# Operationalizing Agentic AI

## A Technical Leader's Guide

**The biggest barrier to agentic AI isn't the technology. Here's a quick guide on how to turn agents into production capability.**

Based on the two-part series by Nav Bhasin & Sri Elaprolu ·
Read Part I - Read Part II

> "Ask 'Are we investing enough in AI?' and the answer is almost always yes.
> Ask 'Which specific workflows are materially better today because of AI agents, and how do we know?' and the room gets quiet."

Sri Elaprolu
Director, AWS Generative AI Innovation Center

**BY THE NUMBERS**

## The operating-model gap, by the numbers

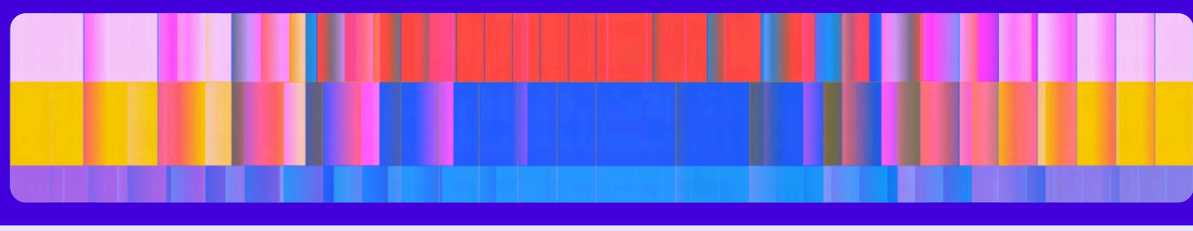
Enterprises are adopting agents fast, yet most stall before production. The constraint is architecture, governance, and evaluation maturity, not model capability.

**88%**
of AI agent pilots never reach production
IDC, 2025

**150,000+**
agents will operate inside an average Fortune 500 enterprise by 2028
Gartner, 2026

**>40%**
of agentic AI projects will be canceled by end of 2027
Gartner, 2025

**54%**
of organizations cannot see the full scope of their AI spend
FinOps Foundation, 2026



## A SHARED FOUNDATION

# Three traits of organizations generating real value

Before role-specific work begins, every organization getting value from agents shares these habits.

### Define workflow precisely
Specify inputs, steps, tools, and what “done” means.

### Set clear limits on autonomy
Decide where the agent acts alone and where a human stays in the loop.

### Optimize
A regular review of how agents behaved and what to change.

# Four signs the work is ready for an agent

Which work is already shaped like a job an agent could do?

01

### A clear start, a clear end

A defined trigger and a definition of completion someone outside the team can verify.

02

### Judgment, not a script

The agent decides what it needs and which systems to check. The path isn't hard-coded.

03

### Observable, measurable success

Outcomes you can instrument: quality thresholds, completion rates, escalation triggers.

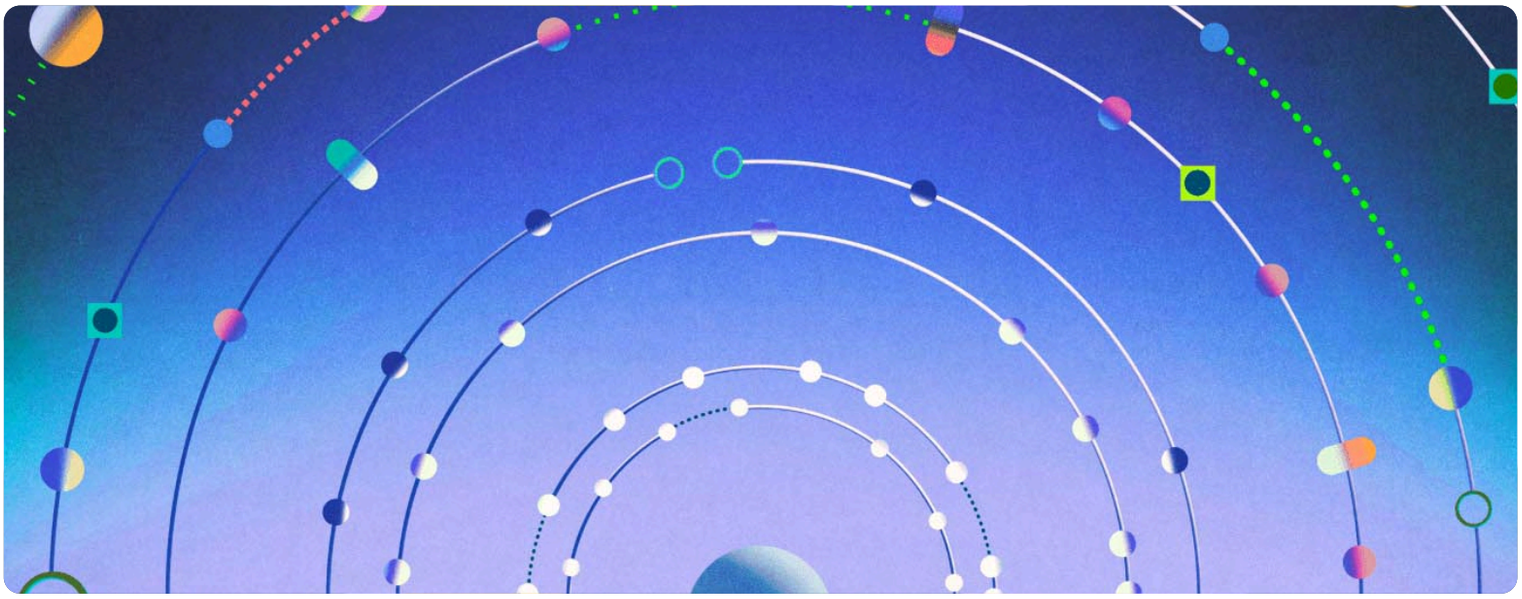
04

### A safe mode when things go wrong

Mistakes are caught quickly, cheap to correct, and never irreversible.

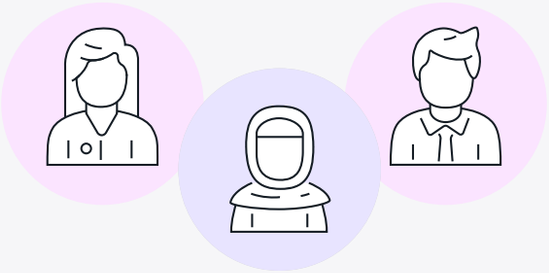
More on signs your work is ready for an agent

## WHO MAKES IT WORK



# Guidance by persona

Find your role. Each carries a different risk and a different piece of leverage.



LINE OF BUSINESS LEADER

### Own agent outcomes

An agent is only useful if it moves numbers you already track: open tickets, days in the cash conversion cycle, compliance exceptions. If you can't measure the workflow today, that's your first project.

**Owns:** A written description for the agent, tied to metrics you already track.

CTO / CHIEF ARCHITECT

### Build the agent platform

Once the first agent lands, every team wants one. If each builds its own framework and access model, you get a zoo no one can monitor as a whole. Build a sturdy floor with identity, policy enforcement, logging, connectors, and evaluation hooks.

**Owns:** A sturdy floor for every team to build on.

CISO

### Shape safe autonomy

Agents make decisions at machine speed, with accounts, roles, and tools of their own. Don't block autonomy but define where it's acceptable, what evidence earns trust, and what happens when trust breaks.

**Owns:** Identity system for agents, approvals enforced in the tools, monitoring for risky behaviors.

CHIEF DATA OFFICER

### Make the data reliable

Agents amplify the data foundation you have. The same question should return the same answer in every region and line of business, or agents will surface your data problems to everyone.

**Owns:** One definition per metric, clear lineage, and a data readiness map

CHIEF DATA SCIENCE / AI OFFICER

### Own evaluation system

Agents fail in ways benchmarks never measure. They get stuck in loops, call the wrong tool, or stop halfway and still report success. Evaluation is what catches that.

**Owns:** An evaluation suite that turns production mistakes into permanent tests.

COMPLIANCE / LEGAL OFFICER

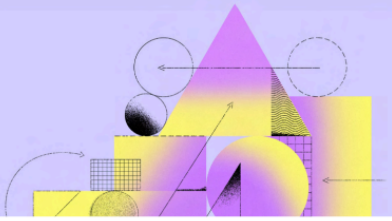
### Encode policy in design

Work backwards from the question you will eventually be asked: on this date, why did this agent take this action? Decide now what evidence you would need to answer it.

**Owns:** A decision trail per agent, and a responsible AI checklist applied before work starts.

Go in depth

## ARE YOU BUILDING A "ZOO" OR A "SYSTEM"?



| | The Zoo: 10 one-off agents | The system: One shared foundation for 100 agents |
|---|---|---|
| Tool integration | Rebuilt for every agent | One per action, reused by all |

| | | |
|---|---|---|
| **Identity & access** | Each team invents its own | One model, applied to every agent |
| **Life cycle** | Scripts that are hard to upgrade | Long-lived services you can upgrade safely |
| **Fleet visibility** | No way to see it as a whole | One view of what is running |
| **Your next agent** | Starts from scratch | Builds on what already exists |

# Evaluation is your real product

Three concrete functions replace vague “measurable success.” This is the AI officer’s real deliverable.

### Turns real work into tests

Every production mistake becomes a permanent test. Your hardest cases become guardrails against regression.

### Runs automatically

Any change to prompts, models, tools, or retrieval triggers evaluation before it goes live. Iterate fast without relying on hope.

### Measures what the business cares about

Technical metrics are not enough. Track the numbers your business already watches.

# Autonomy is earned, not granted

Start where the agent output is a recommendation and mistakes are reversible. Give more autonomy only when evaluation shows the agent has earned it.

### Advisory/ Assisted
### Tier green

Agent collects data & proposes; human decides

100% of outputs human-reviewed

Best first step: collapse handoffs

**Promotion criteria:** Promote after a track record of accurate proposals with low correction rate

### Supervised autonomy
### Tier yellow

Agent acts; humans spot-check via sampling

Automated eval on every change

Monitoring for agents that start behaving differently

**Promotion criteria:** Promote when acceptance rate is high and eval suite is stable across edge cases

### Full autonomy
### Tier blue

Agent closes the loop end-to-end

Audit trails + rollback / flag-and-hold

Never for regulatory red zones (credit, underwriting, employment decisions)

**Promotion criteria:** Reserved for well-instrumented, low-consequence or well-bounded work

# The compliance decision model

Turn “we need responsible AI” into three clear answers. Legal becomes an enabler, not a gate at the end.

## 40%

Gartner (May 2026) predicts that by 2027, 40% of enterprises will **demote or decommission autonomous AI agents due to governance gaps** identified only after production incidents. Decide the boundaries up front, not at launch.

**YES**
Some agents, with clear conditions. Decision trail and success measure defined up front.

**YES, LATER**
Others, with specific prerequisites. Data improvements or controls that must mature first.

**NO**
Say no to a few, with a clear rationale. Regulatory red zones: credit, underwriting, employment decisions.

# Five moves to get started

What separates leaders who move forward: they treat agentic AI as an operating-model challenge, not a technology experiment.

## 01
### Convene the right room

## 02
### Pick one job, not one use case

## 03
### Draw your readiness map

Bring LOB, CTO, CISO, CDO, AI/DS & compliance together for a working session, not a demo. Each names the single biggest blocker to production.

One concrete piece of work with a clear start, end, tools, and a verifiable success measure. Write the job description together.

CDO and CISO jointly sketch which domains are production-ready, which need work, and where the hard boundaries are.

04

### Commit to a cadence

Bring LOB, CTO, CISO, CDO, AI/DS & compliance together for a working session, not a demo. Each names the single biggest blocker to production.

05

### Make governance a design input

Decide now what evidence you'd need if an auditor asks "why did this agent do this?" Build that in before the first line of code.

# Sources and further reading


1. Nav Bhasin & Sri Elaprolu, "Operationalizing Agentic AI Part 1: A Stakeholder's Guide" & "Part 2: Guidance by Persona," AWS Generative AI Innovation Center, 2026.

2. Gartner, "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027," June 25, 2025 (incl. 33% of enterprise software / 15% of decisions by 2028 forecasts).

3. Gartner, "Applying Uniform Governance Across AI Agents Will Lead to Enterprise AI Agent Failure," May 26, 2026 (40% demote/decommission by 2027).

4. Deloitte, "Agentic AI strategy — The agentic reality check," Tech Trends 2026 (11% in production, 14% deployment-ready, 35% no formal strategy).


## Attribution

Adapted from Agentic AI in the Enterprise, Parts I & II — AWS Generative AI Innovation Center, Nav Bhasin & Sri Elaprolu.

Partner with the Generative AI Innovation Center to ground your first agentic pilot in your workflows, your data, and your business outcomes.