

Aegis

Governed AI assurance and adversarial validation.

Agencies and defense programs are fielding AI models and agents faster than they can produce evidence that those systems are safe to authorize. Aegis tests the safeguards around an AI system before it goes live, and turns every result into signed, machine-readable evidence that an Authorizing Official can check independently, including inside a disconnected enclave.

FOR	DELIVERS	RUNS	MATURITY
Authorizing Officials, ISSMs, and Chief AI Officers	Signed evidence pack and OSCAL	Self-hosted on Kubernetes	TRL 5

THE PROBLEM

AI ships ahead of assurance.

Prompt injection, leakage of sensitive data and system prompts, unsafe handling of model output, and over-privileged agents are new attack surfaces that scanners and periodic penetration tests do not cover. AI security tools test the model but do not produce authorization-grade evidence. Governance tools map controls but do not test. Red team results arrive as reports and screenshots that no one can re-verify a year later.

The timing is set by policy. OMB M-25-21 requires minimum risk management practices, including pre-deployment testing, for high-impact AI, and M-25-22 carries testing expectations into AI acquisition. FedRAMP is moving to machine-readable OSCAL packages, DoD is moving toward continuous authorization, and AI is reaching disconnected and classified networks where cloud-only tools cannot follow.

WHAT THIS COSTS A PROGRAM

- ✓ AI fielded with guardrails no one independently tested
- ✓ Agents with more reach than their mission requires
- ✓ Analyst hours rebuilding evidence for every review
- ✓ Findings that never reach a control owner
- ✓ Authorization delays and rework when an assessor rejects artifacts that cannot be verified

WHAT AEGIS DOES

Governed testing in, verifiable evidence out.

Tests AI guardrails safely

Benign, canary-based probes check whether an authorized AI endpoint's guardrails hold, with starter coverage across OWASP Top 10 for LLM Applications categories. No exploit content is used, and every bypass becomes a tracked finding.

Runs governed validation campaigns

Scenarios aligned to MITRE ATT&CK and ATLAS, read-only control checks, approvals under separation of duties, step-up authentication, and a kill switch, with every action recorded in a tamper-evident audit ledger.

Contains agents

Compares what an AI agent can reach in the real dependency graph with what it is authorized to reach, and flags over-privilege before the agent is fielded.

Turns findings into defenses

Routes findings to owners and tickets, compiles policy changes with verified rollback, and generates starter Sigma detection rules for the SOC.

Packages the evidence

Maps results to controls and produces OSCAL assessment results, a POA&M, and an eMASS-aligned export, all signed into an evidence pack that a standalone verifier checks with no network.

OUTCOMES

What the decision makers get.

FOR THE AUTHORIZING OFFICIAL AND ISSM

- ✓ Evidence packs verified on your own machine, with no network and without the vendor present
- ✓ OSCAL assessment results and a POA&M for the system of record, not slides and spreadsheets
- ✓ A record of who approved each test and each change, in an append-only, tamper-evident ledger
- ✓ Artifacts that regenerate to the same digests, so they can be re-checked a year later

FOR THE CHIEF AI OFFICER

- ✓ Pre-deployment guardrail test evidence for high-impact AI, in support of OMB M-25-21 practices
- ✓ Results mapped to the OWASP Top 10 for LLM Applications and MITRE ATLAS
- ✓ Agent containment findings before agents are fielded
- ✓ A consistent, benign test protocol to apply when acquiring AI under M-25-22

DIFFERENTIATORS

Where it is different.

- ✓ **Evidence an assessor can verify offline.** A standalone verifier rejects tampered, incomplete, unsigned, and forged packs and checks a legitimate pack with only the public key.
- ✓ **One step from validation to OSCAL.** Assessment results and POA&M come from the same control results that go into the signed pack.
- ✓ **Guardrail testing without weaponized content.** Benign canaries, run only inside approved scope and behind an allowlisted egress guard.
- ✓ **Agent containment tied to real topology.** Reach is computed from the actual dependency graph.
- ✓ **Built for disconnected and high-side enclaves,** with signed update bundles and no dependency on a hosted model or cloud console.
- ✓ **Determinism as a compliance feature.** The same inputs give the same digests, so artifacts can be regenerated and matched.

CAPABILITY	AEGIS	AI SECURITY AND AI RED-TEAM TOOLS	BREACH AND ATTACK SIMULATION	FEDERAL GRC AND OSCAL TOOLS	TRADITIONAL PENETRATION TESTING
AI guardrail probing	Yes, starter coverage	Yes, typically broad	Emerging	No	Varies by team
Agent containment tied to topology	Yes	Partial	No	No	No
OSCAL results, POA&M, eMASS-aligned export	Yes	No	Rare	Yes	No
Signed, offline-verifiable evidence	Yes	No	No	Partial	No
Disconnected and high-side profile	Yes	Rare	Rare	Rare	Manual

CAPABILITY	AEGIS	AI SECURITY AND AI RED-TEAM TOOLS	BREACH AND ATTACK SIMULATION	FEDERAL GRC AND OSCAL TOOLS	TRADITIONAL PENETRATION TESTING
Runtime guardrail or AI firewall	No, by design	Yes	No	No	No
Category-level comparison based on typical capabilities in each category; individual products vary.					

ADOPTION PATH

Pilot on one AI system, then expand by module.

1

Scope

Agree the AI system or enclave, rules of engagement, approvers, and the controls the evidence should support.

2

Pilot

Run a governed guardrail and containment campaign and deliver a signed pack with OSCAL output your assessor verifies.

3

Expand

Add systems under test and the AI Purple Team module, so findings become detection rules the SOC validates.

4

Operationalize

Add the Federal Evidence module for recurring OSCAL delivery and extend to disconnected and high-side enclaves.

Packaging: a core platform with AI Red Team, AI Purple Team, and Federal Evidence modules. Pricing is scoped per program and per AI system under test.

DEPLOYMENT

Connected self-hosted

Customer Kubernetes, identity provider, storage, and key management

Offline and air-gapped

Signed bundles verified before install, no outbound connections

High-side classified

Deny-egress networking, CAC and PIV through the identity provider

MATURITY

Technology readiness

TRL 5. Components validated end to end in a relevant laboratory environment with synthetic data

Engineering

Automated and negative security testing, and an independent end-to-end review with findings remediated

Authorization

Aegis holds no FedRAMP or Impact Level authorization. It supports an authorization decision and does not replace one

Intellectual property

Patent pending. Wagoner Enterprises, LLC has filed United States patent applications covering technologies in this product. No patent has yet issued.

CONTACT

Point of contact	Telephone	Email	Web	Classified inquiries
Bruce Wagoner, Chief Technology Officer	202-794-3282	sales@wagonerenterprise.com	wagonerenterprise.com	Indicate classification level and preferred secure communication method