Aegis-R: Technical White Paper

Abstract

Aegis-R is a human-governed AI security reasoning infrastructure designed to formalise and scale SOC analyst reasoning.

Rather than detecting anomalies, it evaluates causal feasibility, attack preconditions, and evidence completeness within a specific environment.

Core Principles

Reasoning over detection.

Causality over correlation.

Human authority over automation.

Auditability over opacity.

System Architecture

Aegis-R is composed of an immutable reasoning core, an observation and state-modeling layer, a zero-trust initialization layer, and a governance layer.

Attack logic and causal constraints are never learned from telemetry and cannot be influenced by attackers.

Poison Resistance

Learning is constrained to observation bounds and confidence intervals.

Evidence-gap detection ensures missing corroboration increases suspicion rather than trust.

Human Governance

Analysts authorise reasoning promotions and trust boundary changes.

All decisions are signed, attributed, and auditable.

Conclusion

Aegis-R introduces a new category: Security Reasoning Infrastructure.

It reduces SOC scale requirements while preserving accountability and compliance.