AEGIS | Security & Trust Overview

AUTONOMOUS AI GOVERNANCE - SOVEREIGN BY DESIGN

AEGIS governs AI across 10 pillars using autonomous agents that run inside the client's own infrastructure. Detection happens in place; only redacted findings and metadata are transmitted. This overview answers the questions security, risk and procurement teams ask first.

Data sovereignty

> **Your data never leaves your network.**
Agents inspect data, models and prompts in place. Only governance findings and metadata leave - entity types and counts, metric values, severities, standards mappings. Sensitive values are redacted or one-way-hashed before transmission. Verified end-to-end: zero raw data is stored upstream.

Deployment & network posture

> **Outbound-only, no inbound ports.**
The agent is an HTTPS client - it never opens a listener. It deploys behind your firewall with no inbound rules, through a forward proxy, or fully air-gapped.

> **Least privilege, your keys.**
Read-only by default, using credentials you issue and scope. A revocable run token (issued from your portal) authorises reporting and can be cut off instantly. Passive (observe-only) mode available.

> **Auditable before it sends.**
Dry-run mode performs the full scan and prints the exact payload without transmitting, so your team reviews it before enabling reporting.

Protection in transit & at rest

> **Encrypted and tamper-evident.**
TLS with certificate verification by default; mutual TLS (mTLS) and pinning available for high-assurance deployments. The evidence trail is hash-chained and WORM-anchored so it cannot be silently altered.

> **Hardened build option.**
For high-assurance / air-gapped environments, a compiled, hardware-bound agent build with license attestation runs only where you authorise it.

What leaves your perimeter - and what never does

LEAVES (metadata only)	NEVER LEAVES
Entity types & counts, severities, metric values, OWASP/ATLAS & article mappings, redacted samples, salted fingerprints.	Raw PII/PHI/PCI, prompts, records, model weights, system credentials, and your application data.

Independently benchmarked (reproducible)

> **Detection is measured, not asserted.**
Fairness metrics match IBM AIF360 and Microsoft Fairlearn exactly (0.0000 difference). Drift statistics match SciPy (the engine behind Evidently) to <= 5e-5. PII detection scored F1 1.0 vs Microsoft Presidio 0.886 on a reproducible corpus. AEGIS claims method parity - not '#1'.

Framework alignment

EU AI Act - NIST AI RMF - ISO/IEC 42001 & 42005 - ISO/IEC 27001 - SOC 2 - GDPR - DORA. AEGIS is architected to SOC 2 and ISO/IEC 27001 control objectives and produces the tamper-evident evidence those audits require. Formal attestation reports and penetration-test summaries are available under NDA on request.

Request a security review, DPIA, or pen-test summary -> your AEGIS contact

Confidential - for evaluation by the recipient organisation. AIF360, Fairlearn, Presidio, SciPy, Evidently and Purview are trademarks of their respective owners.