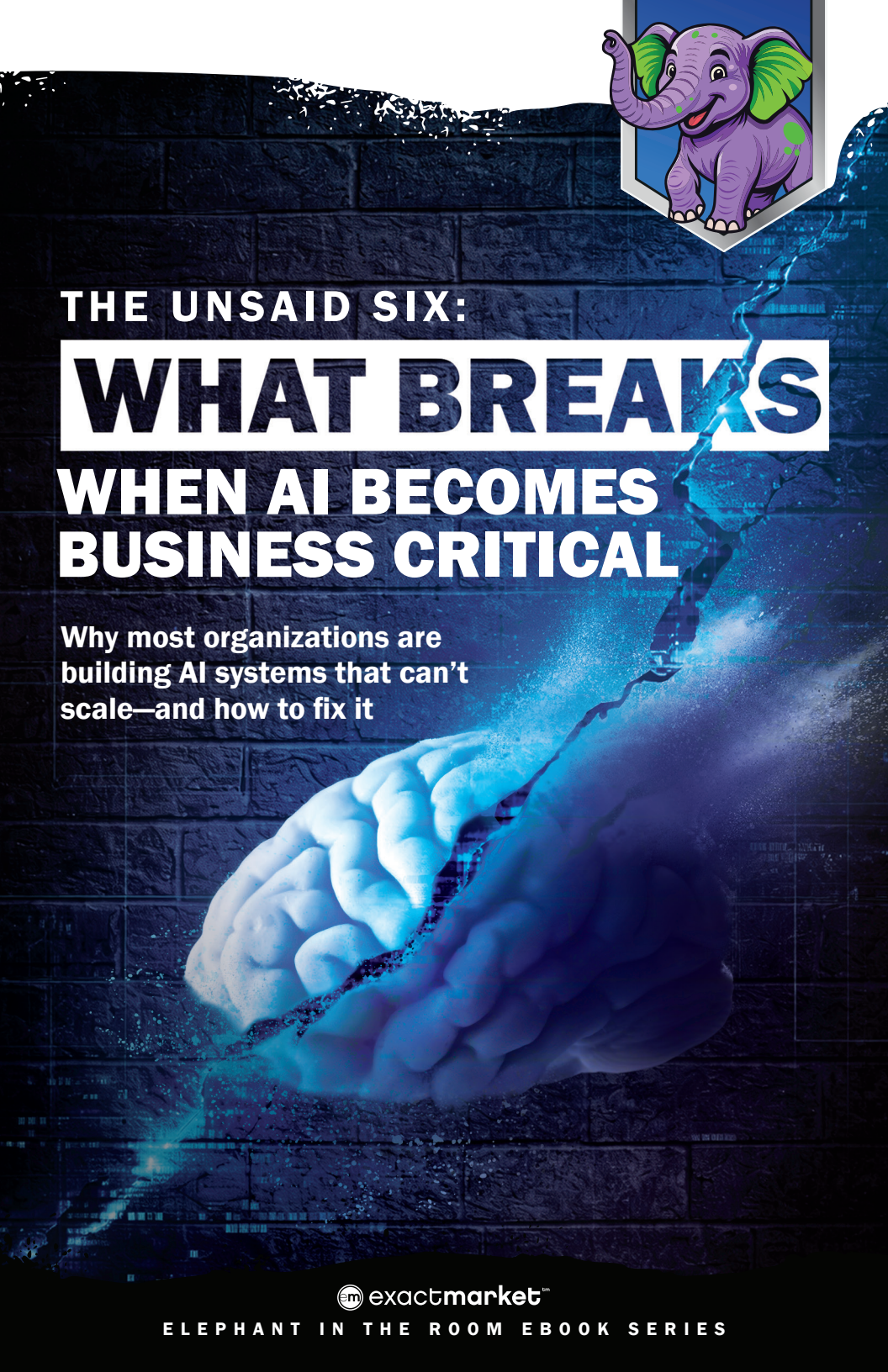


THE UNSAID SIX:

WHAT BREAKS WHEN AI BECOMES BUSINESS CRITICAL

Why most organizations are building AI systems that can't scale—and how to fix it

exactmarket℠

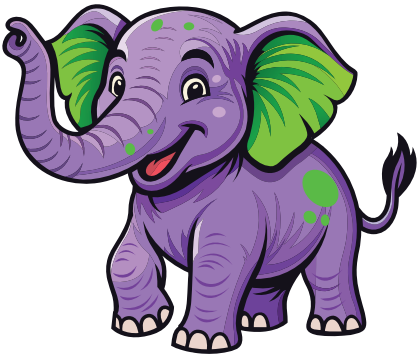
ELEPHANT IN THE ROOM EBOOK SERIES

TABLE OF CONTENTS

Welcome to the Elephant in the Room!	03
Introduction: The Inflection Point No One Prepared For	04
Permissions	05
Auditing	06
Reliability	07
Data Governance	08
Portability	09
Sovereignty	10
Conclusion	11

WELCOME TO THE ELEPHANT IN THE ROOM!

You have in your hands a guide designed to call out the elephant in the room: a topic that's too important to be ignored but isn't getting the attention it deserves.



THE ELEPHANT

The Control Plane Gap: Why AI Pilots Succeed and Production Deployments Fail

Your AI pilot worked brilliantly. Accuracy exceeded benchmarks. Users loved it. Business value was clear. Then production deployment stalled, launched successfully, and failed months later in ways no one anticipated.

The problem isn't the model. The model is fine.

The problem is that no one designed the control plane. No one defined who authorizes the AI to act, how to prove what it did, or how to stop it when things go wrong. These questions tend to feel too early in pilots; rather, they become a concern when AI systems begin making decisions that impact customers, compliance, and, most importantly, revenue.

When AI systems transition from answering questions (chatbots) to taking actions that support or manipulate business processes (such as automated agents), organizations can face six new failure modes. These may appear to be "AI problems," but they are actually control, operations, and governance challenges that ultimately determine whether production deployments succeed or fail under operational pressure.

The elephant in the room is that most organizations are building AI systems that they cannot safely operate at scale. The gaps only become visible after commitment, investment, and dependency.

INTRODUCTION

The Inflection Point No One Prepared For

AI becomes business-critical when organizations depend on AI systems to execute business-affecting decisions, automate workflows, and deliver customer outcomes under time pressure, with regulatory inspection, real financial exposure, and safety implications. The problem is that many businesses are sprinting into AI-driven decisions before they've built the control surface that makes those systems business valid. This gap is now starting to show up in the failure rates.

Gartner reports that by the end of 2025, at least half of generative AI projects were discontinued after proof of concept, with common issues such as poor data quality, insufficient risk management, rising costs, or ambiguous business value.[1] Gartner also forecasts that more than 40% of agentic AI projects will be canceled by the end of 2027 for similar reasons.[2]

These operational gaps are also seen in agentic AI adoption. A Deloitte survey indicated that 30% of organizations are exploring agentic options, 38% are piloting solutions, 14% have solutions ready for deployment, and only 11% are actively using these systems in production.[3] Worryingly, 42% of organizations report they are still developing their agentic strategy roadmap, with 35% having no formal strategy at all.[4]

At this inflection point, the failure modes fundamentally change.

Teams stop worrying about a model's ability to answer correctly, and start confronting harder questions:

- **"Who authorized it to act?"**
- **"What did it change?"**
- **"Can we prove why it made that decision?"**
- **"Can we stop it immediately if needed?"**
- **"Can we operate if it fails?"**

Today, 62% of organizations are experimenting with AI agents, but only 23% have scaled an agentic AI system.[5] The gap between pilot success and production resilience is typically not technical (although there are many technical challenges); it's operational.

SIX KEY BREAKPOINTS SURFACE

01 PERMISSIONS
Collapse as agents introduce delegated action at scale

02 AUDITABILITY
Breaks when teams cannot reconstruct what the system saw, what rules applied, or why it acted

03 RELIABILITY
Breaks as demos encounter real latency, cost, drift, and dependency failure

04 GOVERNANCE
Breaks when retrieval turns into an unintentional policy bypass

05 PORTABILITY
Breaks when hidden coupling and missing inventories make migration and exit strategies theoretical

06 SOVEREIGNTY
Breaks when organizations discover they have never rehearsed operating, isolating, or relocating systems under real jurisdictional and dependency constraints

THIS REPORT HIGHLIGHTS HOW THESE BREAKPOINTS REPEATEDLY EMERGE AS AI MOVES FROM EXPERIMENTAL TO BUSINESS-CRITICAL, AND THE ARCHITECTURAL DECISIONS THAT PREVENT COSTLY REMEDIATION LATER.

BREAK #1
PERMISSIONS

Your AI program becomes a permissions program.

Failure Pattern

Teams deploy agents and automations that deliver measurable value, then encounter friction when security and compliance teams ask fundamental questions:

"UNDER WHAT AUTHORITY DID THIS AGENT MODIFY CUSTOMER DATA?"

"WHAT POLICY GOVERNS ITS ACCESS SCOPE?"

"WHERE'S THE AUDIT TRAIL FOR THIS BATCH OF AUTOMATED DECISIONS?"

Inevitably, workflows revert to manual approval gates, negating the automation's value, or continue operating with accumulated risk that remains invisible until an incident forces an examination.

Root Cause

AI introduces delegated action at scale. When delegation operates without explicit policy boundaries, organizations face privilege creep at lightning speed. The traditional security model, where humans serve as the ultimate decision boundary and accountability checkpoint, erodes without replacement.

Every agent is an identity with permissions. Without deliberate design, these identities accumulate access rights based on convenience rather than principle, creating sprawling attack surfaces and compliance exposure.

The Control Architecture

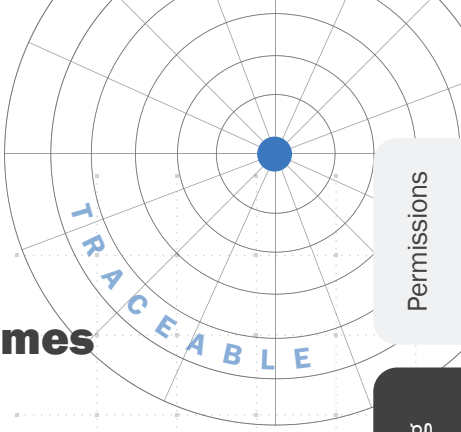
Rather than treating AI agents as an exception, organizations should bring them into standard identity and access management frameworks as first-class entities. Define scoped agent roles with least-privilege access and make an explicit separation between "suggest" and "execute" actions. High-impact operations should require defined approval paths, with human-in-the-loop stop gates for any action that is irreversible or compliance-sensitive.

Critically, these boundaries must be enforced by a control plane programmatically, not left to documentation, convention, or blind trust.

Checklist: How to Ensure You're Ready

- ☑ Documented agent roles with explicit access boundaries and justification for each permission
- ☑ Formal approval models for sensitive actions, with clear escalation paths
- ☑ Complete audit logs capturing prompts, context, tools invoked, actions taken, and outcomes
- ☑ Tested emergency shutdown procedures with defined trigger criteria and accountability

BREAK #2
AUDITING



You can't explain outcomes when it matters.

The Accountability Gap

An AI system's output remains justifiable until it starts to become questionable evidence. During incident investigations, customer disputes, regulatory inquiries, or compliance reviews, teams must be able to reconstruct what the system observed, why it reached a particular decision, or what changes it executed.

The inability to explain such workflows becomes an organizational liability. The consequences can range from failed audits to regulatory sanctions or customer trust erosion.

Why Traceability Fails

Production AI systems are assembled from interdependent components: retrieval mechanisms, tool integrations, prompt templates, policy rules, and complex, often legacy upstream data sources. Each component, hence, evolves independently. Without proper versioning and traceability, you can see the result but cannot reliably trace the steps that led the system to that outcome.

When model versions change, retrieval lookup structures update, or data sources drift, the system's behavior shifts in ways that become visible only in retrospect, often when explaining a failure or defending a decision under scrutiny.

Engineering for Forensics

Organizations should engineer AI workflows for forensic reconstruction from the outset, versioning every component that can influence behavior: prompts, retrieval policies, model endpoints, tool configurations, and data sources. That requires structured traces that capture inputs, retrieved evidence, policy decisions, tool calls, and outputs, enabling the controlled replay of the decision path.

A mature forensic capability should be able to answer:

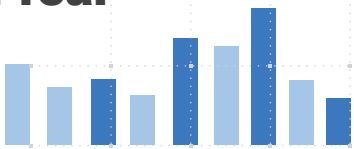
"On this date, when this system made this decision for this user, what exactly did it see, what rules applied, and what would it decide if we replayed the scenario today?"

COMPLIANCE INDICATORS

- Version control for prompts, policies, and configurations, with immutable release artifacts
- End-to-end traceability linking outputs to specific data sources, tools invoked, and model versions
- Reproducible replay capability for regulated workflows, with controlled test environments
- Change management documentation showing what changed, why, who approved it, and what testing validated the change

BREAK #3
RELIABILITY

Great demos fail under real operational load.



Production Reality

Initial successes often degrade in production: latency spikes unpredictably, costs exceed projections, tool integrations fail intermittently, retrieval quality drifts, and quality regressions accumulate, sometimes unnoticed. Users eventually lose confidence and develop workarounds, routing critical work back to manual processes.

Systemic Complexity

Business-critical AI is not a model; it's a distributed system with unique latency characteristics, quota limits, third-party dependencies, and lifecycle management challenges. Traditional site reliability engineering practices are typically applied after problems emerge rather than designed in from the start.

AI-driven systems can exhibit inconsistent behavior, in which agents make probabilistic decisions, adapt to context, and follow different execution paths for similar inputs.[6] Failures often emerge gradually rather than presenting as discrete errors, making typical monitoring of threshold-based alerts insufficient to capture AI system behavior.[7]

Operational Discipline

Organizations should treat AI workloads as top-priority production services from day one, with explicit service-level objectives for latency, quality, and cost. Because these systems are probabilistic and operate in changing environments, they should be engineered for resilience with clear fallback modes and graceful degradation paths such as **"answer with citations," "answer without external tools," "route to a human specialist,"** and **"temporarily freeze automated actions."**

Following a best practice or progressive deployment strategy—canary releases, shadow traffic, and A/B testing—allows operators to learn how the system functions and enables earlier regression detection before user impact.

RESILIENCE METRICS

- Service level objectives with error budgets for latency, quality consistency, and cost efficiency
- Implemented fallback modes with user-visible confidence signals when operating in degraded states
- Automated regression testing against representative production workloads
- Incident runbooks covering model behavior, retrieval systems, and tool integration failures
- Real-time monitoring dashboards with alerts tied to SLO thresholds

BREAK #4
DATA GOVERNANCE

RAG and open data create an inadvertent policy bypass.

The Access Control Problem

Teams connect AI systems to internal knowledge repositories to enhance services for accuracy and relevance. Organizations may discover that access patterns enabled by retrieval-augmented generation (RAG) don't align with enterprise data governance policies. Users see information they shouldn't be able to access, and AI systems use data in ways that violate privacy commitments or regulatory requirements.

The Governance Shift

RAG, whilst different from agentic AI, provides agents with enterprise-specific knowledge to reason and act with.

RAG fundamentally changes the unit of access control from "dataset or table" to "snippet or embedding." Traditional governance enforces permissions at storage boundaries (file share, folder, buckets, etc.). RAG systems query across boundaries, retrieve fragments, and synthesize answers, potentially exposing information that individual components are properly restricted from.

By 2027, 60% of organizations will fail to realize the value they expected from AI use cases because their governance is incohesive.[8] If governance isn't enforced at query time including user identity context, organizations inadvertently create high-speed policy bypass mechanisms that appear to work correctly while systematically violating data access rules.

Identity Aware Retrieval

Organizations need to move from static, storage-centric governance to dynamic, identity-aware controls that are enforced at retrieval time, not just at rest. That means defining explicit AI data contracts: which sources are in scope for which user populations, what redaction or masking applies to sensitive fields, and which topics remain out of bounds regardless of the system's technical capabilities.

In practice, retrieval should be treated as a governed operation, with the same rigor applied to direct database access in regulated environments.

POLICY ENFORCEMENT EVIDENCE

- Identity-aware retrieval with real-time policy enforcement based on user roles and data classification
- Auditable source attribution for every response, showing which documents contributed
- Automated redaction and classification controls that operate before information reaches the model
- Periodic adversarial testing against sensitive information repositories to validate policy enforcement
- Clear documentation of which data sources are accessible to which AI systems under which conditions

BREAK #5
PORTABILITY

You can't move what you can't inventory.

The Dependency Problem

Leadership requests operational resilience: multi-region deployment, multi-cloud optionality, vendor exit strategies, or regulatory-driven data portability. Delivery teams often face tangible challenges: they struggle to precisely list the AI system's dependencies. This difficulty hampers their ability to accurately estimate resiliency effort, determine costs, or evaluate risks confidently.

Compounding the AI system's dependency issues are the use of proprietary APIs, custom prompt logic embedded in application code, unmanaged vector databases, and data movement costs that weren't visible during pilot phases.

Hidden Coupling

AI solutions often accumulate technical debt through "convenience" decisions that were rational during rapid experimentation. Proprietary model APIs, vendor-specific tooling, embedded prompt templates, and point-to-point integrations create tight coupling that only becomes obvious when teams try to swap, upgrade, or relocate components.

The deeper issue is usually an inventory gap. Many organizations cannot answer basic questions: which models run where, which prompts govern which behaviors, which data sources feed which workflows, and which runtime dependencies are truly critical versus remnants of earlier experiments.

Portability as Capability

Organizations should consider creating and maintaining an "AI bill of materials" that inventories the models, prompts, tools, data sources, and runtime dependencies shaping system behavior, and makes coupling points and portability constraints explicit. Also consider periodic "move tests" to rehearse component swaps and relocations, even when migration is not planned, to validate portability rather than assume it.

Portability should be treated as an architectural quality attribute with measurable criteria, not as a negotiation point in procurement.

INDEPENDENCE INDICATORS

- Living dependency map showing models, data sources, tools, runtimes, and integration points
- Documented portability constraints with estimated migration effort and cost drivers
- Quarterly simulation of migration or failover scenarios with measurable outcomes
- Procurement agreements aligned to operational reality with clear exit provisions
- Standardized interfaces for model access, tool integration, and data retrieval where vendor-neutral options exist

BREAK #6
SOVEREIGNTY

You can't claim control if you've never rehearsed it.

The Capability Gap

Organizations assert sovereignty, data residency commitments, jurisdictional control, and operational independence, but demonstrating these capabilities under stress is often overlooked. When faced with identity provider outages, regional network isolation, loss of third-party dependencies, or legal constraints on data movement, the asserted sovereignty evaporates.

The gap between architecture documents and operational reality becomes a crisis during exactly the scenarios sovereignty was meant to address: regulatory enforcement actions, geopolitical disruptions, or disputes with critical vendors.

Operational Rehearsal

Organizations should consider operationalizing sovereignty through disciplined rehearsal. Controlled scenarios can be designed to test whether AI systems can continue to operate, be isolated, be relocated, or cease gracefully, while still preserving audit evidence and strengthening response playbooks.

They should treat sovereignty drills with the same rigor as disaster recovery exercises: scheduled and documented, evaluated against defined criteria, and used to drive concrete architectural and operational improvements.

Sovereignty vs. Assertion

Sovereignty is not a statement or a contractual clause; it's an operational capability that must be built, maintained, and regularly validated. Most AI programs stop at architectural diagrams and vendor assurances, without proving their ability to operate independently when circumstances demand it.

Genuine sovereignty requires the ability to operate AI systems with degraded or severed dependencies, to relocate workloads across jurisdictional boundaries under time pressure, and to produce evidence of control when auditors or customers challenge it.

DEMONSTRABLE CONTROL

- Repeatable drill program with defined scenarios covering dependency failures, regional isolation, and forced migration
- Evidence packages documenting sovereignty capabilities for auditors and customers
- Tested isolation and recovery procedures with measured recovery time objectives
- Clear accountability framework: who decides to invoke sovereignty measures, who executes, who validates, and who communicates to stakeholders
- Post-drill reviews that drive continuous improvement of both technical capabilities and operational procedures

CONCLUSION

Control Planes Matter More Than Models

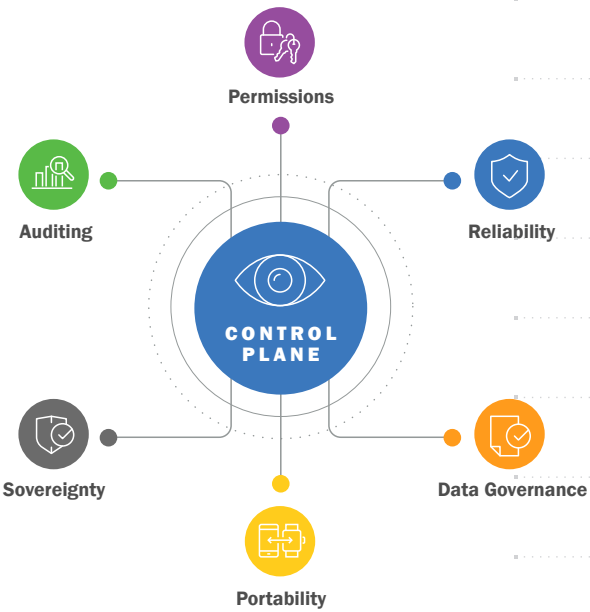
Organizations that successfully take AI from pilot to business-critical service typically share one discipline: they prioritize the operational control plane—governance, identity, safety, observability, and change management—before optimizing AI model performance (accuracy, safety, calibration), which, whilst a gating factor, is one of several quality attributes to meet.

The percentage of companies abandoning most of their AI initiatives before they reach production has surged from 17% to 42% year over year.[9] Analysis confirms that over 80% of AI projects fail, which is twice the failure rate of non-AI technology projects.[10]

Whilst model accuracy improves with better training data and techniques, control plane maturity requires deliberate organizational design: clear accountability, enforced boundaries, operational discipline, and the foresight to test capabilities before businesses depend on them.

These six breakpoints are not theoretical risks: they are patterns observed repeatedly across industries as organizations cross from experimentation to operational dependence. Fixing issues after deployment is significantly more expensive than designing with them in mind from the start.

The question is not whether your organization will encounter these challenges. The question is whether you'll address them before they become incidents, or after.



SOURCES

Permissions

Auditing

Reliability

Data Gov

Portability

Sovereignty

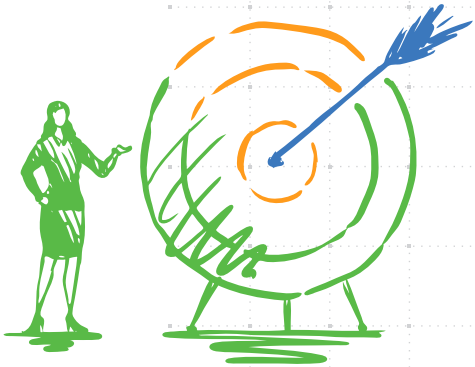
SOURCES

1. Gartner, Why 50% of GenAI Projects Fail — And How to Beat the Odds, Jan 2026
2. Gartner, Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027, June 2025
3. Deloitte, The agentic reality check: Preparing for a silicon-based workforce, Dec 2025
4. Ibid.
5. McKinsey, The state of AI in 2025: Agents, innovation, and transformation, Nov 2025
6. Efficiently Connected, 2026 Predictions: Observability Becomes the Control Plane for AI Operations, Dec 2025
7. Ibid.
8. Gartner, Gartner Predicts Defensive Spend to Derisk IP Loss and Copyright Infringement Will Slow GenAI Adoption and Diminish Returns, Mar 2024
9. S&P Global, Generative AI experiences rapid adoption, but with mixed outcomes – Highlights from VotE: AI & Machine Learning, May 2025.
10. RAND, The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed, Aug 2024

ABOUT

exactmarket[tm]

A WBENC-CERTIFIED WOMAN-OWNED BUSINESS ENTERPRISE



Exact Market
is here to help your organization understand and capitalize on technology disruptions.

We bring together market insights, content strategies, and digital execution to help enterprises, technology providers, and independent software vendors share complex transformation launches and stories clearly and compellingly.

Our Elephant in the Room series captures the meaningful conversations leaders are already having today—about cloud, cybersecurity, AI, and infrastructure—and transforms them into practical frameworks for making informed decisions.

Whether you are refining your cloud strategies, launching new products, or preparing for the next wave of innovation, Exact Market's specialized marketing resources can help align your business, technology, and brand.

WE DON'T JUST EXPLAIN TRANSFORMATION.
WE GUIDE YOU IN SHAPING IT, OWNING IT, AND MOVING IT FORWARD.

Find out more at www.exactmarket.com

exactmarket℠

ELEPHANT IN THE ROOM EBOOK SERIES