PROPRIETARY FRAMEWORK

SENTINEL

Securing AI agents, from access to consequence.

Abhinav — AI Strategy & Governance

Every permission looked fine on its own

1

No boundary

An agent doesn't stay inside the job it was built for. Each new tool connected looked like one small, reasonable yes.

2

No visibility

Security teams can show every permission granted, one at a time. Almost none can show what those permissions add up to together.

3

No timing

By the time the combined effect is visible, the agent has already acted. Not recommended, acted.

The question was never what this system can access. It's what it can cause, what we can prove, and what we'll change.

Every individual permission can be reasonable. The danger only shows up once you look at what the agent can do with all of them at the same time, and by then it's already happened.

Sentinel doesn't replace access control. It finishes it.

TRADITIONAL SECURITY

— Who has access

— What data can it see

— What authority it has

— Who approved it

SENTINEL ADDS

— What it can see, and whether it can be trusted (Context)

— What it can change, scaled to consequence (Action)

— Whether we can reconstruct what happened (Evidence)

— What changes when it teaches us something new (Adaptation)

The Sentinel chain

01

Identity

Who is acting

02

Authority

What it's authorised to do

03

Context

What it can see, and can that be trusted

04

Action & Consequence

What it can cause

05

Evidence

What it leaves behind

06

Adaptation

What changes when we learn

Action deserves the most attention. Adaptation is what stops the same failure happening twice.

THE METHOD

A structured read, not a network scan

1

Map the chain

Every agent's identity, authority and context, documented once.

2

Restrict the action layer

Read, write, send, delete, trigger, scoped to what's actually needed, not what's convenient.

3

Prove it, and teach it

Reconstructable within hours, and every incident feeds a governance-reviewed update to the baseline.

No new access, no new integrations. Delivered in weeks.

Four things land on the table

The agent inventory

Every agent in the organisation, its identity, its owner, named and mapped, not assumed.

The action risk register

Every agent ranked by what it can cause, not by what it can access.

The evidence architecture

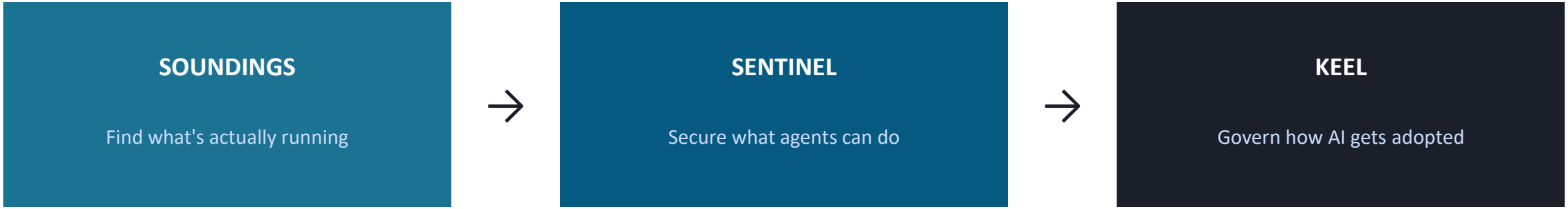
What gets logged, where it lives, and how fast it's reconstructable when something goes wrong.

The adaptation log

What was learned from each incident or near miss, and what governance changed because of it.

THE CONNECTION

Three frameworks, one architecture.



Soundings finds what's running and who knows about it. Sentinel secures what it's allowed to do once it's found. Keel governs how the whole programme scales from there. Skip the first two and governance gets written for agents nobody had actually mapped.

HOW IT'S DELIVERED

Same instrument, two ways in

INTERNAL

Run inside a company that employs you.

The first ninety days of a security or AI leadership role. Maps the real chain before an incident forces the question.

EXTERNAL

Sold as a fixed, scoped engagement.

Agent inventory, action risk register and evidence architecture, delivered as one document. Not ongoing monitoring, that's the next conversation.

ABOUT THE AUTHOR

Abhinav

AI strategy and governance leader, working across FSI, retail, luxury and financial services on AI adoption, security and responsible-use frameworks.

- AI security and governance build across regulated and unregulated sectors
- Frameworks that hold up to risk, compliance and information security review
- A cyberpsychologist's read on why people build workarounds when the process doesn't keep up

WANT TO KNOW WHAT'S ACTUALLY RUNNING?

Let's secure the chain.

A short, fixed-scope read on every AI agent running in your organisation, scored for what it can access, what it can cause, and whether it can learn. Delivered in weeks.

[Email placeholder]

[LinkedIn placeholder]