GOVERNMENT AI IN PRACTICE

Research and analysis from the ThinkCapital GIAG Initiative

ISSUE 5 · MAY 6, 2026

EARLY SIGNAL: FROM THE RESEARCH

OMB M-25-21 required some 400 federal agencies to file AI governance documentation by April 3. They filed. The compliance dashboard stayed dark. Enforcement sat idle. The OIG issued nothing with compliance authority attached. The framework records completion. It cannot measure outcome.

The compliance record looks identical for every agency that filed. Governance programs that changed deployment decisions and programs that changed nothing are indistinguishable in the official record. That indistinguishability is a structural feature of how the framework was designed. This research is designed to measure the difference.

From the Editor

It's anyone's guess whether the April 3 OMB M-25-21 deadline changed a single AI deployment decision. Agencies filed. The process metrics closed green. Whether governance changed decision behavior is rarely measured and almost never asked. Why? The current framework was not designed to ask it.

This is the FISMA pattern. The Federal Information Security Management Act established compliance requirements in 2002. What took more than a decade to develop was the measurement infrastructure to distinguish agencies genuinely improving their security posture from agencies efficiently producing documentation without posture change. M-25-21 is FISMA applied to AI governance, one month in. The compliance record confirms completion. It confirms nothing about outcome.

Stream One structured interviews are now underway. Early intake confirms the central finding: most agencies cannot identify a production AI deployment their governance program constrained, modified, or halted. That is the question this issue puts at the center. Our analysis examines what the April 3 compliance event produced, why the current framework has no instrument to measure governance outcomes, and what a measurement standard built for that purpose would require.

The Morning After: What M-25-21 Compliance Actually Looked Like

OMB M-25-21 established a clear compliance clock. Agencies were required to complete AI use case inventories, designate a Chief AI Officer, update AI governance policies, and file required documentation by April 3, 2026. For agencies with functioning administrative operations, meeting a documentation deadline is not a difficult governance problem. Most did.

What M-25-21 did not establish was any verification mechanism for whether the documentation reflected operational governance. The compliance dashboard went unpublished. The OIG issued no

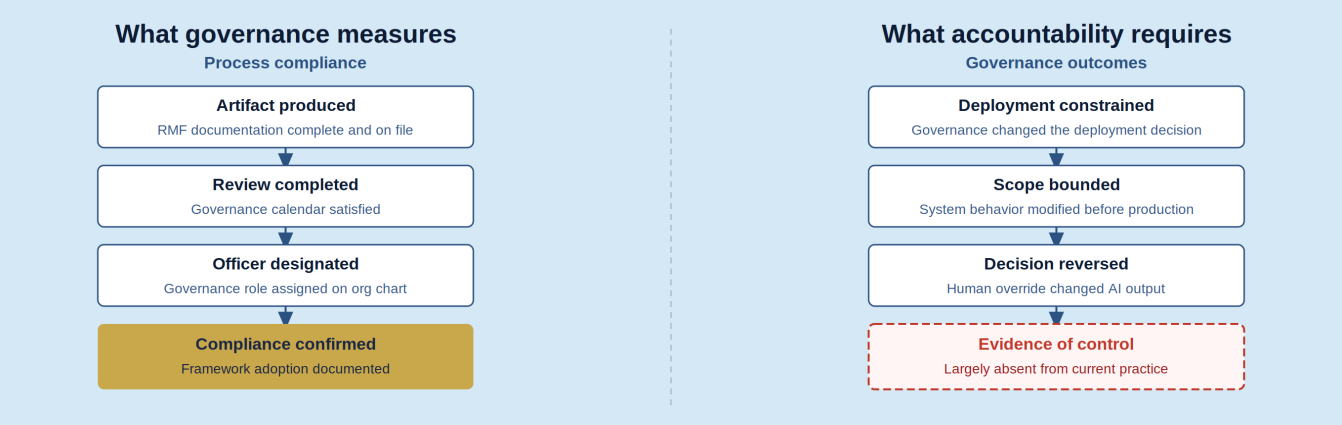
audit program with compliance authority attached to it. The deadline carried no enforcement consequence. M-25-21 is a directive with a compliance clock and no accountability infrastructure.

This is the FISMA pattern. The Federal Information Security Management Act established compliance requirements and regular reporting obligations in 2002. What it took more than a decade to develop was the measurement infrastructure to distinguish agencies that were genuinely improving their security posture from agencies that were efficiently producing the required documentation without security posture changes. The early FISMA years generated enormous documentation activity and modest security improvement. The first FISMA evaluation reports were explicit about this gap. M-25-21 is FISMA applied to AI governance, one month in.

The Framework Documents Completion. It Cannot Measure Outcome.

An agency that designated a Chief AI Officer, filed a use case inventory, and updated its AI governance policy by April 3 has met the M-25-21 requirement. An agency that used the CAO role to change how deployments are approved, scoped, and monitored produced a governance outcome. The compliance process cannot detect it. Both agencies file identical compliance reports. The M-25-21 verification architecture provides no basis for distinguishing them. This is a failure in measurement design, not implementation.

The framework M-25-21 sits within was not designed to ask whether required activities produced governance outcomes. It was designed to ask whether required activities were completed on schedule. Those are different questions with different answers.



*The compliance framework scores the left column. Accountability **requires** the right.*

The Identical Score Problem

Consider two agencies. Agency A has completed every M-25-21 requirement: use case inventory filed, CAO designated, governance policy updated. In the four months since deployment of its benefits routing AI system, the governance program has held scheduled reviews, produced required reports, and satisfied all compliance checks. The system has never been delayed, scoped down, or modified because of a governance review.

Agency B has the same M-25-21 documentation. It also has a CAO with defined decision authority over new AI deployments, a governance calendar that triggers operational review when production system behavior diverges from baseline, and a documented record of three deployment decisions that governance changed over the same period.

Under current M-25-21 measurement practice, these two agencies are indistinguishable. Both accurately report full compliance. An auditor using current assessment criteria cannot tell them apart. Agency A has governance outputs. Agency B has governance outcomes. The compliance architecture measures only the former.

The research question Stream One is designed to answer is not which agencies completed the M-25-21 requirements. It is which implementations are changing decision behavior. Those are different questions with different answers. Currently almost nobody is asking the second one.

What a Functional Measurement Standard Would Require

A measurement standard that distinguishes governance documentation from governance outcomes requires specifying what governance is supposed to accomplish, not what activities governance is supposed to perform. Victor Basili and colleagues at NASA developed exactly this discipline for software quality measurement in the 1970s and 1980s, when organizations were producing quality process outputs with no way to verify that those outputs reflected actual quality.

The Goal-Question-Metric approach works backward from outcomes: define the goal, derive the questions that would confirm whether the goal is being met, then identify the metrics that answer those questions. Applied to AI governance, the goal is not to produce a use case inventory or designate a CAO. The goal is to ensure that AI deployments operate within defined risk boundaries and that human accountability is preserved where it is required.

A CIO who cannot answer the following questions for their active AI deployments already has the answer to the more fundamental one: whether their governance program is producing outcomes or merely documentation.

- Has a governance review changed a deployment decision in the past 12 months?
- Is there a named individual with defined authority to halt or modify a production system outside the normal governance calendar?
- When the system encounters a situation outside its training distribution, what triggers a review?
- What proportion of escalation events result in a human override?

These questions have answers that distinguish Agency A from Agency B. The current M-25-21 compliance architecture asks none of them.

Early Signal: What Stream One Research Is Finding

Research Status: *Stream One structured interviews are in progress. Findings reported here reflect early-stage patterns from practitioner intake conversations and the public deployment record. These are directional observations, not conclusions.*

Pattern 1: Governance Is Front-Loaded, Not Operational

The most consistent early-intake pattern is that governance investment concentrates at the front of the deployment lifecycle. Risk assessment, authorization documentation, and policy development receive substantial organizational attention before deployment. Post-deployment operational oversight receives substantially less, and in many agencies is thin and unassigned. M-25-21 compliance activity reinforced this pattern: the deadline required documentation production, not operational monitoring design. The compliance event was itself front-loaded governance.

Pattern 2: The CAO Role Has Compliance Authority, Not Decision Authority

Multiple agencies designated Chief AI Officers in direct response to M-25-21. What is less common is a CAO with defined authority over production AI deployment decisions. The CAO role in most early-intake conversations holds compliance responsibility: ensuring required documentation is produced, ensuring the use case inventory is filed, ensuring governance procedures are documented. What CAOs typically

do not hold is the authority to delay or modify a deployment that the program office wants to proceed. Compliance designation and operational authority are different things. Most CAO positions reflect the former.

Pattern 3: Calendar Triggers Dominate Behavioral Triggers

Governance reviews occur when the compliance calendar requires them. They rarely occur when the production system generates a signal that warrants review. A system that encounters distributional shift, where it is processing cases outside its training distribution, does not trigger a governance review in most deployments. The calendar that drove M-25-21 compliance activity is the same calendar that drives operational oversight: it is keyed to the audit cycle, not to system behavior.

Pattern 4: Human-in-the-Loop Language Remains Undefined

Human-in-the-loop appears in M-25-21 and in virtually every agency governance document. The term is almost never operationally defined in the documents reviewed. It does not specify which decisions require human review, who conducts the review, what information the reviewer is given, what the review interval is, or what the documentation standard is. It functions as a policy commitment that satisfies a compliance requirement. It does not create an operational control.

Framework Comparison: EU AI Act vs. NIST AI RMF

M-25-21 references the NIST AI RMF as the governance foundation for federal AI programs. Government practitioners navigating compliance are therefore operating within a voluntary framework that establishes a risk management process without specifying what that process must accomplish. Understanding what each major framework measures, and where each is silent, is practical preparation for the accountability questions that compliance deadlines cannot answer.

M-25-21 VS. THE NIST AI RMF: A SHARED MODEL

Note that both M-25-21 and the NIST AI RMF were designed as compliance models:

- Define required activities
- Establish reporting obligations
- Verify completion of activities

Neither was designed to verify that activities produced outcomes. That shared design assumption is the structural source of the identical-score problem. An agency cannot distinguish itself through genuine governance outcomes when the compliance architecture only measures process completion.

Stream One research is examining what governance architecture looks like in organizations where oversight has demonstrably changed decisions. The findings will be the first comparative empirical data on this question across federal and state agency types.

It’s instructive to consider five dimensions for comparing the European and US risk frameworks. For the commercial CIO in a multinational context, the table below summarizes an asymmetry with immediate operational consequences. The EU AI Act is a binding legal instrument with named obligations, external verification, and enforcement. The NIST AI RMF is a voluntary process framework with self-assessment and no external verification. A US-based commercial organization with EU-facing deployments, European data subjects, or vendors operating across jurisdictions is already inside the EU Act's scope

whether its domestic governance program accounts for it. The compliance gap is not theoretical and does not wait for federal harmonization.

The CAO / accountability role row makes the sharpest practical point. The EU Act specifies what operators must be able to do: intervene, override, or shut down. A US commercial CIO who designated an AI officer to satisfy internal governance requirements may have created a compliance position with no operational authority. That organization's EU-facing deployments require something structurally different: named individuals with defined technical capabilities and documented intervention authority. The designation alone does not satisfy the requirement.

The governance outcome verification row has direct board-level implications. Post-market monitoring and incident reporting to a national authority means EU regulators will accumulate data on system behavior over time. US organizations operating under self-assessment alone are making a bet that their internal records will hold up if an EU regulatory action surfaces their deployment. That is a legal exposure question, not a governance philosophy question.

Dimension	NIST AI RMF	EU AI Act (High-Risk)
What is measured	Process completion: artifacts produced, roles filled, reviews conducted on schedule. M-25-21 compliance maps directly to this layer.	Conformity assessment and post-market monitoring: measurable evidence of compliance with specific technical requirements.
Oversight standard	Framework-defined but operationally undefined. GOVERN and MANAGE functions reference oversight; neither specifies what oversight must accomplish.	Explicit and legally binding for high-risk systems. Technical measures must enable human oversight throughout the system lifecycle.
CAO / accountability role	No defined authority standard for the role. M-25-21 mandates designation; neither document specifies what operational authority the role must carry.	Named operators with defined obligations. The regulation specifies what operators must be able to do, not just that a role must be designated.
Governance outcome verification	Agencies self-assess. No external verification standard for framework claims. Documentation is the evidence of compliance.	Post-market monitoring and incident reporting to national authority. External verification exists; documentation alone is insufficient.
Practitioner implication	Agencies can satisfy every RMF and M-25-21 requirement with no evidence that governance changed a deployment decision. The framework does not require that evidence.	US agencies with EU-facing deployments or cross-jurisdictional vendor relationships face binding requirements that exceed current federal guidance. The compliance gap is operational now.

For government CIOs and governance leadership, the comparison exposes what the US framework is missing by showing what a framework looks like when it requires outcomes rather than activities. Three takeaways carry the most weight.

First, the oversight standard row names the core problem precisely. GOVERN and MANAGE reference oversight without specifying what oversight must accomplish. The EU Act requires that technical measures enable human oversight throughout the system lifecycle. This standard implies monitoring, not just policy. A government CIO reading this should ask whether their agency's human oversight commitment is documented as a policy intention or built as a technical capability. Those are different things and the compliance record cannot tell them apart.

Second, the governance outcome verification row suggests what a stronger US framework would need to add: external verification with defined standards, not self-assessment. FISMA eventually developed that infrastructure. The AI RMF has not. Government CIOs who want their governance programs to go beyond compliance and to be credible. They have an interest in advocating for that infrastructure, because without it their strong programs look identical to weak ones in every official record.

Third, the practitioner implication row for the NIST AI RMF column contains the most actionable sentence in the table: agencies can satisfy every requirement with no evidence that governance changed a deployment decision, and the framework does not require that evidence. A government CIO whose program has changed deployment decisions should be documenting those instances now, independent of any compliance requirement. That record is the only thing that distinguishes their program when external verification eventually arrives. The EU Act's trajectory suggests it will.

Practitioner Diagnostic: Three Questions

The following three questions are drawn from the Stream One research design and from M-25-21 compliance practitioner conversations. They function as a rapid self-assessment for government IT leaders and CAOs responsible for AI governance programs.

1	**Has your governance program changed a production AI deployment decision in the past 12 months?**	If you cannot identify a specific example, a deployment delayed, a scope constrained, a system modified as a result of a governance review, your program may be producing documentation rather than control. M-25-21 compliance does not require this evidence. Accountability does.
2	**Does your CAO hold decision authority over deployment decisions, or compliance authority over documentation requirements?**	If your CAO can require documentation but cannot delay a deployment the program office wants to proceed, you have a compliance designation, not an accountability role. The distinction is operational, not definitional.
3	**What triggers a governance review: a scheduled calendar date, or a signal from the production system?**	Calendar-only triggers satisfy M-25-21 and audit requirements. Behavioral triggers, distributional shift monitoring, threshold-based review, anomaly detection, are the operational standard for governance that responds to what systems actually do. Both are required. Most agencies have only one.

STREAM TWO WORKING PAPER

The five-characteristic framework for human oversight in agentic AI deployments, irreversibility, consequence transfer, distributional novelty, value conflict, and legal or regulatory significance, is developed in full in *When Humans Must Intervene: A Decision-Grounded Framework for Human Oversight in Government and Commercial Agentic AI Deployments*, available at thinkcapital.org/publications.html.

Our Participation Ask

Both GIAG research streams are actively recruiting government and public sector technology practitioners for structured research interviews. Stream One examines NIST AI RMF implementation fidelity and M-25-21 compliance outcomes across agency types. Stream Two examines human oversight quality in agentic AI deployments.

Participation in either stream involves a 45-minute structured conversation conducted under your choice of attribution terms. Findings are reported in aggregate and without attribution unless participants choose otherwise. The research is designed to capture the actual range of implementation experience, including programs that have not produced expected outcomes.

WHO THIS RESEARCH NEEDS

Stream One:

- Federal or state agency IT leaders, Chief AI Officers, or AI program managers with direct experience implementing AI governance programs under M-25-21 or the NIST AI RMF.
- Government technology advisors or consultants who have supported agencies through M-25-21 compliance, RMF implementation, or governance architecture design.
- Practitioners who have encountered the gap between governance documentation and governance outcomes firsthand. Programs that have not produced expected results are as important to this research as programs that have.

Stream Two:

- Government IT leaders and AI program managers with active or planned agentic AI deployments.
- Practitioners who have worked on human oversight design for AI systems that execute actions, not just produce recommendations.

To express interest or schedule a conversation:
thinkcapital.org/research.html
calendly.com/thinkcapital

What's Ahead

Issue 6 will examine the Chief AI Officer role in practice. M-25-21 created a named accountability position. The NASCIO *Evolving Role of the State CIO as Change Leader* report (April 2026) describes a three-layer model for executive AI leadership. The analytical question: is the CAO role actually that, or is it a compliance designation sitting alongside existing IT leadership without changing decision authority? The FISMA ISSO analogy is instructive. A named role with a compliance clock is a compliance mechanism. Decision authority makes it governance.

Issue 7 will examine what the NIST AI Agent Standards Initiative tells us about the limits of the current RMF. NIST's Collaborative AI Safety Institute launched the Agent Standards Initiative in February 2026 to develop interoperable technical standards for autonomous AI agents. The analytical argument: the NIST AI RMF was designed for systems that predict and recommend, not for systems that chain decisions and execute actions autonomously. Measurement frameworks built for advisory AI produce systematically wrong readings when applied to agentic systems, for the same structural reason that

lines-of-code metrics fail when applied to object-oriented code. The RMF itself needs a different starting point, and the Agent Standards Initiative signals that NIST knows it.

Issue 8 will be the first data-bearing issue from Stream Two. By early June, 4 to 6 structured practitioner interviews should be complete. The framing discipline will be explicit: preliminary observations, N=5, patterns presented for discussion, not conclusions. That epistemic standard is what distinguished Capers Jones's early SPR work from consulting white papers with the same data.

Working Paper 1, *Implementation Fidelity in Government AI Governance*, has been published and is available at thinkcapital.org/publications.html.

Government IT practitioners with direct experience governing AI deployments in federal or state contexts are encouraged to reach out now. Both streams are open.

Government AI in Practice is a practitioner research letter published by the Government IT and AI Governance Initiative (GIAG), a research program of ThinkCapital LLC, Belmont, California. GIAG examines AI governance implementation in government and public sector contexts.

Michael Bragen, Principal, ThinkCapital LLC | michael.bragen@thinkcapital.org | thinkcapital.org | thinkcapital.substack.com