.

Source: Precisely and Drexel University research

Poor data quality imposes material financial cost on enterprises

Analyst estimates indicate that poor data quality costs organizations millions of dollars annually through inefficiency, rework, and missed opportunity.

Source: Gartner data quality research

Enterprise data environments grow continuously and degrade without governance

Analyst research shows B2B data volumes expand rapidly, compounding operational debt when governance and standardization do not keep pace.

Source: SiriusDecisions research on data growth and RevOps performance

Organizational and systems constraints

Fragmented enterprise systems create inconsistent operational context

Analyst research highlights that disconnected CRMs, marketing platforms, data warehouses, and operational systems often produce conflicting definitions and signals.

Sources: Gartner; HubSpot; Forrester-referenced analysis

Organizations with stronger data alignment outperform peers

Studies indicate organizations with higher data integrity and cross-functional alignment demonstrate stronger revenue and operational performance.

Source: SiriusDecisions research

Illustrative production AI examples

The following examples are included to demonstrate common patterns among organizations that have embedded AI into production workflows. These examples are illustrative and do not imply universal results.

IBM

Use of governed data foundations to support generative AI in marketing and content operations.

Source: Axios reporting, 2024

Salesforce

Deployment of trust and governance layers to support AI assistants within CRM environments.

Source: Salesforce product documentation and disclosures, 2025

Verizon

Use of generative AI to support customer service and retail operations, with stated goals of reducing churn and improving service efficiency.

Source: Reuters reporting, 2024

Microsoft (GitHub Copilot)

AI integrated directly into developer workflows to augment productivity.

Source: Microsoft public disclosures and GitHub research summaries

Adobe, Intuit, ServiceNow, SAP

Embedded AI assistants operating within core production systems.

Sources: Public product documentation and analyst coverage

Interpretation and use

All statistics are used directionally to illustrate common enterprise patterns.

No data point should be interpreted as a guaranteed benchmark or outcome.

Company examples reflect reported use cases, not audited performance claims.

Conclusions are based on convergence across multiple independent sources rather than any single study.