# SECURITY OPERATIONS AT THE NEXUS OF AI

## When to Use AI SOC Tools, AI-Enabled MDRs, or Both

Sponsored by Daylight Security


**Oliver Rochford**
Lead Analyst, Cyberfuturists

Sponsored by Daylight Security

# Executive Summary

AI fundamentally changes the value equation in security operations, but not in the way most vendor narratives suggest.

Organizations that have built AI into the core of how they detect, investigate, and respond can now deliver a quality of service that was previously out of reach for most organisations: expert-grade analysis, personalised to any environment, at a price point that is realistic. Security organizations that have bolted AI onto legacy architectures cannot.

This paper makes that case plainly. We assume AI delivers real capability gains, in context assembly, triage, investigation enrichment, response recommendation. And we ask a more useful question: who is best positioned to realise those gains for you?

For those with the staff and expertise to govern AI decisions directly, an internal AI SOC may be the right model. For other organisations, especially those without a mature internal SOC, the answer may be an AI-native MDR. For many, a hybrid approach may be most realistic.

What this paper argues is that the operating model decision, MDR, internal AI SOC, or hybrid, is now the most important security operations decision a CISO makes.

Most organisations are making it by default, through tool purchases and MDR renewals, without the framework to make it deliberately. This paper provides that framework.

## Three Core Shifts

1

### Shift 1: AI transforms MDR economics.

AI makes expert analysts more efficient and more effective, acting as force multipliers, but not replacements. AI-enabled MDRs can now deliver greater cost efficiency and more personalised service simultaneously, a combination previously impossible to sustain at a realistic price point.

2

### Shift 2: AI shifts the governance question.

The traditional MDR decision was "who runs the SOC?" With AI in the loop, it becomes "who owns decisions when machines are making them?" Decision ownership and not headcount or SLA  is the new baseline evaluation criterion. This makes it more a governance question, and less purely a staffing question, and most evaluation frameworks have not caught up.

3

### Shift 3: AI blurs the tool/service boundary.

AI blurs the tool/service boundary. The clean distinction between buying tools and buying services no longer holds, and deploying an AI SOC tool is not just a technology decision. It is a whole operating model commitment. And AI forces CISOs to make this operating-model decision earlier than they may be used to, with wrong choices creating technical debt that is expensive to unwind.

## What this means for CISOs

If you are considering an internal AI SOC: be honest about whether you have or can build the governance expertise AI decisions require. It is a different capability from running a SOC.

Context engineering capability - AI SOC requires several context dimensions

In either case: make the operating model decision explicitly. Do not let it be made for you by a vendor selection.

Ask different questions: not "how many alerts can you handle?" but "who is accountable when the AI is wrong?"

Treat the AI supply chain as a first-class risk: your MDR's dependency on foundation model providers is your dependency, even if indirect.

## 1.1 The Context Gap AI Claims to Solve

Security operations centres have struggled for years with a problem that does not reduce to alert volume alone: context. SANS 2023[1] data identified a "lack of context related to what we are seeing" as the number one SOC challenge, displacing alert fatigue from prior years. In addition, secondary pressures including the lack of skilled staff, inadequate enterprise visibility, and insufficient automation all compound the same underlying failure, that analysts are having to make crucial decisions without a complete understanding of the situation.

AI SOC vendors position themselves as solving all four challenges simultaneously: AI provides context, compensates for skill gaps, unifies visibility across tools, and automates routine work. The pitch is compelling. AI does deliver real capability gains in each area. But as AI SOC adoption estimates (1%-5%)[2] and AI project failure rates (80%-95%)[3] show, only in the right hands and conditions.
The harder question then is not whether AI works, but who is best positioned to deliver those gains in your environment. For many organisations, the answer is an AI-enabled MDR rather than a direct tool deployment, because realising AI's potential requires operational infrastructure, training data at scale, and governance expertise that most internal teams are not yet equipped to provide.
This paper makes the case that AI has shifted the value equation for managed security services in a way that favours providers who have built AI into their operations from the ground up and provides a framework for evaluating whether a given provider has actually done so....

## 1.2 The Central Argument of This Paper

This paper does not attempt to crown a winner between AI SOC tools and AI-enabled MDRs. Both are legitimate operating models; the right choice depends on your organisation's maturity, risk appetite, and internal capability. What this paper argues is that the decision is now an operating model choice, not a technology preference, and that most organisations are making it by default rather than by design.

The criteria that should drive this decision, including decision ownership, explainability, failure behaviour, and supply chain resilience, are almost entirely absent from traditional MDR evaluation frameworks. This paper provides those criteria and the questions to ask against them.

## 1.3 What This Paper Is (and Is Not)

This paper provides a framework for evaluating AI-enabled security operations as an operating model decision. It is grounded in practitioner experience with security leaders, SOC operators, and vendors, rather than vendor claims.

It is not a vendor ranking. It does not recommend a single operating model. It does not assume one-size-fits-all maturity. It uses Daylight Security's approach as a concrete illustration of AI-native MDR design, not as an exclusive endorsement.

---

[1]SANS 20231 https://www.sans.org/white-papers/2023-sans-soc-survey

[2]AI SOC adoption estimates https://www.gartner.com/en/documents/6625402

Gartner, Hype Cycle for Security Operations, 2025, Jonathan Nunez, Darren Livingstone, 23 June 2025

[3]AI project failure rates https://hbr.org/2025/08/beware-the-ai-experimentation-trap

# 2. Understanding the Choice: AI SOC Tools vs. AI-Enabled MDR

The distinction between AI SOC tools and AI-enabled MDR services may appear straightforward but has become genuinely blurry. Many AI SOC vendors now sell primarily through MSSPs or have become MDRs themselves. Many MDRs bundle proprietary tooling that, in an earlier era, would have been sold directly to enterprise SOCs. But while the boundary is dissolving, the underlying decision has not disappeared. If anything, it has become more consequential.

### WHAT PROBLEMS EACH MODEL WAS BUILT TO SOLVE

| | AI SOC TOOL | VS | AI-ENABLED MDR |
|---|---|---|---|
| **Primary buyer** | Enterprise with existing SOC (11+ analysts) | | Organisation without dedicated SOC capacity |
| **Core problem addressed** | Analyst efficiency — deliver broader coverage with the same team | | Capability gap — expertise and coverage you do not have internally |
| **Who governs AI decisions?** | You own configuration, tuning, and outcomes | | MDR owns day-to-day decisions; you retain oversight rights |
| **Staff requirement** | Detection engineering, AI governance expertise | | Vendor management and contractual oversight capability |
| **Time-to-value** | Slower<br>*Tool + operating procedure alignment required* | | Faster<br>*MDR brings existing baselines, SOPs and expertise* |
| **Customisation ceiling** | High<br>Customisation ceiling | | Variable<br>*Depends on MDR flexibility and your contract* |

## 2.2 How the Models Now Converge and Why It Matters

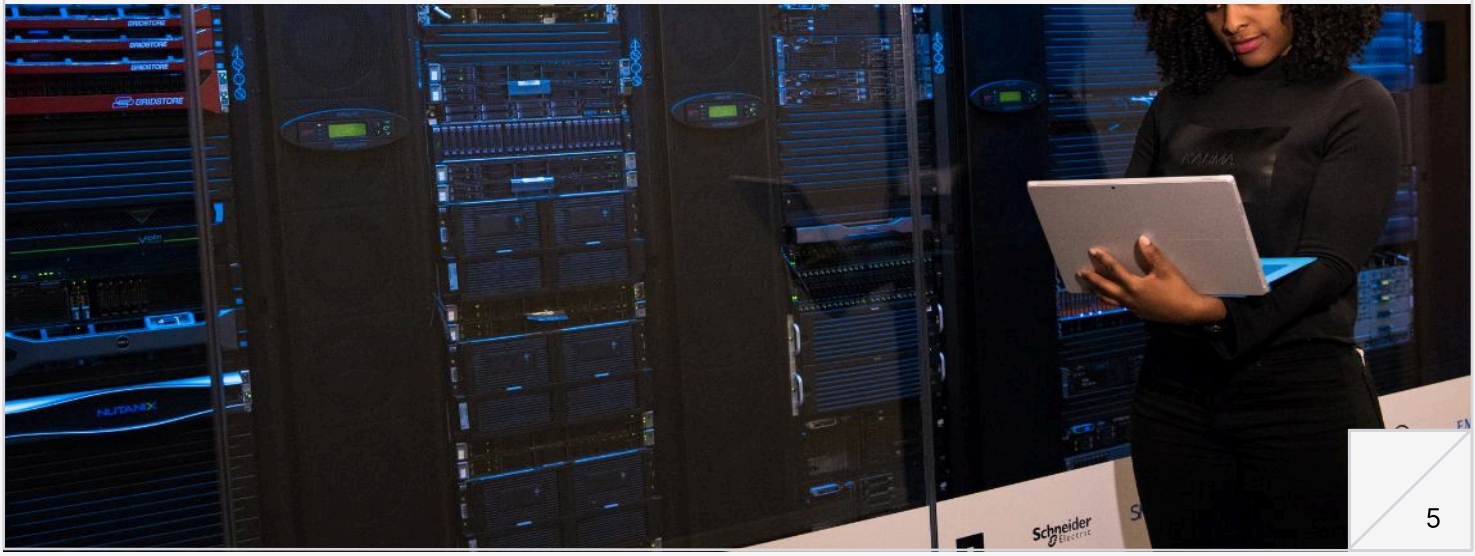
AI has introduced genuine overlap. Both models now involve AI-driven triage and prioritisation, probabilistic verdict generation rather than deterministic rule-firing, human oversight of machine-generated conclusions, and most critically, questions of accountability when AI is wrong.

This convergence is why the "tool vs service" framing is insufficient. The more useful question is: who makes which decisions, and who owns the outcomes? That question applies whether you have bought a tool or contracted a service that relies heavily on AI, and the answer has direct implications for incident post-mortems, regulatory compliance, and contractual liability.

### Why many AI SOC firms are selling through MSSPs, not to enterprises directly

AI models improve with scale. MSSPs accumulate richer, more diverse training data across hundreds of customer environments than most single enterprises SOC can generate. The feedback loop between analyst decisions and model improvement is also faster, especially as security is the primary product and not a support function.

And the operational infrastructure for AI governance including confidence calibration, failure monitoring, and supply chain management, is easier to amortise across a managed service than to build and maintain internally. This is not intended to be a sales pitch, but rather reflects economic reality. CISOs evaluating AI SOC tools should ask whether their organisation has the volume and expertise to benefit from them on the same timeline and efficiency as an MSSP would.



## AI-AUGMENTED vs. AI-NATIVE: WHERE DECISIONS HAPPEN

| AI-AUGMENTED MDR | VS | AI-NATIVE MDR |
|---|---|---|
| *AI added onto existing architecture* | | *AI built into foundational structure* |
| **Telemetry ingested**<br>*Logs, alerts, endpoints* | | **Telemetry + org context ingested**<br>*Assets, relationships, behavioural norms* |
| **Static rule matching**<br>*Predefined queries fire against logs* | | **Knowledge graph updated**<br>*Context encoded continuously* |
| **ML reduces alert volume**<br>*Pattern-matching suppresses noise* | | **AI derives verdict**<br>*Benign · Suspicious · Ambiguous* |
| **Alert queue**<br>*Analyst works through prioritised list* | | `High confidence` **Auto-resolved**<br>*No human required*<br>`Below threshold` **Human review**<br>*Full evidence package assembled* |
| **Human investigates & decides**<br>*Gathers own context manually* | | |
| **Verdict + response**<br>*Human-authored, human-executed* | | **Verdict + response**<br>*Auditable reasoning chain retained* |
| **Outcome: reduces volume** | | **Outcome: improves accuracy** |

### Case Example: Daylight Security's AI-Native Approach

Daylight Security illustrates what AI-native MDR design looks like in practice. Rather than relying primarily on predefined detection queries, the platform builds a knowledge graph encoding organisational context (assets, relationships, behavioural norms) and uses that graph to enrich and evaluate every event.

AI derives a verdict for each event: benign, suspicious, or ambiguous. The confidence-based handoff is the key to the design. When AI confidence is high, events are resolved automatically; when confidence falls below threshold, events are surfaced to a human analyst with the full evidence package assembled. Critically, the handoff comes with evidence, observable artifacts that support the AI's classification, rather than a score alone. This enables genuine analyst evaluation rather than mere ratification.

Context bootstrapping and sustainability are explicit design concerns: the system captures knowledge continuously, so the MDR's understanding of a customer environment deepens over time. DLP policies and ambiguous classifications that require contextual judgment are explicitly retained as human responsibilities, a recognition that not all decisions should be automated, and that some require policy ownership by the customer.

# 3. What Changes When AI Enters the MDR Model

AI does not simply make MDRs faster or cheaper. It changes what MDRs are responsible for, and who is responsible for what. Understanding these changes is prerequisite to evaluating any AI-enabled offering.

## 3.1 The Force Multiplier Effect

The most significant operational impact of AI in the MDR context is the amplification of expert judgment. An analyst who previously managed a fixed number of customer environments, spending significant time on triage and enrichment, can now focus almost entirely on cases that require human reasoning. AI handles the volume; the human handles the judgement.

The practical consequences are measurable: response times improve because AI identifies priority cases faster than human queuing; investigation quality improves because AI assembles context that analysts would otherwise gather manually; and customer service improves because analysts freed from routine work have more time to understand specific environments and tailor recommendations.

This is the architecture that makes the new MDR economics possible: more personalised, more context-aware service at a price point previously available only at the high end of the market. It is also the architecture that raises the governance questions at the centre of this paper.

## ANALYST TIME ALLOCATION: BEFORE AND AFTER AI

| BEFORE AI | | WITH AI-NATIVE MDR | |
|---|---|---|---|
| Alert triage & sorting | ~40% | Complex judgment & exceptions | ~45% |
| Context gathering | ~30% | Customer tailoring | ~25% |
| Report writing | ~15% | AI oversight & verdict review | ~20% |
| Investigation | ~10% | Context (AI assisted) | ~7% |
| Complex | ~5% | | |

*Most time on routine, repeatable work*

*Most time on judgment, relationship, oversight*

**Analyst role: Worker**

*Reviewing alerts · gathering context · writing up findings*

**Analyst role: Supervisor**

*Reviewing AI decisions · validating verdicts · managing exceptions*

*Illustrative. Proportions based on practitioner interviews and analyst research.*

## 3.2 The Human Role Is Changing, Not Disappearing

AI alone is not enough. But the role of the human is changing fundamentally, and organisations that do not understand this change will govern it badly....

| Dimension | Before AI | With AI-Native MDR |
|---|---|---|
| Primary analyst role | Worker: reviewing alerts, gathering context, writing up findings | Supervisor: reviewing AI decisions, validating verdicts, managing exceptions |
| Time allocation | Most time on routine triage and enrichment | Most time on genuine ambiguity, complex investigation, and customer relationship |
| Output focus | Reviewing outputs (alert by alert) | Improving inputs: detection quality, log coverage, context accuracy |
| Feedback loop | Ad hoc; knowledge lives in individual analysts | Distributed: human reviews AI verdict and owns the escalation decision |
| Skill requirement | Alert triage, SIEM proficiency | AI oversight, judgment calibration, governance of probabilistic systems |

The supervisor role is not a downgrade. It demands higher-order judgment: knowing when to trust the AI, when to override it, and how to interpret evidence in context. But it is a different role — and MDRs that have not redesigned their analyst workflows for AI supervision rather than alert review are not AI-native in any meaningful sense.

## 3.3 How AI Alters Specific Decision Points

### Triage and Prioritisation

AI changes triage from a rule-based queue into a probabilistic prioritisation engine. Alerts are no longer simply “fired” or “not fired”; they receive confidence scores, are enriched with contextual evidence, and are ranked for human attention. This is more sophisticated than traditional triage, but it introduces new failure modes: a poorly calibrated confidence model can systematically deprioritise real threats, and that failure is often invisible until a breach occurs.

### Alert Suppression

Suppression is where the stakes are highest and visibility is lowest. When AI suppresses an alert, no human sees the original signal. There is no audit trail of “we saw this and decided it was not important.” If the suppression was wrong, discovery typically happens during incident response, by definition too late.

This is qualitatively different from a human analyst triaging an alert as low-priority. The human decision is traceable and explainable. The AI suppression is statistical and often opaque. A mature AI-enabled MDR addresses this directly: suppression decisions must be visible, confidence-scored, and auditable.

### Investigation Framing and the Cascade Problem

Even when humans make the final decision, AI shapes how they see the problem. When AI assembles the context package for an analyst’s investigation, what AI includes defines the investigation scope. What AI omits may never be considered. AI’s initial classification anchors human judgment, a well-documented cognitive bias, and AI-generated investigation summaries become the official record relied upon in post-incident reviews and regulatory disclosures.

Governance frameworks that treat human review as a sufficient safeguard are addressing the right instinct but underestimating what “meaningful oversight” actually requires. Effective human-in-the-loop design gives the human enough independent context to genuinely evaluate the AI’s conclusions, not simply confirm them. The difference between evaluation and ratification is where governance succeeds or fails.

### Escalation Thresholds and Response Recommendations

The threshold at which AI confidence triggers an escalation to human review, or an autonomous response action, is itself a governance decision, and one that most customers never explicitly make. Vendors set defaults; customers inherit them. In a conventional MDR, escalation logic is documented and negotiated. In an AI-enabled MDR, it is embedded in model configuration and may change without notice when models are updated.

The balance between probabilistic AI judgment and deterministic enforcement is a policy question. CISOs should be able to set and adjust confidence thresholds, review what is happening at each threshold, and understand the trade-off between false positive volume and missed detection risk.

## 3.4 Accountability: MDR vs. Direct Deployment

If you deploy an AI SOC tool directly, you own the configuration decisions, the tuning decisions, and the outcomes of AI-driven triage and response. The vendor provides the capability; you provide the judgment.

If you engage an AI-enabled MDR, the MDR owns, in theory, the configuration and tuning, and bears accountability for AI-driven decisions made on your behalf. You own the decision to trust the MDR, and you still own the ultimate business outcome.

The “in theory” qualification matters. Most MDR contracts were written for human-centric service delivery. They do not explicitly address AI decision accountability. Key questions that frequently go unanswered: if the MDR’s AI suppresses an alert that turns out to be a breach, what is the contractual liability? If the MDR changes AI models mid-contract, do you have approval rights? If the MDR’s AI is trained on aggregated customer data, what are the confidentiality implications?

# 4. How MDR Selection Criteria Must Evolve

Traditional MDR evaluations focus on coverage hours and response SLAs, technology stack and integration capabilities, analyst certifications, incident volume metrics, and price per endpoint. These criteria remain relevant but are insufficient. They do not tell you how much of the response is AI versus human, whether you can see and influence AI decision-making, what happens when AI fails systematically, or who is accountable for AI-driven outcomes.

The following evaluation dimensions address what traditional scorecards miss. They apply regardless of vendor or delivery model, but naturally favour providers that treat AI as an operational responsibility.

## 4.1 Decision Ownership

Process ownership clarity, who owns which decisions and when, is the single most important criterion for AI-enabled MDR evaluation. Many providers cannot answer this question precisely, which is itself diagnostic.

### Questions to Ask

- Show me the decision tree for how an alert becomes an escalation to my team. At each step, is the decision AI-made, AI-recommended, or human-made?
- What AI-driven decisions happen before a human sees anything?
- Can I see what was suppressed or auto-closed, and why?
- How do I know if the AI's confidence was high or low on a given decision?
- Where does your AI act without human approval? Can I adjust those boundaries?
- If your AI suppresses an alert that turns out to be a breach, what is the contractual position?
- Do your SLAs apply to AI-driven decisions or only to human decisions?

## 4.2 Explainability

Explainability in the AI SOC context has two distinct requirements that are often conflated. Real-time explainability means an analyst can understand why the AI is flagging something as suspicious, and can push back on that classification if it does not make sense in context. Post-hoc explainability means you can reconstruct the AI's reasoning after an incident for audit, compliance, or root-cause analysis.

An evidence-based approach, where AI classifications are supported by observable artefacts rather than opaque scores, provides both. When an AI says something is suspicious, it should be able to show what it saw. In regulated industries, this is increasingly a compliance requirement.

### Questions to Ask

- Can I audit the AI's reasoning on a closed case?
- How do you distinguish AI-generated findings from human analysis in your reports?
- How do I explain to a regulator why an alert was deprioritised?
- What AI models do you use, and what are they trained on?
- How do you validate that your AI is not making systematically biased decisions?

### Questions to Ask

- Can I audit the AI's reasoning on a closed case?
- How do you distinguish AI-generated findings from human analysis in your reports?
- How do I explain to a regulator why an alert was deprioritised?
- What AI models do you use, and what are they trained on?
- How do you validate that your AI is not making systematically biased decisions?

## 4.3 Control Boundaries: Where AI May Recommend vs. Act

There is a meaningful difference between AI that recommends actions and AI that takes them. The former preserves human control at the point of consequence; the latter accelerates response but removes the human from the decision chain. Both have legitimate use cases. What matters is that the boundaries are explicit, customer-configurable, and documented in your contract.

Current agentic AI systems, those capable of multi-step autonomous action, are in an early state. Even providers building toward autonomous response acknowledge that confidence thresholds for unattended action must be set conservatively, with clear rollback mechanisms. Any vendor claiming their agentic AI is ready for unattended production response in high-stakes environments without qualification should be scrutinised carefully.

### Questions to Ask

- What actions can your AI take autonomously, without human approval?
- How are confidence thresholds set for autonomous actions, and can I adjust them?
- What is the notification and approval process when you want to expand AI autonomy?
- How do you handle situations where AI confidence is broadly low, for example during a novel attack campaign?

## 4.4 Failure Behaviour

A defining characteristic of mature AI deployment is the ability to fail gracefully. In security operations, this means maintaining human coverage when AI confidence is low, avoiding alert floods when AI systems are stressed or manipulated, and having clear rollback mechanisms when AI model changes degrade performance.

Vendors who cannot answer questions about failure modes have not thought seriously about production operations. The most revealing question you can ask any AI-enabled MDR is: tell me about your last significant AI failure and how you handled it. The answer is more diagnostic than any feature demonstration.

### Questions to Ask

- What happens when your AI confidence is broadly low? Do humans absorb the volume automatically?

## 4.5 The AI Supply Chain

Most AI-enabled MDRs depend on third-party AI infrastructure: foundation models, cloud AI services, and specialised security AI from various vendors. This creates supply chain risks that traditional MDR evaluations ignore entirely. An MDR whose core detection capability depends on a specific foundation model API is exposed to pricing changes, deprecation risk, and upstream model updates that can alter behaviour without notice. These are not hypothetical concerns. API pricing changes have already forced several AI-native security vendors to revise their economics mid-contract. Model updates from major providers have introduced unexpected behaviour changes in downstream applications. The volatility of the current AI supply chain is a first-class operational risk.

### Questions to Ask

- What third-party AI models or services does your platform depend on?
- What happens if your foundation model provider deprecates an API or changes pricing?
- How do you manage upstream model updates before deploying them to production?
- If AI token costs double, does my contract price change?
- Do you have SLAs with your AI infrastructure providers?

## 4.6 Adaptability Over Time

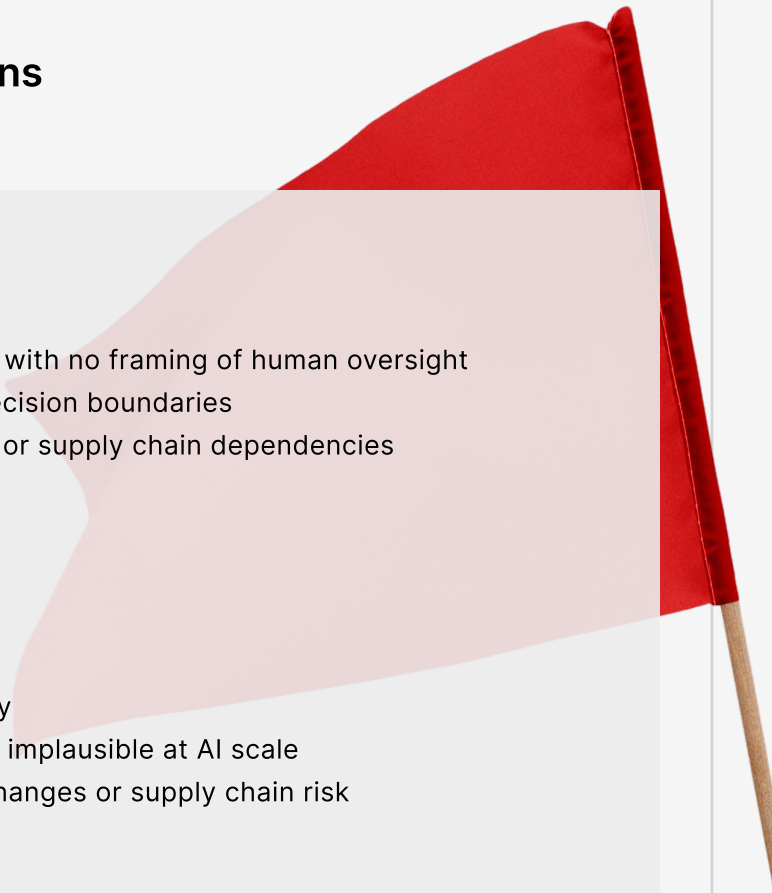
A criterion almost entirely absent from traditional MDR scorecards is adaptability: whether the MDR's understanding of your environment improves as the environment changes, as threats evolve, and as you provide feedback on AI decisions.

This matters more with AI than with previous-generation tooling because AI systems can both improve and degrade over time. A model that is well-calibrated for your environment today may perform worse after a major infrastructure change, a shift in attacker tactics, or an upstream model update. The MDR's processes for detecting and correcting this degradation, and for incorporating your feedback, are as important as their initial detection capability.

### Questions to Ask

- How does your AI adapt to my environment versus generic baselines?
- How does AI behaviour improve as it learns my environment over time?
- What feedback mechanisms exist for me to influence AI behaviour?
- How do you detect and correct model drift, that is, degradation in performance over time?
- What happens to the AI's knowledge of my environment if I switch providers?
- How do your analysts' findings feed back into model improvement?

## 4.8 Red Flags in Vendor Conversations



**When evaluating AI SOC tools:**

- “Our AI handles everything automatically” with no framing of human oversight
- Inability to explain confidence levels or decision boundaries
- Resistance to discussing AI failure modes or supply chain dependencies

**When evaluating AI-enabled MDRs:**

- SLAs that do not mention AI accountability
- “Our analysts review everything,” which is implausible at AI scale
- Contracts that do not address AI model changes or supply chain risk

**Universal red flags:**

- Promises of analyst replacement rather than augmentation
- No clear answer to "tell me about your last AI failure"
- Pricing that does not account for AI infrastructure costs, which signals unsustainable economics

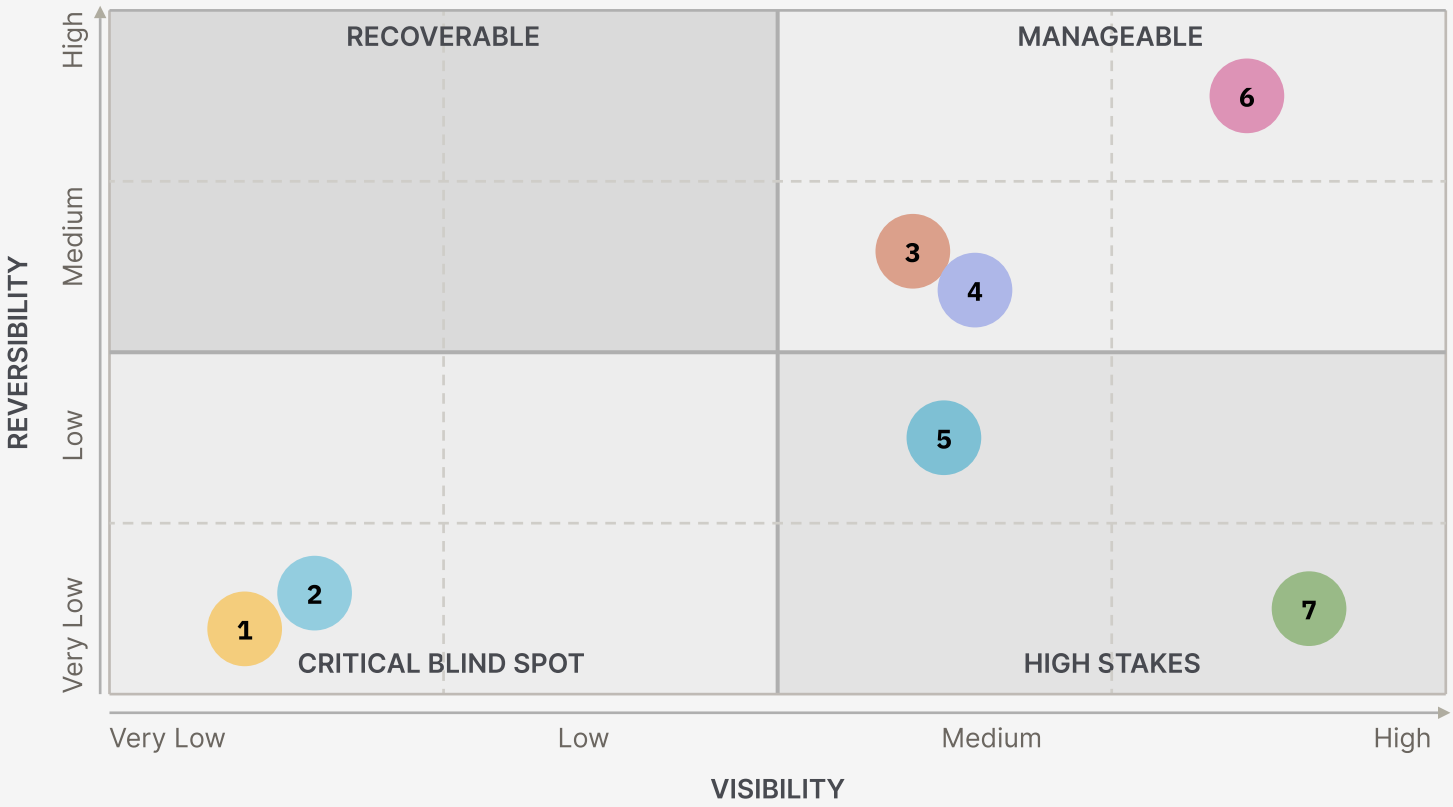
# 5. Decision Ownership: The Governance Question at the Heart of AI SOC

AI does not simply make MDRs faster or cheaper. It changes what MDRs are responsible for, and who is responsible for what. Understanding these changes is prerequisite to evaluating any AI-enabled offering.

## 5.1 Where Decisions Get Made, and Hidden

Every AI-enabled security tool makes decisions. Most of these decisions are invisible unless you specifically look for them. The most consequential AI decisions are often the least visible.

## ANALYST TIME ALLOCATION: BEFORE AND AFTER AI




### Decision Points

**1 Ingestion filtering**
*What telemetry is kept*

**2 Alert suppression**
*What never gets human attention*

**3 Alert prioritisation**
*What gets human attention first*

**4 Investigation framing**
*What context is assembled*

**5 Correlation / attribution**
*What is linked, who is blamed*

**6 Response recommendation**
*What action is suggested*

**7 Autonomous response**
*What action is taken without asking*

**⚠️ Governance gap**

*Points 1, 2 and 7 sit in the highest-risk zone:*
*invisible and irreversible — no human sees these decisions unless explicitly surfaced*

Alert suppression is where the stakes are highest and visibility is lowest. When AI suppresses an alert, no human sees the original signal. A mature AI-enabled MDR addresses this directly: suppression decisions must be visible, confidence-scored, and auditable. Anything less is a governance gap.

## 5.2 Probabilistic vs. Deterministic: A Fundamental Shift

Traditional security logic operates deterministically. If a specific pattern matches, a specific alert fires, and a specific playbook runs. The cause-and-effect chain is transparent, reproducible, and straightforward to audit.

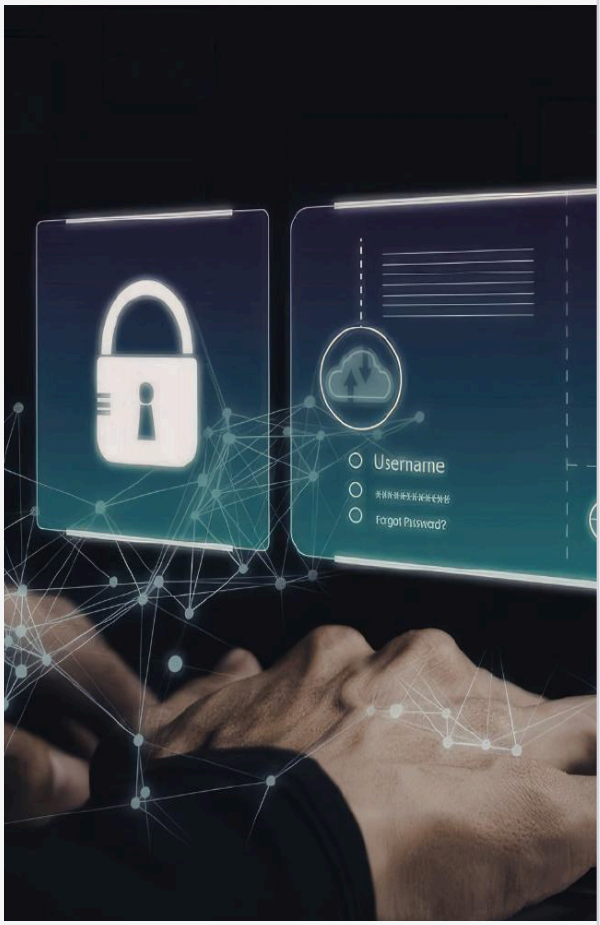
AI-augmented logic operates probabilistically. Rather than a rule that fires or does not, an AI assigns a confidence score, and the threshold at which that score triggers action is itself a governance decision. The operational implications are significant: “why did this happen?” becomes harder to answer; consistency is not guaranteed across different time periods or queuing conditions; and calibration is ongoing, since a model that is 85% accurate today may degrade if the threat landscape shifts.

AI-augmented decisions require a different kind of oversight than rule-based systems. You cannot audit an AI by reading its configuration file.

## 5.3 The Cascade Problem: AI Framing Shapes Human Judgment

Even when humans make the final decision, AI shapes how they see the problem. When AI assembles the context package for an analyst’s investigation, what AI includes defines the investigation scope. What AI omits may never be considered. AI’s initial classification anchors human judgment, and analysts are more likely to confirm an AI assessment than to contradict it, a well-documented cognitive bias. AI-generated investigation summaries become the official record relied upon in post-incident reviews and regulatory disclosures.

Effective human-in-the-loop design gives the human enough independent context to genuinely evaluate the AI’s conclusions. Leading MDR providers are beginning to engineer for this explicitly, using evidence-based verdict systems that separate the AI’s observable findings from its conclusions, allowing analysts to engage with both.



# 6. Making the Operating Model Decision

## 6.1 Why You Cannot Defer This Decision Anymore

The traditional approach to security tooling allowed for staged decision-making: buy the SIEM, figure out staffing later. AI removes this option. When you deploy AI-enabled security tools or services, you are making implicit commitments that compound over time. Changing them later means retraining models, rebuilding workflows, and losing institutional knowledge embedded in the AI's configuration. The switching costs are higher than traditional tooling, and they are often invisible at purchase time.

## 6.2 Three Operating Models

| Model | Description | Best Suited For | Key Requirement |
|---|---|---|---|
| **Model A: Internal AI SOC** | You deploy AI SOC tools and own configuration, tuning, decision governance, and the AI vendor relationship. | Mature SOC (11+ analysts); strong detection engineering; appetite for AI governance complexity; unique requirements poorly served by generic MDR. | The expertise to govern AI decisions, which is a different and more demanding capability than operating security tools |
| **Model B: AI-Enabled MDR** | You engage an MDR that uses AI; accept their AI governance framework; maintain visibility and oversight rights; hold them accountable for AI-driven outcomes. | Organisations without dedicated SOC capacity; preference for predictable costs; environments well-served by combined generic and tuned MDR approach. | Strong vendor management capability and clear contractual protections, especially around AI model changes and decision accountability. |
| **Model C: Hybrid** | Combine tools and services with clear boundaries: AI SOC tools for domains you understand well, MDR for specialised expertise, explicit handoff protocols, unified visibility. | Uneven maturity across security domains; complex multi-stack environments; transition periods when building internal capability while maintaining coverage. | Boundary clarity. Hybrid failures almost always come from ambiguous ownership at handoff points. |

## 6.3 Maturity Alignment Framework

| Current State | Recommended Model | Key Risk |
|---|---|---|
| **No dedicated SOC** | AI-Enabled MDR | Over-reliance without verification |
| **Part-time security (1–2 FTE)** | AI-Enabled MDR with active oversight | Losing visibility into AI decisions |
| **Small SOC (2–10 FTE)** | Hybrid: MDR + selective AI tools | Unclear accountability at boundaries |
| **Medium SOC (11–25 FTE)** | AI SOC with MDR backstop | Internal AI expertise development lag |
| **Large SOC (26+ FTE)** | Internal AI SOC | AI supply chain and talent retention |

## 6.4 The Questions That Actually Matter

| Instead of asking... | Ask instead... |
|---|---|
| How many alerts can you handle? | Who owns decisions when AI is in the loop? |
| What is your analyst-to-customer ratio? | How does AI change what your analysts do for me specifically? |
| What are your SLAs? | Do your SLAs cover AI decisions or only human decisions? |
| What technology do you use? | What AI decisions does your technology make, and can I see them? |
| What is your price per endpoint? | What happens to my price if AI infrastructure costs change? |
| Can you integrate with my stack? | What operating model commitment am I making by choosing you? |

# 7. Conclusion: The Operating Model Decision You Are Already Making

## 7.1 Revisiting the Thesis

AI makes expert analysts more effective, enabling MDRs to deliver personalisation at scale. But this only works if you choose an MDR that uses AI to go deeper, not merely to cut costs, and if you build governance around AI decisions with the same rigour you apply to human ones.

Most organisations are not yet asking who owns AI-driven decisions about their security. If you cannot see those decisions, explain them, or override them, you do not own them, regardless of what your contract says.

The tool/service boundary has already collapsed and the human's role is already changing. The question is whether you are managing these transitions deliberately or by default.

## 7.2 What Good Looks Like

### For AI-enabled MDRs

- Clear accountability for AI-driven decisions in contracts and SLAs
- Customer visibility into AI behaviour, confidence levels, and suppression decisions
- Evidence-based explainability: observable artefacts, not scores alone
- Demonstrated adaptability: AI that improves as it learns your environment
- Honest communication about AI limitations, failure modes, and supply chain dependencies
- Evidence that AI is making experts more effective rather than replacing them

### For direct AI SOC deployments

- Transparency about what decisions AI makes and where humans are required
- Controls that allow you to adjust AI confidence thresholds and autonomy boundaries
- Audit trails that distinguish AI-generated content from human analysis
- Clear information about AI supply chain dependencies and change management
- Feedback mechanisms that allow you to influence AI behaviour over time

## 7.3 Final Recommendations

The organisations that will navigate this transition successfully are those that recognise AI changes the governance question alongside the capability question, and plan accordingly.

Make the operating model decision explicitly. Do not let tool purchases or MDR selections make it for you by default.

Evaluate AI-enabled offerings on decision governance as much as capability. Can you see, understand, and influence how AI triages alerts in your environment?

Treat the AI supply chain as a first-class risk. Your MDR's dependence on foundation model providers is your dependence, even if indirect.

Expect force multiplication, not replacement. If a vendor promises analyst replacement, they are either overselling or underdelivering on quality.

Build verification into your operating model. AI makes oversight harder; engineer for it explicitly.

# Appendix A: CISO Evaluation Checklist

Use the following checklist when evaluating any AI-enabled MDR or AI SOC platform. Answers of "no" or "unclear" warrant deeper scrutiny before proceeding.

### Decision Ownership

- Can you show me every AI-driven decision that occurs before a human sees anything?
- Is there a visible audit trail for suppressed or auto-closed alerts?
- Do your SLAs explicitly cover AI-driven decisions?
- Does your contract address liability when AI makes an incorrect decision?
- Do I have approval rights if you change your AI models mid-contract?
- Can I adjust AI confidence thresholds or autonomy boundaries?
- Can I override AI decisions without breaking the workflow?
- Who owns the AI's learned knowledge of my environment if I switch providers?

### Explainability

- Can I audit the AI's reasoning on a closed case after the fact?
- Is evidence preserved in a form that supports post-incident review?
- Are AI-generated findings clearly distinguished from human analysis in reports?
- Can you explain a deprioritised alert to a regulator or auditor?
- What AI models do you use, and what data were they trained on?
- How do you validate against systematic model bias?

### Failure Behaviour

- What is the fallback when AI confidence is broadly low?
- How do you detect systematic AI failures?
- How do you handle adversarial attempts to manipulate AI behaviour?
- How do you notify customers when AI behaviour changes significantly?

### Failure Behaviour

- Can you share an example of a significant AI failure and how you addressed it?
- What is your process for rolling back AI changes that degrade performance?

### AI Supply Chain

- What third-party AI models or services does your platform depend on?
- How do you manage upstream model updates before deploying to production?
- Do you have contingency plans if a key AI supplier restricts usage?
- How does AI infrastructure cost variability affect my contract price?
- Do you have SLAs with your AI infrastructure providers?
- How do you validate that supplier changes do not alter detection behaviour?

### Adaptability and Force Multiplier Evidence

- How much dedicated analyst time does my account receive?
- What is the ratio of AI-automated work to expert-driven work?
- Can you demonstrate how your AI adapts to my environment specifically?
- How does AI behaviour improve as it learns my environment over time?
- How does your platform detect and correct model drift?
- What feedback mechanisms allow me to influence AI behaviour?
- Can I see accuracy metrics and how they change over time?

# Appendix B: Glossary

Use the following checklist when evaluating any AI-enabled MDR or AI SOC platform. Answers of "no" or "unclear" warrant deeper scrutiny before proceeding.

| Term | Definition |
|---|---|
| AI SOC | A security operations centre architecture in which AI plays a primary role in alert triage, investigation, and (in some cases) autonomous response. Distinct from a traditional SOC with AI tools bolted on. |
| Agentic AI | AI systems capable of executing multi-step tasks autonomously. In security operations, this may include automated containment, investigation steps, or external system communication. |
| Alert Suppression | An AI-driven decision to not surface an alert to a human analyst. High-risk because it is invisible: if wrong, discovery typically happens during incident response. |
| Autonomous Response | AI-initiated action taken without explicit human approval at the point of execution. Distinct from AI-recommended response, where a human approves before action is taken. |
| Confidence Threshold | The minimum AI confidence score required to trigger a particular action. A governance-critical parameter that should be customer-configurable. |
| Context Bootstrapping | The process of capturing sufficient organisational knowledge to enable AI to reason accurately about a specific environment from the outset of a service engagement. |
| Force Multiplier | AI that increases the effectiveness and reach of expert analysts without replacing them. Contrasted with AI deployed primarily to reduce headcount. |
| Force Multiplier | A large-scale AI model (such as those from OpenAI, Anthropic, or Google) used as the basis for downstream applications. MDRs that depend on third-party foundation models carry supply chain risk. |

| Term | Definition |
| --- | --- |
| Human-in-the-Loop | What happens to my price if AI infrastructure costs change? |
| Investigation Framing | The set of contextual information assembled by AI to present to an analyst. Because framing shapes conclusions, AI-generated investigation frames can introduce anchoring bias even when the analyst believes they are making an independent judgement. |
| MDR | Managed Detection and Response. A managed security service providing continuous monitoring, threat detection, and incident response. Distinguished from traditional MSSP by emphasis on detection quality and active response. |
| Model Drift | The degradation of AI model performance over time as the threat landscape, customer environment, or upstream model characteristics change. A first-class operational risk in AI-enabled security services. |
| MSSP | Managed Security Services Provider. A managed security service provider typically focused on device management, monitoring, and compliance. MDR is a more specialised category within the broader MSSP market. |
| Probabilistic Decision-Making | Decision logic that assigns confidence scores to outcomes rather than deterministic true/false rules. AI systems are inherently probabilistic; traditional security detection rules are deterministic. |
| SOAR | Security Orchestration, Automation and Response. A category of security technology that automates and orchestrates security operations workflows. Precursor to AI SOC approaches; differs in that SOAR automation is rule-based rather than AI-driven. |

# Appendix C: Methodology

This whitepaper is informed by a combination of direct industry engagement and open-source research.

- Over 4,000 interactions with security leaders, practitioners, and vendors across enterprise SOCs, MDRs, and security technology providers, conducted by Oliver Rochford over multiple years.
- Open-source intelligence from public reports, vendor documentation, analyst research, conference talks, and incident disclosures.
- Perspectives gathered through invite-only practitioner communities where security leaders discuss operational challenges outside formal marketing environments.
- Synthesis of AI SOC go-to-market strategy analysis and vendor positioning research conducted for clients including multiple MSSP and MDR providers.
- SANS 2023 SOC Survey data for practitioner challenge benchmarks.
- Structured discussions with SOC practitioners used to stress-test frameworks and ensure recommendations reflect real-world operational constraints.

The development of this whitepaper followed a structured approach: Research and Synthesis (Weeks 1–3) covering industry interviews, practitioner input, OSINT review, and synthesis of prior engagements; Drafting (Week 4) covering core narrative development, frameworks, decision models, and evaluation criteria; and Review and Final Edits (Weeks 5–6) covering feedback integration, language refinement, and editorial polish.

# Appendix D: Case Studies

### A. Methodology

This paper draws on three evidence streams:

1. OSINT corpus. Public practitioner commentary collected from Reddit cybersecurity communities (r/cybersecurity, r/Information_Security). All posts were less than one year old at the time of collection. Contributors self-identified as SOC analysts, MSSP operators, detection engineers, and staff-level security engineers. Posts were coded inductively against the paper's core themes.

2. Semi-structured interviews. Two extended interviews with practitioners who have deployed AI SOC platforms in production (not pilots). Informants were selected through theoretical sampling to span different market segments. Both are anonymised below.

3. Vendor interactions. 30+ vendor briefings conducted over 18 months, referenced in the body text but not reproduced here.

Gartner's 2025 Hype Cycle for Security Operations places AI SOC agents at the Innovation Trigger stage with 1–5% market adoption (Nunez & Livingstone, 2025, ID G00825402). Finding production-deployed practitioners to interview was difficult in itself, which corroborates the low adoption figure and explains the small interview N. The two informants were selected to maximise contrast across organisation size, team structure, and market segment.

### B. Informant Profiles

1. OSINT corpus. Public practitioner commentary collected from Reddit cybersecurity communities (r/cybersecurity, r/Information_Security). All posts were less than one year old at the time of collection. Contributors self-identified as SOC analysts, MSSP operators, detection engineers, and staff-level security engineers. Posts were coded inductively against the paper's core themes.

| | INF-01 | INF-02 |
|---|---|---|
| Role | CISO | Sole security practitioner |
| Organisation | European mobility company, 60+ franchise and corporate locations, multi-cloud | US-based conservation NGO, ~1,000 users, 20–30 locations |
| Team | Dedicated SecOps team (multiple analysts) | One-person operation (policy through to operations) |
| AI deployment | AI-driven MDR service running in production alongside and then replacing previous human MDR | Design partner with AI SOC startup since May 2025; running in production parallel with $60K/yr MDR |
| Prior MDR experience | External human MDR; team would "do the investigation again because they wouldn't trust them" | Pass-through alerting: MDR ingested data but performed no correlation or investigation, just forwarded alerts |
| Background | Enterprise security leadership | Detection engineering (ex-Crowd Strike internal detection team) |

## C. Interview Evidence Summarised by Theme

### C.1 The "80% vs 0%" argument

INF-01 framed AI accuracy against the alternative of not investigating at all:

> *"80% is better than 0% accuracy, right? So if you don't look at something, I would rather have something look at it with the 80% accuracy and not just ignore it."*

He argued that human analysts also make mistakes, are not auditable in the same way, and suffer fatigue and churn. The AI's advantage was consistency at scale, not perfection.

### C.2 Both informants draw the same line at no autonomous response

INF-01 restricted AI to detection and investigation only:

> *"We don't do any automated write access stuff on anything. Not remediation, not patching, not anything. The potential to break stuff is so large that you could end up with more damage — self-inflicted damage."*

His team operated in "hybrid mode": AI handles scale (triage, enrichment, investigation steps), humans handle response, containment, and edge cases.

INF-02 drew the same boundary independently, framing it differently:

> *"You can't hold a computer accountable, right, like IBM says? [...] If it comes down to that point in investigation, it rises to the point where a human needs to be in that loop."*

Response actions existed in the platform but required a human to execute them. Both informants arrived at this constraint from different starting points (enterprise risk management vs. accountability principle) but reached the same architecture.

### C.3 Explainability as trust infrastructure

INF-01 described explainability as his key requirement from day one:

> *"I want to see why conclusions were derived. I want to be able to see all the reasoning flow for every incident."*

He drew a direct parallel to the old MDR problem: his team used to receive verdicts from external analysts and then redo the investigation because they did not trust the reasoning. The AI's auditable chain solved this. When the platform recommended aggressive actions (factory reset on a VP's laptop), the team could present the full reasoning as justification. Trust came from transparency, not from vendor assurances.

INF-02 described a related but distinct mechanism: the platform's built-in AI chat let non-technical stakeholders query investigations directly and get explanations at their level. Explainability was not a reporting feature; it was how people outside the SOC could participate in decisions.

### C.4 Honest uncertainty means a system that says "I don't know"

INF-02 described the platform returning "inconclusive" verdicts:

> *"There's scenarios where it comes back with either inconclusive or that it should be something that should be monitored. And I think nine times out of ten, the reason why it comes back that way is because it doesn't have enough data to make a decision."*

He treated this as a legitimate output, not a failure. The system acknowledged its own limits rather than forcing a binary call. This maps to the "tri-state or multi-state outcome" pattern identified in the OSINT corpus as a marker of well-designed architecture.

### C.5 The agent as attack surface

INF-02, drawing on his detection engineering background, raised a concern absent from vendor marketing: prompt injection through attack data.

> *"If a malicious actor is prompt injecting through their attack [...] the system is pulling in that data [...] and it has the ability and the capability to do actions on its own with no human in the loop — well, then you basically provided your own insider risk at that point."*

The tighter the coupling between AI inference and automated action, the greater the blast radius of adversarial manipulation. This extends the workflow coupling failure mode into an adversarial dimension.

### C.6 Scope expansion is where the real value showed up

INF-01 described a case where the AI deployed detection rules his team had previously considered impractical due to false positive rates. The system correlated HR data, LinkedIn profiles, and authentication logs across franchise locations to identify systematic credential sharing between branches, a pattern no human analyst would have investigated at scale.

> *"My team would never be able to go to that depth of investigation for such a low-risk topic. To me, this is the real value — because we can tremendously expand the scope and the depth of investigations we are currently doing."*

### C.7 Detection engineering is unlocked

A second-order effect in INF-01's environment: deploying new detection rules used to require "two to three months of testing to see that not too many false positives are generated." With AI-driven triage absorbing the false positive load, his team could deploy experimental detections without that overhead:

> *"Right now we can just try. We can just throw it to the thing, see what comes back. And if it doesn't work, it's an agent. It's not like I have to waste two FTEs to take care of some false [positives]."*

### C.8 The one-person SOC

INF-02 was a sole practitioner responsible for everything from policy to operations across ~1,000 users and 20–30 locations. His previous MDR ($60K/yr) was pass-through alerting: it ingested data from Defender, AWS, and GCP but performed no correlation. An analyst might spend a few minutes reviewing an alert before forwarding it.

> *"For me, that's not a very good approach to it, right? Especially as being a one-person team, I need to have a lot of that already done so that when it comes to me [...] I can just make a quick decision based off of all the information, the context, the correlation."*

The AI platform cost 50% less than the MDR and freed budget to invest in vulnerability management (Tenable).

### C.9 Workforce disruption is the surprise neither informant planned for

INF-01 described organisational impact he did not anticipate:

> *"One of the things that surprised us [...] it required our SecOps team to go through some kind of evolution. People suddenly had no job."*

Roles focused on phishing triage, header analysis, and DMARC verification were automated within weeks. Staff were retrained into cloud threat investigation. Threat hunting plans were reconsidered. The transition was reactive:

> *"I would have invested much sooner in creating new positions or new titles in the team [...] because what we did is we incorporated AI, got it to a stage where it's really working, and then said, okay, now what do we do with the people."*

His hiring had changed: "I think it's going to be a while before I hire somebody new."

# ABOUT CYBERFUTURISTS

Oliver Rochford is the Lead Analyst at the Cyberfuturists, a boutique research and advisory firm providing expertise and intelligence to decision-makers and investors on cybersecurity market and industry trends. A former Research Director at Gartner, Securonix, and Tenable, he co-defined the SOAR (Security Orchestration, Automation and Response) market category and has published extensively on security operations strategy. He has conducted over 4,000 engagements with security practitioners, vendors, and enterprise security leaders over his career.

**Oliver Rochford**
Lead Analyst, Cyberfuturists

Sponsored by Daylight Security

# ABOUT DAYLIGHT SECURITY


Daylight is a security services company delivering Managed Agentic Security Services (MASS), including MDR, threat hunting, incident response, and more, through a fundamentally different architecture than traditional security services providers. Daylight's architecture combines an agentic platform that runs the full cycle from detection to response with security experts from IR and threat hunting backgrounds. The platform integrates deeply across your environment - cloud, identity, SaaS, endpoints - and collects identity and business context to investigate alerts the way a senior analyst would. It continuously learns your environment to make better decisions over time. Security experts validate decisions, feed insights into the platform, optimize detections, and take over in case of an incident.  The result: security teams move from firefighting mode to strategic work that improves their security posture.