

AI Governance for Federal Programs

Acquisition, Oversight, and Mission Assurance

A practical guide for OMB M-25-21 and M-25-22 implementation, contractor AI oversight, and Inspector General audit readiness.

Governance is enforced at the authorization boundary: the point where an output becomes an effect-bearing action, and where evidence, not explanations, must exist.

PUBLISHER

FERZ, Inc.

VERSION

1.1 · June 2026

LICENSE

CC BY 4.0

FOR

Agency Chief AI Officers · Program Managers · Contracting Officers · Contracting Officer's Representatives ·
Inspector General Staff · Policy Officers

FRONT MATTER

License and Terms of Use

Copyright and License. Copyright © 2026 FERZ, Inc. This work is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). You are free to share and adapt this material for any purpose, including commercial purposes, with appropriate attribution.

Attribution. FERZ, Inc., "AI Governance Executive Guide: Federal Programs," 2026. Available at ferz.ai.

Public Reference Framework. FERZ, Inc. publishes this as a public, vendor-neutral reference framework based entirely on public sources. It does not use nonpublic government information, does not speak for or on behalf of any federal agency, and does not imply government endorsement.

Framework Scope and Citation Currency. This guide is a foundational governance framework, not an exhaustive policy tracker or current-law reference. It cites statutes, regulations, memoranda, standards, and guidance only where they materially affect the general governance model, acquisition posture, evidence requirements, oversight structure, or audit-readiness framework. Use-case-specific, agency-specific, jurisdiction-specific, or rapidly changing requirements may apply and should be verified separately before reliance or implementation.

Not Legal, Procurement, or Ethics Advice. This document is provided as a general reference. The clause language, checklists, and procedures in this guide are illustrative and are not legal, procurement, or compliance advice. Agencies should adapt any contract language to their own authorities, regulations, and counsel review. This document is provided "as is," without warranties of any kind.

Commercial Scope Notice. Commercial services, tooling, and implementation support derived from this framework are offered separately by FERZ, Inc. or its authorized partners.

Limitation of Liability. To the maximum extent permitted by law, FERZ, Inc. is not liable for any direct, indirect, incidental, consequential, special, exemplary, or other damages arising from use of, reliance on, adaptation of, or implementation based on this publication. Users are responsible for obtaining appropriate legal, regulatory, procurement, ethics, technical, and compliance review before applying any framework, checklist, clause language, or procedure described in this guide.

EXECUTIVE AT-A-GLANCE

What This Guide Does

Audience

Agency Chief AI Officers, Program Managers, Contracting Officers (COs), Contracting Officer's Representatives (CORs), Inspector General staff, and policy officers responsible for AI systems acquired, developed, or operated under federal contracts.

Core Thesis

Governance is enforced at the authorization boundary: the point where an output becomes an effect-bearing action, and where evidence, not explanations, must exist.

Your Challenge

Your agency remains accountable for AI systems contractors build and operate on its behalf. But contractors often know more about the AI than you do. This creates an accountability gap: your agency owns the risk, but may lack visibility into the system. This guide helps you close that gap.

Two Governance Tracks

TRACK	PRIMARY RISK	YOUR FOCUS
Decisioning AI	Mission risk: safety, rights, security	Require validation evidence, enforce human oversight, and place a pre-execution authorization boundary on effect-bearing actions
Communications AI	Records and FOIA exposure	Require records management and disclosure controls

Four-Tier Classification

Tier 1 corresponds to high-impact AI as defined by OMB M-25-21. Tiers 2 to 4 are an oversight-calibration scheme for AI that falls below the high-impact threshold.

TIER	RISK LEVEL	WHAT YOU REQUIRE
Tier 1	Critical (high-impact)	Full validation evidence, test and evaluation integration, human oversight proof, pre-execution authorization boundary
Tier 2	Elevated	Proportionate validation, annual review, monitoring data
Tier 3	Standard	Operational controls, FedRAMP or applicable agency security-baseline alignment, basic documentation

TIER	RISK LEVEL	WHAT YOU REQUIRE
Tier 4	Minimal	Acceptable use policy acknowledgment only

90-Day Implementation

Weeks 1 to 4 (Foundation): Inventory AI across your contracts, establish tiering criteria, draft PWS and SOW clauses.

Weeks 5 to 8 (Contractor Oversight): Require evidence packages from contractors, run gap analysis on Tier 1 systems.

Weeks 9 to 12 (Operationalization): Implement monitoring, compile IG-ready documentation, conduct a mock audit.

EXECUTIVE SUMMARY

Accountability Without Visibility

Federal programs increasingly rely on AI systems that contractors develop, deploy, and operate. Your agency remains accountable for AI systems contractors build and operate on its behalf. But accountability without visibility is a trap.

The challenge is asymmetric. Contractors often understand the AI better than the government program office. They built it. They trained it. They know its limitations. Without the right acquisition structures and oversight mechanisms, you are accepting risk you cannot see or manage.

This guide provides the framework to govern AI that contractors operate on your behalf: what to require in contracts, what evidence to demand, how to verify compliance, and how to survive an IG audit.

The Core Problem

OMB M-25-21 makes agencies responsible for AI governance. But when AI is contractor-operated:

- You may not know what AI systems are running on your contracts.
- You may not have access to training data, validation results, or performance metrics.
- You may learn about problems after they affect citizens or operations.
- You bear the accountability, but the contractor holds the information.

The Solution: Governance Through Acquisition

If you cannot mandate internally, you can require contractually. OMB M-25-22, the federal AI acquisition guidance, reinforces this directly: it directs agencies to write contract terms for ongoing monitoring and evaluation, data and intellectual property rights, vendor lock-in prevention, independent evaluation access, and disclosure of unanticipated vendor AI use. This guide provides:

Illustrative PWS and SOW clause language: contract terms that require AI governance.

Evidence requirements: what to demand from contractors, scaled by risk tier.

Verification procedures: how to confirm contractors are actually complying.

IG audit preparation: documentation that survives Inspector General scrutiny.

Where Authorization Fits

For high-impact AI, requiring evidence is necessary but not sufficient. The governing control for an effect-bearing AI action is the point at which that action is permitted to release. Logs record what an AI did after the fact. An authorization boundary determines, before release, whether the action was permitted under the applicable policy, authority chain, and oversight condition. This guide treats that boundary as the enforcement point for the minimum practices M-25-21 requires.

Two Governance Tracks

AI governance operates across two distinct tracks with different risks.

Decisioning AI: eligibility determinations, threat detection, resource allocation, and predictive analytics that affect mission or citizen outcomes. You must require validation evidence and human oversight mechanisms, and you should place a pre-execution authorization boundary on effect-bearing actions.

Communications AI: citizen correspondence, caseworker copilots, report drafting, and FOIA response assistance. You must require records management controls and clear authorship indication.

While communications AI may not directly execute decisions, its risk can be equal or higher due to Federal Records Act obligations, FOIA exposure, and public trust. Do not assume communications AI is lower risk.

PART I

The Regulatory Landscape

CHAPTER 1

Your Obligations Under OMB M-25-21

OMB Memorandum M-25-21, "Accelerating Federal Use of AI through Innovation, Governance, and Public Trust," issued April 3, 2025, establishes AI governance requirements for federal agencies. It rescinds and replaces M-24-10. These requirements apply to AI that contractors operate on your behalf.

REGULATORY CURRENCY

M-25-21 consolidates the prior rights-impacting and safety-impacting categories into the current high-impact AI framework. Existing rights and safety concepts remain relevant, but the operative category for this guide is high-impact AI. The acquisition counterpart, M-25-22, "Driving Efficient Acquisition of Artificial Intelligence in Government," rescinds and replaces M-24-18. Both memoranda implement Executive Order 14179, "Removing Barriers to American Leadership in Artificial Intelligence." Executive Order 14148 separately revoked Executive Order 14110; Executive Order 14179 directed review of actions taken under Executive Order 14110 and revision of prior AI policy actions inconsistent with the current federal AI policy.

Agency Responsibilities

M-25-21 requires your agency to:

Designate a Chief AI Officer: accountable for agency AI governance, including contractor-operated AI.

Maintain a public AI use-case inventory: including AI systems operated by contractors under federal contracts.

Publish an agency AI strategy and compliance plan: on OMB's implementation timeline.

Implement minimum risk-management practices: for high-impact AI, regardless of who operates it.

Key principle: contractor operation does not transfer your accountability. If a contractor operates AI that meets the high-impact threshold, your agency remains responsible for ensuring it meets OMB requirements.

High-Impact AI

M-25-21 defines high-impact AI by whether the AI output serves as a principal basis for decisions or actions with a legal, material, binding, or significant effect on matters such as:

- civil rights, civil liberties, or equal opportunity;
- access to government benefits, services, or programs, and critical government resources;
- human health and safety, or public safety;

- critical infrastructure or strategic assets.

For high-impact AI, M-25-21 directs agencies to implement a set of minimum practices before and during use. Where a high-impact use cannot meet those practices, the agency must safely discontinue it, and must cease use entirely if proper mitigation is not possible.

Minimum Practices for High-Impact AI

The minimum practices include pre-deployment testing, AI impact assessment, ongoing monitoring, adequate human training and oversight, additional human oversight and intervention, accountability, consistent remedies or appeals for affected individuals, and appropriate fail-safes. These are the obligations that contract requirements and oversight must operationalize.

THE AUTHORIZATION BOUNDARY

OMB M-25-21 requires agencies to implement minimum practices for high-impact AI, including pre-deployment testing, impact assessment, ongoing monitoring, human oversight, intervention, accountability, and appropriate fail-safes. The authorization boundary is where those requirements become enforceable. A high-impact AI action releases only on an ALLOW verdict. If the authorization artifact cannot be produced, the action fails closed.

Agency-specific requirements: M-25-21 establishes government-wide expectations. Individual agencies (including DoD, DHS, HHS, and VA) may impose stricter controls through agency-specific AI policies, directives, or guidance. Always check for agency supplements that may exceed the government-wide baseline.

NIST AI Risk Management Framework

While voluntary, the NIST AI Risk Management Framework (AI RMF 1.0) provides the operational structure for meeting OMB expectations. Require contractors to demonstrate alignment with its four functions:

Govern: policies, roles, and accountability structures for AI risk management.

Map: documentation of AI system context, capabilities, and limitations.

Measure: methods for assessing and tracking AI risks.

Manage: procedures for responding to and mitigating AI risks.

NIST is revising the AI RMF pursuant to America's AI Action Plan; the four-function core is stable. For generative AI, the NIST AI 600-1 Generative AI Profile is the companion resource.

OMB M-26-04: Unbiased AI Principles for LLM Procurement

OMB M-26-04, "Increasing Public Trust in Artificial Intelligence Through Unbiased AI Principles," adds a procurement layer for large language models. For procured LLMs, agencies must address compliance with the Unbiased AI Principles: truth-seeking and ideological neutrality. The memorandum complements M-25-21 and M-25-22; it does not replace the high-impact AI minimum practices. For federal AI acquisition, this means LLM vendors may need to provide documentation, evaluation evidence, feedback mechanisms, and contractually material commitments sufficient for the agency to assess compliance.

CHAPTER 2

The Contractor Oversight Challenge

Government IT is largely outsourced. This creates structural challenges for AI governance that you must address through acquisition and oversight.

The Information Asymmetry Problem

Contractors typically have advantages you do not:

Technical depth: they built the AI and understand its architecture, training data, and failure modes.

Operational visibility: they see performance data daily; you may see quarterly reports.

Incentive alignment: their incentive is to deliver on contract terms, and transparency about AI limitations may not serve that goal.

The Authority Fragmentation Problem

Your governance authority is often fragmented:

CO authority: can modify contracts, but may lack technical AI expertise.

COR authority: monitors performance, but may lack authority to require changes.

Program office: understands the mission, but may not control procurement.

Agency CAIO: has the AI governance mandate, but may lack contract-level enforcement.

The Solution: Governance Through Acquisition

You cannot mandate what is not in the contract. But you can:

- include AI governance requirements in new contracts from the start;
- modify existing contracts to add AI oversight provisions;
- require evidence packages as contract deliverables;
- make AI governance compliance a factor in performance evaluation;
- exercise audit and independent evaluation rights to verify contractor assertions.

M-25-22 supports each of these. It directs agencies to use performance-based acquisition, to retain rights to government data and to address intellectual property and data ownership, to prevent vendor lock-in, and to require ongoing monitoring and evaluation across the AI lifecycle.

What Contractors May Resist Providing

Contractors may resist providing training data demographics, validation results by subgroup, incident history, limitation documentation, and audit access. Resistance to providing these items is itself a signal. Make them contractual requirements.

PART II

Risk Framework

CHAPTER 3

AI System Classification

What Counts as AI Under Your Contracts

Contractors may not volunteer that a system includes AI. You must define AI broadly in contracts:

- Include machine learning models, neural networks, expert systems, automated decision systems, predictive analytics, natural language processing, computer vision, and robotic process automation with adaptive logic.
- Include embedded AI: AI features within commercial products, such as cloud AI services, productivity tool copilots, and search with AI ranking.
- Include subcontractor AI: AI provided by subcontractors as part of prime contractor performance.

The Four-Tier Framework

Not all AI requires the same oversight. Tier 1 corresponds to high-impact AI under M-25-21. Use this framework to calibrate your requirements:

TIER	CRITERIA	EXAMPLES	WHAT YOU REQUIRE
Tier 1	High-impact: principal basis for effect-bearing decisions or actions	Benefits eligibility, threat detection, autonomous systems, citizen triage	Full evidence package, T&E results, bias testing, human oversight proof, incident history, pre-execution authorization boundary
Tier 2	Operational impact, federal employee-facing, significant data	Caseworker tools, procurement assist, document processing	Validation summary, annual review, monitoring data, access for verification
Tier 3	Internal operations, no citizen impact, low sensitivity	Meeting transcription, internal search, scheduling	Basic documentation, security controls, acceptable use compliance
Tier 4	Personal productivity, no government data	General research tools, personal drafting	Acceptable use policy only

AUTOMATIC TIER 1 TRIGGERS (NON-NEGOTIABLE)

- 1. The output is a principal basis for decisions or actions affecting civil rights, civil liberties, or equal opportunity.
- 2. The output is a principal basis for decisions or actions affecting human health, safety, public safety, or critical infrastructure.
- 3. The system produces citizen-facing determinations, even if the contractor claims the output is human reviewed.
- 4. The system involves classified information, controlled unclassified information (CUI), or sensitive personally identifiable information in decision logic.
- 5. The system has autonomous or semi-autonomous action capability.

CHAPTER 4

Contractor Evidence Requirements

You cannot verify what you do not require. This chapter specifies what evidence to demand from contractors, by tier.

Tier 1 Evidence Package

For every Tier 1 (high-impact) AI system, require contractors to provide:

EVIDENCE ITEM	WHAT IT CONTAINS	WHY YOU NEED IT
AI System Description	Architecture, intended use, decision authority, limitations, contraindications	Establishes scope and boundaries of the system
Training Data Documentation	Sources, demographics, selection criteria, known gaps, provenance chain	Enables assessment of bias risk and representativeness
Validation Results	Performance metrics, demographic subgroup analysis, false positive and negative rates	Shows the system works as claimed, across populations
Civil Rights, Safety, and Fairness Testing	Methodology, performance results across relevant populations or operational contexts, adverse -impact analysis where applicable, and remediation steps	Supports M-25-21 high-impact AI review, including privacy, civil rights, civil liberties, safety, and unlawful -discrimination risk
LLM Unbiased AI Principles Evidence, where applicable	Vendor documentation, model/system/data cards where available, evaluation methodology, prompt -pair or comparable testing for relevant bias risks, feedback process, and enterprise controls	Supports EO 14319 and M-26-04 requirements for procured LLMs and applicable generative AI
Human Oversight Procedures	Who reviews, with what authority, what training, escalation paths	Shows human oversight is meaningful, not a checkbox

EVIDENCE ITEM	WHAT IT CONTAINS	WHY YOU NEED IT
Incident History	Past failures, complaints, near-misses, corrective actions taken	Reveals problems the contractor may not volunteer
Monitoring Procedures	What metrics, how often, drift detection, alert thresholds	Ensures ongoing performance, not just initial validation

Tier 2 Evidence Package

For Tier 2 AI systems, require: an AI system description and intended-use statement; a validation summary (which may be less detailed than Tier 1); annual performance review results; quarterly monitoring data; and access for government verification.

Tier 3 Evidence Package

For Tier 3 AI systems, require: an AI system description; security and data-handling documentation; and an acceptable use policy compliance attestation.

What “Human Oversight” Actually Means

Contractors often claim “human in the loop” without meaningful review. Require contractors to demonstrate:

Authority: the human reviewer has authority to override or block the AI output.

Training: reviewers are trained on AI limitations and what to look for.

Time: reviewers have adequate time to actually review, not just click “approve.”

Accountability: reviewers are accountable for the decisions they approve.

Checkbox acknowledgments without substantive review do not satisfy human oversight requirements.

PART III

Acquisition Integration

CHAPTER 5

Illustrative PWS and SOW Clause Language

The clause language below is illustrative model text that agencies may adapt for Performance Work Statements and Statements of Work, subject to counsel review and agency authorities. It is organized around the contract requirements M-25-22 directs agencies to address. It is not legal or procurement advice. Agencies should adapt this language through normal acquisition channels, including applicable FAR provisions, agency supplements, program-specific authorities, and counsel review.

AI System Definition

Include in every contract to establish scope.

CLAUSE: AI SYSTEM DEFINITION

(a) "AI System" means any software that uses machine learning, neural networks, natural language processing, computer vision, or other techniques to generate outputs, predictions, recommendations, or decisions that would otherwise require human judgment, including statistical, algorithmic, or rule-based decision systems that materially influence outcomes. This includes standalone AI tools, AI features embedded within commercial products, AI services accessed by API or cloud, and AI systems provided by subcontractors.

(b) "Governed Unit" means an AI System performing a defined use case within a business process, including its inputs, outputs, integrations, and decision authority. The same AI System deployed in different use cases constitutes separate Governed Units, each subject to independent classification and governance.

(c) The Contractor shall notify the Contracting Officer before first use of any AI System in contract performance when reasonably known before use. If AI use is introduced through a product update, subcontractor, or other unanticipated mechanism, the Contractor shall disclose it promptly upon discovery and before relying on it for any effect-bearing function. The Contractor shall identify each distinct Governed Unit.

AI Inventory and Classification

CLAUSE: AI INVENTORY AND CLASSIFICATION

- (a) The Contractor shall maintain a complete inventory of all AI Systems used in contract performance, including systems provided by subcontractors.
- (b) For each AI System, the Contractor shall provide to the Contracting Officer: system name and description; intended use and decision authority; data types processed; proposed risk tier with justification; and whether the system meets the high-impact threshold under applicable OMB guidance.
- (c) The Government reserves the right to reclassify any AI System to a higher tier.

High-Impact AI Governance Requirements

CLAUSE: HIGH-IMPACT AI GOVERNANCE REQUIREMENTS

- (a) For any AI System classified as Tier 1 (high-impact), the Contractor shall provide: training data documentation including sources, demographics, selection criteria, and known limitations; validation results including performance metrics across demographic subgroups; bias and fairness testing methodology and results; human oversight procedures demonstrating meaningful review capability; incident history and corrective actions; and ongoing monitoring procedures and performance data.
- (b) The Contractor shall obtain Contracting Officer approval before deploying any Tier 1 AI System.
- (c) The Contractor shall notify the Contracting Officer within 48 hours of any incident affecting Tier 1 AI System performance or citizen outcomes.

Government Data and Model Use Restrictions

CLAUSE: GOVERNMENT DATA AND IP RIGHTS

- (a) The Government retains rights to government data used in or generated by the AI System through all phases of development, operation, and transition.
- (b) The Contractor shall not use government data to train, fine-tune, or improve any model for the Contractor's own purposes or for other customers without the express written consent of the Contracting Officer.
- (c) The contract shall address ownership and licensing of models, weights, and outputs developed using government data, consistent with applicable OMB guidance on data and intellectual property rights.

Vendor Lock-In and Portability

CLAUSE: PORTABILITY AND LOCK-IN PREVENTION

- (a) The Contractor shall provide documentation and reasonable transition support sufficient to enable the Government to migrate the use case to an alternative provider or to in-house operation.
- (b) The Contractor shall deliver, on request and at contract close-out, government data and any government-owned models, configurations, and documentation in a non-proprietary or interoperable format to the extent practicable.

Independent Evaluation and Audit Rights

CLAUSE: AI AUDIT AND INDEPENDENT EVALUATION RIGHTS

- (a) The Government, including the Inspector General, reserves the right to audit, inspect, test, and independently evaluate any AI System used in contract performance. This includes access to AI system documentation, training data descriptions, and validation records; performance monitoring data and incident logs; personnel involved in AI development and operation for interview; and test environments for independent validation.
- (b) The Contractor shall provide requested AI documentation within 5 business days of Government request.
- (c) The Contractor shall flow down equivalent audit and independent evaluation rights to subcontractors providing AI Systems.

AI Subcontractor Flow-Down

CLAUSE: AI GOVERNANCE FLOW-DOWN

- (a) The Contractor shall flow down all AI governance requirements to subcontractors providing AI Systems or AI-enabled services.
- (b) The Contractor is responsible for subcontractor AI governance compliance and shall maintain evidence of compliance available for Government inspection.
- (c) The Contractor shall notify the Contracting Officer within 48 hours of any subcontractor AI System incident, and the Government may require direct access to subcontractor AI documentation and personnel.

Disclosure of Unanticipated Vendor AI Use

CLAUSE: UNANTICIPATED AI-USE DISCLOSURE

The Contractor shall promptly disclose to the Contracting Officer any use of AI in contract performance that was not identified at award, including AI newly introduced into a commercial product or service used to perform the contract, and shall classify each such use under the tiering framework before relying on it for any effect-bearing function.

Generative AI Controls

CLAUSE: GENERATIVE AI CONTROLS

- (a) For any generative AI used in contract performance, the Contractor shall: prevent input of classified, CUI, or sensitive PII without appropriate controls; require human review of AI-generated content before external release or entry into federal records; clearly indicate when content is AI-generated in any document that becomes a federal record; and retain AI-generated drafts as well as final versions in accordance with applicable NARA General Records Schedules and agency-specific records schedules.
- (b) The Contractor shall not use generative AI for autonomous citizen-facing communications without explicit Contracting Officer approval.
- (c) For any LLM procured, provided, configured, or materially used in contract performance, the Contractor shall provide documentation sufficient for the Government to assess compliance with applicable Unbiased AI Principles, including truth-seeking and ideological neutrality. Documentation may include model, system, or data cards where available; evaluation methodology and results; prompt-pair or comparable testing for relevant bias risks; mechanisms for end-user feedback; and disclosure of enterprise-level controls, system instructions, filters, or other configurations that materially affect outputs.
- (d) Where the solicitation or contract identifies Unbiased AI Principles requirements as material to eligibility, payment, or continued performance, failure to provide required documentation or to correct identified noncompliance may support contractual remedies, subject to the contract terms and applicable law.

AI System Suspension and Fail-Closed Authority (Tier 1)

For high-impact AI, consider including this suspension authority. It is the contractual expression of the fail-closed principle: when governance cannot be established, the effect-bearing action stops.

CLAUSE: AI SYSTEM SUSPENSION AUTHORITY

- (a) The Government may direct the Contractor to suspend operation of any AI System, in whole or in part, pending investigation of: reported or suspected incidents affecting citizen outcomes or mission performance; evidence of unlawful discrimination, material adverse impact, or material output bias under applicable contract requirements; material deviation from validated performance parameters; or failure to provide required documentation or access.
- (b) The Contractor shall suspend AI System operation within 24 hours of Government direction and shall not resume operation without written Contracting Officer authorization.
- (c) Suspension under this clause does not constitute a termination for convenience or default.

PART IV

Oversight and Verification

CHAPTER 6

Verifying Contractor Compliance

Requiring evidence is not enough. You must verify that contractors are actually complying, not just providing paperwork.

Verification Activities by Tier

ACTIVITY	TIER 1	TIER 2	TIER 3
Review evidence package for completeness	Before deployment	Before deployment	Annual
Validate contractor tier classification	Before deployment	Before deployment	As needed
Spot-check performance monitoring data	Quarterly	Semi-annual	N/A
Independent testing (sample cases)	Annual	As needed	N/A
Interview contractor AI personnel	Before deployment and annual	As needed	N/A
Review incident reports and corrective actions	Within 72 hours of incident	Monthly	N/A

Logs Versus Evidence

A monitoring log and an authorization artifact answer different questions. A log records that an event occurred and what the system did. An authorization artifact records whether an effect-bearing action was permitted before it released, under the applicable policy, authority chain, provenance condition, and human oversight requirement. For Tier 1 systems, require the second, not only the first.

OBSERVABILITY IS NOT AUTHORIZATION

Monitoring, logging, and after-the-fact review show that an AI system did something. They do not establish that the effect-bearing action was permitted before it released. The governing control is the authorization boundary: the point at which the action is allowed to proceed, recorded in a tamper-evident authorization artifact. If that artifact cannot be produced, the action fails closed.

Combining monitoring logs, contractor status reports, performance dashboards, telemetry, and independent evaluation results may strengthen oversight, but it does not itself produce a determination,

made before release, of whether an effect-bearing AI action was permitted. For effect-bearing AI actions, require the authorization artifact, not only monitoring data, logs, or contractor reports.

Red Flags That Require Escalation

Escalate to your CO, agency CAIO, or IG if you observe: a contractor unable to produce required documentation within the specified timeframe; an evidence package with obvious gaps or inconsistencies; a contractor that resists or delays audit and inspection requests; performance monitoring data showing unexplained degradation; incidents not reported within required timeframes; a contractor that claims AI is “proprietary” to avoid providing documentation; or tier classification that appears inconsistent with actual system risk.

When to Conduct Independent Testing

Do not rely solely on contractor-provided validation. Conduct independent testing when deploying a new Tier 1 AI system, when contractor validation methodology appears weak, when user complaints suggest performance issues, when an IG or GAO audit is anticipated, or when significant changes to an AI system are deployed.

CHAPTER 7

IG Audit Readiness

Your Inspector General is increasing focus on AI. This chapter helps you prepare for and survive AI-related audits.

What IG Will Ask For

Based on emerging IG audit practices, expect requests across three areas. **Inventory and governance:** a complete inventory of AI systems on your contracts; evidence of risk tiering and classification rationale; policies and procedures for AI oversight; and roles and responsibilities for AI governance. **Contractor oversight:** evidence that you required AI governance in contracts; evidence packages received from contractors; records of your verification activities; and incident reports and your response. **Compliance with OMB M-25-21:** how you identified high-impact AI; evidence that minimum practices are in place for high-impact AI; and coordination with your agency CAIO.

Building Your IG-Ready Documentation

For each Tier 1 AI system, maintain a ready-reference file containing:

TAB	CONTENTS
1	Contract excerpt showing the AI governance clauses you included
2	AI inventory entry with tier classification and your rationale
3	Contractor evidence package (complete)
4	Your verification activities and findings

TAB	CONTENTS
5	Incident reports and corrective actions
6	Your approval decision and sign-off for deployment

Common IG Findings to Avoid

IG audits frequently find: no AI inventory (you did not know what AI was running on your contracts); incomplete contracts (AI governance requirements were not in the PWS or SOW); missing evidence (you could not produce contractor documentation); no verification (you accepted contractor assertions without checking); tier shopping (high-impact AI was classified as low-tier to avoid oversight); and incident gaps (you did not know about AI problems until the audit). This guide helps you avoid each of these findings.

PART V

Governance Structure

CHAPTER 8

Organizational Model

Effective AI governance requires clear roles, authority, and coordination across your organization.

Key Roles and Responsibilities

ROLE	AI GOVERNANCE RESPONSIBILITIES
Agency Chief AI Officer	Enterprise AI policy, OMB reporting, agency-wide inventory, coordination across programs
Program Manager	AI oversight for program contracts, tier classification, evidence review, verification activities
Contracting Officer	Include AI clauses in contracts, approve modifications, enforce contract AI requirements
COR	Day-to-day contractor oversight, evidence collection, verification, incident monitoring
IG Liaison	Audit preparation, documentation maintenance, IG coordination

Governance Authority Requirements

AI governance fails when the oversight function lacks authority or competence. Ensure contract leverage (AI governance must be in the contract, and the CO must be willing to include and enforce AI clauses); tier-determination authority (the Government, not the contractor, has final authority on tier classification); approval gates (Tier 1 AI cannot deploy without your explicit approval); and a clear escalation path to stop contractor AI deployment when governance is inadequate.

Competence Requirements

Program staff must understand enough AI to evaluate contractor claims, must know what AI governance requirements are in their contracts, and must have access to technical AI expertise for complex validation.

Coordination with Agency CAIO

Your program-level AI governance should coordinate with agency-level governance: report your AI inventory to the agency CAIO office, align tier classifications with agency standards, escalate novel or high-impact AI for agency review, and participate in agency AI governance working groups.

PART VI

Implementation Roadmap

CHAPTER 9

90-Day Implementation Plan

Phase 1: Foundation (Weeks 1 to 4)

Weeks 1 to 2, Inventory and Assessment: identify all contracts that may involve AI; request an AI inventory from existing contractors; review existing contracts for AI governance gaps; and assess your current oversight capabilities.

Weeks 3 to 4, Governance Structure: establish AI governance roles and responsibilities; develop tier classification criteria; draft standard AI governance clauses for contracts; and coordinate with the agency CAIO.

Phase 2: Contractor Oversight (Weeks 5 to 8)

Weeks 5 to 6, Evidence Collection: request evidence packages from contractors for Tier 1 and Tier 2 systems; validate contractor tier classifications; identify gaps in contractor documentation; and develop remediation requirements for non-compliant contractors.

Weeks 7 to 8, Verification Procedures: establish verification activities by tier; conduct initial verification for Tier 1 systems; and develop incident reporting and response procedures.

Phase 3: Operationalization (Weeks 9 to 12)

Weeks 9 to 10, Contract Integration: include AI governance clauses in new solicitations; modify existing contracts to add AI oversight where feasible; and train COs and CORs on AI governance requirements.

Weeks 11 to 12, Audit Readiness: compile IG-ready documentation for Tier 1 systems; conduct a mock IG review; remediate gaps identified; and establish an ongoing monitoring and reporting cadence.

CHAPTER 10

Maturity Progression

LEVEL	NAME	WHAT IT LOOKS LIKE
1	Ad Hoc	No systematic AI inventory or oversight; reactive to problems
2	Aware	AI inventory exists; contracts may lack governance clauses; limited verification
3	Managed	AI clauses in contracts; evidence packages required; verification activities conducted

LEVEL	NAME	WHAT IT LOOKS LIKE
4	Measured	Systematic oversight; performance metrics tracked; continuous improvement
5	Independently Verifiable	Evidence-based governance; effect-bearing AI actions are designed to produce tamper-evident authorization artifacts; oversight by record rather than trust

Level 5: Independently Verifiable Governance

At Level 5, the agency is designed to produce tamper-evident authorization artifacts for consequential AI actions, with evidence sufficient for independent replay. The maturity endpoint is not better monitoring. It is an independently reconstructable verdict: whether the action was authorized before it executed.

TECHNICAL FOUNDATION

The Five Tests Standard

The doctrine in this guide rests on a small, testable standard for what an authorization boundary must establish. The Five Tests Standard (5TS) states the requirements that an authorization artifact for an effect-bearing AI action should satisfy.

TEST	REQUIREMENT
Stop	The system can be halted before side effects occur.
Ownership	Each consequential decision maps to a named accountable authority.
Replay	The decision can be reproduced at the boundary.
Escalation	Control transfers at defined policy boundaries.
Provenance	The inputs grounding a verdict have an established origin. Origin, not truth.

The verdict space is ALLOW, DENY, and ABSTAIN. An ABSTAIN blocks execution pending authorized human override; escalation is the consequence of ABSTAIN, not a fourth verdict. Provenance is a test, not a verdict.

OPEN SPECIFICATION

The Five Tests Standard is published as an open, vendor-neutral specification.

Five Tests Standard (5TS) · Open standard DOI: 10.5281/zenodo.21040295 · Explanatory foundation DOI: 10.5281/zenodo.21048661

The standard is independent of any single implementation.

APPENDICES

Templates and Reference

APPENDIX A

AI Inventory Template

Use this template to inventory AI systems on your contracts.

FIELD	DESCRIPTION
Contract Number	Contract or task order reference
Contractor	Prime contractor name
AI System Name	System identifier
Description	What the AI does
Governance Track	Decisioning, Communications, or Both
High-Impact?	Yes or No, with rationale
Tier Classification	Tier 1, 2, 3, or 4, with rationale
Evidence Package Status	Received, Pending, or Gap identified
Last Verification Date	Date of last government verification
Next Review Due	Date of next scheduled review

APPENDIX B

Tier 1 (High-Impact) Approval Checklist

Complete before approving deployment of any Tier 1 AI system.

TIER 1 (HIGH-IMPACT): APPROVAL CHECKLIST

- AI System in contract inventory with tier classification documented.
- Automatic Tier 1 triggers reviewed and confirmed.
- Complete evidence package received from contractor.
- Training data documentation reviewed for representativeness.
- Validation results reviewed, including demographic subgroup analysis.
- Human oversight procedures verified as meaningful, not a checkbox.
- Incident history reviewed; no unresolved issues.
- Monitoring procedures in place; thresholds defined.
- Pre-execution authorization boundary configured for effect-bearing actions; action fails closed if the authorization artifact cannot be produced.
- Incident reporting path confirmed (contractor to COR to CO).
- IG-ready documentation compiled.
- Agency CAIO notified, if required by agency policy.

Approval Signature: _____ Date: _____

APPENDIX C

Clause and Regulatory Quick Reference

PWS and SOW Clause Quick Reference

CLAUSE	WHEN TO INCLUDE	KEY PROVISIONS
AI System Definition	Every contract that may involve AI	Broad definition including algorithmic and statistical systems; defines Governed Unit; notification requirement
AI Inventory and Classification	Every contract with AI	Inventory requirement; tier classification; government reclassification right
High-Impact AI Requirements	Contracts with Tier 1 AI	Evidence package; CO approval; 48-hour incident notification
Government Data and IP Rights	Every contract with AI	Government data rights; restriction on model training; ownership of models and outputs
Portability and Lock-In Prevention	Contracts with AI services	Transition support; data and model return in interoperable format
Audit and Independent Evaluation	Every contract with AI	Government and IG access; 5-day documentation response; independent evaluation; subcontractor access

CLAUSE	WHEN TO INCLUDE	KEY PROVISIONS
Flow-Down	Prime contracts with AI	Subcontractor AI compliance; prime responsibility
Unanticipated AI-Use Disclosure	Every contract with AI	Disclosure of AI introduced after award; classify before effect-bearing use
Generative AI Controls	Contracts with generative AI	Data input controls; human review; NARA-aligned records retention
AI Suspension Authority	Tier 1 and effect-bearing AI	Government stop authority; 24-hour suspension; fail-closed pending investigation

Regulatory Reference

AUTHORITY	APPLIES TO	WHAT IT REQUIRES
OMB M-25-21	Federal agencies (covered AI)	CAIO, AI strategy, public use-case inventory; minimum practices for high-impact AI (pre-deployment testing, impact assessment, monitoring, human oversight, fail-safes). Rescinds M-24-10.
OMB M-25-22	Federal AI acquisition	Performance-based acquisition; data and IP rights; vendor lock-in prevention; ongoing monitoring and independent evaluation; disclosure of unanticipated vendor AI use; compliance with M-25-21 minimum practices. Rescinds M-24-18.
OMB M-26-04	Procured LLMs and applicable generative AI	Unbiased AI Principles: truth-seeking and ideological neutrality; vendor documentation; evaluation evidence; end-user feedback mechanism; procurement-policy update and contract-materiality treatment where applicable.
EO 14179	Federal AI policy	Removing Barriers to American Leadership in AI; directed the AI Action Plan and review of prior AI policy actions. EO 14148 separately revoked EO 14110.
NIST AI RMF 1.0	Voluntary framework	Govern, Map, Measure, Manage; technology-neutral; under revision pursuant to America's AI Action Plan. NIST AI 600-1 is the Generative AI Profile.

For the public reference framework and related materials: ferz.ai. Verify the current status of OMB and NIST guidance before relying on any citation, as federal AI policy continues to evolve.