EXECUTIVE DECK

The Agentic Enterprise Is Here

What Leaders Must Redesign Before AI Can Scale

> **Agents are moving from copilots to coworkers, and this executive brief shows the 5 redesigns and blueprint leaders need to scale Enterprise AI safely, fast, and ahead of competitors.**

Prepared for: CEO, COO, CMO, CIO, CFO, CISO, Head of Transformation & Head of Growth

*By: **Logan Sivanasen** | 15 Apr 2026*

EXECUTIVE SUMMARY

Five Redesigns Stand Between Today's Pilots and Enterprise-Scale AI

Enterprises are moving rapidly from AI copilots that assist to agents that execute multi-step work across systems and data. The difference is consequential. Scaling agents is not a technology deployment. It is a redesign of how work gets done, who owns outcomes, and how risk is governed.

1. Operating Model

Assign ownership of agent workflows across business and technology

2. Workflow Architecture

Redesign processes for orchestration, not just automation

3. Integration

Remove data silos and establish clean API contracts

4. Governance

Define action tiers, permissions, audit trails, and kill switches

5. Measurement

Shift from usage metrics to cost per outcome and eval-gated releases

Action Priority: Start with 2 to 3 thin-slice workflows tied directly to a measurable cost-per-outcome target. Build the agent platform lane in parallel with clear controls and KPIs before expanding.

MARKET SIGNAL

Why Agents, Why Now

Market readiness signals have converged. Leaders across industries are moving from AI experimentation to agentic execution, and the data confirms this is no longer a niche initiative.

Key Data Points

81% of leaders

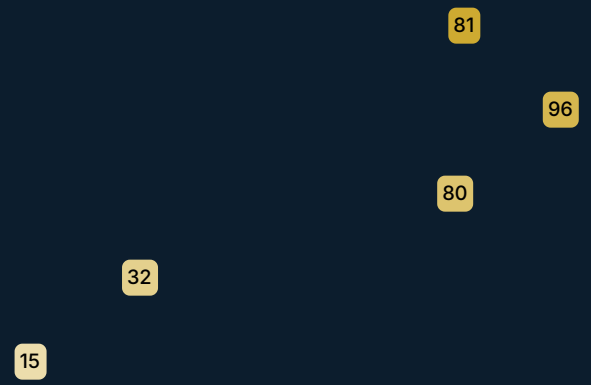
expect agents to be moderately or extensively integrated into their AI strategy within 12 to 18 months

96% of IT leaders

agree agent success depends on seamless data integration across all systems

4 in 5 organizations

are in or past the pilot stage; 32% are preparing to scale, 15% revisiting strategy after early tests



*Sources: **Microsoft Work Trend Index 2025 (Key Findings section)**; **Salesforce 2026 Connectivity Report** announcement (opening paragraph); **Microsoft WorkLab "Agents are here"** (Findings section)*

FOUNDATIONAL DEFINITION

What "Agentic" Means in the Enterprise

Precision matters. "Agentic AI" is not a marketing label for a smarter chatbot. It describes systems that plan, call tools, use context across sessions, escalate when uncertain, and maintain auditable logs of every action taken.

Copilot (Assist Mode)	Agent (Execute Mode)
Surfaces recommendations for a human to act on	Plans and executes multi-step tasks autonomously
Single-turn or session-scoped interaction	Persistent context across sessions and systems
Human drives each step	Agent selects tools, sequences steps, and loops until done
Error impact: human catches it before action	Error impact: action may already be taken; logs required
Example: draft an email, summarize a document	Example: resolve a support case end-to-end, execute a procurement approval
Governance: prompt hygiene, access scoping	Governance: action tiers, permission gates, audit trails, kill switch

1

Read

Access data, surface context

2

Recommend

Propose actions for human review

3

Draft

Generate artifacts pending approval

4

Execute

Commit actions in live systems with controls

VALUE POOLS

Where Agents Create Near-Term Value

The highest-confidence near-term value pools sit in structured, high-frequency workflows where "done" is unambiguous, data is accessible, and the cost of a wrong action is bounded. Prioritize breadth of coverage, not depth of AI sophistication.

Revenue

Account planning, proposal generation, renewal risk scoring, and territory-aware opportunity workflows

Operations

Procurement triage, policy-aware approval routing, and exception handling with audit trails

Customer

End-to-end case resolution orchestration and knowledge base maintenance loops

Finance

Close support, variance explanation drafting, and audit evidence collection workflows

IT and Security

Runbook automation, access request triage, and incident response support with escalation paths

SELECTION RUBRIC

How to Pick the Right First Workflows

Most failed agent pilots share a common root cause: the wrong workflow was chosen first. Use this rubric to score candidates before committing engineering and governance resources. A score of 4 or 5 out of 5 indicates a strong starting candidate.

Selection Criterion	What to Look For	Sample Score (1-5)
High frequency and clear "done"	Workflow runs daily or weekly; completion is unambiguous and measurable	5 - Invoice triage (100s/day)
Policy-heavy with guardrail value	Rules-based logic where agent guardrails reduce human review burden	4 - Procurement approval
Data and actions accessible via API	Source systems expose clean APIs; no manual re-keying required	3 - CRM + ERP connected
Human-in-the-loop acceptable initially	Stakeholders accept a review gate before final action during phase one	5 - Legal review step retained
Measurable by cost per outcome	Can calculate current cost per completed case, approval, or record	4 - Cost per resolved ticket known

Avoid starting with: Workflows that touch unstructured judgment calls, require real-time multi-party consent, or lack an established baseline cost metric. These are phase two or phase three candidates.

CORE FRAMEWORK

The 5 Redesigns Leaders Must Drive

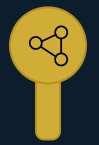
Agentic scale does not fail because models underperform. It fails because the surrounding enterprise infrastructure, ownership structures, and governance layers were designed for a world where humans executed every step. These five redesigns close that gap.

1. Operating Model and Ownership

Define who owns agent workflows. Assign business process owners, not only IT owners. Establish an AgentOps function to manage runtime behavior.

2. Workflow Architecture and Orchestration

Decompose end-to-end processes into agent-executable steps. Design for handoffs, retries, and escalations from day one.



3. Integration and Data Contracts

Establish clean API contracts for every system an agent touches. Eliminate shadow integration and data silos before deploying at scale.

4. Governance and Risk Controls

Implement action tiers with approval gates, least-privilege permissions, mandatory audit logs, escalation paths, and executive risk dashboards.

5. Measurement: Evals, Observability, and Cost per Outcome

Replace usage dashboards with eval suites, production monitors, and cost-per-outcome scorecards. Gate every release with regression checks.

REDESIGN 1 OF 5

Operating Model and Ownership

Agentic AI fails when ownership is ambiguous. Every production agent needs a named business process owner, not just an IT ticket owner. And a dedicated AgentOps function to manage runtime behavior, escalations, and continuous improvement.

What Must Change

Business Process Owner

Each agent workflow must have a named business owner accountable for outcomes, not just the engineering team that built it.

AgentOps Function

A standing team responsible for monitoring agent runtime behavior, handling escalations, and iterating on prompts and tools.

RACI for Agents

Explicit responsibility matrix covering who approves new agent capabilities, who reviews audit logs, and who can trigger kill switches.

Cross-functional Steering

Monthly review cadence with business, legal, security, and IT to align on agent scope expansion.

Common Failure Modes

IT-Only Ownership

Agents built and owned solely by IT drift from business intent within weeks of deployment.

No Runtime Accountability

Without AgentOps, production issues surface only after downstream damage is done.

Shadow Agent Sprawl

Teams deploy agents without central registration, creating ungoverned automation at scale.

Scope Creep Without Gates

Agents accumulate permissions over time without formal review, violating least-privilege principles.

REDESIGN 2 OF 5

Workflow Architecture and Orchestration

Agents do not slot into existing workflows, they require workflows to be redesigned from scratch. End-to-end processes must be decomposed into discrete, agent-executable steps with explicit handoff points, retry logic, and escalation paths built in from day one.

01

1. Map the Full Process

Document every step, decision point, and exception path in the current human workflow before touching agent design.

02

2. Identify Agent-Executable Steps

Isolate steps that are deterministic, data-accessible, and have unambiguous completion criteria.

03

3. Design Handoff Contracts

Define exactly what data passes between steps, what triggers escalation to a human, and what constitutes a retry condition.

04

4. Build Orchestration Layer

Implement a durable orchestration framework (e.g., LangGraph, Temporal, or equivalent) that manages state, retries, and parallel execution.

05

5. Test Failure Paths First

Simulate edge cases, timeouts, and bad data before testing the happy path to ensure resilience.

Orchestration Patterns

Single Agent

A single agent handles a linear, well-defined task from start to finish.

Multi-Agent Pipeline

Multiple agents perform sequential tasks, with explicit handoffs between them.

Parallel Agent Swarm

Multiple agents work concurrently on independent subtasks, results are aggregated.

Human-in-the-Loop

Approval gates or review stages where a human intervenes before agent proceeds.

Anti-Patterns to Avoid

Monolithic Agent

One overly complex agent attempts to handle too many disparate functions.

Stateless Design

Lack of memory or context persistence between agent steps, leading to inefficiency.

Hardcoded Escalation

No dynamic routing or configurable triggers for human intervention when issues arise.

Skipping Retry Logic

A single transient failure halts the entire workflow, lacking robustness.

REDESIGN 3 OF 5

Integration and Data Contracts

Agents are only as capable as the systems they can reach. Fragmented APIs, undocumented data schemas, and shadow integrations are the most common reason agent pilots fail to scale. Clean integration contracts are not a nice-to-have, they are the prerequisite for any production agent.

The Integration Readiness Checklist

- **API Contract Exists** Every system the agent touches must expose a documented, versioned API. No screen-scraping, no undocumented endpoints.
- **Data Schema is Governed** Field definitions, data types, and null handling are documented and enforced at the source system.
- **Auth and Permissions are Scoped** Agent credentials follow least-privilege; no shared service accounts with broad access.
- **Rate Limits and SLAs are Known** Agent orchestration must account for downstream system throttling and latency.
- **Failure Modes are Documented** What does the API return on error? Timeout? Partial success? Agents must handle all cases.

Integration Debt Risk Matrix

67%

Siloed Data

of enterprise data is siloed. Gartner, 2024

3x

Longer Deployment

Average time to deploy agents in organizations with poor API governance vs. those with clean contracts.

#1

Top Blocker

Integration complexity cited as the top barrier to agent scale. McKinsey's 2024 AI survey

40%

Agent Failures

of agent failures traced to data quality or integration issues, not model errors.

The MCP Standard: Anthropic's Model Context Protocol (MCP) is emerging as the open standard for agent-to-tool connectivity. Enterprises adopting MCP-compatible tooling now will reduce integration rework as the ecosystem matures.

REDESIGN 4 OF 5

Governance and Risk Controls

Governance for agentic systems is categorically different from software governance. Agents act — they write, send, book, approve, and delete. A governance model designed for passive software will fail when applied to systems that can cause irreversible downstream effects at machine speed.

Action Tier Framework

Tier 1: Read

Query data, retrieve records, generate drafts. No approval required. Full audit logging.

Tier 2: Recommend

Surface recommendations to humans. Human confirms before any action is taken.

Tier 3: Draft & Stage

Agent prepares actions (emails, orders, updates) for human review before execution.

Tier 4: Execute

Agent acts autonomously. Requires explicit pre-authorization, scoped permissions, and mandatory post-action audit.

Core Control Requirements

Least Privilege

Permissions scoped to the specific workflow, granted just-in-time, revoked on completion.

Mandatory Audit Trail

Every tool call, decision branch, and outcome logged with timestamp and agent identity.

Kill Switch Protocol

Documented circuit-breaker for every production agent; tested quarterly.

Escalation Paths

Defined triggers for human escalation; no agent should be able to loop indefinitely without human review.

Executive Oversight

Risk Dashboard

Anomaly rates, escalation counts, and policy violations visible at board level weekly.

Postmortem Cadence

Weekly review of agent failures and near-misses; findings fed back into policy.

Regulatory Mapping

Each agent workflow mapped to applicable regulations (**GDPR**, **SOX**, **HIPAA**, etc.).

Third-Party Agent Risk

Vendor-supplied agents subject to same governance standards as internally built agents.

REDESIGN 5 OF 5

Measurement: Evals, Observability, and Cost per Outcome

Usage dashboards are not measurement. Knowing how many times an agent ran tells you nothing about whether it ran correctly, safely, or efficiently. The measurement redesign replaces vanity metrics with three interlocking systems: eval suites, production observability, and cost-per-outcome scorecards.

The Three Measurement Systems

Eval Suites (Pre-Deployment)

Automated test batteries covering tool call accuracy, trajectory correctness, edge case handling, and policy adherence. No release ships without passing evals. Source: **Anthropic Engineering, "Demystifying Evals for AI Agents"**

Production Observability (Runtime)

Real-time monitoring of agent trajectories, tool call latency, error rates, and escalation frequency. Surfaces drift that pre-deployment evals cannot catch.

Cost per Outcome (Business)

The unit economics of agent work: cost per resolved ticket, cost per processed invoice, cost per qualified lead. Replaces token-count dashboards with business-relevant scorecards.

Metrics That Matter vs. Metrics That Mislead

API calls per day
Volume without quality

Tokens consumed
Cost without outcome

Tasks initiated
Activity without completion

Uptime %
Availability without accuracy

Task completion rate
% of workflows completed without human intervention

Escalation rate
% of tasks requiring human override (target: declining over time)

Cost per outcome
Business unit economics of agent work

Regression rate
% of releases that degrade established eval benchmarks

REFERENCE ARCHITECTURE

Agent Platform Reference Architecture (Executive View)

This layered architecture is the minimum viable platform for enterprise agent deployment. Each layer must be deliberately designed and owned. Gaps in any layer create operational, security, or compliance risk that will surface at the worst possible time.

Platform principle: The Controls Layer is not optional and cannot be retrofitted cheaply after deployment. Identity, policy, and audit logging must be designed in before production traffic is routed to any agent.

INTEGRATION REALITY CHECK

The Connectivity Gap Will Kill Your Agent Program

The single most underestimated barrier to agent scale is not model capability. It is the state of enterprise integration. Fragmented data, ungoverned APIs, and siloed agent deployments create cascading failure points. The data from the 2026 Connectivity Benchmark Report is unambiguous.

96%

Integration Mandate

of IT leaders agree AI agent success depends on seamless, debt-free integration across all systems

54%

Governance Gap

only 54% of organizations have a framework for centralized integration governance in place

50%

Silo Problem

of AI agents operate in isolation outside cohesive multi-agent systems, undermining enterprise coordination

What This Means for Leaders

- 96% mandate vs. 54% governance capability = a critical execution gap that must be closed before scaling
- Isolated agents cannot hand off context, share memory, or enforce consistent policy
- Integration debt is agent-scale debt: address it in the first 45 days

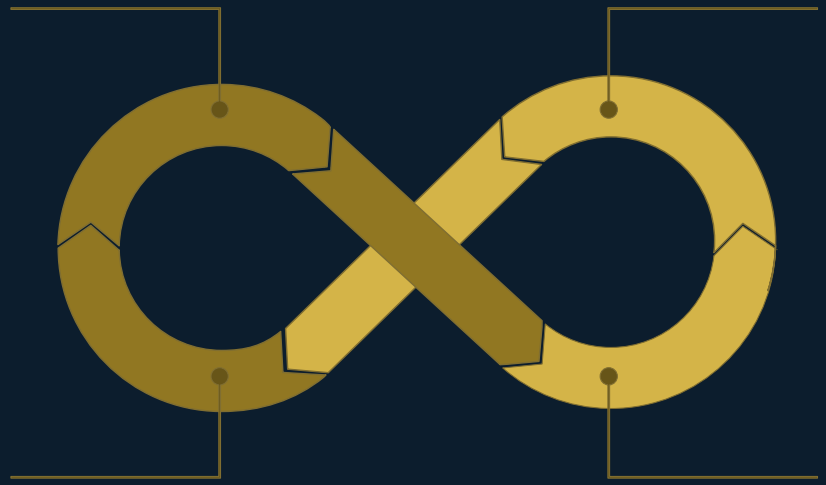
Required action: Appoint an Integration Platform Owner with authority to enforce API contract standards and deprecate shadow integrations before agent workflows go to production.

*Sources: **MuleSoft 2026 Connectivity Benchmark Report** (Key Findings section); **Salesforce 2026 Connectivity Report** announcement (opening statistics paragraph)*

GOVERNANCE

Governance for Systems That Can Act

Agents that can execute in live systems require a governance model that is more rigorous than anything currently applied to software. The risk is not that agents are malicious. The risk is that they are fast, tireless, and wrong in ways that scale before anyone notices.



Action Tiers and Gates

Read, recommend, draft, and execute tiers each require distinct approval thresholds

Least Privilege

Just-in-time permissions scoped to the specific workflow, revoked after task completion

Mandatory Audit Trail

Every tool call, decision branch, and outcome logged with timestamp and agent identity

Kill Switch and Escalation

Defined escalation paths and a documented circuit-breaker protocol for every production agent

Executive Risk Dashboard

Weekly postmortem cadence; anomaly rates, escalation counts, and policy violations visible at board level

EVALUATION AND OBSERVABILITY

Evals and Observability: The Reliability Engine for Agent Scale

Agents cannot be tested the way traditional software is tested. They produce trajectories of actions, not single outputs. A robust evaluation framework, grounded in the approach described by Anthropic Engineering, is the prerequisite for any production deployment.

Three Grader Types (Anthropic Engineering)

→ Code-Based Graders

Deterministic checks on structured outputs: did the agent call the correct tool, in the correct order, with valid parameters?

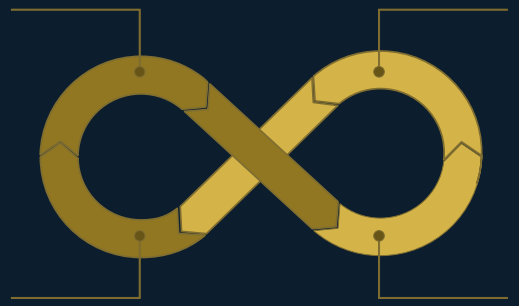
→ Model-Based Graders

LLM-as-judge evaluation for response quality, tone, and adherence to policy in ambiguous scenarios

→ Human Graders

Targeted human review for high-stakes trajectories, edge cases, and calibration of model-based graders

Eval Lifecycle



- Evaluate trajectories and tool use, not only final answers
- Pre-deployment eval suites must pass before any release
- Regression gates block releases that degrade established benchmarks
- Production monitoring surfaces drift that eval suites did not catch

Sources: ***Anthropic Engineering, "Demystifying Evals for AI Agents" (Grader Types section; Trajectory Evaluation section)***

EXECUTION ROADMAP

The 90-Day Agentic Roadmap

The gap between pilot and scale closes in 90 days — not through a big-bang transformation, but through a disciplined sequence of decisions, infrastructure investments, and governance commitments. This roadmap is designed for leadership teams ready to move from experimentation to enterprise deployment.

Days 1–30: Foundation

- Audit integration readiness: identify the top 10 systems agents will need to touch and assess API contract quality
- Select 2–3 pilot workflows using the selection rubric (high frequency, unambiguous completion, low cost of error)
- Stand up AgentOps function: assign business process owners and define RACI
- Establish baseline eval suite for each pilot workflow before any agent is deployed
- Define governance tier for each pilot: what actions are permitted, what requires human approval

Days 31–60: Pilot Deployment

- Deploy first agent in production with full audit logging and kill switch active
- Run weekly postmortems: review escalation rates, error patterns, and cost-per-outcome
- Expand eval suite based on production observations; add regression gates
- Begin MCP-compatible tooling assessment for integration layer
- Identify second workflow for parallel deployment based on pilot learnings

Days 61–90: Scale Preparation

- Publish internal agent registry: all production agents documented with owner, scope, and governance tier
- Present cost-per-outcome scorecard to executive steering committee
- Finalize platform architecture decision: build vs. buy vs. hybrid for orchestration layer
- Expand governance framework to cover third-party and vendor-supplied agents
- Set 12-month agent portfolio targets: workflows, headcount impact, and cost-per-outcome goals

SOURCES & REFERENCES

Sources and Further Reading

This document draws on primary research, engineering publications, and market data from the following sources. All claims are grounded in publicly available material as of Q1 2026.

Primary Sources

- Anthropic Engineering - "Demystifying Evals for AI Agents"
- Anthropic - Model Context Protocol (MCP) Documentation
- McKinsey & Company - "The State of AI in 2024"
- Gartner - "Top Strategic Technology Trends 2025"
- LangChain - LangGraph Documentation
- Temporal Technologies - Temporal Workflow Orchestration

Recommended Reading

- Anthropic - "Claude's Model Specification"
- NIST - "AI Risk Management Framework (AI RMF 1.0)"
- MIT Sloan Management Review - "The New Rules of AI Governance"
- Harvard Business Review - "Your Company Needs an AI Policy"
- a16z - "The Enterprise AI Stack"
- Sequoia Capital - "AI's $600B Question"

This document was prepared for executive and senior leadership audiences. It is intended as a strategic orientation, not a technical implementation guide. For implementation support, engage your enterprise AI platform team or a qualified systems integrator.