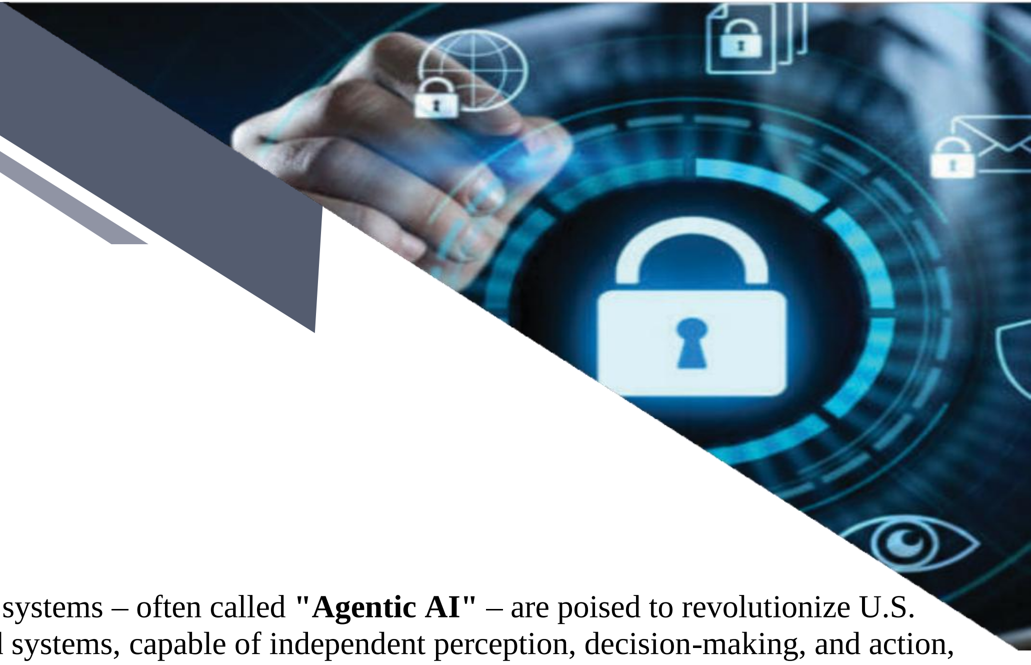
ANTEAN TECHNOLOGY

TRUSTED AUTONOMY: A FRAMEWORK FOR SECURELY DEPLOYING AGENTIC AI



Executive Summary

Autonomous Artificial Intelligence (AI) systems – often called **"Agentic AI"** – are poised to revolutionize U.S. Government operations. These advanced systems, capable of independent perception, decision-making, and action, promise unprecedented gains in efficiency, analysis, and mission effectiveness across defense, civilian, and intelligence domains.

However, this very autonomy introduces new and complex cybersecurity risks – from potential manipulation and unpredictable behavior to data misuse and governance gaps – that traditional security models were not designed to address.

To harness AI's power safely, agencies need a structured way to understand and manage these unique risks. This white paper introduces Antean Technology's **Agentic AI Cybersecurity Framework (AACF**). AACF provides a clear, layered model that guides organizations through the process of securing Agentic AI—offering a structured view of risks, targeted controls, and alignment with established federal frameworks such as Zero Trust Architecture (ZTA) and NIST AI RMF.

By adopting AACF principles, agencies can gain a holistic view of AI security, align efforts with federal mandates, and responsibly innovate while safeguarding missions and public trust.

1. New Capabilities, New Cyber Risks

Artificial Intelligence is no longer futuristic—it's foundational to how government agencies operate. The latest generation, Agentic AI, takes this to a new level. Unlike older AI systems designed for narrow tasks, Agentic AI agents can:

- **Understand complex goals**: Interpret broad objectives.
- **Act autonomously**: Carry out tasks with minimal human direction.
- **Learn and adapt**: Improve over time using data and feedback.
- **Collaborate**: Work with humans or other AI agents to achieve joint goals.

These abilities offer major benefits—like accelerating threat analysis, automating logistics, or personalizing citizen services—but also introduce **novel risks**:

- **New attack surfaces**: How agents reason, act, and share data becomes exploitable.
- **Unpredictable behavior**: AI may misinterpret intent or operate in unsafe ways.
- **Amplified data risk**: AI needs broad access to sensitive data.
- **Corrupted inputs**: Poisoned training data or compromised tools create embedded vulnerabilities.
- **Oversight challenges**: Ensuring ethical and accountable AI use is difficult without clear rules.

These challenges go beyond the scope of traditional cybersecurity. We need frameworks that address AI's behavioral logic, data dependencies, and autonomy risks—not just its network footprint.

2. The Structured Solution: Introducing AACF

To manage complex systems effectively, you need a structured approach. Just like the OSI model breaks down how networks work, AACF breaks down how to secure Agentic AI.

The **Agentic AI Cybersecurity Framework (AACF)** is a **seven-layer model** designed to:

- **Deconstruct Agentic AI into manageable components**
- **Identify threats at each layer**
- **Align with existing federal security mandates**
- **Guide proactive governance and security design** throughout the AI system lifecycle

AACF makes AI security understandable, actionable, and policy-aligned.

While AACF is structured around five high-level security principles for simplicity, it builds on a detailed seven-layer model—ranging from Human-AI Governance (Layer 1) to AI Agent Interface (Layer 7)—designed to align with familiar models like the OSI stack. This layered architecture supports both practical implementation and deeper technical alignment with federal security standards.

🔄 **The AACF model groups these seven technical layers into five functional domains for ease of understanding.** For instance, the **'Operations' domain spans orchestration (Layer 6) and interfaces (Layer 7)**, where **inter-agent communication (A2A)** and protocols like **Message Context Protocol (MCP)** are addressed.

---

3. Key Principles of the AACF Approach

Rather than diving into technical specifics, AACF simplifies AI security into **key strategic domains**:

- **Strong Governance & Human Oversight (Layer 1)**: Define clear rules, policies, ethical standards, and human oversight responsibilities. Ensure there's always a *"human in the loop"* where appropriate.
- **Secure AI Development & Supply Chain (Layer 2)**: Verify the **source and integrity** of everything used to build your AI. Prevent tampering before it starts.
- **Protecting Data & Knowledge (Layer 4)**: Rigorously control access to data. Prevent leaks, poisoning, or misuse. This is the **fuel behind AI decision-making**.
- **Securing AI Operations (Layers 6 & 7)**: Monitor behavior. Secure all human, system, and agent interactions. Block harmful prompts or actions.
- **Robust Foundational Security (Layers 3 & 5)**: Harden the infrastructure. Use **Zero Trust principles** to protect cloud, endpoints, and core services.

Emerging Imperatives: A2A and the Message Context Protocol

As Agentic AI systems evolve, **Agent-to-Agent (A2A) communication** is becoming increasingly common—enabling AI entities to **collaborate, negotiate, and delegate tasks autonomously**. This introduces a new layer of complexity, where security must account not only for agent behavior in isolation, but also for inter-agent dynamics.

Antean's AACF addresses this through heightened focus on **Layer 7: Interfaces and Protocols**, where novel tools like the **Message Context Protocol (MCP)** play a pivotal role. MCP structures how agents interpret intent, handle ambiguity, and maintain contextual integrity in high-speed, high-stakes environments.

🚨 **This isn't theoretical or a 2030 concern. These architectures are emerging now—in 2025.**

They bring real and immediate risks:

- **Collusion or adversarial influence between agents**
- **Semantic misalignment** due to insufficient context in messages
- **Deceptive behavior** introduced via indirect agent manipulation

Effective governance in the age of A2A means designing systems where message provenance, content boundaries, and oversight are built in from the beginning—not bolted on later. AACF is built to evolve with these realities.

While A2A and MCP are often associated with Layer 7, their logic originates in Layer 6, where agent behavior is orchestrated. These are inherently cross-layer functions—manifesting how autonomous intent is structured, constrained, and safely executed across interacting systems. AACF treats them accordingly, embedding controls at both the behavioral and interaction layers.

4. Aligning AACF with Federal Mandates

AACF doesn’t replace existing frameworks—it enhances them by adding AI-specific context and structure.

- **Zero Trust Architecture (ZTA):** AACF extends ZTA to the AI domain. It helps agencies answer: *Is this agent authorized? Is its action within policy? Should its request be verified every time?* Trust must be **continually validated**, especially when autonomy is involved.
- **NIST AI RMF:** AACF maps directly onto NIST’s AI Risk Management Framework, providing the structure and layers needed to operationalize **“Govern, Map, Measure, Manage”** in real AI systems.
- **FedRAMP, CMMC, etc.:** These protect the infrastructure. AACF ensures the AI workloads running within them—including models, agent behavior, and communications—are also secure. Without AACF, compliance at the infrastructure layer alone can leave **gaps at the AI logic and interaction layers**.

AACF is not another framework—it’s a strategic enhancement layer for AI safety and mission alignment.

---

5. From Theory to Practice: Applying AACF in Real Missions

Let’s look at how AACF applies in mission-critical environments:

🔍 AI Assisting SOC Analysts

An AI agent scans networks for threats in real time.

- **Layer 4 (Data Protection):** How do we ensure the agent only accesses the logs it needs?
- **Layer 7 (Interface Security):** How do we prevent prompt injection or adversarial input?
- **Layer 1 (Governance):** When can the agent auto-block traffic versus requiring escalation?
- **Layer 2 (Supply Chain):** Is the threat detection model authentic and verified?

💡 **AACF turns vague concerns into structured review questions at each layer.**

Emerging Use Case: Autonomous Inter-Agent Coordination

Picture multiple agents across agencies collaborating on a rapid-response plan—one pulling intelligence, another allocating logistics, a third drafting policy impact guidance.

- Who governs inter-agent communication across boundaries?
- Are messages authenticated and contextually valid (MCP)?
- Is there oversight or auditability of agent decisions and coordination?

AACF empowers agencies to ask these questions early, design for them deliberately, and monitor them continuously.

---

6. Recommendations & Call to Action

To realize the promise of Agentic AI securely, federal leaders must act decisively and now. We recommend:

- Adopt a layered view of AI risk. Move beyond firewalls—understand the intent and autonomy of AI systems.

- Use a structured framework like AACF. Embed security across the entire AI lifecycle—from design to deployment to decommissioning.
- Align security investments with AI-specific challenges—like model integrity, agent behavior analysis, and protocol safety.
- Evolve policies. Update acquisition, compliance, and ATO processes to account for Agentic AI risks and cross-layer behavior.
- Foster collaboration. Use AACF as a common language for risk, governance, and mission alignment across agencies.

Agentic AI without structured oversight isn't innovation—it's exposure. AACF helps you move fast *and* secure.

7. Conclusion

Agentic AI isn't just another software deployment. It's a paradigm shift in how agencies think, act, and achieve their missions.

Traditional security models are not built for autonomous decision-makers. We must go further—beyond Zero Trust, beyond infrastructure hardening—to consider the full lifecycle, intent, and communication of AI systems.

The Agentic AI Cybersecurity Framework (AACF) offers a policy-aligned, mission-ready, and technically rigorous foundation for doing just that.

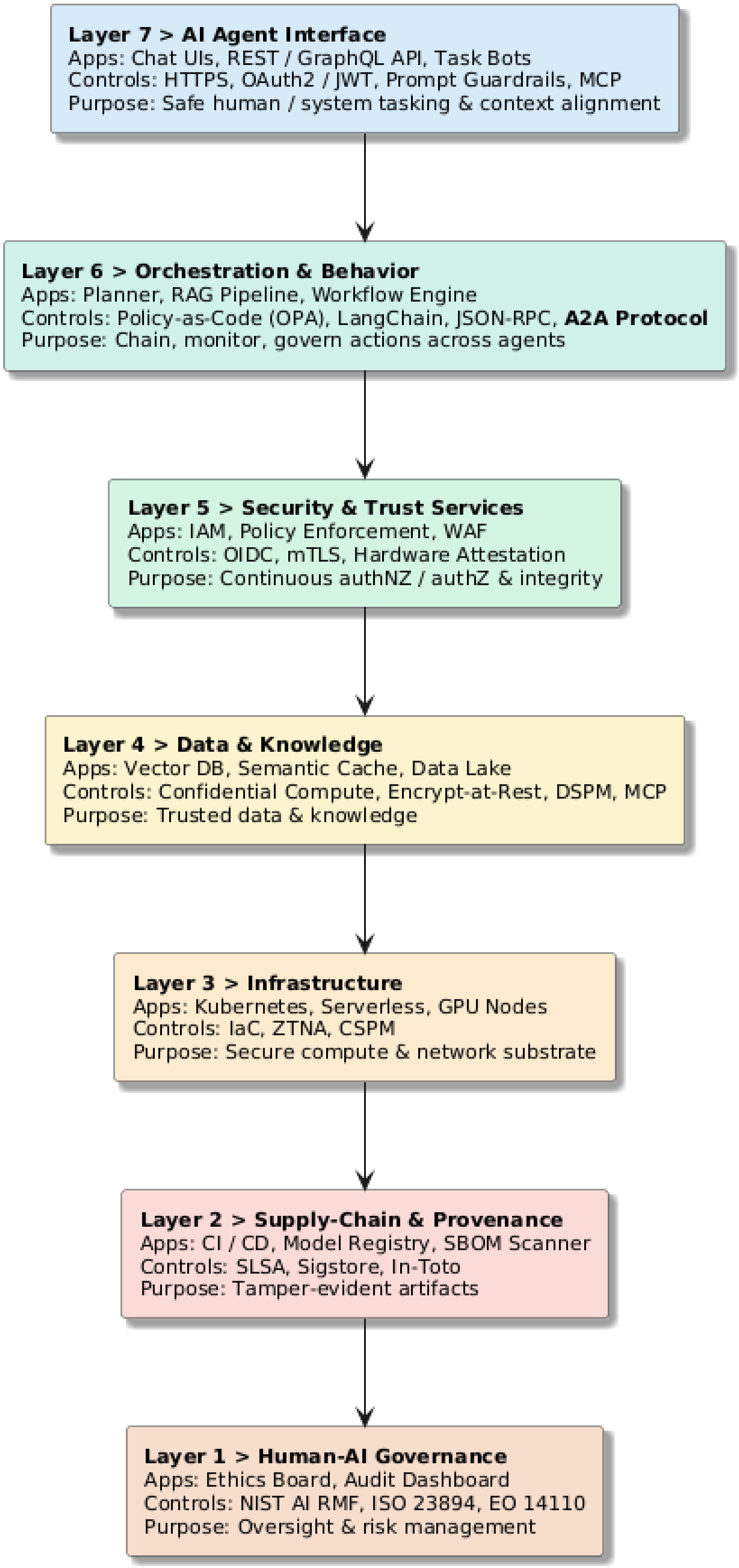
🕰️ The future is no longer 2030. It's now. And AACF helps ensure that future is secure, trusted, and aligned with the public interest.

About Antean Technology

Antean Technology is an SBA-certified 8(a) Woman-Owned Small Business specializing in cybersecurity, Zero Trust, and secure AI integration for the U.S. Federal Government.

We combine deep technical expertise with a mission-first mindset, and our development of the AACF reflects our commitment to responsible innovation, national security, and the ethical advancement of AI capabilities for the government.

AACF Layered Model



Layer 7 > AI Agent Interface
Apps: Chat UIs, REST / GraphQL API, Task Bots
Controls: HTTPS, OAuth2 / JWT, Prompt Guardrails, MCP
Purpose: Safe human / system tasking & context alignment

Layer 6 > Orchestration & Behavior
Apps: Planner, RAG Pipeline, Workflow Engine
Controls: Policy-as-Code (OPA), LangChain, JSON-RPC, **A2A Protocol**
Purpose: Chain, monitor, govern actions across agents

Layer 5 > Security & Trust Services
Apps: IAM, Policy Enforcement, WAF
Controls: OIDC, mTLS, Hardware Attestation
Purpose: Continuous authNZ / authZ & integrity

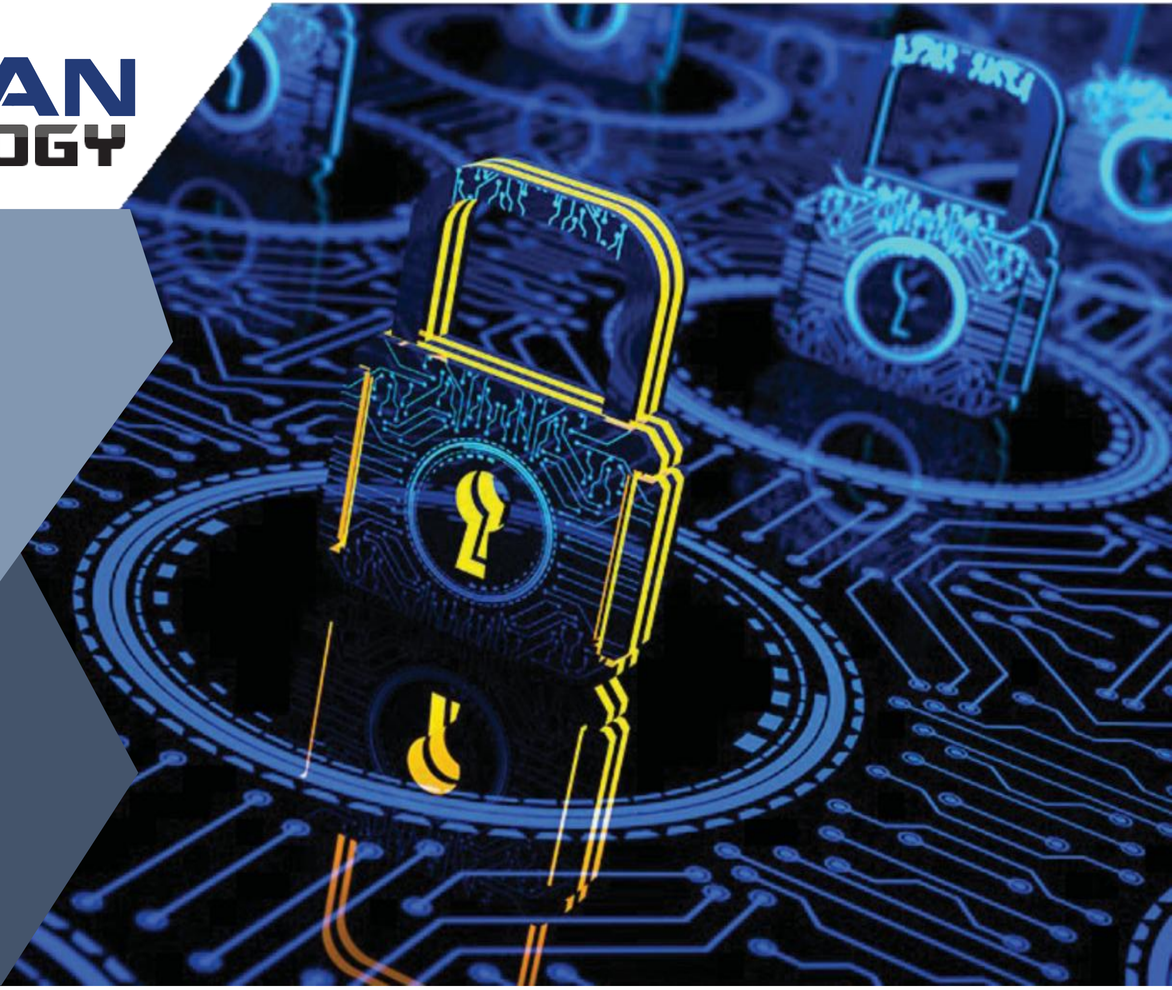
Layer 4 > Data & Knowledge
Apps: Vector DB, Semantic Cache, Data Lake
Controls: Confidential Compute, Encrypt-at-Rest, DSPM, MCP
Purpose: Trusted data & knowledge

Layer 3 > Infrastructure
Apps: Kubernetes, Serverless, GPU Nodes
Controls: IaC, ZTNA, CSPM
Purpose: Secure compute & network substrate

Layer 2 > Supply-Chain & Provenance
Apps: CI / CD, Model Registry, SBOM Scanner
Controls: SLSA, Sigstore, In-Toto
Purpose: Tamper-evident artifacts

Layer 1 > Human-AI Governance
Apps: Ethics Board, Audit Dashboard
Controls: NIST AI RMF, ISO 23894, EO 14110
Purpose: Oversight & risk management

ANTEAN TECHNOLOGY



ABOUT US

Antean Technology is deeply rooted in the design, development, implementation, and delivery of cybersecurity, and artificial intelligence services & solutions. Our understanding of technology as it relates to time, cost, and performance allows us to quickly navigate through the nuances and challenges of organizations to provide bespoke solutions.

Cyber Security

- Security Architecture Design & Implementation
- Zero Trust Planning and Implementation
- Cloud/hybrid Systems Migration & Support
- Assessment & Authorization
- Security Compliance & 3rd Party Assessments
- Penetration Testing and Audit
- ISSM/ISSO Support
- Security Documentation Development System
- Hardening/Vulnerability Management

Artificial Intelligence

- Agentic AI Security Architecture & Design
- AI Governance & Risk Management
- Secure AI Model Evaluation & Supply Chain Integrity
- Ethical AI Oversight & Decision Boundary Engineering
- Agent-to-Agent (A2A) Communication & Protocol Engineering (e.g. Message Context Protocol)
- AI System Integration w/ZTAs
- AI Powered Assessments

CONTACT US

Ernest McCaleb
CTO/AACF Visionary
Email : ernestmccaleb@anteantech.com

SEAN FLOYD
Chief Operating Officer
Email : seanfloyd@anteantech.com

Website: **https://anteantech.com/**
SBA 8(a) and EDWOSB
TOP SECRET Facility Clearance DUNS
Number: 080011379
Cage Code: 7HPX4