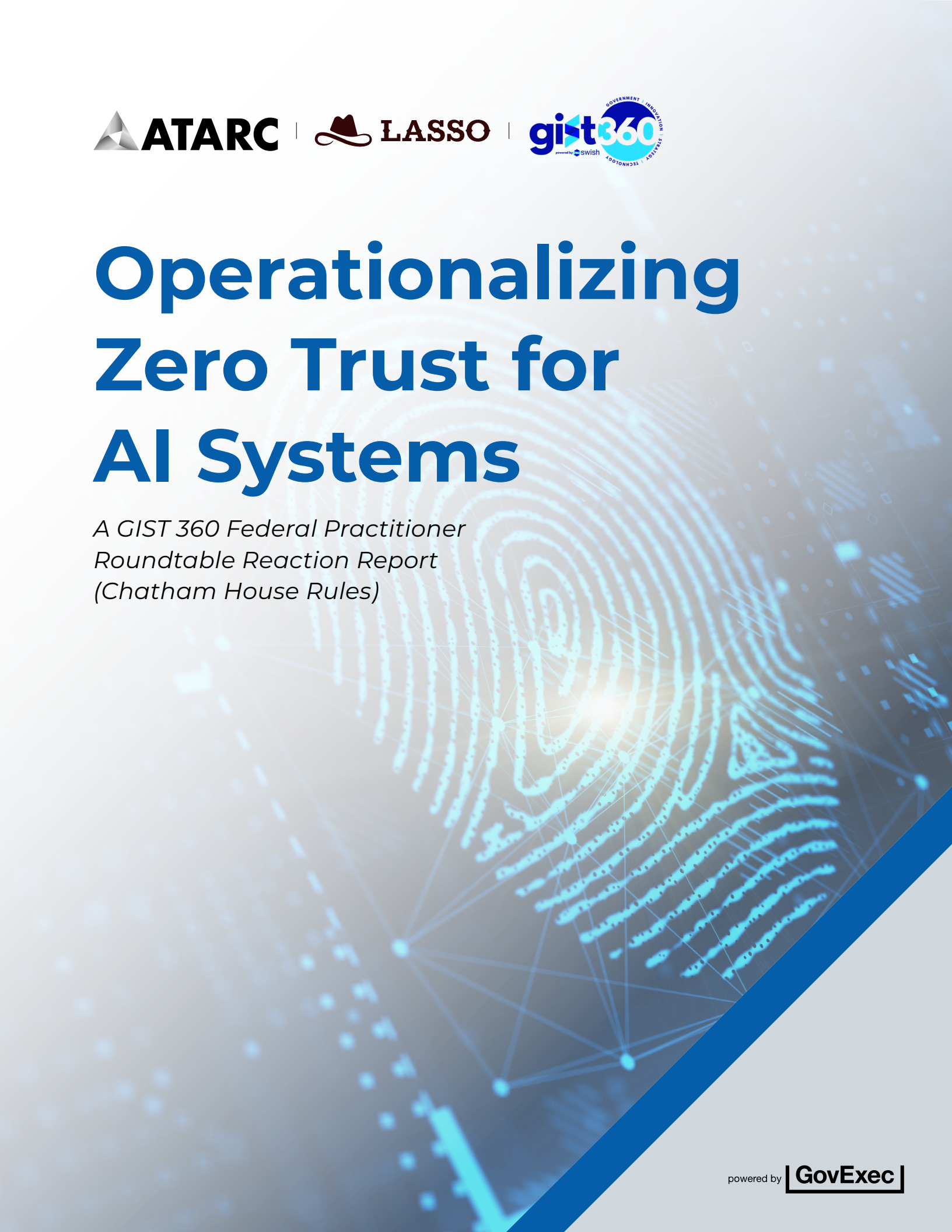
ATARC | LASSO | GIST 360

Operationalizing Zero Trust for AI Systems

A GIST 360 Federal Practitioner Roundtable Reaction Report (Chatham House Rules)



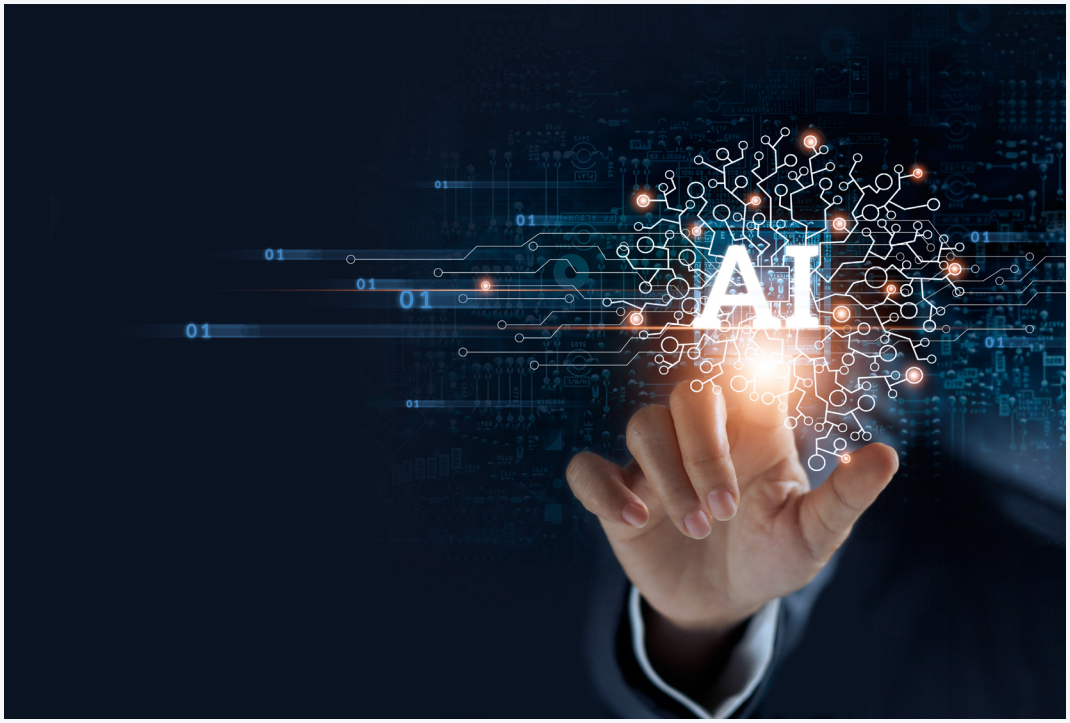
powered by GovExec

Executive Summary

Artificial intelligence—particularly generative and agentic AI—is rapidly reshaping how federal agencies operate, deliver services, and execute mission objectives. Across civilian, defense, intelligence, and research environments, agencies are moving from experimentation to scaled adoption. This acceleration is introducing meaningful gains in efficiency, responsiveness, and analytical capability. At the same time, it is fundamentally altering the cybersecurity landscape.

This whitepaper synthesizes insights from a not-for-attribution federal roundtable involving senior leaders and practitioners. A central conclusion emerges: the nature of risk is shifting. Cybersecurity is no longer limited to protecting data and systems. It now extends to governing the behavior of autonomous and semi-autonomous systems operating at machine speed.

Agencies are confronting a convergence of challenges—data quality, scaling agent platforms, workforce readiness, and evolving threat vectors—while simultaneously being pushed to adopt AI quickly. The path forward requires integrating security into the foundation of AI adoption, not treating it as a downstream control. Those that succeed will unlock mission advantage; those that do not will face new forms of operational and reputational risk.



1. Introduction: A New Operational Paradigm

Federal agencies are entering a period in which artificial intelligence is embedded across mission delivery, enterprise IT, and operational environments. Unlike prior waves of digital transformation, AI systems are not simply tools that execute predefined logic. They generate outputs, influence decisions, and increasingly take action within systems.

As reflected in the roundtable discussion, agencies are applying AI across a wide range of use cases, including scientific research acceleration, financial and agricultural analytics, law enforcement support, citizen-facing services, and military operations. This breadth of application highlights both the opportunity and the complexity of adoption.

Traditional cybersecurity approaches were designed for deterministic systems and well-defined boundaries. AI systems, by contrast, are non-deterministic where context and intent matter immensely.. They introduce ambiguity in how decisions are made and how outcomes are produced. This shift requires a rethinking of security from the ground up.

2. The Evolving Threat Landscape

The introduction of AI into federal environments is changing not just the scale of risk, but its nature. Historically, cybersecurity focused on protecting data—ensuring confidentiality, integrity, and availability. With AI, the focus expands to include what systems do with that data.

Agentic systems can access multiple systems, initiate workflows, and generate outputs that directly influence mission outcomes. As one participant noted, the risk is moving from what systems expose to what systems actually do. This introduces new concerns, including unintended actions, behavioral drift, and misuse of integrated capabilities.

The attack surface is expanding accordingly. Threat vectors now include manipulation of inputs, corruption of training data, exploitation of model behavior, introduction of unvetted tools into secure environments, and excessive agency to execute actions without a human in the loop. The accessibility of AI tooling has created a modern form of shadow IT, where developers and users can rapidly deploy capabilities without full awareness of their implications.

This shift demands that agencies move beyond perimeter-based defenses and toward continuous observability and governance of system behavior and intent.

3. Data as the Critical Dependency

Data remains the foundation of all AI systems, and its quality directly determines the reliability of outcomes. Across agencies, concerns about data integrity, provenance, and usability were consistently raised.

Federal data environments are characterized by scale and complexity. Large volumes of information are stored across disparate systems, often with inconsistent structures and varying levels of trust. Efforts to normalize and standardize this data are ongoing but incomplete.

The implications for AI are significant. Poor-quality data does not simply produce inaccurate outputs; it can propagate errors at scale and influence downstream decisions. In mission-critical environments—such as financial systems, intelligence analysis, and law enforcement—this risk is amplified.

Agencies are also grappling with gaps in data classification and tagging. Without consistent metadata and governance, it becomes difficult to control access, enforce policy, and ensure that sensitive information is handled appropriately within AI systems.

4. Deployment Realities: Cloud, Edge, and Offline Environments

While commercial AI innovation is largely cloud-driven, federal environments often operate under very different constraints. Many missions require operation in disconnected, low-bandwidth, or classified environments where reliance on external services is not feasible.

As a result, agencies are investing in offline and on-premises AI capabilities. These include air-gapped environments, locally hosted models, and specialized hardware configurations designed to operate without external connectivity. This approach enhances control over data and reduces exposure to external threats, but it also introduces challenges related to scalability, operations, and tooling.

Edge computing is becoming increasingly important, particularly in defense and intelligence contexts. Processing data at the edge enables real-time decision-making and supports operations in contested or bandwidth-constrained environments. However, securing these deployments requires rigorous controls on both inputs and outputs, as well as mechanisms to validate and govern model behavior.

5. Integrating Zero Trust with AI

Agencies are increasingly applying Zero Trust principles to AI systems, recognizing that AI-enabled applications, models, and agents must be continuously authenticated, monitored, and validated like any other critical enterprise asset. This includes enforcing strong identity and access management, least-privilege access, segmentation of AI workloads, data governance controls, and continuous monitoring of interactions between users, models, APIs, and downstream systems. As AI capabilities become more deeply integrated into mission environments, agencies are also beginning to treat AI pipelines, training data, and inference processes as part of the broader attack surface requiring ongoing assessment and protection.

At the same time, AI introduces a new class of security risks that extend beyond traditional Zero Trust architectures. Organizations must account for threats such as prompt injection, data poisoning, insecure plugin or agent behavior, model theft, sensitive data leakage, hallucinations, and adversarial manipulation of outputs. To address these risks, agencies are incorporating AI-specific security frameworks and testing methodologies into their cybersecurity strategies, including resources such as the OWASP Top 10 for LLM Applications, MITRE ATLAS, and emerging federal guidance including AI security control frameworks like AIUC-1. Security teams are also expanding red teaming and adversarial testing efforts to evaluate how AI systems behave under unexpected or malicious conditions, while implementing monitoring capabilities to detect model drift, anomalous behavior, unauthorized data access, and output manipulation over time.

Ultimately, the goal is not only to secure access to AI systems, but to establish confidence that those systems operate safely, reliably, and transparently in support of mission objectives. This requires agencies to continuously validate both the security posture and operational behavior of AI systems throughout their lifecycle, from development and deployment to ongoing monitoring and governance.

6. Governance, Policy, and Organizational Culture

A recurring theme in the discussion was the importance of building security into AI from the outset rather than attempting to layer it on after deployment. This requires close collaboration between organizational leaders such as the CISO, CAIO, CIO, and mission owners, as well as day-to-day coordination between developers, data scientists, security teams, and mission stakeholders to ensure AI and Zero Trust controls are implemented with the proper intent and continuously adapted as operational needs, technologies, and threats evolve.

Policy frameworks are evolving, but often lag behind the pace of technological change. Agencies are working to define acceptable use, establish guardrails, and clarify responsibilities. However, policy alone is insufficient without corresponding cultural change.

Organizations must foster an environment in which innovation is encouraged but bounded by clear expectations. This includes reducing reliance on unsanctioned tools, increasing awareness of risks, and embedding security considerations into everyday workflows.

7. Workforce Dynamics and Readiness

The workforce dimension of AI adoption is both an enabler and a risk factor. Participants noted a growing perception among employees that failing to adopt AI could negatively impact their roles, creating pressure to use AI-enabled tools quickly, sometimes without fully understanding the associated security, privacy, or operational implications.

Agencies are responding by increasing investments in workforce training, education, and AI readiness initiatives. This includes foundational AI literacy programs designed to help employees understand the capabilities, limitations, and risks of AI systems, as well as more advanced training in areas such as prompt engineering, model evaluation, data governance, AI security, human oversight, and responsible AI implementation. Participants emphasized that training efforts must extend beyond technical teams to include leadership, acquisition personnel, legal and privacy offices, and mission operators who will ultimately interact with or oversee AI-enabled systems.

Many organizations are also leveraging commercially available and government-supported learning resources to accelerate workforce readiness, including platforms such as O'Reilly Learning Platform, federal AI training initiatives, NIST guidance, and role-specific cybersecurity and AI coursework. Participants noted that ongoing education will be necessary as AI technologies and associated risks continue to evolve.

Maintaining a human-in-the-loop remains essential. Even as AI systems become more capable and autonomous, accountability for decisions and outcomes must remain with people. This requires transparency into how systems operate, clear governance over how outputs are used in decision-making, and the ability to trace recommendations, actions, and data sources back to their origin when needed.

8. Agentic AI and the Governance Challenge

Agentic AI represents a significant evolution in capability. These systems are not limited to generating content; they can take actions, interact with multiple systems, and operate with a degree of autonomy.

This increases both the potential value and the potential risk. Governing these systems requires new approaches that focus on intent and behavior rather than static rules. Agencies are beginning to explore methods for monitoring how systems reason, detecting deviations from expected patterns, and constraining actions within defined boundaries.

This area remains nascent, but it is likely to become a central focus of federal cybersecurity efforts in the coming years.

9. Strategic Considerations for Federal Agencies

Agencies should approach AI adoption as a transformation that spans technology, policy, and people. Security must be integrated from the beginning, with particular attention to data governance, deployment models, and workforce readiness. There is a need to expand existing frameworks, such as Zero Trust, to address AI-specific risks, while also developing new capabilities to monitor and govern system behavior.

Equally important is controlling the introduction of tools and components into federal environments. The rapid proliferation of AI capabilities increases the risk of unvetted software being used in sensitive contexts. Establishing clear controls and oversight mechanisms is critical.

10. Conclusion

Artificial intelligence is redefining how the federal government operates. It offers significant opportunities to improve mission outcomes, but it also introduces new and complex risks. The insights from this roundtable highlight that the challenge is not simply technical. It is organizational, cultural, and strategic.

The agencies that succeed will be those that treat AI security as a foundational element of adoption rather than an afterthought. By focusing on data quality, system behavior, workforce readiness, and integrated governance, the federal government can harness the benefits of AI while maintaining trust, accountability, and mission integrity.

This reaction report is based on a not-for-attribution federal government roundtable discussion and reflects aggregated insights rather than individual viewpoints.