December 2025

ATARC

WHITE PAPER

Navigating the AI Regulatory Landscape:

A Strategic Framework for Government Organizations

Acknowledgements

Christopher Stephenson, *Nuix*

John Sprague, *National Aeronautics and Space Administration*

Natalie Buda Smith, *Library of Congress*

powered by GovExec

Table of Contents

Figures and Tables

Disclaimer: This document was prepared by the members of the ATARC Agentic AI Working Group in their personal capacity. The opinions expressed do not reflect any specific individual nor any organization or agency they are affiliated with and shall not be used for advertisement or product endorsement purposes.

Executive Summary: With the rapid advancement of artificial intelligence, government organizations are navigating a challenging landscape where innovation must be balanced with regulatory compliance. This paper outlines a strategic approach to effectively manage AI regulations while ensuring operational efficiency.

Introduction: The Current State of AI

Artificial intelligence is no longer a future concept; it is actively reshaping global dynamics, national security, and agency operations. Today, as government entities incorporate AI into critical systems, it is essential for leaders to grasp not only the technologies' potential, but also its regulatory and operational challenges. This calls for a strategic approach to encourage innovation while safeguarding ethics, oversight, and public trust.

Our objective is to guide senior government officials and program managers in responsibly integrating AI, aligning agency goals with regulatory compliance and best practices for long-term resilience. To do so, we first explore the current global and national AI landscapes.

1.1 Global Context

Rapid Advancement of AI Capabilities

AI has evolved from specialized applications to more versatile and powerful systems. Technologies like large language models, autonomous systems and generative AI are now integrated into critical infrastructure and defense platforms. This fast-paced innovation outstrips many organizations' policy frameworks leading to increased risk if not addressed.

Agentic AI Emerging Quickly: Increasing Autonomous Actions and Decision-Making

Autonomous systems are making critical decisions in national security, legal, and financial sectors, often without transparency. Unlike traditional AI that relies on human intervention, autonomous AI adapts and operates independently, raising significant governance challenges. Without clear explainability and transparency, trust and accountability are at risk.

A Shift Toward Edge Intelligence

There is a significant move from centralized AI systems to edge computing. This shift allows real time processing and improved privacy but introduces new security challenges. For government agencies, this means new opportunities and risks requiring adaptive security frameworks and governance models.

Emerging International Regulatory Frameworks

Countries and global organizations are quickly establishing AI regulations. The EU's AI Act is comprehensive, imposing strict obligations based on risk. U.S. agencies operating internationally must navigate complex compliance across borders, making this a strategic priority.

Cross-border Implications for Government Agencies

Federal agencies operate in multi-jurisdictional environments, whether through alliances or joint research. Systems must be designed for interoperability and legal defensibility across borders. Missteps in one area can lead to broader diplomatic and operational issues.

Prolonged Regulatory Uncertainty

Federal agencies face an extended period of regulatory uncertainty as hundreds of sometimes conflicting AI regulations make their way through a multi-year maturity cycle across multiple jurisdictions. This sustained ambiguity means agencies must make critical AI deployment decisions now without clear regulatory guidance, driving the need for adaptive governance practices as outlined below.

1.2 National Landscape

U.S. Leadership in AI Innovation and Regulation

The US leads in AI research and development, driven by federal investments from agencies like DARPA and NSF. However, with leadership comes the responsibility to maintain standards. The NIST's AI Risk Management Framework provides direction, and new policies like the AI Action Plan emphasize a pro-innovation approach. These frameworks set the baseline for government agencies.

Critical Role of Government Agencies in AI Adoption

U.S. Federal agencies are at the forefront of AI adoption across diverse missions, from climate research to national security. These agencies don't just use AI; they set standards for its implementation. This calls for governance frameworks that ensure fairness, reliability, and safety throughout the AI life cycle.

Balance Between Security and Technological Advancement

Federal agencies face the dual challenge of advancing technology while protecting civil liberties and national security. Overemphasizing speed can lead to failures, while too much caution can leave the U.S. behind adversaries. Agencies must integrate regulation and governance into their AI strategies for systems that are responsible, reliable, and trustworthy.

The following sections provide a detailed blueprint for embedding regulatory navigation, strategic alignment, and AI governance into government agencies' digital transformations.

2. Regulatory Framework Overview

As AI integrates into society, a complex web of regulations emerges. For U.S. government leaders, understanding this landscape is crucial. It's about fostering innovation responsibly while protecting national security, economic stability, and civil liberties.

This regulatory environment is multi-layered and evolving. For instance, California's new Transparency in Frontier Artificial Intelligence Act focuses on AI safety. Domestically, there are executive orders and guidance from agencies like NIST and the FTC. Internationally, legal frameworks coexist with ethical principles from organizations like OECD and UNESCO.

Regulations evolve through a multi-year process of legislative groundwork, passage, enforcement precedent, and judicial clarification, which can take 7 years or more for each regulation. With hundreds of AI regulations in various stages across federal, state, and international jurisdictions, agencies face sustained regulatory uncertainty through at least 2032. Leaders must build adaptive governance frameworks that can evolve alongside these shifting requirements.

Organizations that navigate these guidelines effectively will be better positioned to:

- **Build Trustworthy AI Systems: Ensuring fairness, transparency, and accountability.**
- **Mitigate Risks Proactively: Addressing bias, privacy, and security issues before they arise.**
- **Ensure Compliance and Reduce Legal Exposure: Avoiding penalties and reputational harm.**
- **Foster Interoperability and Collaboration: Aligning with national and international standards.**
- **Drive Responsible Innovation: Using frameworks as guardrails for ethical AI advancement.**

Understanding these standards is foundational for any agency's AI strategy, allowing leaders to anticipate regulatory changes and cultivate a culture of innovation and responsibility.

For a detailed breakdown of the specific US guidelines, international standards, and industry frameworks currently shaping the AI regulatory environment, please refer to *Addendum A: AI Regulatory Reference Guide.*

3. Strategic Implementation Framework

3.1 Core Principles

- **Explainability:** AI systems should produce outputs that are understandable to analysts, auditors, and decision makers. It is crucial for sensitive areas like public services and law enforcement to ensure trust in AI decisions.
- **Transparency:** Agencies must make AI use visible and understandable to both stakeholders and the public, promoting accountability and trust.
- **Fairness:** AI systems should not perpetuate bias, especially in areas affecting people's lives. Continuous monitoring is necessary to ensure equitable treatment.
- **Safety:** AI systems must be reliable and secure, particularly in high-stakes applications. This involves stress testing, fail-safes, and human oversight.

AI Core Principles – Quick Guide

Explainability	• Ensure AI outputs can be explained in plain language. • Citizens, auditors, and oversight bodies should understand the "why" behind decisions. • *Checkpoint:* Can you justify this AI decision in a hearing or to the public?
Transparency	• Be clear about where and how AI is used. • Disclose data sources, intended use, and known limitations. • *Checkpoint:* Would the public know they are interacting with AI in this process?
Fairness	• Test for and prevent bias that harms protected groups. • Monitor systems continuously for discriminatory outcomes. • *Checkpoint:* Does this AI system treat all populations equitably?
Safety	• Ensure systems are reliable, secure, and resistant to failure or manipulation. • Put in place fail-safes and maintain human oversight for high risk uses. • *Checkpoint:* What's the plan if this AI system makes an error or is attacked?

3.2 Governance Structure

- **Policy Development:** Establishes guidelines for AI use, ensuring consistency and legal soundness.
- **Risk Management:** Addresses AI specific risks like bias and cybersecurity using technical and organizational safeguards.
- **Authorization Requirements:** Evaluates AI systems for ethical, legal, and mission specific risks before deployment.
- **Compliance monitoring:** Ensures ongoing adherence to standards, maintaining public trust in AI systems.

4. Practical Implementation Guidelines

4.1 Risk Management

Assessment Methodologies:

Establish evidence-based methods to identify and prioritize AI risks.

1. Scope & system profiling defines the AI system boundary (data sources, models, prompts, third party services, deployment channels), and classifies by mission use (public services, law enforcement, critical infrastructure, and decision criticality).
2. Algorithmic Impact Assessment (AIA) covers stakeholder mapping, use-case analysis, impact estimation, and governance controls.
3. Measurement plans (eval strategy)

   Define fit-for-purpose metrics: e.g., task performance, calibration, robustness, fairness, (group, individual, interactional), privacy leakage, security posture, latency, cost (sample table below).

Task Performance	Calibration Assessment	Quantify Robustness	Fairness Metrics	Privacy & Security
Use standardized scoring (0-1 scale or %) with clear thresholds	Measure Expected Calibration Error (ECE), e.g., predicted vs actual	Stress testing with perturbation (e.g., 10, 25, 50% data corruption)	Demographic Parity: Track selection rates across protected groups	Implement differential privacy with epsilon values (e.g., ε ≤ 1.0 for sensitive apps)
Use confusion matrices for classification, precision, recall and F1 scores across groups	Use reliability diagrams to track confidence across ranges	Measure degradation under adversarial conditions using standardized attacks	Equalized Odds: Track true & false positive rates across groups	Measure info leakage thru membership inference attacks with success rate thresholds
Establish statistical significance testing with confidence levels and min. sample sizing	Set acceptable calibration thresholds (e.g., ECE <0.05 for high stakes apps)	Set minimum acceptable performance floors under various stress conditions	Individual Fairness: Use distance-based metrics to ensure similar individuals or cases receive similar outcomes or decisions	Conduct pen testing with quantified vulnerability scores and remediation timelines
			Set fairness thresholds tied to legal and ethical requirements.	

4. Security & supply-chain risk assessment involves modeling threats, capturing AI components, and aligning with relevant controls.
5. Privacy assessment ensures data handling complies with regulations and best practices.
6. Human-in-the-loop (HITL) analysis defines points where human oversight is essential, validating training and ergonomics.
7. Risk register & Authority To Operate (ATO) alignment log risks and tie them to authorization artifacts.

Deliverables: System Profile, AIA report, Evaluation Plan, Threat Model, Privacy Analysis, Risk Register, Governance RACI, ATO crosswalk.

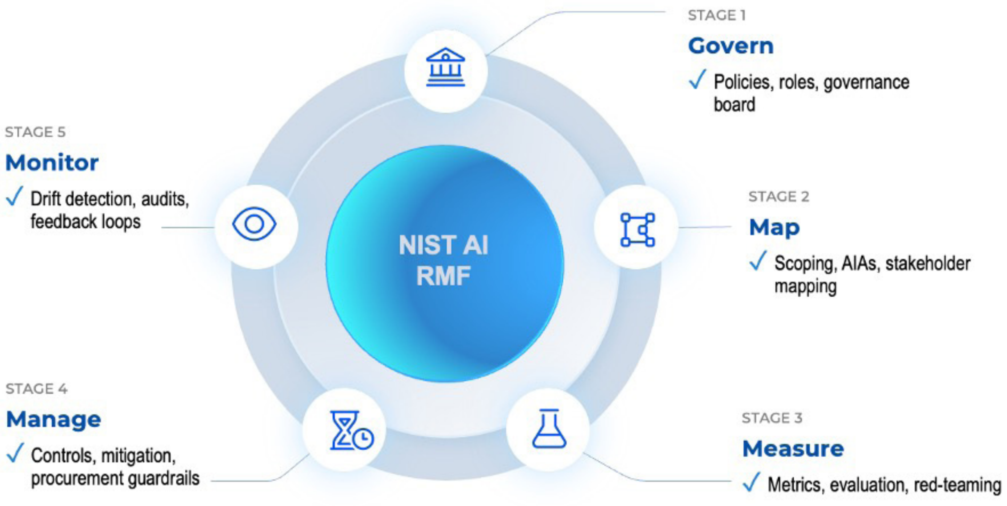

STAGE 1
Govern
✓ Policies, roles, governance board
STAGE 2
Map
✓ Scoping, AIAs, stakeholder mapping
STAGE 3
Measure
✓ Metrics, evaluation, red-teaming
STAGE 4
Manage
✓ Controls, mitigation, procurement guardrails
STAGE 5
Monitor
✓ Drift detection, audits, feedback loops
NIST AI RMF

Fig 1. A continuous cycle of governance, measurement and improvement ensures that AI remains trustworthy throughout its lifecycle.

Mitigation Strategies:

Purpose: Convert identified risks into practical controls and operational conditions.

1. Data governance & quality

- Track data lineage; ensure consent documentation and collection authority.
- Mitigate bias at the source (through re-sampling or re-weighting) and in post-processing (using threshold and decision rules).
- Use gold-standard labeling with quality assurance, check inter-rater reliability, and maintain drift-resilient refresh cycles.

2. Model-level controls

- Robustness: Implement adversarial testing suites, set guardrails for prompt injection and toxic output, and fine-tune safety filters to match the context.
- Fairness and explainability: Employ constraint-aware training or post hoc adjustments; use interpretable surrogates for audits and conduct counterfactual analysis for high-stakes applications.
- Privacy: Utilize differential privacy, secure enclaves for sensitive inference, and emphasize minimization and on-device processing where possible.
- Security: Ensure model isolation, apply egress filtering, moderate content, secure pipelines, enforce strict dependency pinning, and use code-signing.

3. Process/governance controls

- Gate reviews: Transitioned from concept to sandbox, then to controlled pilot, limited production and full production with exit criteria and executive approval at each stage.
- HITL Protocols: Established documented approval thresholds, sampling regimes for secondary review, escalation paths, and kill-switches.
- Policy alignment: Map mitigations to NIST AI RMF functions, 800-53 controls, Consult 800-53 Control Overlays for AI (when available), civil rights/consumer protection obligations, records management, and FOIA discoverability.

4. Operational safeguards

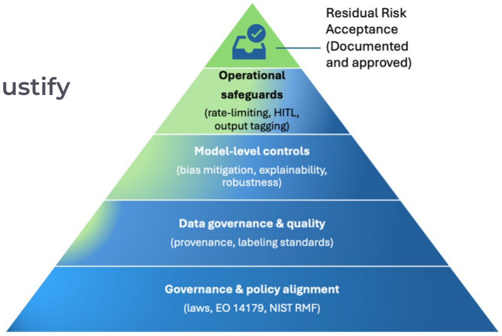
- Output controls: Implement confidence tagging, use-constraint banners and provenance/watermarking for generated content.
- Context controls: Enforce retrieval-augmentation whitelists, data classification, and least-privilege API keys.
- Rate-limiting & anomaly detection: Apply throttling for unusual query patterns, set cost caps, and identify abuse signatures.

5. Procurement & third-party risk

- Include contract clauses for model change notifications, evaluation, transparency, data set disclosures, vulnerability reporting, uptime/SLOs, incident response, and data locality.
- Require evaluation artifacts, (like red team results and bias studies) and independent verification rights.
- Ensure FedRAMP/FISMA alignment where hosting is involved, and plan for portability and exit strategies.

6. Residual Risk & Risk Appetite

- Document remaining risks after controls and justify their acceptability relative to mission value.
- Define compensating controls and conditions for pausing or rolling back deployment.


Residual Risk Acceptance (Documented and approved)
Operational safeguards
(rate-limiting, HITL, output tagging)
Model-level controls
(bias mitigation, explainability, robustness)
Data governance & quality
(provenance, labeling standards)
Governance & policy alignment
(laws, EO 14179, NIST RMF)

Effective mitigation relies on layered defenses — from foundational data governance up to executive-level risk acceptance.

Continuous Monitoring:

Purpose: Maintain trustworthiness over time as data, models, users, and threats evolve/drift (model/data drift).

1. Metrics and telemetry

- KPIs: Measure task success rate, latency, operator workload, user satisfaction, and mission outcomes.
- KRI: Track error rates by subgroup fairness, drift, hallucination/toxicity rates, privacy/ security events, and model/data drift indicators.
- Quality gates: Set rolling thresholds with alerting; link error budgets to auto-throttle or rollback.

2. Drift, Bias, and Robustness Monitoring

- Use data/label drift detectors and periodically re-score on holdout and stress test sets.
- Deploy shadow models or ‘canary’ deployments to compare pre- and post-behavior.
- Conduct scheduled fairness audits (e.g., quarterly) with intersectional analysis; publish audit summaries where appropriate.

3. Security and Abuse Monitoring

- Utilize threat intelligence feeds for model-specific exploits; conduct red team exercises, and “chaos days” to probe defenses.
- Implement continuous dependency scanning and SBOM diffs; use prompt injection and exfiltration detectors on inputs/outputs.
- Apply automated containment: session quarantine, credential rotation, and policy re-evaluation on signals.

4. Human Oversight-in-Production

- Conduct sampling reviews: recheck a percentage of determinations by trained reviewers; log errors, taxonomy, and corrective actions.
- Establish feedback loops: implement one-click operator flags, a user appeals process and structured incident tickets in a central AI Issue Tracker.

5. Change Management & Re-approval

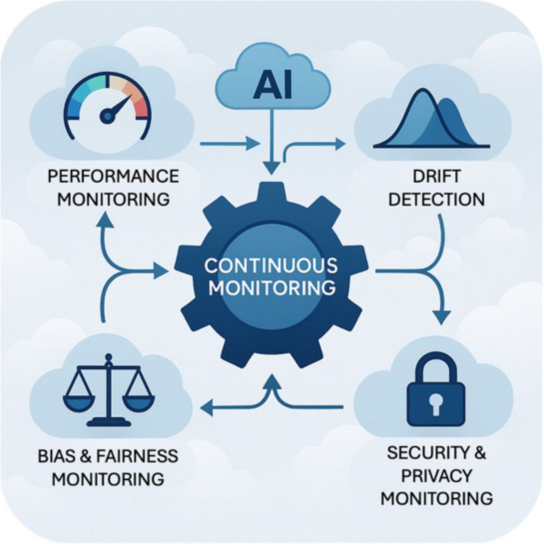
- Maintain versioned model release notes (weights, prompt datasets, hyperparameters); classify changes by risk tier.
- Trigger re-evaluation on material changes (data swap, major weight update, domain shift, new user group).
- Conduct periodic governance reviews (e.g., semiannual) to renew risk posture and ATO artifacts.

6. Incident response and learning

- Develop AI-specific incident response playbooks: containment, external communications, legal/privacy coordination, and post-incident causal analysis.
- Conduct after-action reviews to update evaluations, controls, and training; track closure in Plan of Action and Milestones.

7. Transparency and Reporting

- Create operator dashboards with real-time KRIs; executive scorecards; red/yellow/green status by system.
- Provide public-facing transparency reports when appropriate; ensure records management and audit-ready evidence retention.



Deliverables: Monitoring Plan, Live Dashboards and Alerts, Audit Logs, Change Records, Incident Playbooks, Transparency Reports.

4.2 Authorization Process

The authorization process ensures AI systems are thoroughly evaluated before going into production. It acts as a formal gatekeeper to verify systems are secure, ethical, and compliant with U.S. government regulations, mirroring traditional federal IT security reviews while addressing AI-specific risks.

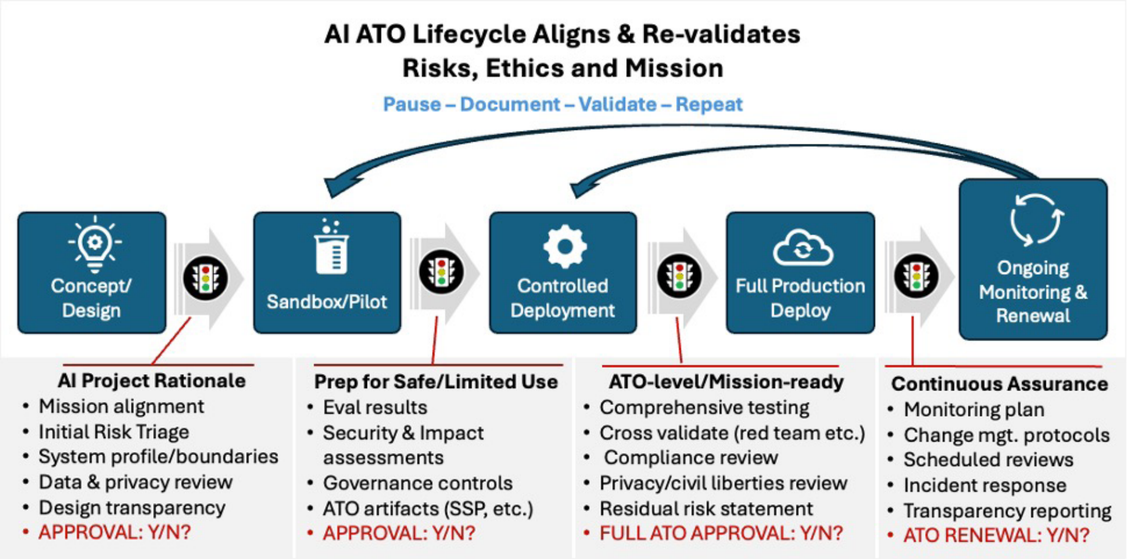
1. ATO Requirements (Authority to Operate)

The authority to operate (ATO) is the formal approval needed for any AI system in a government environment. While similar to traditional IT systems, AI ATO evaluations focus on algorithmic transparency, bias, and continuous risk management.

- **Expanded Evaluation Scope:**
  - **ATO Set reminder, reviews must evaluate ethical risks, model transparency, privacy impacts, and mission alignment alongside standard cyber security assessments.**
  - **Traditional controls (aligned with NIST 800-53) are supplemented by AI-specific factors such as explainability, fairness, and robustness.**
  - **AI-specific threat models (e.g., data poisoning, prompt injection, model inversion) are mandatory.**
- **Lifecycle-Based Authorization:**
  - **AI systems evolve rapidly, making a single point-in-time authorization insufficient.**
  - **Agencies should adopt continuous authorization models with risk reviews at each stage of the AI lifecycle:**

    Design → Pilot → Limited Production → Full Deployment → Ongoing Monitoring
- **Documentation Alignment:**

  Tie all risk assessments, evaluation plans, and mitigation strategies back to ATO artifacts such as:
  - **System Security Plan (SSP) – detailing boundaries, components, and controls.**
  - **Security Assessment Plan/Report (SAP/SAR) – outlining evaluation methods and findings.**
  - **Plan of Action & Milestones (POA&M) – tracking remediation tasks.**
  - **This creates traceability between AI risks and compliance obligations, ensuring decisions are defensible under audits or external review.**



AI ATO Lifecycle Aligns & Re-validates Risks, Ethics and Mission

Pause – Document – Validate – Repeat

Concept/ Design

Sandbox/Pilot

Controlled Deployment

Full Production Deploy

Ongoing Monitoring & Renewal

AI Project Rationale

- Mission alignment
- Initial Risk Triage
- System profile/boundaries
- Data & privacy review
- Design transparency
- APPROVAL: Y/N?

Prep for Safe/Limited Use

- Eval results
- Security & Impact assessments
- Governance controls
- ATO artifacts (SSP, etc.)
- APPROVAL: Y/N?

ATO-level/Mission-ready

- Comprehensive testing
- Cross validate (red team etc.)
- Compliance review
- Privacy/civil liberties review
- Residual risk statement
- FULL ATO APPROVAL: Y/N?

Continuous Assurance

- Monitoring plan
- Change mgt. protocols
- Scheduled reviews
- Incident response
- Transparency reporting
- ATO RENEWAL: Y/N?

The AI ATO lifecycle is the iterative, staged process of moving an AI system from conception to deployment.

2. Compliance Documentation

Compliance documentation records due diligence, demonstrating that the agency has identified, assessed, and mitigated risks before deployment. It provides legal defensibility and public accountability, streamlining external oversight.

- **Required Documentation Types:**
  - Algorithmic Impact Assessments (AIA): Document the system's purpose, potential harms, affected populations, and alternatives considered.
  - Data Cards and Model Cards: Detail data lineage, provenance, and intended model use to support transparency and audit readiness.
  - Privacy Impact Assessment (PIA): Required for systems collecting, processing, or generating personally identifiable information (PII) or sensitive data.
  - Security and Supply Chain Reports: Identify dependencies, third-party models, licensing terms, and update cadences.
- **Version Control Auditability:**
  - Version all documents and store them in a centralized repository accessible to both internal reviewers and external oversight bodies.
  - Ensure documentation is audit-ready, supporting rapid responses to inquiries from GAO, OMB, or congressional committees.
- **Cross-Agency Interoperability:**

  Standardized documentation formats should align with government-wide frameworks such as:
  - NIST AI Risk Management Framework (AI RMF)
  - FedRAMP and FISMA requirements for hosting environments
  - Sector-specific regulatory overlays (e.g., DoD RMF, HIPAA for health data)

This ensures consistency across agencies and promotes collaboration among federal partners.

3. Security Protocols

AI systems introduce new attack surfaces requiring specialized security measures. Traditional cyber security approaches are necessary but insufficient alone. Security protocols must address conventional IT threats and AI-specific vulnerabilities, especially as edge AI becomes more prevalent.

- **AI-Specific Threat Modeling:**

Establish a comprehensive threat model addressing:

  - Data poisoning - malicious inputs corrupting model training.
  - Prompt injection and manipulation - influencing outputs in unintended ways.
  - Model inversion and data leakage - exposing the sensitive training data through model queries.
  - Adversarial attacks - exploiting weaknesses in model decision boundaries.

- **Technical Safeguards:**
  - Segmentation: Isolate models and datasets from other systems to reduce blast radius.
  - Secure development pipelines: code signing, dependency pinning, and reproducible builds to prevent supply chain compromise.
  - Content and output filtering: guardrails to detect and block malicious or inappropriate outputs before reaching the end users.
  - Telemetry and anomaly detection: Real-time monitoring for suspicious activity such as excessive query rates or abnormal usage patterns.
  - Hardware-based security modules and encrypted local processing environments including:
    - ◊ Trusted Execution Environment (TEE) methodologies
    - ◊ Hardware Security Modules (HSM) at the edge endpoints
- **Operational Protocols:**
  - Continuously scan third-party libraries and pre-trained models for vulnerabilities.
  - Implement credential management policies, including automated rotation and least-privileged access enforcement.
  - Conduct regular red team exercises, “chaos days,” and penetration tests focused on AI-specific exploits.
  - Establish comprehensive local access controls and multi-factor authentication systems to ensure only authorized personnel can access edge-processing intelligence. Include real-time monitoring and anomaly detection to identify potential security breaches at distributed processing points.

- **Integration with Broader Cybersecurity Governance:**

Align these protocols with:

- NIST SP 800-53 for system controls.
- Zero Trust Architecture (ZTA) principles for access control.
- CISA directives for federal incident response coordination.

By integrating AI-focused defenses with established federal cybersecurity frameworks, agencies create a layered adaptive defense posture.

The authorization process ensures that AI systems deployed in government contexts are not only secure, but also ethically sound and legally defensible. By combining rigorous ATO reviews, comprehensive documentation, and AI-specific security protocols, agencies can achieve responsible innovation while protecting both public trust and national interests.

4.3 Quantifying AI Metrics: Practical Guidance

- **Start Simple, Scale Systematically:** Begin with core performance and fairness metrics, then gradually add sophistication as organizational capability matures. Use automated monitoring tools to reduce manual overhead and ensure consistent measurement.
- **Build Interpretable Dashboards:** Create executive-level dashboards that translate technical metrics into business and mission impact terms. Include trend analysis and early warning indicators for proactive management.
- **Establish Metric Governance:** Implement formal processes for metric selection, threshold setting, and evolution. Document rationale, stakeholder approval processes, and conduct regular effectiveness reviews.

The key to successful metric quantification is balancing statistical rigor with operational practicality, ensuring that measurement systems drive better AI outcomes rather than becoming bureaucratic obstacles to innovation.

5. Maintaining Operational Agility

In today's fast-paced AI landscape, keeping up with change means being ready to adapt both legally and technically, without losing sight of mission goals and regulatory responsibilities. For government entities, this is no small feat. Federal agencies are designed for stability and consistency, not rapid shifts. Their systems are built to favor predictability over experimentation.

For US agencies entering the age of AI, agility can't just come from policy statements; it needs to be cultivated intentionally. This involves weaving flexibility into governance structures, ensuring oversight is integrated across different functions, and nurturing a culture of ongoing learning where innovation and accountability go hand in hand. Achieving true agility in government means updating processes and mindsets while retaining the discipline and public trust that are the bedrock of these institutions. Essentially, it calls for the bravery and foresight to rethink how agencies operate and innovate with AI. This requires exploring new approaches that balance innovation with compliance, evolve your workforce, and integrate long term strategic thinking that is resilient to the rapid and relentless changes that AI ensures.

5.1 Balancing Innovation and Compliance

- **Adaptive regulatory frameworks**

  Compliance should be seen as an evolving discipline rather than a fixed requirement. Adaptive frameworks enable agencies to quickly adjust to new mandates and guidelines without starting from scratch. This involves creating flexible governance structures where documentation and control systems can be updated in line with changing standards by embedding regulatory intelligence and forward-looking practices. Agencies can anticipate changes instead of merely reacting. The aim is to make adaptability a core feature, allowing compliance to keep pace with innovation. Setting UA regulatory sandbox is a useful approach, as demonstrated by some international agencies.

  One example of a government agency, albeit a foreign organization, using the regulatory sandbox approach can be found here.

- **Integration strategies**

  Operational agility hinges on aligning AI governance with existing IT and mission operations, rather than creating separate bureaucracies. Integration ensures consistency and reduces compliance friction. Key strategies include interoperable data architectures, shared (Cyber, AI, Data and Program) risk registers, and standardized documentation that connects policy and engineering with mission functions. This approach allows AI projects to integrate seamlessly into existing operations, maintaining accountability and transparency.

Successful approaches include:

- **Data Mesh/Domain-centric Architecture:** By organizing data ownership by domain (like border security or health surveillance), each team manages and shares data with standardized APIs and metadata. This structure allows AI systems to access domain-specific data without needing a centralized repository.

  The CDAO's Data Mesh Reference Architecture (DMRA), for example, encourages a decentralized, interoperability approach captured in this document.

- **Shared Risk Registers Across Domains:** A central risk register acts as a coordination tool, creating a communication backbone between IT, mission, and governance teams. NIST paper on cybersecurity and enterprise risk management can be found here.

- **Standardized Documentation & Template Libraries:** Having a shared template library helps maintain consistency across the agency and streamlines reviews and audits. The GSA's USAi platform serves as a helpful example.

- **Additional approaches include:** embedding governance in existing processes, establishing cross-functional AI Centers of Excellence (CoE), and running pilot integration projects to test concepts.

- **Operational Flexibility**

  Cultivating a culture of controlled experimentation is crucial for innovation within defined limits. Agencies should enable teams to test and scale AI tools in secure environments before full deployment. Policies should support risk-based authorization models that allow for iterative improvements without disrupting missions. Cross-functional collaboration ensures decisions are informed by real-world conditions. Operational flexibility is about balancing precision with adaptability, maintaining mission integrity while adapting to technological changes.

  In practice, operational flexibility comes from creating safe spaces for experimentation, where teams can rapidly iterate without compromising compliance or mission goals.

- **Metric Evolution and Adaptive Frameworks**
  - Continuous Learning Integration:
    - Start with industry benchmarks and refine based on operational experience.
    - Use statistical tools to detect when metrics deviate from expectations.
    - Adjust thresholds using Bayesian methods as more data becomes available.
  - Stakeholder Feedback Integration:
    - Use user satisfaction scores and complaints as indicators of metric adequacy.
    - Regularly review and assess metric relevance.
    - Use A/B testing to validate metric improvements.
  - Regulatory Adaptation:
    - Proactively adjust metrics in response to regulatory changes.
    - Track metric changes over time for audit purposes.
    - Harmonized metrics across jurisdictions for systems operating in multiple regions.

5.2 Balancing Innovation and Compliance

To stay agile with AI while ensuring mission assurance and compliance, government agencies need a workforce that combines traditional expertise with emerging AI skills. This shift requires rethinking recruitment, hiring, and staff development, as well as upskilling existing employees.

Considerations include:

- These individuals bridge the gap between technical AI concepts and policy needs, serving as crucial links between engineering and compliance.
- Adaptive Governance Specialists: Focused on creating flexible governance frameworks rather than rigid structures, these specialists help evolve compliance systems with changing regulations.
- Cross-functional Integration Teams: Agile teams that work across silos, combining expertise from various domains to enable a culture of controlled experimentation.
- Regulatory Intelligence Analysts: Dedicated to monitoring the AI regulatory landscape, anticipating changes, and translating requirements into organizational actions.
- AI-Literate Mission Specialists: Domain experts who understand both their field and AI capabilities, ensuring AI projects are relevant and connected to operational needs.
- Continuous Learning Facilitators: Focused on knowledge transfer and adaptive practices, ensuring insights from experiments feed into organizational learning and policy refinement.

5.3 Future-proofing

In a world where AI technology and governance evolve together, future proofing means building the ability to detect changes early and adapt smoothly. For U.S. government agencies, it's about being ready to adjust without disruption. This approach includes:

- **Regulatory Evolution Monitoring:** Keeping up with AI regulation requires continuous monitoring and proactive adaptation to changes. Agencies need to maintain situational awareness to anticipate compliance obligations.
- **Strategic Adaptation:** Turning awareness into action requires the ability to adjust policies, workflows, and investments in response to new risks and technologies without disrupting missions. It's about embedding adaptive thinking into strategy cycles, treating each deployment as both a tool and a learning opportunity.
- **Emerging Capabilities Considerations**

  a. Edge Intelligence: With AI moving to edge computing, real-time processing, enhanced privacy, and improved resilience are possible. This shift presents opportunities and challenges requiring adaptive security frameworks and distributed oversight models.

b. **AI in Telecom:** Concomitant with edge intelligence, the transition from 4G to 5G (and eventually 6G) is heavily AI-driven. U.S. can win the innovation race (with China, Europe, SE Asia) in 5G if it concentrates more on software, and less on hardware. Cloud-native 5G architecture is well-suited for the U.S. Ecosystem. Edge AI/computing is one of the many components of this multi-pronged innovation race. Eventually, the transition to 6G is expected to be primarily AI-based.

c. **AI and Quantum Computing:** The convergence of artificial intelligence (AI) and quantum computing (QC) holds transformational potential across the economy. AI has evolved since its inception in the 1950s and now includes a wide range of approaches and an even wider range of application areas. QC, on the other hand, is still in the early days of a long-term research and development (R&D) path but has enormous future potential that would rival what is currently unfolding for AI. QC is projected to dramatically increase the scale, complexity, and scope of problems that can be solved computationally, while AI has already demonstrated its capacity to produce value in solving problems across numerous domains. As these two fields continue to develop, their combined use may offer opportunities to go well beyond the current limits of either technology.

d. **AI and Robotics:** New robotics labs are exploring AI's role in scientific workflows, focusing on hazard mitigation, knowledge transfer, and discovery. Research is beginning with mobile robots and will likely expand to humanoid platforms.

6. Recommendations

6.1 Immediate Actions

- **Establish a Centralized AI Governance Structure**

  As emphasized in OMB Memorandum M 25-21, effective AI governance serves as the foundation for responsible innovation across the federal enterprise. The memorandum underscores that strong governance empowers personnel at all levels to align policies, streamline processes, promote accountability, and maintain a registry of all AI systems and products.

- **Develop Control Matrices**

  Map specific requirements to different impact levels (minimal, moderate, high, very high) allowing the agency to apply proportionate controls without over-engineering solutions for low-risk applications.

- **Build Cross-functional Implementation Teams**

  Establish working groups that combine technology, legal and policy expertise for balanced implementation. Provide role-based training programs to equip stakeholders with the knowledge they need.

- **Leverage Existing Standards as Building Blocks**

  Use a combination of frameworks (like NIST AI RMF, EU AI Act and ISO 42001) to create robust foundation that can adapt to changing requirements. Use voluntary frameworks to build capability before regulations take effect.

- **Implement AI RMF Key Functions:**

  1. Govern: Create a risk aware culture.
  2. Map: Understand deployment context.
  3. Measure: Assess risks and benefits.
  4. Manage: Decide whether to use the model results.

Federal leaders and program managers should recognize AI systems as IT systems 1st and apply traditional cyber security measures. Determine access to training data carefully, consult relevant NIST publications, and use internal models to achieve agency goals. Stay informed with AI-specific overlays and frameworks as they become available.

There is a need for guardrails to protect AI systems, including AI systems to protect other AI systems, in addition to AI in Cyber Defense (at Machine Speed in some cases).

Considering the above, Federal Leaders and Program Managers should:

1. Realize that AI systems (Models, Training data…) are IT systems first, so they must apply traditional Cybersecurity defense mechanisms. (These are in place in most Federal systems.)
2. Carefully determine who gets access to training data sets and models.
3. Install measures to ascertain success (as enumerated in OMB Memorandum 25-21).
4. Consult NIST Special Publication 800-53 Revision 5 (NIST SP 800-53, Revision 5 Crosswalk), NIST SP 800-218A (for Machine Learning, Decision Tree type AI) and SP 800-218, NIST-AI-600-1: AI RMF Generative AI Profile, and the **AI RMF** (3. Secure and Resilient from the 7 Principles of AI Risk and Trustworthiness).
5. Use Internal Models – **NIST** and many other Federal Agencies use their Internal Models to achieve agency specific goals. People are authorized internally to use these models, with Classified data in some cases as needed.
6. Consult 800-53 Control Overlays for AI (Simplified View of the key differences for AI), when available.
7. Consult Cybersecurity Framework (CSF) Profile for AI.

- **Develop Compliance Frameworks**

  Compliance needs a shift from a simple checklist approach to a dynamic framework that continuously aligns agency operations with new AI guidelines and legal requirements. Agencies should start by examining their current compliance programs like FISMA, FedRAMP, and the Privacy Act, and compare them with AI-specific risks such as data origins, algorithmic bias, and model drift. This comparison helps identify where existing controls are effective, and where new AI related responsibilities arise.

  Each agency should create a Compliance Playbook, to serve as an authoritative guide outlining documentation standards (e.g., model cards and assessments like AIAs and PIAs). These should be integrated into procurement and ATO processes early on, reducing the need for later revisions and ensuring clear documentation from design to deployment. Viewing compliance as a tool for building trust, rather than a hindrance, also allows agencies to move swiftly while ensuring each system withstands scrutiny and audits.

- **Emphasize Documentation & Traceability**

  - Develop thorough documentation standards that create audit trails for AI development and deployment decisions. This includes keeping records of impact assessments, risk evaluations, and implemented mitigation measures.
  - Implement version control for AI governance policies and ensure all stakeholders have access to the latest guidance and templates.
  - Document decision-making rationale, not just outcomes. To address extended regulatory uncertainty and conflicting regulations, record the reasoning behind key AI governance decisions. This demonstrates good faith compliance efforts when choosing between competing requirements, and provides defensible justification as regulatory boundaries clarify over time.

- **Engage Proactively with the Regulatory Community**
    - Participate in industry workshops and regulatory consultations to stay updated on new requirements, gaining early insight into regulatory trends and opportunities to shape policy development.
    - Build relationships with regulatory bodies before you need them. Proactive engagement shows good faith efforts and can lead to smoother interactions when formal compliance is necessary.

- **Leverage Specialized Tools**

  To make governance practical, agencies should consider adopting specialized toolkits and frameworks designed for AI risk, security, and compliance. One emerging option is MAESTRO (Multi-Agent Environment, Security, Threat, Risk and Outcome), a threat modeling framework tailored for agentic AI systems. It offers a structured approach for analyzing vulnerabilities across various components like foundation model, data operations, and deployment, identifying cross-layer attack vectors.

  Beyond MAESTRO, agencies should explore AI governance platforms that support policy creation, compliance enforcement, drift detection, audit trails, and integration with existing IT systems to reduce manual work and enhance consistency. This link offers a "Best of" list of AI Governance Platforms in 2025.

  By grounding compliance and monitoring in tools rather than relying solely on manual processes, agencies can achieve repeatability, traceability, and scalable oversight across various AI projects.

- **Implement Monitoring Systems**

  Monitoring should be viewed as ongoing assurance, not just periodic checks. Agencies must develop real-time monitoring systems that track performance, bias, and security throughout the AI lifecycle. This involves setting up telemetry dashboards linked to mission outcomes and risk indicators like fairness, drift, anomaly detection, or system reliability thresholds.

  Practically, agencies can enhance existing cybersecurity and IT monitoring systems by adding AI-specific metrics rather than creating separate tools. For instance, integrating AI performance data into a Security Operations Center (SOC) feed or enterprise risk dashboard can provide immediate visibility without redundancy. Automated alerts can trigger targeted reviews, ensuring issues such as ethical or technical concerns are caught and addressed properly.

  The goal is not constant oversight but balanced vigilance: a feedback loop that builds public trust, informs model improvements, and reinforces accountability as AI systems evolve within real-world missions.

6.2 Long-term Strategy – Sustained Momentum

As artificial intelligence continues to transform operations, agencies must avoid viewing compliance and governance as one-off tasks. True AI maturity involves institutionalizing regular reflection and renewal, routinely reviewing policies, assessing technologies, and engaging stakeholders to ensure governance stays in line with evolving missions and societal expectations. These efforts turn AI governance from a reactive stance into a continual cycle of improvement and trust building.

Three key approaches can help ensure that your agency's AI regulatory stance will be strong in the long run.

1. **Regular Policy Review**
   AI governance frameworks must be dynamic, revisited at least annually or whenever new legislation, executive guidance, or mission changes occur. A structured policy review cadence ensures that lessons learned (from audits, pilot projects, and incidents) are captured and acted upon. Agencies should convene multidisciplinary review boards, bringing together policy, legal, technical, and operational leaders to assess whether current policies still reflect the agency's risk appetite and public accountability standards. Treating policy review as an ongoing governance ritual rather than a compliance event fosters institutional agility and prevents outdated rules from constraining innovation.

2. **Technology Assessment**
   Because AI technologies evolve faster than procurement cycles, agencies must establish formal mechanisms to evaluate both the capabilities and risks of emerging tools. Regular technology assessments (conducted through structured pilot and sandbox testing mechanisms) allow agencies to explore innovation without operational disruption. These reviews should consider not only performance and security, but also explainability, interoperability, and ethical implications. Embedding assessment checkpoints into acquisition and deployment pipelines ensures that technology decisions remain evidence-based and mission aligned rather than trend driven.

3. **Stakeholder Engagement**
   AI systems serving the public must be guided by trust, transparency, and inclusivity. Ongoing stakeholder engagement within agencies across the federal ecosystem and with the public anchors AI development and real-world needs and expectations. Agencies should maintain dialogue with employees, oversight bodies, academic partners, and affected communities to gather feedback and identify emerging risks early. Establishing advisory boards, open consultation sessions, or transparency reports can transform engagement from a procedural step into a source of legitimacy and shared ownership. The more openly agencies communicate about how and why AI is used, the stronger the foundation for enduring public confidence.

Conclusion

Success in the AI era will not come from chasing every new capability, nor from retreating behind rigid compliance. It will come from learning to live in the tension between cautious urgency and deliberate governance, moving forward quickly enough to stay relevant, yet carefully enough to remain trustworthy.

No one in government has this fully figured out. Every agency leader and practitioner is building the plane while flying it, trying to reconcile mission demands with emerging regulations, evolving risks and public expectations. That's not a failure of preparation; **it's the nature of this moment.**

AI is reshaping more than technology. It is reshaping how we think, organize, and decide. It challenges longstanding habits of stability and hierarchy, asking institutions designed for predictability to adapt to constant change. The good news is that government already knows how to operate under pressure, balancing competing priorities and act in the public interest. Those instincts are exactly what this next chapter requires.

The goal isn't perfection; it's progress with integrity. By working together (sharing lessons, testing new approaches, and keeping transparency at the core), leaders can guide their organizations through uncertainty with confidence and care. This paper is one small example of the knowledge sharing and collaboration that will help us all succeed. None of us have all the answers, but by staying curious, grounded, and committed to the public trust, we can ensure that AI strengthens, rather than erodes, the values that define public service.

We may not control the pace of change, but we can choose how we meet it—together, with purpose and humility.

Glossary of Terms

5G - Fifth-generation wireless technology that is transforming internet infrastructure and fueling advances across sectors, with cloud-native architecture well-suited for AI integration.

6G - Sixth-generation wireless technology expected to be primarily AI-based, representing the next evolution in telecommunications infrastructure.

A/B Testing - A controlled experimental method that compares two versions of a system to determine which performs better, used to validate metric improvements in AI systems.

Adaptive Governance Specialists - Staff who specialize in building and maintaining modular, flexible governance frameworks rather than rigid bureaucratic structures, focusing on creating living compliance systems.

Advisory Boards - Formal groups established to provide ongoing stakeholder engagement and feedback on AI system development and deployment.

Agentic AI - Autonomous AI systems that exhibit goal-driven adaptability and behavior, capable of making decisions and taking actions without human intervention, unlike traditional AI models that operate within predefined guardrails.

AI Centers of Excellence (CoE) - Cross-functional organizational units that provide expertise, best practices, and coordination for AI initiatives across an agency.

AI Governance Platforms - Specialized tools that support policy definition, compliance enforcement, drift detection, audit trails, and integration with existing IT systems.

AI SBOM/BOM (Software Bill of Materials) - Documentation capturing AI system components including models, weights, datasets, libraries, APIs, license terms, and update cadence for supply chain risk assessment.

AI-Literate Mission Specialists - Domain experts in core agency missions who also understand AI capabilities and limitations, preventing AI initiatives from becoming isolated technical projects.

Algorithmic Complexity - The computational resources required to solve problems, which can be substantially reduced through hybrid AI and quantum computing approaches.

Algorithmic Impact Assessment (AIA) - A comprehensive evaluation process that analyzes the potential benefits and harms of AI systems, including stakeholder mapping, use-case analysis, impact estimation, alternatives analysis, and governance controls.

America's AI Action Plan - A federal policy framework released in July 2025 that signals a "pro-innovation," deregulatory approach to federal AI policy with substantial implications for government agencies.

Anomaly Detection - Real-time monitoring systems that identify suspicious activity such as excessive query rates or abnormal usage patterns in AI systems.

Authority to Operate (ATO) - A formal authorization process within the federal environment that evaluates AI systems for IT security, ethical, legal, and mission-specific risks before deployment.

Automation Bias - The tendency for humans to over-rely on automated systems, potentially leading to reduced vigilance and critical thinking in human oversight roles.

Bayesian Updating - A statistical method for refining acceptance criteria and thresholds as more data becomes available over time.

California SB53 - State-level legislation representing evolving AI regulatory requirements that agencies must adapt to incrementally.

Cautious Urgency - A strategic posture that embraces AI innovation while maintaining ethics, oversight, and public trust; balancing rapid advancement with responsible implementation.

CDAO (Chief Digital and AI Office) - The Department of Defense office responsible for AI strategy and implementation, including the Data Mesh Reference Architecture.

CISA - Cybersecurity and Infrastructure Security Agency, responsible for federal incident response coordination and cybersecurity directives.

Cloud-Native 5G Architecture - A software-focused approach to 5G implementation that leverages cloud computing principles and is well-suited for AI integration.

Compliance Playbook - A lightweight but authoritative reference that defines documentation standards, including model cards, algorithmic impact assessments, and privacy impact assessments.

Continuous Authorization - An ongoing approval model for AI systems that includes risk reviews at each stage of the AI lifecycle rather than a single point-in-time authorization.

Continuous Learning Facilitators - Staff focused on institutionalizing knowledge transfer and adaptive practices to ensure organizational learning from AI deployments.

Control Matrices - Frameworks that map specific requirements to different impact levels, allowing agencies to apply proportionate controls without over-engineering solutions.

Controlled Innovation Sandboxes - Secure environments where teams can test AI tools under real-world conditions within predefined ethical, legal, and security boundaries.

Cross-border Implications - The regulatory and operational challenges faced by government agencies operating in multi-jurisdictional environments, requiring systems designed for interoperability, legal defensibility, and ethical resilience across different countries' regulations.

Cross-Functional Integration Teams - Small, agile teams that coordinate across organizational silos, combining mission domain experts, AI technologists, legal counsel, and security professionals.

Cybersecurity Framework (CSF) for AI - A specialized profile of cybersecurity controls adapted for AI systems and applications.

Data Cards - Documentation detailing data lineage, provenance, and characteristics to support transparency and audit readiness.

Data Drift - Changes in the statistical properties of input data over time that can affect AI system performance and require monitoring and response.

Data Mesh/Domain-Centric Architecture - A decentralized approach where data ownership is structured by mission domain, with each team curating and sharing data through standardized APIs and metadata schemas.

Data Poisoning - Malicious inputs designed to corrupt model training and compromise AI system integrity.

Differential Privacy - A mathematical framework that provides quantifiable privacy guarantees by adding controlled noise to data or query results.

DoD RMF - Department of Defense Risk Management Framework, a sector-specific regulatory overlay for defense applications.

Edge Intelligence/Edge AI - The architectural shift from centralized cloud-based AI to distributed computing where AI capabilities are deployed to endpoint devices, vehicles, sensors, and local infrastructure for real-time processing.

Enterprise Risk Management - The integration of AI risk management into existing organizational risk management practices and governance structures.

EU AI Act - The European Union's comprehensive legal framework for artificial intelligence that classifies AI systems by risk level and imposes strict regulatory obligations.

Expected Calibration Error (ECE) - A metric that measures how well an AI system's confidence scores align with its actual performance by comparing predicted versus actual accuracy across confidence ranges.

Explainability - The ability to clearly describe how an AI system reaches its outputs or decisions in plain language that analysts, auditors, and decision-makers can understand and trust.

Fairness - The principle ensuring AI systems don't perpetuate or amplify bias, particularly in areas with real-world impacts such as hiring, policing, or access to healthcare and benefits.

FedRAMP - Federal Risk and Authorization Management Program, a government-wide program providing a standardized approach to security assessment and authorization for cloud products and services.

FISMA - Federal Information Security Management Act, legislation that defines a framework for protecting government information and information systems.

GAO - Government Accountability Office, a legislative branch agency that provides auditing, evaluation, and investigative services for Congress.

Generalizable Robotic Skill Abstraction - Research focused on developing transferable capabilities that can be applied across different robotic platforms and applications.

Hardware Security Modules (HSM) - Physical computing devices that safeguard and manage digital keys and provide hardware-based security at edge endpoints.

Hazard Mitigation - Applications of AI and robotics systems to identify, assess, and respond to dangerous conditions or threats.

HIPAA - Health Insurance Portability and Accountability Act, providing sector-specific regulatory requirements for health data protection.

Human-in-the-Loop (HITL) - An approach that maintains human oversight and intervention capabilities in AI systems, including defined decision points, override capabilities, and time-to-intervention requirements.

Humanoid Platforms - Advanced robotic systems designed to mimic human form and capabilities for complex task execution.

Hybrid AI-Quantum Approach - Integration of classical AI methods with quantum computing algorithms to leverage the strengths of both technologies.

Hybrid Technical-Policy Professionals - Personnel who can translate between technical AI concepts and policy requirements, serving as bridges between engineering teams and compliance officers.

ISO 42001 - International standard for AI management systems that provides requirements for establishing, implementing, maintaining, and continually improving AI management systems.

Jailbreaking - Techniques used to bypass AI system safety measures and constraints to elicit prohibited or harmful outputs.

Key Performance Indicators (KPIs) - Metrics that measure the success and effectiveness of AI systems in achieving their intended objectives.

Key Risk Indicators (KRIs) - Metrics that provide early warning signals of potential problems or risks in AI system performance.

Large Language Models (LLMs) - Advanced AI systems capable of understanding and generating human-like text, representing a significant advancement from narrow AI applications toward more general and powerful systems.

Living Crosswalks - Dynamic mappings between evolving laws and agency policies that are continuously updated to maintain regulatory alignment.

MAESTRO (Multi-Agent Environment, Security, Threat, Risk and Outcome) - A threat modeling framework tailored for agentic AI systems that provides layered architecture analysis for vulnerabilities.

Model Cards - Documentation that captures an AI system's purpose, limitations, and provenance as part of system profiling and risk assessment.

Model Drift - Performance degradation in AI systems over time due to changes in data patterns or environmental conditions.

Model Inversion - An attack technique that attempts to reconstruct training data or extract sensitive information from AI models.

Modular Governance - A flexible approach where compliance artifacts, documentation, and control libraries can be updated incrementally as standards evolve.

Multi-Arm Robot - Robotic systems with multiple manipulator arms designed for complex task execution and research applications.

NIST 800-53 - A publication by the National Institute of Standards and Technology that provides a catalog of security and privacy controls for federal information systems.

NIST AI Risk Management Framework (AI RMF) - The National Institute of Standards and Technology's framework providing structured guidance for identifying, assessing, and mitigating AI-specific risks while fostering trust in AI implementation.

NIST AI RMF Community of Interest - An interagency working group focused on AI risk management framework implementation and best practices.

NIST SP 800-218 - NIST Special Publication providing secure software development framework guidance.

NIST SP 800-218A - NIST Special Publication specifically addressing machine learning and decision tree AI security considerations.

NIST-AI-600-1 - AI RMF Generative AI Profile providing specific guidance for generative AI systems.

OMB - Office of Management and Budget, responsible for overseeing federal agency operations and regulatory compliance.

OMB Memorandum M-25-21 - Federal guidance emphasizing the importance of AI governance structures and senior official designation for AI strategy leadership.

Open Radio Access Networks (O-RAN) - 5G innovation that enables AI integration in radio access networks through virtualization and standardization.

Operational Flexibility - The ability to balance precision with adaptability, preserving mission integrity while staying responsive to technological change through controlled experimentation.

Optimization Tasks - Computational problems involving finding the best solution from available alternatives, where quantum computing can provide efficiency advantages.

Overfitting - A modeling error where AI systems perform well on training data but poorly on new, unseen data.

Packet Processing - Network data handling that is expected to become AI-dominated in 5G and 6G telecommunications systems.

Phased ATOs - Risk-based authorization models that allow for iterative system improvement through staged approvals without halting mission execution.

Plan of Action & Milestones (POA&M) - A document that identifies tasks needing to be accomplished to correct deficiencies noted during security control assessments.

Privacy Impact Assessment (PIA) - A required evaluation for systems that collect, process, or generate personally identifiable information (PII) or sensitive data.

Privacy Threshold Assessment (PTA) - An initial screening to determine whether a full Privacy Impact Assessment is required.

Probabilistic Tasks - Computational problems involving uncertainty and probability calculations where quantum computing can provide advantages.

Prompt Injection - An attack technique where malicious inputs are crafted to manipulate AI systems into producing unintended or harmful outputs.

Quantum Computing (QC) - An emerging computational technology that uses quantum mechanical phenomena to process information, with potential to dramatically increase computational capabilities.

RACI Matrix - A responsibility assignment chart that clarifies roles (Responsible, Accountable, Consulted, Informed) for AI governance activities.

Radio Access Network (RAN) - The part of telecommunications infrastructure that connects individual devices to the core network, increasingly incorporating AI capabilities.

Red Teaming - Systematic testing of AI systems using adversarial approaches to identify vulnerabilities, biases, and potential failure modes.

Regulatory Evolution Monitoring - The continuous function of tracking and analyzing changes in AI regulations, standards, and policy developments to anticipate rather than react to compliance obligations.

Regulatory Horizon Scanning - The systematic monitoring of emerging regulatory trends and developments to anticipate future compliance requirements.

Regulatory Intelligence - The continuous function of monitoring and analyzing evolving AI regulations, standards, and policy developments to anticipate changes rather than react to them.

Regulatory Intelligence Analysts - Personnel dedicated to continuous monitoring of the evolving AI regulatory landscape, capable of anticipating changes and translating emerging requirements into actionable organizational adjustments.

Regulatory Sandbox - A controlled environment that allows agencies to test AI innovations under relaxed regulatory constraints while maintaining oversight and learning opportunities.

Residual Risk - The level of risk that remains after security controls and mitigation measures have been implemented.

Robotic Skill Abstraction - The development of transferable capabilities that allow robots to apply learned skills across different tasks and environments.

Safety - The principle focusing on ensuring AI systems are reliable, secure, and resilient against errors, adversarial attacks, or misuse, particularly critical for government applications involving critical infrastructure or sensitive data.

Security Assessment Plan/Report (SAP/SAR) - Documentation outlining evaluation methods and findings for security assessments of information systems.

Security Operations Center (SOC) - A centralized facility for monitoring and managing cybersecurity, which can be enhanced with AI-specific metrics and monitoring.

Shadow Models - Alternative AI models run in parallel to production systems for comparison and monitoring purposes.

Shared Risk Registers - Centralized coordination vehicles that track cyber, AI, data, and program risks across domains, serving as communication backbones between IT, mission, and governance functions.

Statistical Process Control Charts - Tools used to monitor AI system performance over time and identify when metrics deviate from expected ranges.

Strategic Adaptation - The disciplined ability to pivot policies, workflows, and investments in response to new risks, technologies, or mandates without derailing mission continuity.

System Security Plan (SSP) - A document that details system boundaries, components, and security controls for federal information systems.

Technology Assessment - Formal mechanisms to evaluate both capabilities and risks of emerging AI tools through structured pilot and sandbox testing.

Telemetry Dashboards - Real-time monitoring systems that track performance, bias, and security posture across the AI lifecycle.

Transparency - The principle of making AI use visible and understandable to both internal stakeholders and the public, including disclosure of when and how AI is applied, what data it uses, and what limitations exist.

Transparency in Frontier Artificial Intelligence Act (TFAIA) - California state legislation focused on AI safety, representing an example of the evolving multi-layered regulatory landscape.

Transparency Reports - Public-facing documents that communicate how and why AI systems are used by government agencies.

Trusted Execution Environment (TEE) - Secure areas within processors that provide hardware-based security for sensitive data processing and code execution.

Trustworthy AI - AI systems that adhere to principles of fairness, transparency, accountability, and robustness, designed to be reliable, defensible, and worthy of public trust.

Underfitting - A modeling error where AI systems fail to capture underlying patterns in data, resulting in poor performance on both training and new data.

USAi Platform - GSA's platform that serves as an example of standardized documentation and template libraries for AI governance across government agencies.

Virtualization - The creation of virtual versions of computing resources, a key component of O-RAN architecture enabling AI integration.

Zero Trust Architecture (ZTA) - A security framework that requires verification for every user and device trying to access systems, regardless of their location or previous access history.