The Autonomous SOC Manifesto

A Framework for Classifying Levels of Security Operations Autonomy

Faiz Shuja
SIRP Labs | faiz@sirp.io
April 2026

Abstract

Security operations faces a scaling crisis driven by workforce shortages, analyst burnout, and alert overload. While AI and automation have improved parts of detection, triage, and response, the industry still lacks a broadly adopted, vendor-neutral framework for classifying degrees of SOC autonomy — leading to vendor confusion, misaligned buyer expectations, and unfocused research investment. This paper introduces the SOC Autonomy Framework (SAF), defining six levels of security operations autonomy (L0 through L5), analogous to the SAE J3016 standard for automated driving. For each level, we specify the decision scope of the AI system, the governance requirements, the human role, and proposed operational metrics for classification. We analyze the critical transitions between levels — particularly the L2-to-L3 reasoning boundary and the L3-to-L4 trust boundary — and take the normative position that L5 (full autonomy), while potentially technically achievable, raises fundamental questions about the appropriate role of human moral reasoning in proportional security response. We present SIRP OmniSense as a reference architecture designed toward L3/L4 implementation. The SOC Autonomy Framework is offered as a common vocabulary for vendors, analysts, researchers, and regulators to evaluate, compare, and advance autonomous security operations.

Keywords: Autonomous SOC, Security Operations, AI Autonomy Levels, Self-Driving SOC, SOAR, Agentic AI, Governed Autonomy

1. Introduction

For three decades, the Security Operations Center has operated on a fundamentally human-centric model: analysts receive alerts, investigate threats, make decisions, and execute responses. This model faces structural limits that AI alone has not yet resolved.

The evidence is substantial. A 2025 industry survey of 739 cybersecurity leaders reported that full investigation of a single alert takes approximately 70 minutes on average, with 56 minutes passing before initial action [1]. The alert volume problem is severe: a Ponemon Institute study found that organizations receive nearly 17,000 malware alerts per week on average, of which only 19% are deemed reliable and only approximately 4% are actually investigated [2]. The global cybersecurity workforce gap reached 4.8 million positions in 2024 [3], and analyst burnout is pervasive: a 2025 industry report found that 76% of SOC analysts experience burnout, with 64% likely to change jobs within a year [4]. These are not temporary pressures. They represent a structural ceiling on human-dependent security operations.

AI has been proposed as the solution, and meaningful progress has been made in threat detection, alert prioritization, and investigation assistance. However, the industry currently lacks a broadly adopted, vendor-neutral framework for classifying the degree of autonomy that AI systems exercise in security operations. This absence creates three problems:

Vendor confusion. Dozens of companies claim "AI-powered SOC" capabilities ranging from simple alert correlation to autonomous incident response. Gartner has warned of "agent washing" — vendors rebranding existing products with AI labels without substantive agentic capability [5]. Without a classification framework, buyers cannot meaningfully compare solutions.

Misaligned expectations. Organizations deploying AI in their SOCs lack vocabulary for specifying what level of autonomy they require, what governance must accompany it, and what metrics would demonstrate success.

Unfocused research. A recent survey of 189 papers on AI for security operations found that the literature is heavily weighted toward detection and triage, with minimal formal treatment of autonomous decision-making, response execution, or security reasoning under uncertainty [6].

This paper addresses these gaps by introducing the SOC Autonomy Framework (SAF) — a six-level classification system for security operations autonomy, modeled on the structure of the SAE J3016 standard for automated driving [7].

2. Related Work

2.1 The SAE J3016 Precedent

The Society of Automotive Engineers published J3016 in 2014, defining six levels of driving automation (L0-L5). This framework became a widely adopted vocabulary for the autonomous vehicle industry, used by regulators, insurers, manufacturers, and researchers worldwide. Its power lies not in technical innovation but in definitional clarity — it gave an entire industry a common scale. No broadly adopted, vendor-neutral autonomy framework currently exists for cybersecurity operations.

2.2 Existing Security Automation Taxonomies

SOAR platforms, introduced circa 2017, represented the first systematic attempt to automate SOC workflows through human-authored playbooks. However, SOAR operates on a fundamentally prescriptive model that cannot adapt to novel attack patterns without human intervention. Recent work has explored LLMs, RAG, and multi-agent architectures for security operations [6][8], but these contributions are primarily empirical, lacking formal frameworks for classifying degrees of autonomy.

Gartner's December 2024 report "Predict 2025: There Will Never Be an Autonomous SOC" [9] argues that full security automation is aspirational. We respectfully disagree with this conclusion while acknowledging the legitimate concerns it raises — concerns this framework directly addresses.

2.3 The Agentic AI Wave

Agentic AI systems have accelerated interest in autonomous security operations. A 2026 survey found that 87% of security teams prioritize agentic AI adoption [10], and global AI-in-cybersecurity spending is projected to exceed $219 billion by 2034 [11]. However, of the thousands of claimed agentic AI vendors, Gartner has warned that a significant majority lack genuine agentic capability [5]. A formal classification taxonomy is urgently needed.

3. Classification Methodology

Classification is based on four measurable dimensions:

Decision Scope: What categories of security decisions the AI system makes without human initiation. Measured by categorizing all SOC decision types and determining which the AI performs autonomously.

Autonomous Action Rate: The percentage of total security incidents resolved end-to-end without human intervention. The denominator is all incidents ingested; the numerator is incidents where the AI completed detection, investigation, decision, response, and documentation without human approval or modification.

Governance Requirements: The oversight mechanisms required at each level, from basic model tuning (L1) to comprehensive policy frameworks (L4-L5).

Human Role: The nature of human involvement. The transition from "human decides" (L0-L2) to "human supervises" (L3) to "human monitors" (L4) to "human sets policy" (L5) is qualitative, not merely quantitative.

A system's SAF level is determined by the **lowest** dimension at which it operates. A system with L3-level decision scope but L1-level governance is classified as L1, because governance constrains the safe operating envelope.

4. The SOC Autonomy Framework (SAF)

Note: The operational metrics specified for each level are proposed operational thresholds based on the author's operational experience and available industry data. They are not industry-established benchmarks and are offered as starting points for discussion and empirical validation.

***Table 1:** SOC Autonomy Framework — Summary (metrics are proposed operational targets)*

Level	Name	AI Decision Scope	Human Role	Proposed Auto. Action Rate
L0	Manual SOC	None	Everything	0%
L1	Assisted Detection	Surface, prioritize alerts	Investigate, decide, respond	0%
L2	Automated Triage	Triage, enrich, correlate, filter false positives	Validate, investigate, respond	0-10%
L3	Conditional Autonomy	Investigate, recommend, execute low-risk actions	Approve high-impact, supervise	20-50%
L4	High Autonomy (Self-Driving SOC)	Full lifecycle within governed boundaries	Monitor, exceptions, policy updates	70-90%
L5	Full Autonomy	Entire SOC lifecycle	Set policy only	99-100%

4.1 Level 0: Manual Security Operations

All detection, investigation, and response decisions are made by human analysts. MTTD is entirely dependent on analyst availability (typically 4-24 hours). Characteristic technology: manual SIEM dashboards, email alerting, spreadsheet case management.

4.2 Level 1: Assisted Detection

The AI assists by surfacing alerts, reducing noise (approximately 20-40%), and providing basic prioritization. All investigation and response decisions remain human. **Illustrative market examples:** CrowdStrike Charlotte AI, Microsoft Copilot for Security, SentinelOne Purple AI, Google SecOps Gemini. Based on publicly described capabilities, the majority of current "AI SOC" products appear to operate at this level.

4.3 Level 2: Automated Triage

The AI automatically triages alerts — classifying severity, enriching indicators, correlating events, and filtering false positives (50-80% noise reduction). Investigation recommendations may be generated but

not executed without human approval. **Illustrative examples:** Dropzone AI, Intezer, D3 Morpheus, SOAR platforms with AI augmentation.

4.4 The L2-L3 Boundary: The Reasoning Gap

The transition from L2 to L3 represents the most significant architectural leap. L2 systems follow predefined or learned logic through pattern matching. L3 systems must **reason** about novel situations — correlating evidence never correlated before, forming hypotheses about attacker intent, recommending actions outside any playbook. This requires: (a) **contextual reasoning** — understanding organizational context; (b) **causal inference** — distinguishing correlation from causation; and (c) **uncertainty quantification** — calibrated assessment of what the system does and does not know.

4.5 Level 3: Conditional Autonomy

The AI investigates alerts end-to-end, correlates evidence across data sources, forms hypotheses, and recommends specific response actions. Low-risk actions may execute within defined boundaries; high-impact actions require human approval. Governance requires confidence thresholds, auditable decision traces, action boundary definitions, and continuous accuracy monitoring. All recommendations must include a reasoning chain. Proposed targets: 20-50% autonomous action rate, 80-95% investigation coverage, 100% explainability rate.

Illustrative: SIRP OmniSense is designed toward L3, with architecture targeting L4 (see Section 5). Prophet Security describes autonomous investigation capabilities consistent with aspects of L3. Verified L3 operation requires independent measurement against the metrics above.

4.6 The L3-L4 Boundary: The Trust Threshold

The transition from L3 to L4 is primarily a **trust** challenge. At L3, humans approve high-impact actions. At L4, the system acts autonomously within governance boundaries. This requires: (a) **calibrated confidence** — tightly calibrated against measured accuracy; (b) **governed boundaries** — formal policy specifications enforced architecturally; (c) **auditable decision traces** — evidence-bound, policy-validated action paths; and (d) **graceful degradation** — recognition of operation outside competence, with appropriate escalation.

4.7 Level 4: High Autonomy (Self-Driving SOC)

The AI handles the complete investigation-to-response lifecycle for the majority of incidents. Proposed targets: 70-90% autonomous action rate, near-real-time MTTD, less than 10% human escalation rate. Requires comprehensive governance policy framework, real-time audit logging, confidence-gated execution, and regular adversarial testing. **No system has achieved independently verified L4 in production as of April 2026.**

4.8 Level 5: Full Autonomy — A Normative Position

We include L5 for taxonomic completeness but take the explicit normative position that L5, while potentially technically achievable, raises fundamental questions about the appropriate role of human judgment in security. Security decisions involve proportional response judgments, privacy considerations, business impact assessments, and moral reasoning about acceptable collateral impact.

> *"Autonomy is not about automating everything. It is about knowing what should never be automated."*
> *The goal of this framework is L4 — not L5. The Self-Driving SOC is one where AI handles the*

operational majority while humans retain authority over decisions requiring moral and strategic judgment. This is not a limitation. It is a design principle.

5. Reference Architecture: SIRP OmniSense

To ground the framework in a concrete implementation, we present SIRP OmniSense — a reference architecture designed toward L3 with an engineering path targeting L4. OmniSense is not presented as the only possible implementation, nor as having achieved verified L3 status by the metrics defined in Section 3. It is presented as evidence that the L3-L4 boundary is an engineering challenge, not a theoretical impossibility.

5.1 Agentic Mesh Architecture

Rather than a single monolithic AI, OmniSense employs an **agentic mesh** — a network of specialized AI agents, each responsible for a distinct SOC function (triage, enrichment, investigation, response recommendation, compliance verification). These agents collaborate through an orchestration layer — mirroring effective human SOC teams where specialists collaborate on complex cases.

5.2 OmniSec LLM: Domain-Trained Security Model

General-purpose LLMs demonstrate concerning limitations in security contexts — a 2025 study found 45% of AI-generated code introduced OWASP Top 10 vulnerabilities [12]. OmniSec is a cybersecurity-domain-trained language model **intended to reduce** hallucination risk by constraining the model's domain while preserving natural language reasoning capabilities for security investigation.

5.3 OmniMap: Knowledge Graph + RAG for Contextual Memory

OmniMap addresses the persistent context limitation through a dynamic knowledge graph connecting incidents, assets, vulnerabilities, threat actors, and threat intelligence. Retrieval-Augmented Generation queries this graph at inference time, grounding responses in organizational reality and solving the context window problem by retrieving precisely relevant context for each decision point.

5.4 OmniFlex: Reinforcement Learning for Adaptive Response

Static playbooks cannot adapt to novel attack patterns. OmniFlex introduces reinforcement learning that adapts response strategies based on incident outcomes. Published research has demonstrated 27.3% faster resolution and 31.2% higher defense effectiveness compared to rule-based approaches [13]. OmniFlex aims to apply RL in production SOC environments with continuous feedback loops.

5.5 OmniCollective: Federated Learning for Collective Intelligence

Individual SOCs observe limited threat patterns. OmniCollective addresses this through federated learning — sharing defensive patterns across participating SOCs without exposing sensitive data. The biological analogy is precise: the immune system shares antibody patterns, allowing the entire organism to benefit from any cell's encounter with a pathogen. OmniCollective aims to create an analogous digital immune network.

5.6 Governed Autonomy: Confidence-Gated Execution

Every autonomous action passes through a confidence gate: **high confidence + low impact** → execute autonomously; **high confidence + high impact** → execute with audit trail and notification; **low**

confidence → escalate to human with full reasoning chain; **out-of-scope** → refuse to act. This ensures the system never takes an action it cannot justify.

6. Illustrative Market Positioning

The following placements are illustrative assessments based on publicly described capabilities as of April 2026. They are not independently validated benchmarks. Vendors may possess capabilities not reflected in public documentation.

Table 2: *Illustrative Market Positioning (based on publicly described capabilities, April 2026)*

Level	Illustrative Examples
L0 — Manual	Legacy SOCs, organizations without automation tooling
L1 — Assisted Detection	CrowdStrike Charlotte AI, Microsoft Copilot for Security, SentinelOne Purple AI, Google SecOps Gemini
L2 — Automated Triage	Dropzone AI, Intezer, D3 Morpheus, SOAR platforms with AI augmentation
L3 — Conditional Autonomy	SIRP OmniSense (designed toward; see Section 5), Prophet Security (aspects of L3)
L4 — High Autonomy	No independently verified production deployment (April 2026)
L5 — Full Autonomy	Theoretical; normative concerns raised (Section 4.8)

7. Key Research Challenges

7.1 Neurosymbolic Security Reasoning. Combining neural pattern recognition with symbolic rule enforcement, causal reasoning, and temporal reasoning represents the most promising path to reliable security reasoning. The World Economic Forum has endorsed neurosymbolic AI as a path toward auditable, grounded AI outcomes [14]. Cybersecurity-specific application remains largely unexplored.

7.2 Adversarially Robust Confidence Calibration. Autonomous action requires accurate self-assessment. Adversaries will specifically target confidence estimation mechanisms. Adversarially robust calibration is an open research problem.

7.3 Federated Collective Intelligence at Scale. Federated learning prototypes for cybersecurity exist [15], but no production-scale deployment has been demonstrated. Challenges span model convergence and organizational trust frameworks.

7.4 Formal Governance Specification. L4 requires governance policies precise enough for machine enforcement yet flexible for organizational adaptation. No formal specification language for SOC governance currently exists.

7.5 End-to-End SOC Benchmarks. The field lacks a comprehensive, vendor-neutral benchmark spanning the full incident lifecycle. Existing benchmarks address narrow tasks [16]; a community-owned benchmark is urgently needed.

8. Conclusion

The SOC Autonomy Framework offers the cybersecurity industry a common vocabulary for evaluating and advancing security automation — analogous to what SAE J3016 provided the automotive industry. By defining six levels with criteria for decision scope, autonomous action rate, governance, and human role, we aim to enable meaningful comparison, focused research investment, and informed adoption decisions.

The path to Level 4 — the Self-Driving SOC — requires advances in neurosymbolic reasoning, calibrated confidence estimation, federated intelligence, and formal governance. We invite the research community, vendor ecosystem, and practitioner community to adopt, extend, critique, and improve this framework.

The goal is not to claim ownership of a category but to offer the vocabulary the industry needs to build the future of security operations — responsibly, transparently, and with appropriate humility about what machines should and should not decide.

> *"True autonomy is not proven when everything goes right. It is revealed when the system encounters something it has never seen — and responds with competence, transparency, and the wisdom to know when to ask for help."*

References

[1] Gurucul & Cybersecurity Insiders, "Pulse of the AI SOC Report," 2025. Industry survey of 739 cybersecurity leaders.

[2] Ponemon Institute, "The Cost of Malware Containment," 2016 (sponsored by Damballa). Survey of 630 IT security professionals. Finding: ~17,000 malware alerts/week; 19% deemed reliable; approximately 4% investigated.

[3] ISC2, "2024 Cybersecurity Workforce Study," October 2024. 4.8M global workforce gap. Note: the 2025 study did not include a workforce gap estimate.

[4] Tines, "Voice of the SOC Analyst," 2022. Survey of 468 SOC analysts. 71% report burnout; 64% likely to change jobs within a year.

[5] Gartner, "Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027," press release, June 25, 2025. Introduced "agent washing" concept.

[6] "AI-Augmented SOC: A Survey of LLMs and Agents for Security Automation," MDPI, 2025. Analysis of 189 papers.

[7] SAE International, "J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems," 2021.

[8] "A Survey of Agentic AI and Cybersecurity: Challenges, Opportunities and Prototypes," arXiv, January 2025.

[9] Gartner, "Predict 2025: There Will Never Be an Autonomous SOC," December 2024.

[10] Ivanti, "2026 State of Cybersecurity: Bridging the Divide," February 2026. 87% of security teams prioritize agentic AI.

[11] Polaris Market Research, "AI in Cybersecurity Market Forecast," May 2025. $25.4B (2024) to $219.5B (2034), 24.1% CAGR.

[12] Veracode, "2025 GenAI Code Security Report." 45% of AI-generated code samples introduced OWASP Top 10 vulnerabilities.

[13] "ARCS: Adaptive RL Framework for Automated Cybersecurity Incident Response," Applied Sciences, MDPI, January 2025.

[14] World Economic Forum, "Neurosymbolic AI: Real-World Outcomes," December 2025.

[15] "TrustFed-CTI: A Trust-Aware Federated Learning Framework for CTI," Future Internet, MDPI, 2025.

[16] CrowdStrike & Meta, "CyberSOCEval: Benchmarks for Evaluation of AI in Security Operations," September 2025.

About the Author

Faiz Shuja is the Co-Founder and CEO of SIRP Labs, where he created the OmniSense Autonomous SOC platform — a system designed to autonomously understand signals, reason in real-time, and take action based on evolving context.

His career in cybersecurity spans two decades. He founded Rewterz in 2006 from a small room on a rooftop in Karachi, Pakistan, with a single goal: build something meaningful in cybersecurity. That company grew into one of the Middle East's leading cybersecurity firms, now protecting 50+ enterprises across the globe with a 200-member team and a state-of-the-art SOC in Riyadh.

He served as CEO of The Honeynet Project (2016-2021), the international non-profit dedicated to investigating cyber attacks and developing open-source security tools. He holds CISSP, GCIH, and GSEC certifications.

Contact: faiz@sirp.io | Medina, KSA.