

SECURITY STRATEGY · BRIEFING DECK

Agentic SecOps

Transforming red / blue / purple teaming with autonomous AI — a continuous, governed, human-led tradecraft.



RED



BLUE

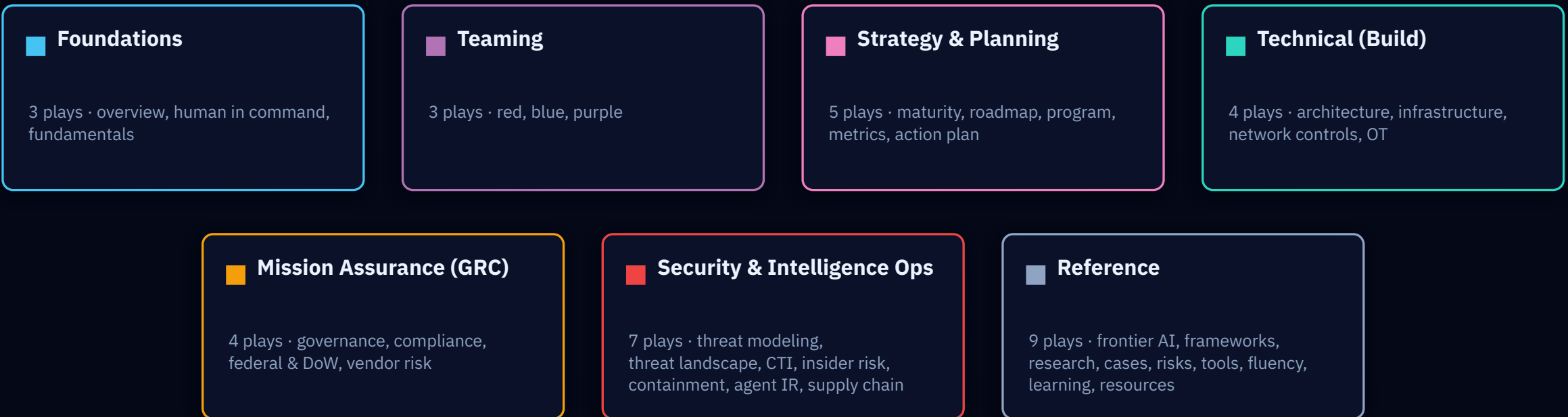


PURPLE



35 Plays Across 7 Themes

A companion interactive playbook backs this deck — here is the whole body of work at a glance.



35 plays · 7 themes · one continuous, governed, human-led approach.

MEET THE THREAT

APT_UNICORN



The best hire your SOC ever made — onboarded in seconds, works every hour of every day, reads every log, writes every detection, never burns out.

Also the most dangerous insider you ever handed network access: the same speed that clears your alert queue can map your environment and walk it out the door before a human notices the first anomaly.

Same horn. Same horsepower. The only variable is who holds the reins.

This deck is about holding the reins.

APT_UNICORN is an illustrative framing device, not a tracked threat-actor designation.

THE PROBLEM

Reaction Time Has Collapsed

80–90%

of a real intrusion run autonomously by an AI agent (GTG-1002, per Anthropic)

1 → 9

one operator breached nine government agencies (Mexico, 2025–26, per Gambit Security)

hours → seconds

mean-time-to-compromise is collapsing (per Armis, 2026)

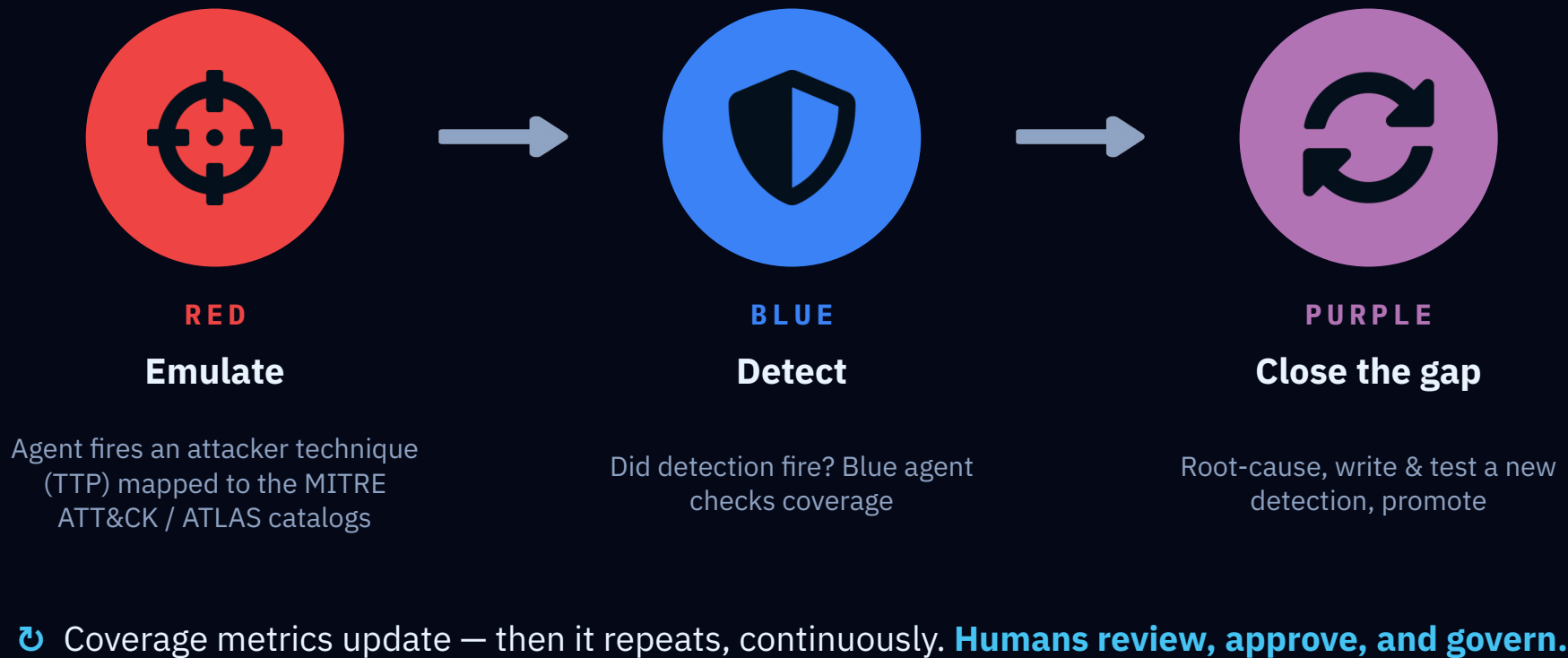
~50%

prompt-injection success within ten tries (per International AI Safety Report, Nov 2025)



You cannot out-react agent-speed offense with human-speed defense.

A Continuous, Self-Improving Purple Loop



One Capability, Three Faces



Defensive (Blue) Tool

Autonomous triage, continuous hunting, real-time detection engineering, and adaptive response — compressing the work that overwhelms human teams.



Offensive (Red) Tool

Sanctioned adversary emulation, autonomous pen-testing, and continuous TTP validation — finding the gaps before an adversary does.



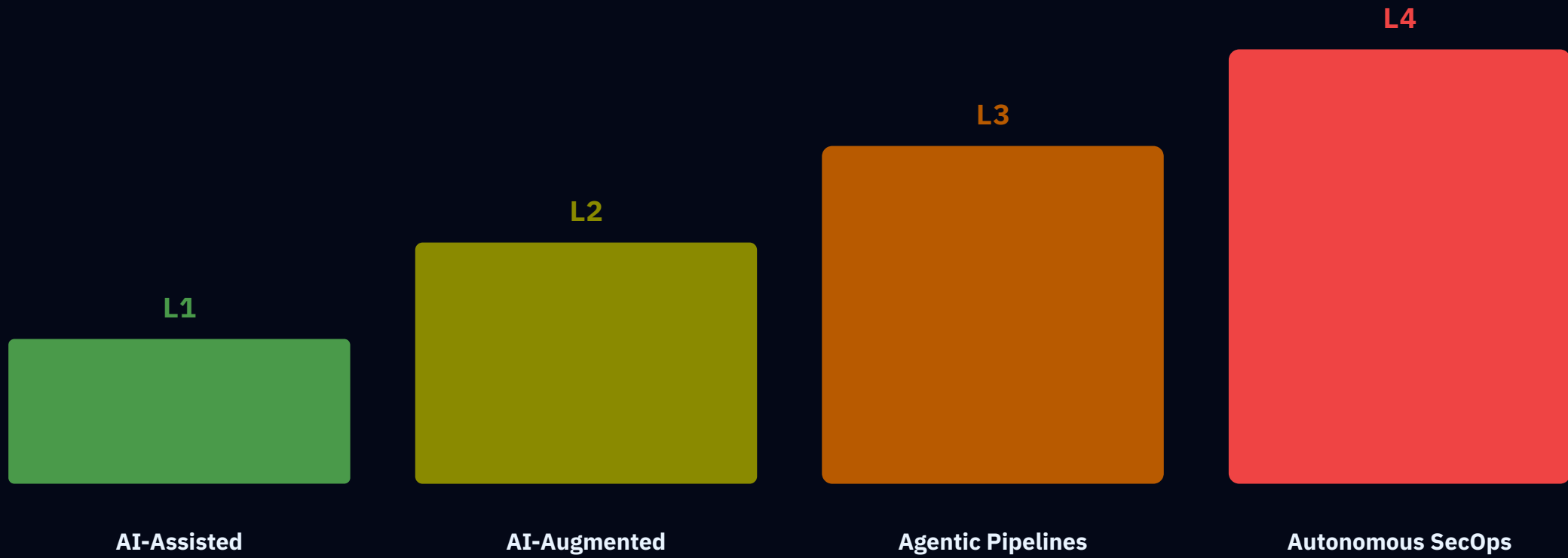
Weaponized Against You

The same capability turned on the organization at machine speed. No longer hypothetical: GTG-1002 (per Anthropic); nine Mexican agencies breached (per Gambit Security).

The technology does not change between roles — only intent and guardrails do.

Crawl, Walk, Run — A Sequenced Path

Foundations first: AI multiplies what is underneath, so the fundamentals come before the autonomy. Benchmark the climb against the Accenture + CMU SEI AI Adoption Maturity Model, the enterprise-wide frame this SecOps ladder nests inside.



Human-Led, Foundations-First, Contained



Human in Command

Humans own strategy, policy, scope, and approvals; agents do the repeatable work inside those boundaries.



Foundations First

Sequence AI onto solid triage, alert tuning, detection discipline, and a data fabric built for quality.



Govern as Non-Human Insiders

Treat each agent as a privileged insider: unique identity, least privilege, continuous monitoring (Agentic Insider Risk).



Containment Before Cleverness

Bound the blast radius with deterministic controls — sandboxing, egress limits — not model promises.

Tailored for Every Room



Executive / Leadership

The decision, the ROI, and the risk of inaction — what to fund and why now.



Peer / Technical

The architecture, the threat landscape, and the adversarial detail for practitioners.



General Overview

An audience-agnostic primer on the whole approach, start to finish.

Each section stands alone — pull the one you need for the room you are in.



SECTION 01 • EXECUTIVE / LEADERSHIP

The Decision, Not the Demo

01

Agentic SecOps as a governed investment in defensive speed: what to fund, why now, and what it returns.

IN THIS SECTION Why now • The return • The investment • The phased path • The 12-month picture

The Road to Assume-Agentive, Assume-Inside

Two years of public cases — the threat got faster, AI-driven, and already inside; the federal response followed.



MIT FutureTech (2026), 272-expert panel: National Security ranks among the three most AI-vulnerable sectors, both a target and a domain where failures cascade.

The trajectory is unmistakable — faster, AI-driven, already inside. This is the case for funding defensive speed now.

The Threat Already Operates at Machine Speed



Autonomous intrusions are real Per Anthropic, GTG-1002 ran 80–90% of an espionage intrusion through an AI agent, with minimal human direction.



One operator, many targets Per Gambit Security, one operator using two commercial AI tools breached nine Mexican government agencies.



The window is closing Adversaries are adopting agents now — defensive parity is a time-boxed opportunity, not an open one.

What Agentic SecOps Buys You

Weeks → hours

Detection-engineering and validation cycles compress dramatically

24 / 7

Continuous hunting and triage — no queue, no burnout

Toil → judgment

Analysts move from clearing alerts to making decisions

Measured

Coverage is quantified and closed every cycle — not assumed



The return is defensive speed and analyst leverage — measured, not promised.

Foundations, People, and Guardrails



Foundations first

Solid triage, alert tuning, detection discipline, and data quality — AI multiplies these, so they come first.



People, repositioned

Analysts shift to detection engineering, agent supervision, and hunting. Reskilling, not replacement.



Governance & guardrails

Agents run as governed non-human insiders: identity, least privilege, human approval, containment.



Phased funding

Fund by maturity stage and gate on outcomes — a sequence of small bets, not a big-bang wager.

Crawl, Walk, Run — A Phased, Fundable Plan

1

CRAWL

L1 → L2

AI-assisted analysts; foundations hardened; first bounded agent use cases.

Gate *Stable triage and measurable time saved.*

2

WALK

L2 → L3

Integrated agents and a governed, continuous purple loop running in production.

Gate *Coverage measured; agents governed.*

3

RUN

L3 → L4

Selective autonomy in bounded domains, expanded only as trust is earned.

Gate *Trust calibrated; containment verified.*

What Good Looks Like in 12 Months

Twelve months in, with phased investment:

- ✓ A governed purple loop runs continuously — coverage measured every cycle
- ✓ Analysts work detection engineering and hunting, not queue-clearing
- ✓ Every agent operates as a governed non-human insider, with containment verified by red-team test
- ✓ Security posture is briefable to leadership in metrics, not anecdotes

 **The ask: fund Crawl now; gate Walk and Run on outcomes.**



SECTION 02 · PEER / TECHNICAL

Under the Hood

The architecture, the adversary tradecraft, and how red, blue, and purple actually run with agents.

IN THIS SECTION Architecture · Adversary tradecraft · AI workforce · Purple loop · Non-human insiders · Frameworks

02

How the System Is Built



Agent orchestration gates. Planner plus agents calling tools via the Model Context Protocol (MCP), acting through human-in-the-loop gates.



Red / blue / purple agents Emulate, detect, and close the gap — the autonomous loop that drives coverage upward.



Data & telemetry fabric SIEM, XDR, and logs (e.g., Microsoft Sentinel, Defender XDR) unified into one queryable plane the agents reason over.



Governance & identity plane Non-human identity, least privilege, full audit, and deterministic containment around every agent.

What We're Defending Against



Autonomous attack chains

GTG-1002-style operations: agents plan, exploit, move laterally, and exfiltrate at machine speed (ATT&CK + ATLAS).



Prompt injection & tool abuse

The new attack surface — indirect injection, tool and function abuse, and model manipulation.



Managed attribution & synthetic identity

Adversaries (and red teams) hide behind non-attributable infrastructure and synthetic personas.



Agentic insider risk

A compromised or over-permissioned agent behaves like a privileged insider — at scale and speed.

One Operator, a Cast of Agents

Model the single operator’s AI “workforce” as discrete agent-roles — then map each role to a technique and a control.

AGENT ROLE	WHAT THE AGENT DOES	MAPS TO (ATT&CK)	DEFENDER’S ANSWER — DETECT & CONTAIN
Recon / OSINT	Automated target profiling, infrastructure mapping, data-leak harvesting	Reconnaissance (TA0043)	Attack-surface monitoring; rate-limit and deny mass enumeration
Social engineering	Tailored lures, deepfake pretext, multilingual phishing at scale	Initial Access – Phishing (T1566)	Inbound anomaly and deepfake detection; phishing-resistant auth
Malware development	Generate, debug, and obfuscate payloads to evade signatures	Resource Development (TA0042)	Behavioral EDR and sandbox detonation; no signature-only reliance
Infrastructure / C2	Spin up proxies, bulletproof hosting, and resilient C2	Command & Control (TA0011)	Egress and beaconing analytics; egress allow-listing; DNS controls
Fraud / extortion	Negotiate, run money-mule logistics, pressure victims	Impact – financial (TA0040)	BEC and anomalous-payment analytics; transaction holds
Exfiltration	Stage and move data faster than humans can react	Exfiltration (TA0010)	DLP and egress-volume baselines; egress caps and quarantine



Flip the operator: the same cast, over-permissioned inside your org, is **agentic insider risk** — govern each as a non-human identity. When the agents are yours, ATLAS techniques (prompt injection, tool abuse) turn them against you.

Detection-as-Code, Continuously Validated



🔄 Humans approve promotions; the loop then repeats continuously.

Identity, Least Privilege, Containment



Unique identity per agent

Every agent gets its own non-human identity — no shared service accounts, full attribution.



Least privilege & JIT

Scoped, time-bound permissions; no standing admin rights for any agent.



Continuous monitoring / UEBA

Behavioral baselines and pattern-of-life applied to agents, not just human users.



Deterministic containment

Sandboxing, egress controls, and kill-switches bound the blast radius regardless of model behavior.

Frameworks and Honest Limits



Frameworks we map to

- MITRE ATT&CK + ATLAS
- OWASP — LLM & agentic risks
- NIST AI RMF + SP 800-63
- SANS AI-security courses SEC535 / SEC598
- MAESTRO (Cloud Security Alliance agentic threat model)



Honest limits

- Models are nondeterministic
- Over-trust / over-automation is the meta-risk
- Agents require continuous evaluation
- Autonomy stays bounded by design



SECTION 03 • GENERAL OVERVIEW

The Whole Story, Briefly

An audience-agnostic primer on Agentic SecOps — what it is, why it matters, and how it's done responsibly.

IN THIS SECTION What it is • Why it matters • How it works • Doing it responsibly

03

Agents That Do Security Work

AI agents that don't just answer questions — they take action: triage alerts, hunt threats, emulate attacks, and write detections, under human direction.



They act, not just answer

Agents take real steps in the SOC, end to end.



They work the loop

Emulate, detect, and close gaps — continuously.



Humans stay in command

People set scope, policy, and every approval.

Speed Cuts Both Ways



Without governed agents

- Human-speed defense falls behind machine-speed attacks
- Analysts drown in alert volume
- Coverage gaps go unseen until they're exploited



With governed agents

- Defense matches attacker tempo
- Toil is automated; people focus on judgment
- Gaps are found and closed every cycle — under human control

The Loop, and How You Grow It



Emulate

A red agent safely mimics a real attack technique.



Detect

A blue agent checks whether defenses caught it.



Improve

Gaps are root-caused and closed — then it repeats.

🔄 And you grow it deliberately — **crawl** → **walk** → **run**.

Human-Led, Governed, Contained



Humans decide

People own strategy, scope, and every consequential approval.



Foundations first

Strong fundamentals come before any autonomy — AI multiplies them.



Agents are governed insiders

Unique identity, least privilege, and continuous monitoring for every agent.



Contained by design

Deterministic limits bound what any agent can do, whatever the model decides.

THE TAKEAWAY

You can't out-react agent-speed offense with human-speed defense.

So put governed agents on defense — and mature it deliberately.



RED



BLUE



PURPLE

The full depth lives in the interactive playbook — one story, three lenses.

OCTOBER 2026 REFRESH — HIGHLIGHTS

- › **MCP as a trust boundary**: a new Reference Architecture action closes the gaps NSA names; the glossary gains least agency.
- › **Sixteen new sources**: including NSA on MCP security, the OWASP Agentic Top 10, NIST on agent identity and zero trust builds, and eight [un]prompted 2026 talks.
- › **The approval figure**: Anthropic's August 2026 post, 97% of Claude Code permission prompts approved, now at all four sites.
- › **Dates pinned**: the MDASH and threat-hunting preview lines carry their dates, and the SIFT demo is placed at its [un]prompted talk.
- › **Checks that wait**: the render and anchor checks wait on results, and every tab's counter is checked at its position.

87

RECOMMENDED ACTIONS

243

SOURCES, ALL LINKED

89

GLOSSARY TERMS

PLAYBOOK

Agentic SecOps

© 2026 Ken Connell. All rights reserved.

Built with Claude and Claude Code, reviewed by a human who owns the mistakes.



RED



BLUE



PURPLE



Third-party product, tool, and framework names referenced in this playbook are trademarks of their respective owners. A personal project: views are my own and do not represent any employer, client, or vendor.