From Aspiration to Activation

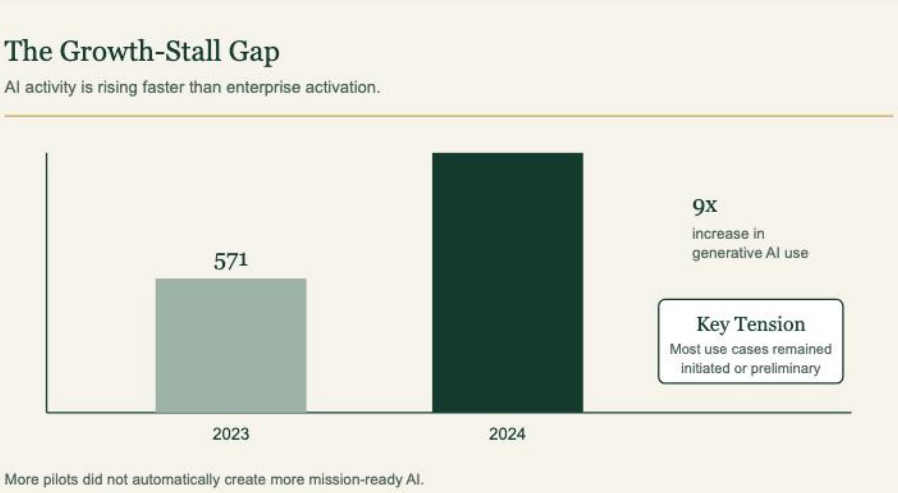
A Practical Guide for Federal Agencies Moving from AI Pilots to Organization-Wide Transformation

The Sainth | AI Enablement Series

Most federal agencies have already crossed the AI threshold. The question is no longer whether to begin. It is whether the work already done — the pilots, the proof of concepts, the isolated use cases — will ever become something the organization actually runs on.

The answer, for most agencies, is not yet.

Documented AI use cases across major federal agencies nearly doubled from 571 in 2023 to 1,110 in 2024. Generative AI usage increased ninefold over the same period.[1] And yet the majority of those use cases remained in an initiated or preliminary phase — meaning the capability existed, but the infrastructure to sustain, govern, and expand it did not.[2] Pilots declared successful. Programs stalled at the boundary of enterprise adoption. The gap between AI aspiration and AI activation remained wide.



This guide is for the leaders and teams responsible for closing that gap. It is not a policy overview. It is a working framework — structured around the decisions, actions, and accountabilities that move an agency from experimentation to embedded, mission-wide transformation.

It is organized into five stages. Each stage includes what to do, how to do it, and the most common place where agencies stall.

Before You Begin: Understand Why Pilots Stall

The most persistent failure pattern in federal AI is not a failed pilot. It is a successful pilot that never scales.

Agencies launch controlled experiments, demonstrate measurable results, and then watch the initiative lose momentum when it encounters the friction of the broader enterprise: procurement constraints, workforce resistance, legacy system incompatibilities, security review timelines, and organizational structures that were never designed to absorb a technology that moves faster than the budget cycle.[3]

The problem is almost never the technology. It is the organizational conditions the technology has to operate within.

Understanding this reframes the work. Activation is not a technology deployment. It is an organizational transformation that happens to involve technology. The sequencing matters. The governance has to be designed before the scaling begins. The workforce has to be prepared before the tools are expanded. The accountability structure has to be established before the compliance reporting is due.

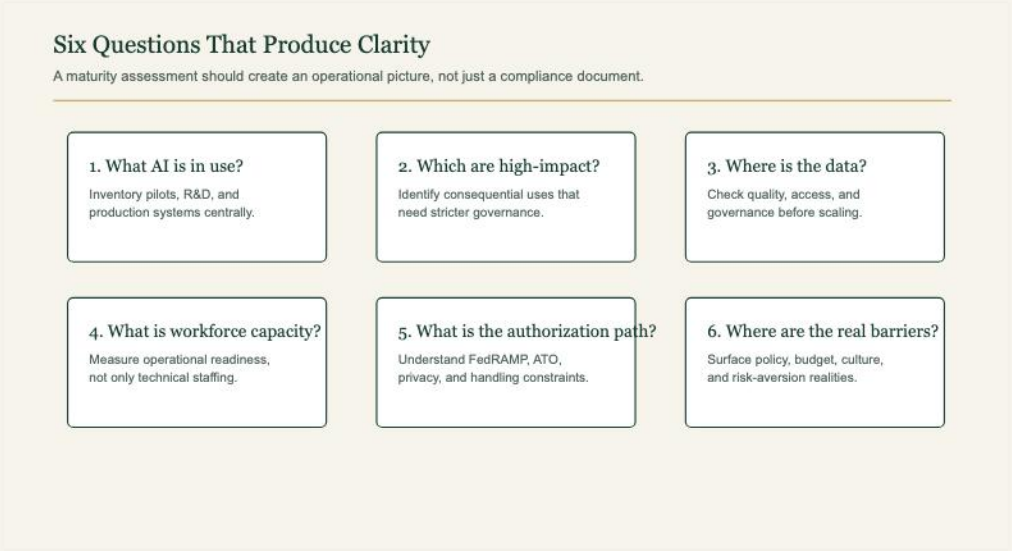
Agencies that try to skip these steps do not fail dramatically. They stall quietly — and expensively.

Stage One: Assess Where You Actually Are

What to do: Conduct an honest, structured assessment of your agency's current AI maturity before making any new investments or expansion decisions.

OMB Memorandum M-25-21 requires covered agencies to assess their current state of AI maturity as part of their published AI Strategy, including a plan to achieve AI maturity goals.[4] Most agencies have completed or are completing this requirement. The question is whether the assessment produced a document or produced actual clarity.

An assessment that produces clarity answers six questions:



1. What AI is actually in use? Every AI use case — whether in pilot, research, or production — must be inventoried. M-25-21 requires annual AI use case inventories, publicly released and submitted to OMB.[5] But the inventory requirement is also a management tool. Agencies that do not know what AI they are operating cannot govern it, cannot prioritize it, and cannot learn from it. GSA's approach is instructive: all AI use cases, whether pilots, R&D efforts, or production systems, must be registered in a central enterprise AI inventory regardless of their maturity stage.[6]

2. Which use cases are high-impact? M-25-21 defines high-impact AI as any system whose output serves as a principal basis for decisions with legal, material, binding, or significant effect on individuals — including decisions about civil rights, government services, health and safety, law enforcement, and benefits.[7] High-impact use cases require pre-deployment testing, risk mitigation plans, human oversight mechanisms, and explicit CAIO approval before deployment.[8] Know which of your use cases meet this threshold. The governance path is different, and the compliance requirements are real.

3. Where is your data? AI systems perform at the level of the data they are built on. Most federal agencies have significant data quality, accessibility, and governance challenges that predate AI entirely. Before scaling any AI capability, assess the state of the data it will depend on: Is it accurate? Is it accessible to the systems that need it? Is it governed appropriately for AI use? Agencies that skip this step build AI on a foundation that produces unreliable outputs — and then blame the model.

4. What is your current workforce capacity? This is not a question about whether your agency has data scientists. It is a question about whether the people doing mission work have the knowledge, skills, and tools to use AI outputs appropriately and to identify when those outputs are wrong. Most workforce assessments undercount this gap because they focus on technical roles and miss the operational dimension entirely.

5. What does your authorization environment look like? Federal AI deployments must operate within cybersecurity, privacy, and acquisition frameworks that were not designed for AI. Understanding the current state of your authorization environment — including FedRAMP coverage, ATO timelines, and data handling classifications — is essential before planning any expansion. One of the most significant barriers to scaling federal AI is the time gap between when a system is ready and when it is authorized to operate.

6. Where are the real barriers? Every agency has formal barriers (policy, regulation, budget) and informal barriers (culture, skepticism, competing priorities, risk aversion). Both matter. The informal barriers are often the harder ones to address and the less likely to appear in a written assessment. Build time into the assessment process for direct conversation with program staff — not just leadership.

Where agencies stall: Producing a compliance document instead of an operational picture. The maturity assessment exists to tell you the truth about where you are. If it tells you what you expected to hear, it probably did not go deep enough.

Stage Two: Define What Activation Actually Means for Your Mission

What to do: Translate AI ambition into specific, mission-anchored outcomes — and establish the criteria that will tell you when you have achieved them.

"Enterprise-wide AI transformation" is not an outcome. It is a description of a direction. Activation requires specificity: which mission functions will AI change, in what ways, measured by what indicators, and by when?

This work happens at two levels.

At the enterprise level, leadership must define the two or three mission outcomes that AI transformation is expected to deliver — not the technology outcomes (models deployed, use cases cataloged) but the mission outcomes. Faster benefits adjudication. Reduced application processing backlog. Improved accuracy in fraud detection. Accelerated permit review. The mission outcome defines what success looks like and anchors every subsequent governance and resource decision to something that matters operationally.

At the program or component level, each unit must identify its highest-value AI opportunities — the places where AI can remove friction, improve accuracy, or free staff capacity for higher-judgment work — and sequence them in order of feasibility and impact. Not every opportunity needs to be addressed simultaneously. The agencies making the most durable progress start with one or two components, build consensus around governance and compliance models, and expand gradually rather than attempting simultaneous enterprise-wide transformation.[9]

A useful framework for this prioritization:

Dimension	Questions to Ask
Mission impact	Does this use case materially advance a mission outcome? Will a senior leader notice if it succeeds — or fails?
Feasibility	Is the data available and reliable? Is the technology mature enough? Can the current workforce use the output appropriately?
Risk level	Is this a high-impact use case under M-25-21? What are the consequences of error? What oversight is required?
Time to value	How long from deployment to measurable outcome? Is that timeline compatible with current budget and planning cycles?
Dependencies	What has to be true for this to work? Are those conditions in place? If not, what does it take to create them?

Where agencies stall: Setting goals at the technology level rather than the mission level. When the measure of success is "use cases deployed" rather than "processing time reduced," the organization optimizes for deployment rather than performance — and pilots proliferate without ever producing durable change.

Stage Three: Build the Governance Infrastructure

What to do: Establish the decision rights, accountability structures, and risk management practices that allow AI to be governed as a mission function — before expanding deployment.

Governance is the stage most agencies try to defer. It is also the stage that determines whether everything that follows will sustain or stall.

Establish or empower your Chief AI Officer. M-25-21 requires covered agencies to designate a Chief AI Officer with authority to develop and publish the agency's AI compliance plan, coordinate AI governance, and approve high-impact AI use cases.[10] The CAIO role only functions if it carries real authority — budget influence, decision rights over use case approval, and direct access to agency leadership. Agencies that have created CAIO roles without organizational backing have created compliance positions, not governance functions.

Establish an AI Governance Board. The governance board is the cross-functional body that coordinates AI oversight across legal, privacy, cybersecurity, acquisition, and mission offices. It should meet regularly — not just when there is a crisis — and should have defined authority to approve, modify, and terminate AI use cases. The Federal Reserve Board's AI governance model is a useful reference: its AI Program team, which spans senior staff and key compliance offices, is responsible for maintaining the use case inventory, evaluating permissibility, routing use cases through the appropriate governance path, and making approval decisions.[11] The structure distributes the work without diffusing the accountability.

Define your risk management tiers. Not all AI use cases carry the same risk. Your governance framework should establish at minimum three tiers: general use (low risk, no significant decisions, no sensitive data), moderate use (some mission consequence, requires monitoring and documentation), and high-impact use (principal basis for consequential decisions, requires pre-deployment testing, independent evaluation, human oversight, and CAIO approval). M-25-21's high-impact definition provides the compliance floor.[12] Your agency's mission context may require a higher standard in specific areas.

Establish a structured pathway from pilot to production. One of the most frequently cited governance gaps in federal AI is the absence of a defined, documented process for moving a use case from pilot status to operational deployment. The Federal Reserve Board addresses this directly, establishing structured technical and governance pathways including model validation, code review, and security risk assessment requirements as formal prerequisites for production deployment.[13] Without this pathway, AI systems exist in a perpetual pilot state — never fully authorized, never fully accountable, never fully governable.

Address the authorization timeline. The FedRAMP 20x program, launched in 2025, introduced automated authorization approaches designed to make cloud authorization simpler and faster for commercial providers.[14] Agencies should understand how FedRAMP 20x affects their authorization environment and whether it creates new pathways for AI deployments that have been waiting on traditional ATO timelines. DHS's shift to a continuous authorization model — a secure-by-design approach that gives developers access to existing secure systems rather than waiting for static approval cycles — is another model worth examining.[15]

Where agencies stall: Building governance frameworks that are designed to satisfy compliance requirements rather than to govern actual AI behavior. A governance structure that exists on paper but is not operationally connected to the people making AI deployment decisions is not governance. It is documentation.

Stage Four: Prepare the Workforce

What to do: Build the human capacity that AI depends on — at every level, not just in technical roles.

Workforce readiness is where most federal AI strategies have the widest gap between intention and execution. Agencies understand that workforce development is required. What they less frequently understand is the scope of what workforce readiness actually means.

It is not primarily a training problem. Training addresses knowledge gaps. Workforce readiness for AI also requires role redesign (what does this person's job look like when AI is part of it?), tool access (can this person actually use the AI systems the agency has deployed?), and cultural permission (does this person feel authorized to experiment, to surface problems when AI outputs are wrong, and to contribute to continuous improvement?).

Build literacy at every level — including leadership. Agency leaders who do not understand AI well enough to make informed decisions about it cannot govern it. This is not a criticism. It is a structural gap that requires a structural response. Leadership AI literacy programs should focus not on technical details but on decision-relevant concepts: what AI can and cannot do reliably, how to read an AI risk assessment, what questions to ask before approving a high-impact use case, and what accountability looks like when an AI system fails.

For mission staff, the priority is operational literacy: how to use AI tools appropriately in their specific work context, how to identify when an output is likely unreliable, and how to escalate concerns. The Department of State's StateChat platform — an enterprise GenAI tool available to approximately 45,000 active users by September 2025 — is an example of broad deployment done with workforce enablement in mind, including terms of service clarity, appropriate use guidance, and access to training resources alongside the tool itself.[16]

Create structured space for experimentation. Agencies that reward only successful outcomes suppress innovation. Staff who try AI approaches that do not work and face negative consequences will stop trying. Build structured sandbox environments where employees at all levels can experiment with AI tools, share what they learn, and contribute to identifying high-value use cases — with the explicit expectation that some experiments will fail and that failure is information, not a problem.[17]

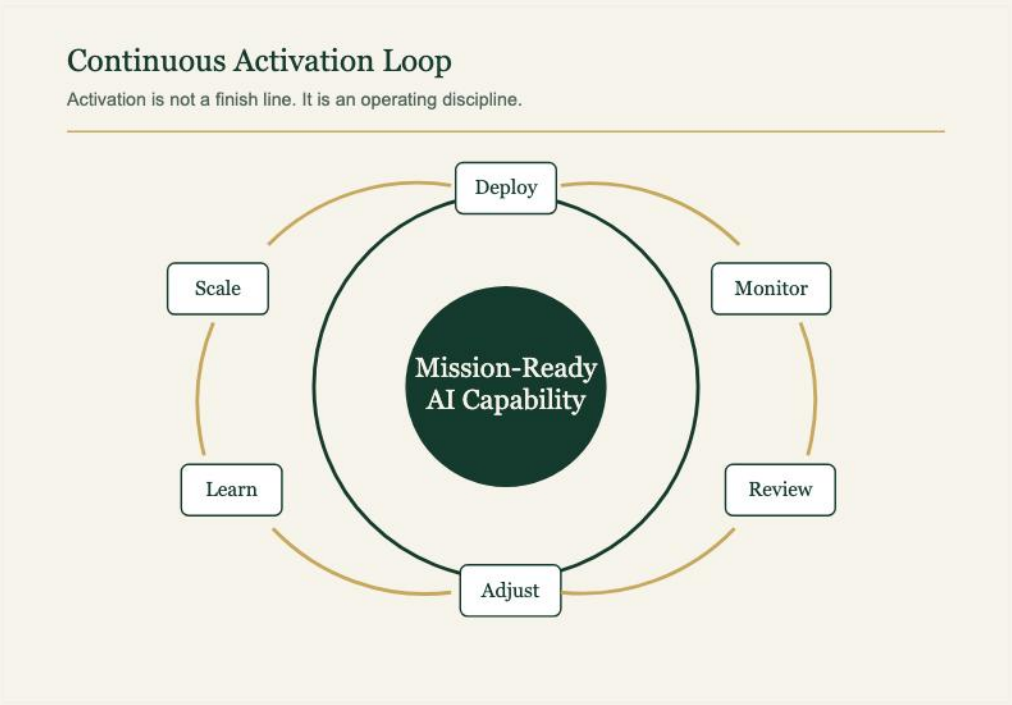
Embed technical capacity in mission-facing roles. Central AI offices cannot reach the mission. Agencies that concentrate technical AI capacity in a central function and expect it to serve the entire enterprise consistently underperform those that embed technical and multidisciplinary capacity in the components that are closest to the mission. DHS's staffing strategy — explicitly designed to identify skill needs and embed them across components — reflects this principle.[18]

Where agencies stall: Treating workforce development as a one-time program cost rather than an ongoing operational requirement. AI capabilities evolve continuously. A workforce prepared for the tools available today will be unprepared for the tools available in eighteen months. Continuous learning is not a nice-to-have. It is a maintenance requirement.

Stage Five: Scale, Monitor, and Sustain

What to do: Expand AI deployment deliberately, with active performance monitoring and a governance model that adapts as conditions change.

Activation is not a destination. It is an operating state — one that requires ongoing management, continuous adjustment, and a leadership posture that treats AI performance as a permanent operational responsibility rather than a project to be completed.



Scale gradually and intentionally. The agencies making the most durable progress start with one or two high-value, feasible use cases, build the governance and workforce muscle around them, and expand from that foundation. This is not timidity. It is discipline. "Big bang" implementations — attempting to deploy broadly across the enterprise before the infrastructure is ready — reliably produce the same outcome: broad deployment of underperforming systems that erode staff confidence and create political liability.[19]

Build monitoring into deployment, not after it. Every production AI system should have defined performance indicators, monitoring protocols, and thresholds for human review before it goes live — not as an afterthought. GSA's approach requires all

high-impact use cases to submit AI Impact Statements, independent evaluation plans, and real-world test results prior to deployment, with approved systems then subject to continuous monitoring, human-in-the-loop validation, and annual re-registration.[20] This is the model: authorization is not a one-time event. Performance is an ongoing responsibility.

Establish feedback loops that are fast enough to matter. Federal program cycles are measured in years. AI performance cycles are measured in weeks. The operating model must create mechanisms — formal and informal — that surface performance signals quickly enough for decision-makers to act on them. This often requires changes to how programs are structured and funded, not just how AI is governed. Where program timelines do not allow for rapid adjustment, agencies need governance-level mechanisms — standing review bodies, escalation protocols, and clear authority to pause or modify a deployment — that operate outside the normal program cycle.

M-25-21 requires ongoing accountability — use it. The OMB framework establishes ongoing monitoring through annual inventories, AI maturity assessments, and compliance reporting.[21] It also requires agencies to maintain the ability to pause or discontinue AI systems that fail to meet performance or safety thresholds, and mandates discontinuation of AI that cannot be made compliant with the memo's requirements.[22] These requirements are compliance floors, not performance ceilings. Agencies that treat them as the full scope of their accountability obligations are not managing their AI. They are managing their reporting.

Treat lessons learned as enterprise assets. A 2026 GAO report on federal AI acquisitions found that agencies were not systematically collecting and applying lessons learned across AI programs — meaning every new program was starting from the same baseline rather than building on accumulated institutional knowledge.[23] This is an operating model failure. Every AI deployment — successful or not — should generate documented learning that is accessible to the next program. Lessons about vendor performance, data quality, workforce readiness, authorization timelines, and governance gaps compound into capability when they are shared. They evaporate when they are not.

Where agencies stall: Treating activation as a finish line. The agencies that have achieved durable AI integration do not treat it as a completed transformation. They treat it as an operating discipline — one that requires continuous investment, continuous adjustment, and continuous leadership attention.

A Note on Authority

Every stage of this guide returns to the same underlying condition: authority.

Governance without authority is documentation. Workforce development without authority to change role structures is training. Scaling without authority to modify or pause a deployment is exposure without accountability.

The agencies furthest along in meaningful AI integration — not just in use case count, but in operational performance — share a common characteristic. Somewhere in the organization, someone has genuine authority over AI outcomes: the standing to set standards, the budget influence to back them, and the organizational access to enforce them when enforcement is required.

This is the condition that the operating model must create before everything else begins. Not a title. Not a charter. Actual authority, recognized and backed by the institution.

Everything else in this guide is executable once that condition is in place. Without it, the work will produce documentation, not transformation.

Using This Guide

This guide is designed to be used, not filed. A suggested approach:

For senior leaders: Read Stages One and Two in full. Determine whether your agency has honest answers to the maturity assessment questions. Determine whether your AI strategy is anchored to mission outcomes or technology outputs. Those two conditions determine the ceiling of everything that follows.

For CAIO and governance functions: Read Stages Three and Five in full. Stage Three defines the governance infrastructure your authority depends on. Stage Five defines the ongoing accountability obligations that M-25-21 establishes and that your governance model must operationalize.

For workforce and program leaders: Read Stages Two and Four in full. Stage Two will help you prioritize which AI opportunities deserve investment and in what sequence. Stage Four defines what workforce readiness actually means for your teams — beyond training completion rates.

For all: Return to Stage One regularly. Maturity assessments are not one-time events. The conditions change. The inventory grows. The workforce evolves. The policy environment shifts. An honest picture of where you are is the most useful tool in this guide — and the one most frequently allowed to go stale.

Notes

1. U.S. Government Accountability Office, *Artificial Intelligence: Generative AI Use and Management at Federal Agencies*, GAO-25-107653, July 29, 2025. https://www.gao.gov/products/gao-25-107653

2. Nextgov/FCW, "Agency AI use doubled in 2024, GAO finds," July 29, 2025. https://www.nextgov.com/artificial-intelligence/2025/07/agency-ai-use-doubled-2024-gao-finds/407067/

3. Federal News Network, "From strategy to structure: How federal agencies can build the organizational engine for AI at scale," May 2026. https://federalnewsnetwork.com/commentary/2026/05/from-strategy-to-structure-how-federal-agencies-can-build-the-organizational-engine-for-ai-at-scale/

4. Office of Management and Budget, Memorandum M-25-21, *Accelerating Federal Use of AI through Innovation, Governance, and Public Trust*, April 3, 2025. https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf

5. Ibid.

6. General Services Administration, *Strategies for OMB Memorandum M-25-21*, 2025. https://www.gsa.gov/technology/government-it-initiatives/artificial-intelligence/ai-guidance-and-resources/strategies-for-omb-m2521

7. OMB M-25-21, Section 4(a).

8. Inside Government Contracts, "OMB Issues First Trump 2.0-Era Requirements for AI Use and Procurement by Federal Agencies," April 8, 2025. https://www.insidegovernmentcontracts.com/2025/04/omb-issues-first-trump-2-0-era-requirements-for-ai-use-and-procurement-by-federal-agencies/

9. Federal News Network, "From pilots to production: How federal agencies are scaling AI for mission success," September 2025. https://federalnewsnetwork.com/federal-insights/2025/09/from-pilots-to-production-how-federal-agencies-are-scaling-ai-for-mission-success/

10. OMB M-25-21, Section 3.

11. Board of Governors of the Federal Reserve System, *Compliance Plan for OMB Memorandum M-25-21*, September 2025. https://www.federalreserve.gov/publications/compliance-plan-for-OMB-memorandum-m-25-21.htm

12. OMB M-25-21, Section 6.

13. Federal Reserve Board, *Compliance Plan for OMB Memorandum M-25-21*, September 2025.

14. General Services Administration, "AI in Action: How GSA is Transforming Federal Services," December 31, 2025. https://www.gsa.gov/blog/2025/12/31/ai-in-action-how-gsa-is-transforming-federal-services

15. Department of Homeland Security, *DHS AI Strategy for OMB Memorandum M-25-21*, September 2025. https://www.dhs.gov/sites/default/files/2025-09/25_0926_cio_dhs_ai_strategy_for_omb_m-25-21_508.pdf

16. Department of State, *Compliance Plan for OMB Memorandum M-25-21*, September 2025. https://www.state.gov/wp-content/uploads/2025/09/DOS-Compliance-Plan-with-M-25-21.pdf

17. Federal News Network, "From strategy to structure," May 2026.

18. Department of Homeland Security, *DHS AI Strategy for OMB Memorandum M-25-21*, September 2025.

19. Federal News Network, "From pilots to production," September 2025.

20. General Services Administration, *Strategies for OMB Memorandum M-25-21*, 2025.

21. Regulations.AI, "OMB Memorandum M-25-21: Accelerating Federal Use of AI Through Innovation, Governance, and Public Trust," 2026. https://regulations.ai/regulations/RAI-US-NA-OMMAFXX-2025

22. Ibid.

23. U.S. Government Accountability Office, *Artificial Intelligence Acquisitions: Agencies Should Collect and Apply Lessons Learned to Improve Future Procurements*, GAO-26-107859, April 2026. https://www.gao.gov/products/gao-26-107859

The Sainth works with federal agencies at every stage of this journey — from maturity assessment and governance design to workforce enablement and sustained operational integration. Our work is not advisory in the conventional sense. We are embedded partners in the design of how agencies operate.

For leaders ready to move from AI aspiration to activation, we are prepared to begin.
