DEF CON 34 / BIC VILLAGE / AUG 2026

Agents on Alert

Building an Agentic Framework for Detection & Response

Agents can investigate. They can't be accountable.

Samson Adewale

Senior Security Engineer · Threat Detection & Response

More detections.

Vendor built-ins. Detection engineers. Threat-hunt output. Every source, one goal: find bad.

More detections.

New rules run hot.

A new rule fires on normal activity long before it fires on a real threat.

More detections.

New rules run hot.

Analysts absorb it.

Dig logs. Forty minutes gone.

More detections.

New rules run hot.

Responders absorb it.

Capacity caps out.

Coverage scales. Investigation capacity doesn't.

AI's report card.

Four stages of work. AI is not equally good at any of them.

SEE - detect the signal

IMMATURE

Data architecture and detection dev - still human

UNDERSTAND - gather context

STRONG

Gather cross-source context - who, what, why

ACT - respond & remediate

DANGEROUS

Real incident response - cross-team collaboration

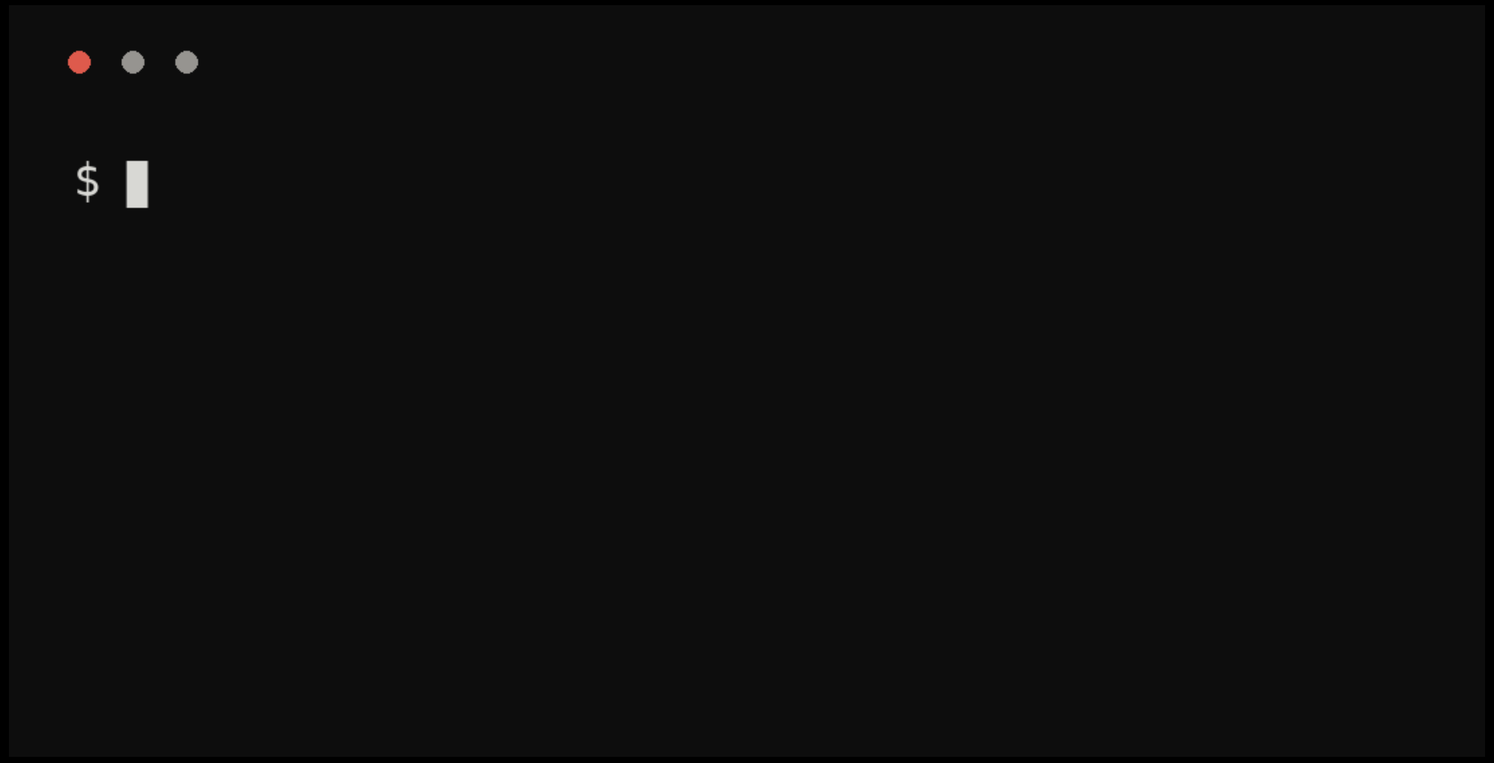
DECIDE - reach a verdict

PLAUSIBLE

Can produce a verdict - human validation required

It's a data problem

before it's an AI problem.



Before an agent can help, four things have to exist.

01	**TELEMETRY** breadth	every plane you can reach

Before an agent can help, four things have to exist.

02	**RETENTION** depth	long enough to BE a baseline
01	**TELEMETRY** breadth	every plane you can reach

Before an agent can help, four things have to exist.

03	**KNOWLEDGE** documents, not logs	tickets · Git · chat · runbooks · policy
02	**RETENTION** depth	long enough to BE a baseline
01	**TELEMETRY** breadth	every plane you can reach

Before an agent can help, four things have to exist.

04	**SEMANTIC MODEL** meaning	declared once, so nothing guesses
03	**KNOWLEDGE** documents, not logs	tickets · Git · chat · runbooks · policy
02	**RETENTION** depth	long enough to BE a baseline
01	**TELEMETRY** breadth	every plane you can reach

Anomaly detection needs 02. "Was this authorised?" needs 03. An agent that doesn't guess needs 04.

04 / HOW IT'S BUILT

What it actually took.

THE CASE

a synthetic attack with a ground-truth answer key

so the agent can be graded, not vibe-checked

THE DATA

9 tables · 4 domains · one semantic view each

deployed with Terraform - it rebuilds from nothing

TRIAL AND ERROR

weeks of it - most early runs were wrong

every guardrail in this build exists because something failed first

One alert. Three agents. One human.

Only the investigator holds tools.

ALERT → **TRIAGE** → **INVESTIGATOR** → **CASE UPDATER** → **HUMAN**

ALERT: webhook

TRIAGE: entities · hypotheses

INVESTIGATOR: asks · routes · correlates

CASE UPDATER: writes it up for a human

HUMAN: decides · acts

SNOWFLAKE CORTEX — one semantic view each, read-only

TINES — OSINT + internal docs

Tines orchestration · Snowflake Cortex over semantic views · scoped read-only role · guardrails in Terraform, not in prose

[LIVE]

Case METADATA

GuardDuty: temporary credentials issued to an EC2 instance are being used from an address outside AWS.

That's the whole alert.

WHICH ONE IS IT?

H1

AN ENGINEER

exported them to debug

H2

A SANCTIONED SYSTEM

does this by design

H3

STOLEN CREDENTIALS

someone else has them

ALT-2026-0805-0447 · ecs-prod-instance-role · i-0b41e9c77a3d5f210 · read-only

If you're going to build this today,

four things I'd tell you first.

01 **START WITH THE DATA**
give your agents the best chance to succeed

02 **MAKE IT PROVE IT LOOKED**
make every query reproducible - check its work, understand it yourself

03 **AGENTIC INFRASTRUCTURE AS CODE**
prompts too - one unintended change shifts how the model reasons

04 **STAY ACCOUNTABLE FOR YOUR AGENTS**
they propose. you answer for it.

Thank you.

Everything from this talk

code · scenario · agent instructions

samsonadewale.com

linkedin.com/in/samson-adewale-seceng