ATARC

DevSecOps Working Group

powered by GovExec

MARCH 2026

The New Logic of Innovation

Scaling Federal DevSecOps with Agentic AI

TABLE OF CONTENTS

Disclaimer: This document was prepared by members of the ATARC DevSecOps Working Group in their personal capacities. The views and opinions expressed herein are those of the authors and do not necessarily reflect the official policy or position of any individual organization, employer, agency, or affiliated entity. This document is released for public use and distribution. It may be shared, cited, and reproduced without restriction, provided it is not used for commercial advertising, marketing, or product endorsement purposes. Nothing in this document should be construed as an endorsement of any specific technology, vendor, product, or service.

ACKNOWLEDGMENTS

ATARC would like to take this opportunity to recognize the following DevSecOps Working Group members for their contributions:

Spence Spencer, U.S. Patent and Trademark Office, ATARC DevSecOps Working Group Government Co-Chair

Graham Baggett, U.S. Census Bureau, ATARC DevSecOps Working Group Government Co-Chair

Steven Terhar, GitLab, ATARC DevSecOps Working Group Industry Co-Chair

Tom Tapley, Sonatype, ATARC DevSecOps Working Group Industry Co-Chair

Hasan Yasar, Software Engineering Institute, Carnegie Mellon University, ATARC DevSecOps Working Group Academic Chair

Alyssa Feola, Seventeen Sierra, LLC

Elena Peterson, Pacific Northwest National Laboratory

Anna Libkhen, U.S. Bureau of Economic Analysis

Joe Boye, Palo Alto Networks

Sameet Nasnodkar, Social Security Administration

Hashim Khan, Zimperium

Sean Applegate, Swish

OVERVIEW

In 2026, the federal government stands at a pivotal inflection point, transitioning from a decade of manual software authorship to a future where DevSecOps gets scaled. This paradigm shift, moving the human from a “Creator” to “Curator”, is not merely a technical upgrade but a fundamental re-imagining of the federal workforce. By evolving through a maturity model that spans from **Ad-Hoc** research to **Assisted** co-design and finally to **Agentic** orchestration, agencies are empowering their personnel to act as strategic stewards. In this high-maturity state, federal professionals no longer write every line of code; instead, they manage digital squads of autonomous agents, ensuring that machine-speed delivery remains anchored to human judgment, mission intent, and the rigorous security standards of the American public.

This transformation is operationalized through the software lifecycle where intelligence is shifted as far left as possible. Autonomous agents handle the heavy lifting of legacy refactoring where they translate millions of lines of outdated code into modern, cloud-native microservices while getting integrated into the shielded Continuous Assurance environment, also monitored by autonomous agents. By leveraging breakthroughs like those demonstrated in the DARPA AI Cyber Challenge (AIxCC), agencies can now identify and patch vulnerabilities across millions of lines of code at machine speed. This creates a self-healing operational environment where network and security operation centers fuse and autonomous remediation allows federal teams to neutralize threats and restore mission-critical systems proactively, shifting the workforce from reactive troubleshooting to high-level strategic oversight.

Securing this agentic frontier requires a new defensive architecture grounded in Zero-Trust for machine autonomy. By treating agents as Non-Person Entities (NPEs) and implementing Zones of Intent Scaffolding, agencies ensure that autonomous reasoning is physically bounded within designated walled gardens, preventing lateral movement and goal drift. This technical rigor is matched by a robust 2026 regulatory framework, led by OMB M-25-21 and the NIST Cyber AI Profile (IR 8596). These mandates transform compliance into a strategic accelerator, providing the intelligent guardrails necessary to scale innovation safely. Ultimately, this roadmap elevates the human role, allowing the federal workforce to orchestrate a more resilient and responsive government with unparalleled precision and security.

THE PARADIGM SHIFT: FROM CREATOR TO CURATOR

The fundamental nature of software development in the federal government is undergoing a transformation as we move from a world where humans are the primary authors of every line of code to one where they act as strategic orchestrators of autonomous systems. This progression loosely follows the **Agentic Software Engineering (SE 3.0) Maturity Model[1]:**

- **Human-Centric (The Baseline): Developer as the Creator.** In this foundational state, every line of code is human-authored, and the system's behavior remains entirely deterministic. Even with the use of basic "tab-to-complete"[2] coding or standard IDE autocompletion, the logic and structural intent remain solely with the individual developer.

- **Ad-Hoc / Consultative (The Gateway): Individual as the Researcher.** In this transitional state, AI exists primarily as an external chatbot used for non-integrated tasks, such as researching requirements or crafting edge cases. Even when used for development, the AI acts as an isolated partner and the developers must perform the manual labor of transferring code snippets and data between the development environment and an external application.[3]

- **Assisted ("Human-in-the-loop"): Developer as the Pilot and Co-Designer.** This is the ideal state for the Federal Government, where the AI tools are integrated directly into the IDE. The technical tools are used to accelerate task-centric goals such as unit test generation. As this maturity increases, AI begins to function as a co-designer that understands broader architectural contexts and suggests refactorings based on established boundaries.[4]

    - **Human-in-the-loop** can be defined as a governance model in which an AI system cannot execute consequential actions without direct human review and approval. The human is an active participant inside the decision cycle, validating outputs before they are finalized or deployed.

[1] https://arxiv.org/html/2509.06216v1
[2] https://www.jetbrains.com/help/ai-assistant/code-completion.html
[3] https://levelup.gitconnected.com/how-to-use-chatgpt-as-your-first-pair-programmer-948531246349
[4] https://www.usda.gov/sites/default/files/documents/fy-2025-2026-usda-ai-strategy.pdf

- **Agentic ("Human-on-the-loop"): Developer as the Curator.** In this advanced operational state, AI agents autonomously plan and execute complex, multi-step workflows. The human professional acts as a high-level supervisor, defining strategic mission goals, managing **"Agentic Personas"**[5] as non-person entities (NPEs),[6] and reviewing architectural outcomes to ensure they align with federal security posture.
    - **Human-on-the-loop** can be defined as a governance model in which an AI system operates autonomously within predefined guardrails while a human supervises overall behavior. The human monitors performance, sets mission intent, and retains the authority to intervene, override, or shut down the system if necessary.

[5] https://csrc.nist.gov/pubs/other/2026/02/05/accelerating-the-adoption-of-software-and-ai-agent/ipd
[6] https://digital.gov/resources/m-19-17-enabling-mission-delivery-through-improved-identity-credential-and-access-management

FEDERAL IMPLEMENTATION ROADMAP

The Federal Implementation Roadmap serves as a strategic blueprint designed to transition agencies from traditional, linear IT deployment to a smooth, transitional model. This roadmap balances the national mandate for rapid innovation with the stringent requirements of federal security and public trust.

Phase 1 Human Foundation

Phase 1 represents the "cognitive re-calibration" of the federal workforce. It ensures that as technology evolves from standalone assistants to autonomous actors, the human's ability to oversee and validate that technology grows in lockstep.[7] Once agencies establish a baseline of trust and literacy,[8] they can shift their focus from manual task execution to strategic stewardship, turning AI into a foundational capability for secure, resilient software delivery.

1. Exposure: From Observation to Orchestration

Exposure is the journey from anecdotal experimentation to the mastery of complex, multi-agent logic. In the Ad-Hoc (Researcher) stage, staff use chatbots for low-stakes prose like feature development and summaries in an individual, anecdotal manner. As teams transition to the Assisted (Pilot) stage, they interact with AI inside technical tools like IDEs and Project Management Tools, turning exposure into a visible, daily collaborative experience. Finally, in the Agentic (Curator) stage, staff oversee multi-agent systems, shifting their focus to observing agent-to-agent handoffs and verifying autonomous logic chains.

2. Literacy: From Fact-Checking to Goal-Setting

Literacy evolves from verifying simple outputs to defining the mission intent that governs autonomous behavior.

- **Ad-Hoc (Generative Literacy):** Establishing a baseline understanding that LLMs are probabilistic, requiring users to identify "hallucinations"[9] in research or prose. The NIST Generative AI Profile (NIST AI 600-1) defines "Confabulation" (colloquially known as hallucinations) as the production of confidently stated but erroneous or false content, which is a primary risk of probabilistic models.

[7] https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf
[8] https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf
[9] https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

- **Assisted (Spec-Driven Literacy):** Maximizing the development velocity by learning to define strict API contracts that force AI to be useful within a codebase. Mastering this stage involves the use of a "Constitution"[10] —a non-negotiable artifact that encodes project-specific standards and security principles that the AI agent must respect during the code generation process.
- **Agentic (Agentic Literacy):** Agentic Literacy represents the transition from manual task oversight to Goal Specification Mastery. As defined in the Model AI Governance Framework (MGF) for Agentic AI,[11] this involves defining a mission so precisely that their loop remains bound by the human's original intent, ensuring that the system has the appropriate permissions and controls without deviating from its defined intent.

[10] https://arxiv.org/html/2602.02584v1
[11] https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf

3. Skill: The Evolution of New Federal Roles

The move toward agentic systems creates a demand for specialized roles that combine technical oversight with mission strategy.

- **Ad-Hoc Stage (The Researcher):** This role focuses on **Generative Discernment**. Federal staff must act as "Critical Reviewers," capable of identifying probabilistic errors and bias in non-integrated AI outputs. **As defined in OPM's Merit-Based Hiring Plan,**[12] this role requires data fluency to verify AI-generated research against authoritative federal repositories, ensuring that anecdotal experimentation does not compromise mission accuracy.

- **Assisted Stage (The Context Engineer):** Context Engineers provide AI agents with architectural "fences" and legacy constraints. AI-generated code is grounded in the specific requirements of the federal mission. This role involves "Mapping" (NIST AI RMF[13]) project-specific facts and security boundaries into the AI's reasoning loop, transforming the AI from a general assistant into a project-aware co-designer.

- **Agentic Stage (The Flow Architect):** Curators with Flow Architecture skills design the **"human-on-the-loop"** checkpoints where autonomous chains must pause for approval. **According to NIST's Agent Identity Framework**[14] this role manages **Agentic Personas** as Non-Person Entities (NPEs). The Flow Architect ensures that autonomous planning remains bounded by "Meaningful Human Accountability," defining the strategic "mission goals" that an agent cannot drift from during multi-step execution.

4. Culture: From Replacement Fear to Curation Pride

Cultural success is found in redefining personal value from effort spent to impact achieved.

- **Ad-Hoc (Curiosity Culture):** Moving from "will this replace me?" to "how can this handle my repetitive paperwork?". The **LSE 2026 Workplace Adaptation Study**[15] notes that initial resistance to AI is often rooted in threats to professional identity. By framing AI as a tool for "micro-efficiencies"—such as handling repetitive administrative tasks—agencies can foster a culture of curiosity where employees view AI as a "connective tissue" that removes friction rather than a replacement for human judgment.

[12] https://www.opm.gov/policy-data-oversight/hiring-information/merit-hiring-plan-resources/
[13] https://www.nist.gov/itl/ai-risk-management-framework
[14] https://www.nccoe.nist.gov/sites/default/files/2026-02/accelerating-the-adoption-of-software-and-ai-agent-identity-and-authorization-concept-paper.pdf
[15] https://www.lse.ac.uk/research/research-for-the-world/economics/remote-jobs-are-transforming-work

- **Assisted (Verification Culture):** Shifting team pride from "I wrote this code" to "I verified this spec," ensuring no code is accepted without validation. According to the **Insight Partners 2026 Engineering Outlook,**[16] senior engineers are increasingly shifting away from writing syntax and toward **"Value Architecture."** In this culture, success is measured by the quality of the "high-conviction decisions" made during the review process, moving the team's pride from the act of production to the act of rigorous verification and alignment.

- **Agentic (Curation Culture):** Redefining success as **Orchestration**—successfully leading a squad of agents to achieve a secure and compliant mission outcome. The **World Economic Forum (2026)**[17] identifies software developers as the "vanguard" of this shift, where their value is found in **"AI-enabled decision-making"** and architecture. In a curation culture, the federal professional acts as a "strategic decision-maker," where the ultimate "win" is the successful orchestration of autonomous agents to deliver complex, resilient mission services at machine speed.

5. Training: From "How-To" to "How-to-Audit"

Training moves beyond teaching employees how to use a tool and starts teaching them how to govern a digital worker. This involves a pedagogical shift from manual execution to rigorous oversight of autonomous logic chains.

- **Ad-Hoc Stage (Acceptable Use):** Initial training focuses on establishing a baseline of "Generative Literacy" and identifying the risks of **"Shadow AI."** The **OPM 2026 AI Training Initiative**[18] mandates SCORM-compliant modules that teach staff to recognize **"Confabulation"** and strictly enforce data-sharing rules. This phase ensures personnel understand what data is permissible for AI ingestion and how to identify errors in research outputs before they are integrated into federal workflows.

- **Assisted Stage (Triaging):** Following the **NIST Cyber AI Profile**[19] developers are trained to use AI to enhance cyber defense at "machine speed." This stage involves learning to treat AI-generated code as a first-class asset that requires automated validation against **Software Bill of Materials (SBOMs),**[20] ensuring no code enters the pipeline without a structured specification check.

[16] https://www.insightpartners.com/ideas/ai-adoption-2026/
[17] https://www.weforum.org/stories/2026/01/software-developers-ai-work/
[18] https://www.opm.gov/ai/2026-ai-training/

- **Agentic Stage (Oversight):** Ultimately, training focuses on **Strategic Oversight**, enabling staff to audit **"Chain of Thought" (CoT)** logs and manage agents as **Non-Person Entities (NPEs)**. Under the **NIST Concept Paper on AI Agent Identity,**[21] federal personnel are trained to maintain CoT audit logs for at least 7 years. Training centers on mastering the lifecycle of NPE identities—ensuring they are "distinguishable, auditable, and traceable"—to ensure mission intent remains bounded by federal security posture.

Phase 2 Infusing AI Across the SWE Lifecycle

Phase 2 transforms the "engine room" of federal IT. It moves beyond individual productivity to create integrated, autonomous workflows that handle the complexity of multi-step tasks and legacy refactoring. By embedding AI across the lifecycle, agencies move toward a self-healing and secure-by-default state.

1. Plan/Code/Build (The Shift-Left)

The infusion of AI into the earliest stages of the lifecycle enables agencies to resolve architectural flaws before a single line of code is committed.

- **Ad-Hoc (Consultative):** Staff use external LLMs to explain legacy code snippets or draft high-level requirements. In this stage, AI lacks "codebase awareness," acting as a general-purpose consultant that requires developers to manually bridge the gap between AI advice and the live repository.
- **Assisted (The Pilot):** AI is integrated into the IDE and CI/CD pipelines to institutionalize **Spec-Driven Development (SDD).**[22] It compares architectural options against concrete targets for cost, performance, and scalability. Human architects still decide, but they do so using AI-generated trade-off analysis.
- **Agentic (The Curator):** Autonomous agents perform repository-wide **Legacy Modernization**. Using specialized automation suites,[23] agents autonomously extract business logic from legacy Assembler or COBOL and refactor it into cloud-native microservices. The curator reviews the **Architectural Decision Records (ADR)** to ensure the "reasoning path" aligns with federal security posture.

[19] https://nvlpubs.nist.gov/nistpubs/ir/2025/NIST.IR.8596.iprd.pdf
[20] https://www.cisa.gov/sbom
[21] https://www.nccoe.nist.gov/sites/default/files/2026-02/accelerating-the-adoption-of-software-and-ai-agent-identity-and-authorization-concept-paper.pdf
[22] https://martinfowler.com/articles/exploring-gen-ai/sdd-3-tools.html
[23] https://cloud.google.com/blog/products/infrastructure-modernization/mlogica-and-google-cloud-partner-on-mainframe-modernization/

2. Test/Release/Deploy (The Shield)

In this stage, AI acts as a protective shield, ensuring that machine-generated output remains compliant with federal security standards.

- **Ad-Hoc (Instructional):** Developers use AI to draft test scaffolds and Adversarial Emulation scenarios. In this stage, the AI suggests "what to test," but a human must still manually execute the tests and verify that the logic captures the specific nuances of the federal mission.
- **Assisted (Integrated):** AI-powered tools provide **Continuous Assurance** by automatically generating unit, integration, and property-based tests as code is written. Following a report on Agentic SDLC[24], these tests are "built-in" rather than "bolt-on," using Policy-as-Code[25] to enforce security rules and architectural guardrails before any code is merged into a production-bound branch.

[24] https://www.boozallen.com/insights/velocity/agentic-software-development-decoded.html
[25] https://www.crowdstrike.com/en-us/cybersecurity-101/cloud-security/policy-as-code/

- **Agentic (Autonomous):** The system shifts to **Autonomous Cyber Reasoning**. Utilizing frameworks proven in the **DARPA AI Cyber Challenge (AIxCC)**[26] autonomous agents move beyond finding bugs to actively generating and applying patches[27]. The "Shield" now operates as a real-time, self-hardening pipeline that conducts its own penetration testing and vulnerability remediation at a fraction of the cost of traditional methods.

3. Operate/Monitor (The Self-Healing)

The final stage of the lifecycle moves federal operations from reactive firefighting to proactive, autonomous remediation. This creates a "self-healing" environment where the system observes, governs, and corrects risk in flight.

- **Ad-Hoc (Intelligent Reference):** Staff utilize Generative AI-powered knowledge bases to accelerate manual incident response. Rather than sifting through siloed dashboards, analysts use chatbots or AI search engines to generate evidence-backed summaries of impact and "human-verified action plans" for long-term resolution.[28]

- **Assisted (Adaptive Monitoring):** Integrated AI performs Event Correlation and NOC (Network Operations Center) - SOC (Security Operations Center) Fusion. By unifying signals across metrics, logs, and traces, the system identifies "unknown unknowns" —novel patterns without predefined labels—reducing alert fatigue and allowing SOC analysts to focus on true behavioral anomalies across the entire hybrid fabric.[29]

- **Agentic (The Curator):** The system enters an Autonomous SOC state. In this phase, the curator's role shifts to "autonomy with control." Using next-generation platforms, agents not only pinpoint root causes but autonomously execute predefined fixes—such as isolation, rollback, or capacity re-scaling—while the curator manages the high-impact checkpoints and "kill switches" required for federal accountability.[30]

[26] https://aicyberchallenge.com/
[27] https://youtu.be/6Dy8BALCy_4?si=OHgOXs4uyb_7UD-F
[28] https://www.splunk.com/en_us/blog/ciso-circle/ai-generative-ui-security-analyst-investigations.html
[29] https://www.crowdstrike.com/en-us/blog/built-for-scale-powered-by-ai-innovation-driving-falcon-exposure-management/
[30] https://www.paloaltonetworks.com/blog/security-operations/2025-the-year-of-the-autonomous-soc-the-year-of-xsiam/

Phase 3 Risk Analysis & Guardrails: Zero-Trust for Machine Autonomy

As agencies transition from localized Generative AI applications to autonomous, agentic workloads, the degree of machine agency increases significantly. This progression requires a balanced strategy that pairs transformative capabilities with aggressive risk management. The central challenge is the migration from managing relatively deterministic vulnerabilities to managing the non-deterministic behaviors inherent in agentic agency. While foundational cybersecurity practices like threat analysis and Zero Trust guidance remain the bedrock of defense, the age of agents demands the adoption of "intelligent guardrails" paired with robust observability to handle threats across inputs, processing, and outputs.

1. The Agentic Threat Landscape

In the age of machine autonomy, guardrails must go beyond traditional signature-based approaches to handle complex, non-deterministic content across several critical categories identified by the **OWASP Agentic AI Threat Navigator:**

- **Agency & Reasoning:** Because agents have a degree of autonomy, a primary risk is that their internal "reasoning" process can be manipulated or drift from the mission goal. This leads to misaligned behaviors or goal manipulation, where the agent executes actions that fulfill a sub-goal while violating the strategic mission. "Repudiation and untraceability" are significant concerns because multi-step reasoning chains can be difficult for humans to audit after the fact. Referenced as **ASI01: Agent Goal Hijack** and **ASI10: Rogue Agents**. This happens when an attacker redirects objectives by manipulating tool outputs or internal instructions.

- **Memory & Context:** Autonomous agents rely on "working memory" to maintain the state of a task and their future plan. This memory is an active attack vector; memory poisoning occurs when an attacker introduces malicious information into the context that corrupts the agent's long-term decision-making or leads to cascading hallucinations, where one incorrect premise triggers a sequence of failed logic. Referenced as **ASI06: Memory & Context Poisoning**. This validates your point about persistent corruption in the agent's "working memory" (RAG stores or vector databases) shaping future behavior.

- **Tools & Execution:** Unlike standard chatbots, agents invoke functions (sending emails, executing code, accessing databases). This introduces risks of **tool misuse** and **privilege compromise**. Because an agent might be granted the ability to generate and run code, they are uniquely vulnerable to **Remote Code Execution (RCE)** if an attacker can manipulate the agent's instructions to execute malicious scripts instead of mission-critical tasks. Referenced as **ASI02: Tool Misuse & Exploitation** and **ASI05: Unexpected Code Execution**. This highlights the risk of "Vibe Coding Runaway" where agents execute unreviewed, malicious shell commands.

- **Identity & Authentication:** In agentic workflows, agents often act as Non-Person Entities (NPEs). If an agent's credentials or "persona" is hijacked, an attacker can perform identity spoofing, acting with the high-level privileges granted to that specific agent within the network. Referenced as **ASI03: Identity & Privilege Abuse**. This treats agents as Non-Person Entities (NPEs) and warns against "Delegated Privilege Abuse".

- **Human Engagement:** A major risk in "human-on-the-loop" systems is **overwhelming the Human-in-the-Loop (HITL)**. At machine-speed execution, the volume of agentic decisions can lead to "alert fatigue," causing a human curator to miss critical errors. Additionally, agents can be used for **human manipulation**, effectively "convincing" their human supervisors to approve risky or malicious actions. Referenced as **ASI09: Human-Agent Trust Exploitation**. This addresses the psychological manipulation of supervisors and "Authority Bias".

- **Multi-agency:** When multiple agents collaborate, the complexity of their interactions creates new vulnerabilities. Agent communication poisoning can occur if one "rogue agent" (either compromised or poorly prompted) provides malicious data to others in the chain, leading to a system-wide failure that is difficult to isolate. Referenced as **ASI07: Insecure Inter-Agent Communication** and **ASI08: Cascading Failures**, where a fault in one agent propagates through the entire workflow.

AI guardrails must be intelligent to handle non-deterministic content across various content categories and scenarios well beyond traditional deterministic signature-based approaches.

2. Maturity-Based Risk Analysis Matrix

The following matrix maps the evolution of threat vectors and the corresponding guardrail techniques required as an agency moves from manual creation to agentic curation.

Adoption State	Primary Threat Vector	Key Guardrail Technique
Human-Centric	**Deterministic Error:** Standard software vulnerabilities (CVEs) and manual misconfigurations.	Ensuring unit correctness for every human-authored line of code.
Ad-Hoc / Consultative	**Data Leakage & Hallucination:** Pasting Controlled Unclassified Information (CUI) into public LLMs or trusting inaccurate summaries.	Verifying all prose outputs against authoritative agency sources.
AI-Assisted (Pilot)	**Prompt Injection & Code Poisoning:** AI suggesting insecure patterns or "hallucinated" packages into the codebase.	Utilizing strict API contracts as **"Walled Gardens"** to fence probabilistic AI suggestions.
AI-Agentic (Curator)	**Goal Drift & Identity Hijacking:** Agents misinterpreting mission intent or compromised Non-Person Entities (NPEs) acting at machine speed.	Auditing **"Chain of Thought" (CoT)** logs to verify the autonomous reasoning behind every action.

3. Strategic Guardrails for the Agentic Frontier

To maintain a Zero-Trust posture for autonomous systems, federal curators must implement high-level structural defenses:

- **Non-Person Entity (NPE) Governance:** Every agent is treated as a high-privilege Non-Person Entity (NPE). Access is controlled through task-scoped, rapidly expiring credentials that prevent lateral movement within the network.
- **Zones of Intent Scaffolding:** By implementing micro-segmentation, agencies ensure that agents are physically incapable of reaching outside their designated architectural boundaries.[31]
- **Reversible Resilience (The "Big Red Button"):** Resilience is maintained through mandatory Standard Operating Procedures (SOPs) that ensure any agentic action can be surgically rolled back with a single click.[32]

Technical guardrails provide the necessary protection for machine autonomy, but long-term success requires alignment with national policy. As we move from the **technical "how"** of risk analysis to the **regulatory "what"** of governance, we ensure that agentic innovations remain fully compliant with the mandates that define secure federal AI.

POLICY, GOVERNANCE & COMPLIANCE

Compliance has evolved into a strategic accelerator. By providing a trusted framework for innovation, these policies allow agencies to scale AI with the confidence that their systems are resilient, auditable, and aligned with national security.

Alignment with National Mandates

To succeed in an agentic future, agencies must synchronize their operations with the definitive mandates governing federal artificial intelligence.

- **OMB M-25-21 (Accelerating Federal AI)[33]:** This directive shifts the federal focus toward removing unnecessary bureaucratic barriers. It empowers agencies to scale AI through mission-aligned strategic plans that prioritize rapid deployment and operational impact.

[31] https://www.zscaler.com/es/resources/brochures/zero-trust-microsegmentation-guide.pdf
[32] https://www.nextgov.com/ideas/2025/10/when-ai-agents-go-rogue-federal-government-needs-reversible-resilience/408757/
[33] https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf

- **OMB M-25-22 (AI Acquisition Standards)[34]:** To protect federal interests, this mandate requires rigorous pre-award testing. It secures the government's rights to its own data and AI outputs while setting strict limits on the use of federal data to train commercial models.
- **NIST Cyber AI Profile (NIST IR 8596)[35]:** This critical profile integrates the **NIST Cybersecurity Framework (CSF) 2.0[36]** with the **AI Risk Management Framework (AI RMF).[37]** It provides the technical foundation to address AI-specific threats, such as data poisoning and model exfiltration, in a unified defensive posture.

[34] https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-22-Driving-Efficient-Acquisition-of-Artificial-Intelligence-in-Government.pdf
[35] https://csrc.nist.gov/pubs/ir/8596/iprd
[36] https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf
[37] https://www.nist.gov/itl/ai-risk-management-framework

Operationalizing the Frameworks by Maturity

Governance must be adaptive. As an agency moves from research to curation, the technical controls transition from static lists to real-time behavioral monitoring.

Maturity Level	Governance Focus	Technical Reference
Ad-Hoc (The Researcher)	Use Case Inventories: Establishing a map of "Shadow AI" and initial risk categorization.	OMB M-25-21 (Inventory Requirements)
Human-Centric (The Creator)	SSDF v1.2 Integration: Traditional secure coding standards applied to human-authored lines.	NIST SP 800-218 (SSDF)
AI-Assisted (The Pilot)	COSAIS Overlays: Customizing SP 800-53 controls for predictive and generative assistants.	NIST COSAIS Framework (2026)
AI-Agentic (The Curator)	Continuous Assurance: Real-time monitoring of Chain of Thought (CoT) and model drift.	NIST AI RMF Core / IR 8596

CLOSING VISION: THE SECURE FRONTIER

AI is no longer a peripheral experiment; it is the **foundational capability** for secure, resilient software delivery. By navigating this roadmap—from the **Cognitive Re-calibration** of the workforce to the **Technical Infusion** of the SWE lifecycle and the **Strategic Implementation** of intelligent guardrails—federal agencies can lead the agentic era with precision.

The shift from **Creator to Curator** does not diminish the human role; it elevates it. The federal workforce is now the orchestrator of a digital squad, leading the future of government service with unparalleled speed and security.