Establishing an Agentic AI Cybersecurity & Defense Organization in Department Of War

March 7, 2026

Amazon Web Services, Inc.
410 Terry Avenue North
Seattle, WA 98109-5210

Michael Medgyessy
Head of Cloud Innovation, DoW
Amazon Web Services
medgyess@amazon.com

Table of Contents

Executive Summary

This document intends to guide DoW Chief AI Officers, their staff and security teams to build and create successful sustainable production Agentic AI capabilities with governance, security enabled by the latest in automation and education.

The modern DoW organization must prioritize continuous learning, adaptability, collaboration, and the integration of AI-specific controls across the entire system lifecycle. Success depends on three critical elements: strategic foundations and governance, specialized personnel and expertise, and technical architecture with external enablement capabilities. Most new Information Technology capabilities are shifting from being built and deployed from the IT departments to outside the IT departments. The ability to have access to powerful Agentic AI partners enables non-technical end-users the ability to have access to the most powerful infrastructure and services through the cloud and create on-demand, tailormade solutions which are also easier to rebuild than sustain. As the human machine interface shifts from apps to AI interfaces to access data and accomplish tasks the security landscape drastically shifts. This also adds new focus from traditional software development lifecycle management to one for AI Operations and Agentic AI lifecycle management. Interestingly, this trend leaves the number of applications which remain as technical debt in the face of AI superseding many productivity style user-facing applications.

AWS has provided the ability to govern and secure agentic AI with Amazon Bedrock AgentCore. This comprehensive service provides enterprise-grade infrastructure for deploying and operating AI agents at scale, with comprehensive security, identity management, observability, and evaluation capabilities essential for production multi-agent systems.

For Department of War a key consideration is the availability of some of these commercially available services in the Amazon Dedicated Cloud regions and GovCloud. This document in conjunction with the partner document on securing multi-agentic AI systems will help DoW prioritize the services for AWS to bring to those regions and provide accreditation priority for availability to the workforce.

Introduction: The AI Security Imperative

Agentic AI Authorization introduces fundamental changes in how traditional security responsibilities are implemented and how stakeholders interact, demanding a highly integrated and collaborative approach. Unlike traditional software systems, AI systems are dynamic, continuously learning, and can exhibit emergent behaviors that require new security paradigms. Your leadership role requires you to frame existing AI guidance into a manageable way forward while ensuring secure deployment of advanced technologies. This involves setting risk tolerance, establishing strategy, and ultimately granting Authority to Operate (ATO)—often with conditions due to AI's evolving nature.

Prioritized Implementation Roadmap

The following roadmap is organized into three delegable Lines of Effort (LOEs), each with clear priorities and ownership.

Line of Effort 1: Strategic Foundations & Governance
Owner: Director/Authorizing official (AO)

Priority 1: Mandate Zero Trust & Continuous Authorization

Formally transition the security paradigm from network-centric to data-centric Zero Trust (ZT) across all DoW seven pillars: User, Device, Applications, Data, Networking, Automation & Orchestration, and Visibility & Analytics. Security cannot rely on perimeter defenses; all hosts must be treated as though they are compromised and hostile.

Mandate Continuous Authorization (event-driven risk assessments) over periodic approvals to match AI's dynamic nature. For continuously evolving AI systems, risk assessments should be performed continuously, with reauthorization triggered by significant changes such as major model updates. Implement CI/CD capability and automated monitoring to bring modern agile software and AI capabilities to the command.

AI has both user interface capabilities and Application Programming Interface capabilities as a non-person entity. Agentic AI must be incorporated into the Zero Trust architecture and framework in what will soon become the primary Human Machine Interface and actor on data in the network.

AWS Implementation with ZT for AI:

AWS can help with your cloud-hosted workloads to fast-track them into a Zero Trust compliant architecture:

- **AWS IAM Identity Center:** Centralized identity management with MFA and federation capabilities
- **AWS Verified Access:** Zero trust access for applications without VPN, with continuous device posture verification
- **AWS VPC Lattice:** Zero trust service-to-service communication with micro-segmentation at Layer 7
- **AWS Network Firewall:** Deep packet inspection with TLS decryption capabilities
- **AWS PrivateLink:** Secure data transfer without traversing public internet
- **AgentCore Identity:** Integrates with existing identity providers and provides secure token vaulting with OAuth 2.0, API keys, and AWS IAM authentication

AI has both user interface capabilities and Application Programming Interface capabilities as a non-person entity. Agentic AI must be incorporated into the Zero Trust architecture and framework in what will soon become the primary Human Machine Interface and actor on data in the network.

Priority 2: Frame AI Risk Management Framework (RMF)

Apply the Risk Management Framework (RMF) tailored for AI systems, using NIST AI Risk Management Framework (NIST AI 100-1), the AI RMF Playbook, and the Generative Artificial Intelligence Profile (NIST AI 600-1), Cyber AI Profile (NIST IR 8596) as core resources.

As Director/AO, you must set the organization's risk tolerance and define the path for granting conditional Authorities to Operate (ATOs) for AI systems. Use the AI RMF Playbook as a core resource for implementation. Balancing mission risk incurred by not progressing capability due to security is as important as enforcing security on the capabilities. Focusing on what is the best that can be done given time and resources versus security at all costs ensures a "Yes If" culture which can continue to move capability quickly in a measured manner.

Priority 3: Establish GRC/Ethical Oversight

Appoint a Governance, Risk, and Compliance (GRC) / Ethical Advisor. Task this role with establishing clear ethical guidelines, ensuring initial legal and regulatory alignment, and creating documentation and audit trails for all AI models. This role is vital for maintaining thorough documentation for long-term compliance and facilitating interdisciplinary collaboration across legal, ethical, and technical teams.

AWS Implementation:
Amazon Bedrock AgentCore Observability provides comprehensive audit trails with Open Telemetry-compatible traces integrated with Amazon CloudWatch. AgentCore Policy (Preview) enables deterministic policy enforcement with fine-grained governance and complete audit trails for compliance requirements.

Especially in Department of War warfighting activities supported by agentic AI systems responsible AI guardrails are of utmost importance. Through AWS Bedrock the systems cannot access various foundation models directly via API, but through this access gateway external guardrails can be uniformly applied and curated across all removing security fears of individual model weaknesses. Guardrails easily applied like automated reasoning ensure mathematically that LLM outputs are not hallucinations and can be trusted. Responsible AI guardrails ensure uniformly the governance policies work the same across all vendor frontier models of choice through AWS. This ultimately raises confidence in what these agents can impact at higher risk levels.

Line of Effort 2: Personnel & Expertise
Owner: AI Security Lead

Priority 1: Recruit Core AI/Cybersecurity Expertise

Hire or upskill the AI Security Leadership and the AI/ML Scientist. This Leader will immediately begin tailoring the Risk Management Framework (RMF) proposals to the Authorizing Official, while the Scientist ensures security controls are integrated during model development. Personnel should have experience comparable to those in Computer Science, Artificial Intelligence, Data Science, Machine Learning, and Cybersecurity. There are a lot of new NIST and government regulation continuing to try to keep current which will require constant vigilance in updating policy and processes in conjunction with what technology is making possible.

AWS Training Resources:
Leverage AWS Skill Builder course "Securing Generative AI on AWS" and AWS Certified AI Practitioner certification to upskill personnel on AgentCore security capabilities.

Priority 2: Establish Adversarial Robustness Team

Recruit an Adversarial Robustness Specialist. Task this role with immediately focusing on AI-specific penetration testing, utilizing frameworks like MITRE ATLAS™ and tools like the Adversarial Robustness Toolbox (ART). This specialist will test the model's responses to adversarial inputs and implement defenses such as adversarial training to make models more resilient.

Priority 3: Launch Continuous Training Programs

Delegate the establishment of an internal training program focused on AI literacy, scenario-based learning to defend against adversarial manipulation, and relevant certifications. Invest in continuous personnel education and upskilling programs to build AI literacy across the workforce. Foster collaborative partnerships with academia, cloud providers, and industry experts to infuse new talent educated on the latest technologies.

Line of Effort 3: Technical Architecture & External Enablement Owner: AI/ML Scientist & AI Security Lead

Priority 1: Build Secure Data Pipelines, Model Registry, Agent Registry

Identify where the data is that the AI models will be referencing for the desired tasks. Data security and governance remain the key to success. Ultimately security of the data is the root requirement in security, and accessibility is the root goal of capability. Garbage in, Garbage out holds true in agentic AI systems, but security through obscurity does not. Concepts like data aggregation requiring higher classification become archaic as the world's knowledge becomes related, contextualized and instantly accessible. Security of data must be further automated and matured rapidly as the agentic AI implementation is outpacing governance and security.

AWS Services which are key to this Implementation include:

- **AWS KMS:** FIPS 140-3 Security Level 3 validated encryption for data at rest with automated key rotation and cross-region replication
- **AWS Glue:** Secure ETL pipelines with encryption at rest and in transit for data transformation **AWS Lake Formation:** Data lineage and governance to track provenance and transformations of datasets
- **Amazon SageMaker Model Registry:** Track model artifacts and create the AI Bill of Materials (AIBOM), providing visibility into the AI supply chain
- **AgentCore Runtime:** Deploy and manage agent registry with version control and deployment tracking
- **AWS DataSync:** Secure, high-performance data transfers with encryption in transit and end-to-end integrity validation
- **AWS Secrets Manager:** Encrypted credential management with automatic rotation, integrated with KMS for envelope encryption

Monitor data inputs continuously for integrity issues and signs of poisoning.

Priority 2: Implement Inference Guardrails

Deploy multi-layered inference protection, including prompt sanitization and output validation.

AWS Implementation:

- **Amazon Bedrock Guardrails:** Enforce safety policies on model inputs and outputs with content filtering, denied topics, word filters, and sensitive information redaction
- **AWS WAF:** Application layer protection with TLS-aware content filtering for web-facing agent interfaces

- **AWS Lambda:** Implement prompt sanitization and output validation logic
- **Amazon Bedrock Automated Reasoning:** Improve accuracy with mathematical checking on model outputs
- **AgentCore Gateway:** Centralized tool access control with MCP protocol support for secure, standardized tool integration
- **AgentCore Policy (Preview):** Deterministic policy enforcement to ensure agents operate within defined boundaries

For Retrieval-Augmented Generation (RAG) architectures, implement access control lists (ACLs) using IAM and Amazon OpenSearch Service in the vector retrieval system to enforce least privilege access to data. Limit agent access to only necessary systems and data sources, using execution isolation (sandboxing/containerization) for agents through AgentCore Runtime's dedicated microVM architecture.

Priority 3: Deploy Continuous Monitoring Stack

Integrate security tools for real-time visibility and automated response.

AWS Implementation:

- **Amazon Security Lake:** Centralized security data lake (SIEM) for aggregating security logs **Amazon GuardDuty:** AI/ML-powered threat detection analyzing CloudTrail logs, VPC Flow Logs, and DNS queries
- **AWS Security Hub:** Centralized security posture management with automated compliance checks against frameworks like NIST, PCI-DSS, and CIS
- **Amazon SageMaker Model Monitor:** Drift and bias detection for deployed models
- **AWS Lambda + AWS Step Functions:** Automated response workflows (SOAR equivalent) for security incident orchestration
- **AgentCore Observability:** Comprehensive tracing and monitoring with OpenTelemetry-compatible traces, integrated with Amazon CloudWatch for unified visibility
- **AgentCore Evaluations (Preview):** Automated quality assessment with continuous monitoring of agent performance, including correctness, faithfulness, helpfulness, and custom metrics
- **AWS Config:** Continuous configuration compliance monitoring with automated remediation **Amazon Macie:** Automated sensitive data discovery and classification
- **AWS CloudTrail:** Comprehensive API activity logging with immutable storage integration

Deploy real-time monitoring tools that detect anomalies such as model drift or security breaches and trigger alerts. Set up automated response workflows for incident management.

Priority 4: Develop External Enablement Resources

Package internal guidance, reference architectures, and tabletop review aids to help external partners adopt the NIST AI RMF and implement secure AI capabilities.

AWS Implementation:

- **AWS CloudFormation:** Infrastructure-as-code templates for repeatable, secure deployments
- **AWS Service Catalog:** Curated catalog of approved AI/ML services and configurations
- **AWS Well-Architected Framework:** AI/ML lens for architectural best practices

- **AgentCore Runtime:** Provide reference architectures for deploying agents with any framework (LangGraph, CrewAI, Strands Agents SDK)

Develop resources to help others simulate and evaluate their AI systems against federal security standards, guiding them through pre-authorization, initial authorization, and continuous monitoring phases.

Organizational Structure and Key Role

Your organization should be structured around a cross-functional team or working group tailored to address unique AI concerns. The following table outlines critical roles and responsibilities:

Table 1 Roles and Responsibilities of AI Office Staff

Role	Key Responsibilities
Director / Authorizing Official (AO)	Frames existing AI guidance into manageable strategy; sets risk tolerance; grants Authority to Operate (ATO) with conditions; ensures AI systems meet security requirements and operate fairly and transparently
AI Security Lead / Cybersecurity Expert	Leads implementation of Risk Management Framework (RMF) tailored for AI; handles technical steps from categorization through continuous monitoring; possesses expertise in anticipating, recognizing, and mitigating AI-specific risks including adversarial AI techniques
AI/ML Scientist or Data Scientist	Focuses on development, integrity, and performance of AI models; implements security controls during model development; conducts bias detection, model validation, and optimization; ensures models operate with high accuracy and reliability
Adversarial Robustness Specialist	Tests model responses to adversarial inputs; uses tools like Adversarial Robustness Toolbox (ART) to simulate attacks (poisoning, evasion); implements additional defenses such as adversarial training
Governance, Risk, and Compliance (GRC) / Ethical Advisor	Evaluates AI systems for potential biases and ethical concerns; aligns AI development with business objectives, regulatory requirements, and ethical standards; facilitates interdisciplinary collaboration; maintains documentation and audit trails for long-term compliance

Key Architectural Components

Data Security and Pipelines

Implement encryption at rest and in transit. Use data lineage tools to track provenance and transformations of datasets. Employ solutions to manage secure data pipelines and monitor data inputs continuously for integrity issues and signs of poisoning. Avoid data commingling, especially mixing datasets with differing security classifications.

AWS Services:

- AWS KMS (FIPS 140-3 Level 3)
- AWS CloudHSM (FIPS 140-2 Level 3, and now 140-3)
- AWS Glue
- AWS Lake Formation
- AWS DataSync
- AWS Certificate Manager
- S3 Dual-Layer Encryption (DSSE-KMS) for NSA CNSSP 15 compliance
- AWS Nitro Enclaves for isolated compute with cryptographic attestation down to custom microchip level

Model Vetting and Supply Chain

For pre-trained models (e.g., from Amazon Bedrock, Huggingface), shift focus from securing the original training pipeline to robust vetting and monitoring. Verify the source, assess training data used for privacy and fairness (Amazon Nova model cards), and audit dependencies and libraries for security vulnerabilities. Maintain an AI Bill of Materials (AIBOM) to provide visibility into the AI supply chain.

AWS Services:

- Amazon SageMaker Model Registry
- Amazon Inspector for vulnerability assessments
- AWS Artifact for compliance documentation.

Continuous Monitoring and Validation

Deploy real-time monitoring tools integrated with Security Information and Event Management (SIEM) systems. Automated systems must detect anomalies such as model drift or security breaches and trigger alerts. Ensure logs are stored in immutable storage.

AWS Services:

- Amazon Security Lake
- Amazon GuardDuty
- AWS Security Hub
- Amazon SageMaker Model Monitor
- AgentCore Observability with OpenTelemetry traces

- AgentCore Evaluations for continuous quality assessment
- AWS CloudTrail with S3 Object Lock for immutable logs

Adversarial Robustness Testing

Perform AI-specific penetration testing simulating adversarial attacks (e.g., data poisoning, input manipulation). Utilize frameworks like MITRE ATT&CK® and MITRE ATLAS™ to map potential adversarial actions and attack points. Implement automated tools for real-time incident management and automated checks.

AWS Services:

- AWS Lambda for automated testing workflows
- AWS Step Functions for orchestration
- Amazon Bedrock Guardrails for input/output filtering

Inference and Agent Security

Implement multi-layered inference guardrails (policy-enforcing filters applied to model inputs/outputs) to prevent harmful or unauthorized responses. Sanitize, validate, and filter inputs/prompts (the most common attack vector) and postprocess all model outputs prior to response.

AWS Implementation:

- **Amazon Bedrock Guardrails:** Content filtering, denied topics, word filters, sensitive information redaction
- **AgentCore Gateway:** Centralized tool access with MCP protocol support, providing unified interface to thousands of tools with granular permissions
- **AgentCore Policy (Preview):** Deterministic policy enforcement ensuring agents stay within defined boundaries
- **AgentCore Runtime:** True session isolation with dedicated microVMs (compute + memory + filesystem) preventing cross-contamination
- **AWS WAF:** Web application firewall for prompt injection protection

For Retrieval-Augmented Generation (RAG) architectures, implement access control lists (ACLs) using IAM and Amazon OpenSearch Service in the vector retrieval system to enforce least privilege access to data. Limit agent access to only necessary systems and data sources, using execution isolation (sandboxing/containerization) for agents.

Multi-Agent System Security

For multi-agent systems, implement comprehensive security across agent orchestration, communication, and coordination.
AWS Implementation:

- **AgentCore Runtime:** Host multiple agents with complete isolation using dedicated microVMs per session
- **AgentCore Gateway:** Centralized tool registry with security assessments and role-based access control (RBAC)
- **AgentCore Identity:** Agent-to-agent authentication with OAuth 2.0 and secure credential management

- **AgentCore Memory:** Tenant-isolated memory with namespace separation for multi-user environments
- **AWS Step Functions:** Multi-agent workflow coordination with state validation
- **Amazon SQS:** Secure asynchronous messaging between agents with encryption
- **AWS PrivateLink:** Secure agent-service communication without internet exposure
- **Amazon GuardDuty:** Detection of unusual API call patterns indicating compromised agents

Multi-Agent Collaboration Protocols: A2A (Agent-to-Agent): Enable inter-agent communication and coordination. MCP (Model Context Protocol): Standardized tool and resource access Encrypt and authenticate all inter-agent communications using TLS 1.3+. Implement attribute-based access control (ABAC) for fine-grained permissions. Use policy-as-code for automated governance enforcement.

Authorization Model

For continuously evolving AI systems, adopt a Continuous Authorization approach, where risk assessments are performed continuously and reauthorization is event-driven (triggered by significant changes like major model updates).

AWS Implementation:

- AgentCore Evaluations (Preview) provides continuous monitoring with automated quality gates, enabling event-driven reauthorization based on performance degradation or policy violations.

Enabling Secure AI Operations for External Partners

Adopt a mindset that facilitates secure operations for others, recognizing that the biggest risk might be not using AI effectively and securely. Ultimately, security providing mission assurance can be at risk in also slowing the adoption of Agentic AI. Mission success is enhanced with the ability to harness the power of the latest technology at rates never seen. Security is an imperative, but in balance with capability deployment rate.

Develop and Share Guidance and Frameworks

As IT capability creation proliferates down to end-users augmented with AI it is important for the Chief AI Officer and security team to focus externally through education and empowerment. Augment and compile key resources and guidance to help instill a guiding mindset for secure AI integration. Provide resources such as tabletop review aids to help others simulate and evaluate their AI systems against federal security standards. Encourage the use of voluntary standards such as the NIST AI Risk Management Framework (AI RMF). Promote the use of AI-specific controls tailored for threats like adversarial robustness testing, bias detection, and model integrity. This coupled with automated functionality ensures speed of delivery with governance and security from design to production and eventual deprecation.

AWS Resources:

- AWS Well-Architected Framework AI/ML Lens
- AWS CloudFormation templates
- AgentCore reference architectures for multi-framework support.

- AWS ProServe consultants

Address Skill Gaps through Training and Development

Invest in continuous personnel education and upskilling programs to build AI literacy across the workforce. Provide training recommendations and focus on hands-on, scenario-based learning to secure data and defend against adversarial manipulation. Foster collaborative partnerships with academia and industry experts to infuse new talent educated on the latest technologies.

AWS Training:

- AWS Skill Builder "Securing Generative AI on AWS"
- AWS Certified AI Practitioner certification
- AWS ProServe consultants help design a learning path and stay updated with available certifications, services and knowledge bases

Champion Ethical and Responsible AI Practices

Focus on the principles of responsible AI, including controllability, privacy and security, safety, fairness, veracity, explainability, transparency, and governance. Establish clear ethical guidelines for AI use in cybersecurity and regularly audit AI systems for bias and ethical concerns to mitigate unintended consequences. Promote transparency and explainability in AI decision-making processes to build trust and accountability.

AWS Implementation:

- Amazon SageMaker Clarify for bias detection
- Amazon Bedrock Guardrails for content filtering
- AgentCore Observability for complete audit trails

Promote Hybrid Approaches and Human Oversight

Advocate for a hybrid approach that leverages the strengths of both AI and human analysts Emphasize maintaining human oversight and intervention capabilities for critical decisions and cannot maintain logical consistency over time without external verification. Encourage the use of automated reasoning (formal verification) alongside generative AI outputs to verify correctness and provide logically accurate guardrails, especially for use cases where sound reasoning is critical. Automated Reasoning guardrails in AWS Bedrock utilize Formal Methods to convert LLM outputs into provable mathematical formulas which reduce accuracy and bring confidence in results to near deterministic automation levels.

AWS Implementation:

- Amazon Bedrock Automated Reasoning for mathematical verification
- AWS Step Functions for human-in-the-loop workflows

Foster Collaboration and Information Sharing

Participate in industry forums and collaborative initiatives, sharing threat intelligence and best practices. Align with established industry security projects like the Open Worldwide Application Security Project (OWASP) Gen AI Security Project (including resources like the Top 10 List for LLM Applications and the Securing Agentic Applications Guide).

Conclusion

Building a secure AI organization requires combining organizational rigor with technical agility and ethical responsibility. This approach ensures that your organization can manage the complexities of AI authorization while maximizing the tool's potential.

Amazon Bedrock AgentCore provides the enterprise-grade infrastructure needed to deploy agentic AI in production with comprehensive security, identity management, observability, and evaluation capabilities. The platform enables organizations to build, deploy, and operate AI agents at scale while maintaining the security posture required for defense and government operations

.

Think of building a secure AI organization like establishing a high-security automated bank vault. The roles (AOs, Ethical Advisors, AI Security Leads) are the combination of the chief security officer, the vault designer, and the certified auditors, all working together. The architecture is the vault itself—it relies on multiple layers of defense (Defense in Depth), using constant, real-time monitoring (Continuous Monitoring) rather than just checking once a month. Crucially, the vault must use dynamic security systems (Adaptive Security Controls) that can instantly adjust if they detect an AI-powered drill or a sophisticated forgery (an Adversarial Attack), assuming the walls might be compromised (Zero Trust). With this secure by design and continuous validation constructs like a single ATO for a prescriptive environment to be modified quickly and easily enable speed and security unlike past flat file and multi-year reassessments. Finally, helping others secure their AI is akin to sharing the engineering blueprints for safe construction and offering training on how to handle new high-tech threats, ensuring common security standards across the financial ecosystem. The reality is that the genie is out of the bottle and reigning in where and who can create agentic AI systems in the name of security is in direct opposition to mission assurance. How the workforce is empowered and enabled with strong automated guardrails and latest available services to do so delivers the outcomes desired.

Next Steps

As the Chief AI Officer, begin with Line of Effort 1 to establish strategic foundations and governance. This creates the framework for all subsequent activities. Simultaneously, begin recruiting for Line of Effort 2 to build your specialized team. Once foundations are in place and key personnel are onboarded, execute Line of Effort 3 to build technical capabilities and external enablement resources.

Leverage Amazon Bedrock AgentCore to accelerate deployment:

1. Start with AgentCore Runtime for secure agent hosting with session isolation
2. Implement AgentCore Identity for authentication and authorization
3. Deploy AgentCore Gateway for centralized tool management
4. Enable AgentCore Observability for comprehensive monitoring
5. Configure AgentCore Evaluations for continuous quality assessment
6. Implement AgentCore Policy for deterministic governance (when available)
7. Utilize AgentCore Memory for context management across sessions

Table 2 AWS Service Mapped to Function

As services continue to be added to regions and accreditations it is important to use the list as a starting point to either prioritize addition of missing services for DoW use or addition of new services as they become generally available in this rapidly evolving landscape.