PCI GS
GOVERNMENT SERVICES

ARTICLE

Rule Ownership is Sovereignty in the Age of Artificial Intelligence

DECEMBER 18, 2025
BY: PAUL D. REMPFER

PCI-GS | A POARCH CREEK INDIAN COMPANY

PCI GS
GOVERNMENT SERVICES

EXECUTIVE SUMMARY

If you are deploying AI inside a federal mission, the hard problem is not picking a model. It is controlling behavior at scale and being able to prove that control under review. The model can come from a vendor, but the rules that shape what it will do, what it will not do, when it must escalate to a human, and what systems it can touch have to be government-owned, versioned, and auditable. The December 2025 executive order pushes toward a single national policy direction, but it does not answer the operational question of internal governance once these systems are live. This paper lays out a practical way to treat the rule layer as critical infrastructure, implement repeatable technical controls, and pair rule ownership with independent oversight so agencies can defend outcomes, maintain trust, and avoid governance drift as vendors and architectures change.

PCI GS
GOVERNMENT SERVICES

Introduction

Artificial intelligence (AI) is now in the operational bloodstream of the United States Government. Federal agencies are deploying large language models (LLMs) in workflows including intelligence analysis, operational planning, regulatory interpretation, and benefits adjudication. In each case, AI outputs now influence how decisions are made, not just how documents are written. That reality raises a question of sovereign authority: who is accountable for how AI systems behave when the stakes are high?

That urgency is being reinforced at the federal level. Recent Office of Management and Budget (OMB) memoranda (M-25-21 and M-25-22) push agencies to move faster on adoption, expand enabling infrastructure, and streamline compliance so AI can scale beyond experimentation. The White House also issued the December 11, 2025, executive order calling for a single national approach to AI policy, warning that a patchwork of state laws can distort AI outputs and slow innovation. The order also argues against requirements that would compel models to alter truthful outputs.

But speed has a side effect. Governance becomes harder to see, harder to coordinate, and harder to prove. In their September 2025 report, the Government Accountability Office (GAO) described a landscape with 94 AI-related requirements spread across laws, policies, and guidance, plus 10 different oversight entities. Requirements clearly exist, but they are distributed. There is no single unified structure that pulls them into one operational playbook for agencies deploying real systems.

Meanwhile, agencies are already building internal tools that look a lot like "mission-ready GPTs." GovCIO's reporting describes deployments such as CISA's internal chatbot work, the Air Force Research Laboratory's NIPR GPT, U.S. Central Command's CENTGPT, and the Army's CamoGPT, along with secure deployments supporting research and admin workflows at the National Institutes of Health.

That creates a problem the federal government has not fully confronted. AI is entering operations faster than agencies are standardizing how these systems are governed, constrained, and independently reviewed. When rules are unclear or outsourced, authority doesn't disappear. It moves.

PCI GS
GOVERNMENT SERVICES

The Core Claim: The Rule Layer Is the Decision Layer

Treating the rule layer as a first-class governance asset is not an architectural preference; it's the only way agencies can reliably explain decisions, reproduce outcomes, and demonstrate that AI systems are operating in line with law, policy, and mission requirements.

Most AI debates still focus on the model: Which foundation model is used? Which cloud? Which benchmarks? That framing misses where authority actually sits once a system is deployed.

Modern AI systems don't act based on model weights alone. Their behavior is governed by an operational "rule layer" that sits above the model and shapes every interaction. System instructions define role and purpose. Refusal logic determines what the system will not do. Escalation thresholds decide when human judgment is required.

This is where interpretation happens. The rule layer decides how ambiguity is handled, what risks are tolerated, and when the system must stop and hand off to a human. Two agencies can use the same underlying model and still get very different outcomes based solely on differences in these rules. In practice, the rule layer defines the AI system's operational identity.

This distinction matters because much of this logic is often embedded in vendor-managed alignment systems or implemented informally through prompts that are not versioned, audited, or mapped to legal authority. When that happens, agencies may technically "use AI" without actually controlling how it reasons, refuses, escalates, or behaves over time.

That sets up the next section. If the rule layer is where behavior is decided, then rule ownership is where federal AI governance either holds or breaks.

Why Rule Ownership Matters for Federal AI Governance

Rule ownership is how agencies keep sovereign authority over outcomes. It is the difference between "we used AI" and "we can prove what governed it" across civil rights, national security, interoperability, and public trust.

LLMs may come from vendors, but the decisions that matter most are shaped by the rules layered on top of those models. The sections below examine the four domains where rule ownership is most consequential for federal AI governance.

Civil rights and constitutional accountability

Federal agencies, including the Department of Veterans Affairs, the Social Security Administration, and the Department of Education, are increasingly using AI-supported tools in benefits adjudication, eligibility screening, case management, and administrative review. In these contexts, the question is not whether AI is being used, but whether agencies can demonstrate that its use complies with due process, nondiscrimination requirements, and long-standing civil rights protections.

Consider a Veterans Affairs (VA) office using an AI system to draft disability claim assessments by reviewing medical records, associating symptoms with rating criteria, and generating structured summaries for adjudicators. If the logic governing that system resides inside a vendor's proprietary alignment layer, the VA cannot reliably show that the system applies evidentiary standards consistently or treats similarly situated claimants equally under Title VI. It cannot reproduce the reasoning path behind a disputed decision. It cannot verify that certain categories of evidence were not discounted because of training data biases or consumer-oriented safety heuristics.

Similar risks apply to other benefit-related decisions. An AI-assisted review of student loan forgiveness applications may rely on vendor-derived heuristics to assess borrower hardship, undercounting certain forms of economic distress or applying discounting logic inconsistent with federal policy. A model supporting Social Security disability determinations may filter out specific categories of evidence because of generic safety constraints rather than statutory requirements. In each case, the agency may be unable to certify lawful operation or defend outcomes when challenged.

National security and mission integrity

AI is embedded in national security workflows across the Department of Defense (DoD) and the intelligence community. Combatant commands and military services are deploying internal LLMs to assist with planning, analysis, logistics, cyber defense, and operational coordination. Predictability is not optional. Military and intelligence operations depend on doctrine-aligned reasoning, strict classification handling, and consistent behavior across time.

Imagine a multi-domain intelligence fusion cell supporting a unified combatant command. Analysts rely on an internal LLM to synthesize signals intelligence, summarize HUMINT reporting, interpret open-source data, and flag emerging patterns. If the system's refusal logic or safety heuristics are governed by vendor-controlled alignment layers, legitimate classified queries may be declined or softened because they resemble restricted content in consumer environments. Even subtle shifts in how uncertainty is handled or how threats are framed can influence analytic confidence and strategic warning timelines during crisis response.

The same pattern appears in operational planning and cyber defense. A model assisting with logistics or force projection analysis may avoid discussing casualty estimates or risk tradeoffs because of generalized safety constraints rather than military doctrine. In defensive cyber operations, generative or agentic AI used for triage or response recommendations may refuse to analyze exploit chains or detection evasion techniques even when the context is defensive, because vendor safety layers alter reasoning to avoid creating potential harmful content. These changes often occur through routine model updates, outside agency control, and without mission-specific validation.

Interoperability and vendor independence

Federal AI systems cannot be designed as one-off deployments tied to a single model provider or platform. Federal systems are built to outlast vendors, contract cycles, and technology fads. AI systems must meet that same standard. As LLMs evolve rapidly and vendors update capabilities on short cycles, federal missions require behavioral consistency even as underlying models change. Without that consistency, agencies face operational risk every time a model is upgraded, replaced, or temporarily unavailable.

Suppose the Department of Homeland Security (DHS) uses a specific vendor model for document review, identity screening, or risk assessment at ports of entry. If the rule logic that governs how documents are interpreted, anomalies flagged, or follow-up actions triggered is embedded inside a specific vendor's alignment stack, the agency becomes dependent on that vendor's internal updates. A forced transition to a new model, whether due to cost, performance, supply-chain risk, or a service outage, would introduce unpredictable behavioral changes. Officers on the ground would see the same inputs produce different outputs without warning, and the agency would struggle to explain why.

Similar challenges apply in highly regulated operational domains such as aviation. If the Federal Aviation Administration (FAA) relies on an AI system to help with air traffic pattern prediction or safety risk forecasting, even subtle shifts in model behavior can have outsized consequences. Vendor-controlled rule layers that evolve independently of federal validation cycles undermine the predictability these systems require. Operational systems cannot afford logic drift driven by external release schedules.

Public trust and institutional legitimacy

Public trust in government depends on the ability to explain how decisions are made, defend the process used to reach them, and correct errors when they occur. As federal agencies integrate AI into administrative, regulatory, and enforcement workflows, that expectation does not change. If anything, it becomes more important. When automated or AI-assisted systems influence outcomes that affect citizens' rights, finances, or legal standing, agencies must be able to show that those systems operate according to law and established policy, not opaque or shifting external logic.

"Public skepticism rises quickly when agencies cannot explain how automated decisions were reached. Coverage of AI-generated benefits determinations and automated enforcement tools shows that trust erodes quickly when affected individuals believe decisions are driven by black-box systems rather than accountable public processes.

If the Internal Revenue Service (IRS), for example, uses a generative model to help draft narratives explaining tax discrepancies, and the model's reasoning that's embedded in those explanations is shaped by proprietary vendor alignment policies, the IRS cannot reliably demonstrate that the narrative reflects the Internal Revenue Code rather than generic language optimization or risk-avoidance heuristics. It also cannot reproduce the model's reasoning when a taxpayer appeals the assessment.

Similarly, if the Department of Justice (DOJ) uses AI-assisted tools to organize investigative materials or identify relevant patterns across large document sets, courts and defense counsel will expect clarity on what reasoning the tool applied. Vendor-controlled rule layers can't meet the evidentiary standards required in court.

What Rule Sovereignty Looks Like in Practice

Rule sovereignty is not a concept. It is a set of controls that let a CIO keep behavior stable at scale and prove, after the fact, what governed the system.

For a CIO, it really comes down to two things: can we keep these systems behaving the way we intend, and can we prove it when someone asks? Rule sovereignty helps, but only if it shows up as real, repeatable controls. The next sections explain what those controls look like in practice.

Government-owned rule control planes

A rule control plane is the government's rulebook for how an AI system is allowed to behave. It isn't a folder of prompts; it's a managed system that stores the rules that shape outputs in real workflows. It defines what the model can do, what it must refuse, when it has to hand off to a human, and what tools it is allowed to use. Each set of rules should be treated like a controlled configuration and include metadata, version history, authorship records, validation states, and digital signatures. If a rule changes, you should be able to see who changed it, when, and why. You should also be able to tie that rule back to policies, authorities, statutes, and mission directives.

The rule control plane also must respect mission boundaries. In other words, DoD rules cannot be visible to civilian agencies (i.e., namespace isolation). Highly classified rule bundles should be stored in separate enclaves with tight access

controls. The system must support granular access control, MFA, encryption, and zero-trust principles.

Deterministic policy compilation and execution

In practice, agencies need a "policy compiler." A user asks a question, and the compiler assembles the exact set of rules that should apply before the model ever responds. It pulls the right rule bundle for that mission or workflow, adds the required system instructions, applies safety constraints and escalation triggers, and sets what tools the model is allowed to call, if any. It also resolves conflicts: if a user request clashes with an agency rule, the agency rule wins. Every time.

This is also how you keep vendors in the right place. Vendors can provide models and a hosting environment, but the government controls the rule stack, the order of authority, and the handoff points where a human must review and approve.

Observability, logging, and reproducibility

Rule sovereignty fails without evidence. That is what observability is for; at minimum, it logs the applied rule bundle ID, key constraints, escalation thresholds, any tools the system was permitted to call, and what it actually did. It also needs to show the boundaries, like what data sources were allowed, what was blocked, and what was redacted or minimized. When an inspector general, FOIA team, oversight body, or program lead asks "why did the system do that," observability is how you answer with facts instead of guesswork.

Reproducibility is the next step. It means an authorized reviewer can replay the interaction later using the same rule bundle and see the same governed workflow, including which rules were in effect at the time.

Governing Agentic AI Systems

Agentic AI without rule sovereignty is not automation; it is delegated authority without accountability. Government-owned rules are what keep agents useful without turning them into mission liabilities.

An AI agent is not just a chat window that generates text. It is a system that can pull files, call APIs, update records, trigger workflows, and take actions that affect real operations. Governance for agentic AI is therefore not about output quality alone. It's about defining hard boundaries around authority, judgment, and accountability. This is where rule sovereignty stops being conceptual and becomes operational.

Alignment logic

Before an agent ever takes an action, it reasons. That reasoning is shaped by alignment logic that determines how the system balances accuracy, safety, helpfulness, and risk. In consumer systems, alignment often prioritizes harm

reduction or politeness. In federal missions, alignment must reflect statutory mandates, regulatory requirements, civil rights protections, classification rules, and mission doctrine. When alignment logic is vendor-defined, agencies lose control over how ambiguity is resolved and how law is interpreted at the system level. Government-owned alignment logic ensures that agents reason according to federal authority, not private philosophies.

Classifier pipelines

Agents do not reason in a vacuum. Classifiers sit around the model and inspect inputs and outputs for risk, sensitivity, or restricted content. These classifiers shape what the agent is allowed to consider before it reasons and what it is allowed to produce afterward. In consumer environments, classifiers are tuned to block harmful content. In federal missions, the same classifiers can unintentionally block legally required analysis or suppress valid operational context. Agencies must be able to see, tune, version, and audit classifier logic so that it enforces mission rules rather than distorting them.

Tool registry and permissions

Once reasoning is governed, action must be constrained. Agencies need an authoritative list of tools an agent is allowed to call. Each tool requires a clear description, a typed schema, parameter limits, and defined authorization levels. The agent should never be allowed to invent tools, change tool definitions, or improvise new access paths. All validation must happen at a government-controlled tool gateway so every action is intentional, authorized, and logged.

Escalation logic

Some decisions should never be automated end-to-end. Escalation thresholds belong in the rule layer, not in informal prompts or after-the-fact review. An agent supporting case management may require human approval before issuing any adverse determination. An agent assisting cyber operations may require multi-person authorization before triggering a system-level response. These handoffs are not failures of automation. They are the points where accountability is preserved.

Routing logic and multi-model orchestration

Federal agentic systems often rely on more than one model. Routing logic determines which model is used for which task and how outputs are combined. If that logic is vendor-controlled or opaque, agencies lose visibility into which systems influenced a decision. Government-owned routing rules ensure that model selection aligns with mission requirements and remains consistent across upgrades or provider changes.

Reproducibility and drift control

Every agentic action must be replayable. Logs must capture reasoning steps, classifier decisions, tool calls, state transitions, and the rule bundle in effect at the time. This is the only way to detect rule drift, the gradual behavior change that occurs through model updates, classifier tuning, or unreviewed rule edits. Drift is rarely obvious in real time, but it can quietly undermine classification handling, legal interpretation, and escalation safeguards. Version control and independent review are the only reliable defenses.

Independent Oversight for Federal AI Systems

Rule ownership only holds if someone outside the agency can inspect it, challenge it, and report on it without fear or favor. Independence is what turns governance into accountability.

Owning the rules that govern AI behavior does not, by itself, guarantee accountability. Federal agencies operate under budget pressure, political scrutiny, and operational urgency, all of which can quietly shape how rules are written, relaxed, or enforced over time. Independent oversight ensures rule sovereignty remains durable, reviewable, and aligned with the public interest as AI systems scale and evolve.

Why agency self-governance is not enough

Federal programs operate under constant pressure to move faster, reduce backlogs, and demonstrate near-term results. In that environment, agency self-governance and rule logic can drift. Escalation thresholds are weakened, logging requirements are narrowed, and transparency is reduced in the name of efficiency.

These risks don't require any bad intent; they arise naturally from workload pressure, political incentives, and vendor dependencies. Over time, small changes to rule bundles or classifier thresholds can materially alter system behavior without clear visibility or external challenge. When agencies are both the rule authors and the rule judges, there is no structural safeguard against this kind of drift.

That is why rule ownership must be paired with independent oversight that combines technical expertise, legal authority, organizational independence, and a mandate to protect democratic legitimacy.

A GAO-centered oversight model

The GAO is uniquely suited to federal AI rule oversight. It's structurally independent of the Executive Branch and reports directly to Congress. It already audits federal programs, investigates technology failures, and evaluates compliance across agencies. It's insulated from political turnover and cannot be directed by agencies or political appointees to alter findings. Importantly, GAO

already maintains expertise in data science, cybersecurity, and federal technology programs.

Under this framework, GAO would establish a new, statutorily mandated **AI Rules and Systems Audit Offic**e. This office should be staffed with:

- Prompt engineers who understand the nuances of generative and agentic behavior.
- Alignment researchers who can identify reasoning distortions.
- Safety scientists trained in adversarial testing.
- Civil rights lawyers familiar with equal protection frameworks.
- Intelligence professionals who understand mission integrity.
- Computer scientists fluent in tool orchestration and agentic architectures.

It would include divisions for rule analysis, agentic behavior evaluation, adversarial red teaming, civil rights and legal compliance, public transparency, and internal integrity to examine every rule bundle across the federal government. This would mean operating sandbox environments to reproduce real-world conditions so that auditors can test rule changes before they take effect. The office would then oversee both unclassified and classified systems, produce public reports and classified reports to Congress, and be protected from political and vendor influence.

Investigative authority and continuous review

Oversight only works if the oversight body is protected and empowered. An auditor that can be defunded, redirected, or blocked cannot govern federal AI in practice. GAO's independence is the point here. It cannot be defunded by agencies or commanded by the Executive Branch. It publishes findings directly to Congress and the public.

However, in this case, that independence needs to come with real investigative access. GAO would need the authority to review rule logic directly, including access to:

- Rule bundles and historical versions
- Compiled instruction sets used in production
- Logs showing how rules were applied at inference time
- Vendor documentation describing embedded safety layers
- Replay environments capable of reproducing system behavior

This oversight cannot be periodic. AI systems evolve continuously, and rule changes occur far more frequently than traditional audit cycles. Agencies would be required to submit material rule updates, classifier changes, tool permission

modifications, and escalation logic adjustments for review in near real time. GAO would be able to evaluate changes before they materially affect operational systems. GAO must also maintain a secure reporting channel for whistleblowers who identify rule misuse or unsafe agency behavior.

Legal and Policy Foundations Supporting Rule Ownership

Rule sovereignty is not a policy preference. It is what due process, civil rights law, records requirements, and classification rules already assume: the government must be able to explain, reproduce, and defend the logic behind its decisions.

Rule sovereignty is grounded in law. Several federal statutes and doctrines already demand that agencies control the rule logic used in their systems.

Administrative law and due process requirements

The Administrative Procedure Act requires transparency and reasoned decision-making. An agency cannot justify an action if it cannot explain how rule logic shaped the outcome. The Due Process Clause requires that individuals receive explanations for decisions affecting their liberty or property. If rule logic resides inside proprietary systems that cannot be inspected or replayed, agencies cannot meet this obligation.

Civil rights, records management, and transparency obligations

Federal civil rights statutes require demonstrable nondiscrimination. Records management laws require preservation of decision logic. The Freedom of Information Act requires disclosure of materials that influence decisions. Opaque rule layers make compliance with these requirements difficult or impossible.

Classification and national security constraints

Records management laws require agencies to preserve complete documentation of decisions. Vendor controlled rule systems obscure the decision pathway. National security and classification laws require strict adherence to information handling protocols. Vendor safety layers cannot be trusted to enforce these boundaries.

The law already aligns with rule sovereignty. The technology must now follow suit.

PCI GS
GOVERNMENT SERVICES

Operational Next Steps for Federal CIOs and AI Leaders

Rule sovereignty only works when it becomes infrastructure, not guidance. If agencies cannot version, enforce, and reproduce their rule logic under audit, they do not truly control AI behavior in federal missions.

Implementing rule sovereignty is not a theoretical exercise. It requires agencies to make uniform choices about how rules are built, owned, tested, and enforced across systems that will be used under real operational pressure. The goal is not perfection on day one. The goal is to ensure that AI behavior is governable, reviewable, and defensible as systems scale.

Establishing rule ownership as infrastructure

Rule governance should be treated as shared federal infrastructure, not as an application feature buried inside individual programs. Agencies need centralized rule repositories with clear ownership, version control, and certification workflows. Rule bundles should be treated like controlled configurations, not informal prompt files.

At scale, this points toward shared federal services. A Federal Rule Registry should store certified rule bundles and their full lineage. A Federal Policy Compiler Service should assemble system messages, alignment rules, classifier settings, and tool permissions into authoritative instruction sequences.

Shared evaluation suites can test for civil rights compliance, classification handling, adversarial resilience, and mission-specific stress scenarios. Where workflows span agencies, rule changes must be coordinated, not improvised.

Integrating governance into procurement and deployment

Procurement is where many governance failures are locked in. AI contracts should require a clear separation between models and rules, with agencies retaining ownership of the rule logic that governs behavior. Contracts should explicitly allow those rules to be applied across vendors, models, and hosting environments.

Governance also has to be enforced at deployment, not retrofitted later. If escalation logic, logging, and replay are not operational on day one, they rarely get added cleanly later. Deployment gates should verify that rule bundles are versioned, mapped to authority, and observable before systems touch real data or real decisions.

Preparing for audits before they are required

Agencies should assume that AI systems will be reviewed by Inspectors General, Congress, courts, and independent oversight bodies. That means systems must be designed to support replay, explanation, and independent review from the outset.

Preparing early reduces both risk and cost. Systems that can reproduce behavior, show which rules were in force, and explain how outputs were generated are easier to defend, easier to correct, and far less likely to become liabilities during investigation or litigation.

As AI becomes more capable and more deeply embedded in mission systems, the rule layer becomes one of the most important governing artifacts in federal operations. Those who write the rules shape how AI behaves. Those who audit the rules protect the legitimacy of government action. Those who control the rule architecture shape the future of federal AI.

Conclusion

The models may be private, but the rules must belong to the public. In the AI era, sovereignty is exercised through who writes, controls, and audits the rules that govern behavior.

Artificial intelligence presents a historic opportunity and a historic risk. The United States can adopt powerful systems without surrendering lawful authority, but only if the government treats the rule layer as a governing asset. Model providers can supply neural engines and infrastructure. That is not where legitimacy lives. Legitimacy lives in the operational rules that shape what the system will do, what it must refuse, when it must escalate, and how it can be audited.

This is the hinge point for federal AI governance. If agencies cannot write, control, version, and audit the rules that govern behavior, they will be left defending outcomes they cannot reproduce and relying on assurances they cannot verify. Rule sovereignty is how the government keeps control of mission execution, protects civil rights, sustains national security discipline, preserves interoperability across vendors, and maintains public trust.

The technology will evolve. Capabilities will expand. The only stable anchor is governance that can be inspected, replayed, and defended. In the digital century, sovereignty resides in the rules that govern AI. Those rules must be written by the government, enforced by the government, and overseen by an independent body accountable to the public interest.

PCI GS
GOVERNMENT SERVICES

Sources

1. U.S. Government Accountability Office. Artificial Intelligence: Federal Agencies Need Better Planning for Oversight and Implementation. GAO-25-107933. Published September 9, 2025. https://www.gao.gov/assets/gao-25-107933.pdf

2. Kennan A, Singh L, Garcia Guevara A, Ahmed M, Goodman J. AI-Powered Rules as Code: Experiments With Public Benefits Policy. Digital Government Hub. Published March 24, 2025. Updated September 4, 2025. https://digitalgovernmenthub.org/publications/ai-powered-rules-as-code-experiments-with-public-benefits-policy/

3. Oakland S, Gianfortune R. How Federal Agencies Are Building Secure, Mission-Ready GPTs. GovCIO Media & Research. Published October 9, 2025. https://govciomedia.com/how-federal-agencies-are-building-secure-mission-ready-gpts/

4. Office of Management and Budget (OMB). Executive Office of the President. Accelerating Federal Use of AI through Innovation, Governance, and Public Trust. Memorandum M-25-21. Published 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. Office of Management and Budget (OMB). Executive Office of the President. Driving Efficient Acquisition of Artificial Intelligence in Government. Memorandum M-25-22. Published April 3, 2025. https://www.whitehouse.gov/wp-content/uploads/2025/04/M-25-22-Driving-Efficient-Acquisition-of-Artificial-Intelligence-in-Government.pdf

6. The White House. Eliminating State Law Obstruction of a National Artificial Intelligence Policy. Washington, DC; December 11, 2025. https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/