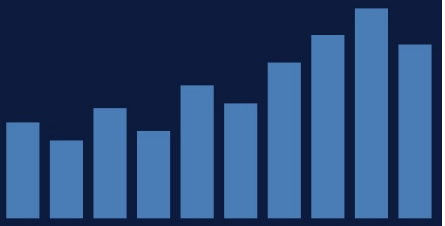
GOVERNMENT AI IN PRACTICE

Research and analysis from the ThinkCapital GIAG Initiative

ISSUE 9 · JUNE 3, 2026



EARLY SIGNAL: FROM THE RESEARCH

A practitioner exchange this week produced a question that cuts to the center of what this research program is trying to measure: how do you distinguish a governance framework that is in place from one that is working? The distinction sounds simple. In practice, it is the difference between an agency that can demonstrate its AI systems are performing as governed and one that can only show its policy documents are current.

That distinction is the practical problem this issue addresses directly. OMB Memorandum M-25-22, the federal AI procurement directive now eight months in effect, gave agencies specific obligations. This issue translates those obligations into a set of concrete actions that IT and AI leadership can take today to determine whether their contracts are producing governance or producing documentation.

From the Editor

The gap between governance on paper and governance in practice has been the central finding of the Government IT and AI Governance Initiative (GIAG) research since January. This week it showed up in a specific, actionable form. Ongoing research comparing AI governance compliance across US and EU frameworks asked how practitioners distinguish between frameworks that are in place and frameworks that are working. The question has no standard answer in any published framework. It does have a practical answer, drawn from our experience conducting hundreds of government and commercial enterprise IT assessments. You'll read about that in this issue.

Office of Management and Budget (OMB) Memorandum M-25-22 established specific AI procurement requirements eight months ago. Federal IT and AI leadership now face a concrete question: are your contracts producing the oversight M-25-22 requires, or are they producing documentation of that requirement? The issue walks through what to check, what to ask vendors, and what to escalate. The five questions at the close are designed to be used in a contract review meeting, filed as a self-assessment.

~ Michael

FEDERAL

What M-25-22 Requires

OMB M-25-22 established four specific obligations for federal AI procurement: performance-based acquisition language in contracts, pre-award testing and evaluation of AI systems before acceptance, ongoing monitoring of system performance after deployment, and disclosure requirements when contractors use AI in ways not specified in the original contract. Eight months in, the practical question for IT and AI leadership is whether the contracting infrastructure is executing those requirements.

Contracting Officer Technical Representatives (COTRs) are the primary mechanism for ongoing surveillance. In practice, COTR workloads frequently prevent systematic review of AI system performance against contracted specifications. The result is a structural gap: M-25-22 compliance is documented at award, and what the system is doing in production goes unverified against what the contract specifies. IT leaders who assume the contracting process is handling this verification are, in most cases, assuming more than the process delivers.

Actions to Take Now

First, pull the AI-related contracts currently in performance and identify whether each has M-25-22 performance language. Specifically, confirm whether the performance work statement defines measurable AI system outcomes. A contract that specifies delivery of an AI tool without specifying how its performance will be measured in production provides no basis for surveillance.

Second, confirm whether pre-award testing was conducted and documented. M-25-22 requires it; Contractor Performance Assessment Reporting System (CPARS) records and contract files should reflect it. Testing that occurred but was not documented in a form that supports ongoing comparison cannot serve its governance function. Testing that happened once and was not designed for repetition is a delivery event, not ongoing monitoring.

Third, identify whether a disclosure mechanism exists for unanticipated AI use. This is the M-25-22 provision least likely to have been operationalized. Contractors integrating new AI capabilities into existing task orders may be doing so without triggering any disclosure requirement, because the mechanism was never specified in the contract. Ask the program office: if a vendor started using an AI model to generate deliverables that was not present at award, would your agency know?

What the Audit Record Should Show

By June 2026, M-25-22 should be visible in contract surveillance records, COTR activity logs, and CPARS ratings for AI vendors. The Government Accountability Office (GAO) has the access and the mandate to assess whether it is. The Small Business Administration (SBA) AI accountability legislation that passed Congress unanimously this year, requiring AI inventories, use-case reporting, and transparency, adds a legislative baseline. IT leadership whose agencies have not yet aligned contract surveillance practice to M-25-22 now have both a regulatory and a legislative exposure to address.

Practitioners in AI procurement, contract management, or vendor oversight willing to describe what surveillance looks like in practice are the participants GIAG's Stream One research needs most. Contact us at research@thinkcapital.org to engage.

STATE

The Colorado Deadline and What It Demands from Contracts

State IT leaders face a parallel pressure with a hard deadline. Colorado Senate Bill 24-205 takes effect June 30, 2026, establishing a “reasonable care” standard for algorithmic systems affecting consequential decisions. “Reasonable care” is an outcome standard. Demonstrating it requires showing that a system performed as expected and that governance mechanisms detected and addressed deviations. That is a materially higher bar than a policy document.

Most state AI contracts were written before SB 24-205 was enacted. State Chief Information Officers (CIOs) and Chief AI Officers (CAIOs) in Colorado, and in states watching Colorado’s compliance experience, should now audit whether their vendor contracts specify the performance evidence needed to demonstrate reasonable care. Contracts that reference responsible use policies without specifying measurable performance obligations will not satisfy the standard.

The National Association of State Chief Information Officers (NASCIO) 2025 State CIO Survey documented that 88 percent of states have responsible use policies in place while 75 percent of state CIOs retain serious concerns about deploying generative AI (GenAI) in direct citizen services. The Colorado deadline requires a resolution of that tension. A responsible use policy that does not specify how vendor performance is measured and enforced will not serve as a defense against a reasonable care claim.

LOCAL

Citizen-Facing AI Without a Specified Accuracy Standard

Honolulu’s Department of Planning and Permitting this week deployed an AI assistant for permit applications, using natural language processing (NLP) to guide applicants and reduce submission errors. The practical benefit is real: processing backlogs driven by incomplete submissions impose costs on both applicants and staff, and real-time guidance addresses that directly.

The deployment announcement leaves three contract governance questions unanswered: the accuracy standard the system is held to, the testing conducted before deployment, and the recourse available to a citizen who follows incorrect guidance. If the system was procured from a vendor, the contract should specify performance standards, testing requirements, and liability for incorrect outputs. Local IT leaders procuring citizen-facing AI should treat those three terms as non-negotiable contract requirements, addressed before deployment.

PRACTITIONER SIGNALS: THREE TESTS THAT REVEAL GOVERNANCE MATURITY

A structured exchange this week with a doctoral researcher comparing AI governance compliance across US and European Union (EU) frameworks produced three diagnostic questions that proved more revealing than any framework alignment checklist. They are presented here as practical tools.

Test One: Can Anyone Describe the System’s Current Scope?

Ask the person responsible for a deployed AI system to describe its current operational scope: what data it accesses, what decisions it informs or makes, and how its outputs are used. An inability to answer accurately without consulting documentation signals an open governance gap. Scope drift, where

systems expand their operational footprint beyond original authorization without a corresponding governance review, is the single most common finding in GIAG's Stream Two intake interviews. The test takes five minutes and requires no framework.

Test Two: Has the Governance Structure Changed Since Deployment?

A system deployed 18 months ago under a governance structure built for its original scope requires updated governance if the system has since added data integrations, expanded user populations, or taken on additional decision functions. Ask whether the governance structure, including the oversight roles, review authorities, and audit mechanisms, has been updated since initial deployment to reflect how the system is operating. In most cases it has not. A governance structure that does not track the system it governs is a documentation artifact.

Test Three: Has a Review Ever Produced a Decision to Constrain the System?

This is the most diagnostic question of the three. Oversight that has never produced a decision to modify, limit, or stop a system is functioning as periodic documentation. Ask whether any governance review of a deployed AI system has ever resulted in a concrete action: a scope restriction, a deployment pause, a vendor requirement, or a contract modification. A consistent answer of no means the review process is generating records rather than governance outcomes. This is the authority gap GIAG's Stream Two research is documenting consistently across intake interviews.

The US-EU Framework Comparison

The researcher's EU background surfaces a structural difference with direct advisory implications. The EU AI Act's risk-tiering mandate requires outcome-based verification for high-risk AI systems, demonstrating that governance produced the outcomes it describes. The National Institute of Standards and Technology AI Risk Management Framework (NIST AI RMF) is process-oriented and voluntary. US practitioners working within the framework must demonstrate governance process. IT and AI leaders who want their governance to be defensible should hold their oversight structures to the outcome standard regardless of what the framework requires.

MEASUREMENT NOTE: WHY COMPLIANCE METRICS MISLEAD

A recurring finding in technology benchmarking research, including work conducted across central banks in multiple countries, is that the metrics organizations use to report performance were built to measure a prior state of technology adoption. AI governance compliance metrics have the same problem at the agency level. They were designed to capture documentation activity: policy adoption, inventory completion, framework alignment. Capturing behavioral change, system scope constraints, and governance outcomes falls outside their design.

An agency that scores well on compliance metrics and poorly on the three behavioral tests above represents the modal case in the GIAG data set. IT and AI leadership should treat compliance scores as a floor and apply the behavioral tests to identify where the gap is.

APPLIED RESEARCH: WP3 AND STREAM TWO

GIAG Working Paper 3 (WP3), “Mandate Translation: How Federal AI Governance Requirements Arrive at State and Local Agencies,” publishes later this month. The paper’s central finding is directly relevant to this issue: most agencies produce governance documentation aligned with federal requirements while leaving the oversight architectures those documents describe operationally unchanged. The M-25-22 contract surveillance gap documented above is a procurement-layer instance of the same pattern. A supplement examining M-25-22 specifically, covering what contract surveillance capacity is required to make the provisions operative and where current practice falls short, will follow the WP3 release.

Stream Two practitioner intake is ongoing. The two questions emerging most consistently from interviews: where in the decision workflow does the human reviewer sit, and what information do they have access to at the point of review? If you have a governance role in a deployed AI system and are willing to answer those questions, or if you have federal or state AI procurement experience to contribute to Stream One, contact research@thinkcapital.org or use the participation form at https://www.thinkcapital.org/research.html.

Five Questions for Practitioners

These questions are designed for use in a contract or governance review meeting. They identify where the gap between M-25-22 requirements and actual practice is most likely to be found.

1. Does each AI contract currently in performance contain measurable outcome specifications, beyond delivery milestones, that a COTR can use to evaluate ongoing system performance?

2. Was pre-award testing conducted before your agency accepted delivery of its AI systems? Is that testing documented in a form that supports comparison to current system performance?

3. Does your agency have a mechanism to detect when a contractor begins using AI capabilities not present at contract award? Has that mechanism been triggered?

4. What is the current operational scope of your largest deployed AI system? Does it match what was authorized at initial deployment? If scope has expanded, was a governance review conducted?

5. Who has the authority to require a contractor to modify or suspend AI system operation based on a performance finding? Has that authority ever been exercised?

Recent Intelligence

The following items from the GIAG Daily Intelligence Digest (May 28-30, 2026) are relevant to the M-25-22 and procurement governance themes in this issue.

NIST Consortium Rebranding

The National Institute of Standards and Technology (NIST) AI Safety Institute Consortium has been renamed the Artificial Intelligence Standards and Innovation Consortium (AISIC), reflecting a shift from safety-first framing toward standards development. Agencies whose governance frameworks cite the prior NIST AI Safety Institute should verify whether the rebranding affects the practical guidance they are calibrated to, particularly in contract specifications that reference NIST standards by name.

USDA Cybersecurity Audit

A United States Department of Agriculture (USDA) Inspector General audit found that the department prioritized rapid AI implementation over cybersecurity and governance controls, creating measurable vulnerabilities. The finding is a direct example of what M-25-22’s pre-award testing provisions are

designed to prevent. IT leaders in agencies under similar deployment pressure should treat the USDA audit as a contract review trigger: are your AI vendors subject to security evaluation requirements, and is compliance documented?

Federal Data Strategy Warnings

Multiple federal sources identified data strategy, rather than model access, as the primary bottleneck to scaling AI. The practical implication for IT leadership: procurement contracts for AI systems that omit data quality requirements, data governance obligations, and data access controls are missing the input layer that determines whether M-25-22's ongoing monitoring provisions can function. Data governance is a contract term, addressed alongside technical specifications.

House National Defense Authorization Act AI Incident Disclosure

The House National Defense Authorization Act (NDAA) includes provisions to establish a protected disclosure program for AI incidents, creating whistleblower protections for those reporting AI-related safety concerns in government contexts. If enacted, this creates an AI-specific incident reporting channel that would surface contract performance failures that current CPARS reporting does not capture.

HHS Predictive Child Welfare Funding

The Department of Health and Human Services (HHS) Administration for Children and Families is offering $6 million to state and local governments for predictive child welfare analytics systems. State IT leaders receiving these funds should treat procurement governance as a condition of the grant. High-stakes AI in child welfare services requires specified accuracy standards, bias testing requirements, and defined human review authority. Omitting those terms from the procurement creates significant legal and operational exposure.

Government AI in Practice

Published weekly by ThinkCapital LLC under the Government IT and AI Governance Initiative (GIAG), a practitioner research program examining AI governance implementation in federal, state, and local government. Research participation, practitioner inquiries, and correspondence: michael.bragen@thinkcapital.org. Archive and publications: thinkcapital.org/publications.html.

WP3 (Mandate Translation): How Federal AI Governance Requirements Arrive at State and Local Agencies. Publication in June 2026 at thinkcapital.org/publications.html