Runlayer

What's Actually Running in Your Coding Agent

September 2026 / The Shadow AI Stack

Agenda

- Problem
- Visibility
- Wrap Up

Alex Frazer

- Chair - AAIF Security & Privacy WG
- Shadow AI Endpoint Security Engineer
- Agentic SOC Security Research
- Threat Intelligence Operations
- Endpoint Security Research

Problem

Anyone can download/use AI artifacts from **anywhere**, without **any** visibility

- MCP servers
- Skills
- Plugins
- Rules/Instructions
- AI Clients
- Custom agents

A new telemetry layer on endpoint

- The risk is **intent** encoded in static **text**
- STDIO MCPs never touch the wire
- Skills/Instruction files never spawn a process

Problem

This means:

- MCP servers that you don’t know about see your data
- Skills, rules, and instruction files you don’t know about steer your agents
- Plugins you don’t know about bring MCP servers and skills you don’t know about that influence your agents and see your data

Problem

It gets worse:

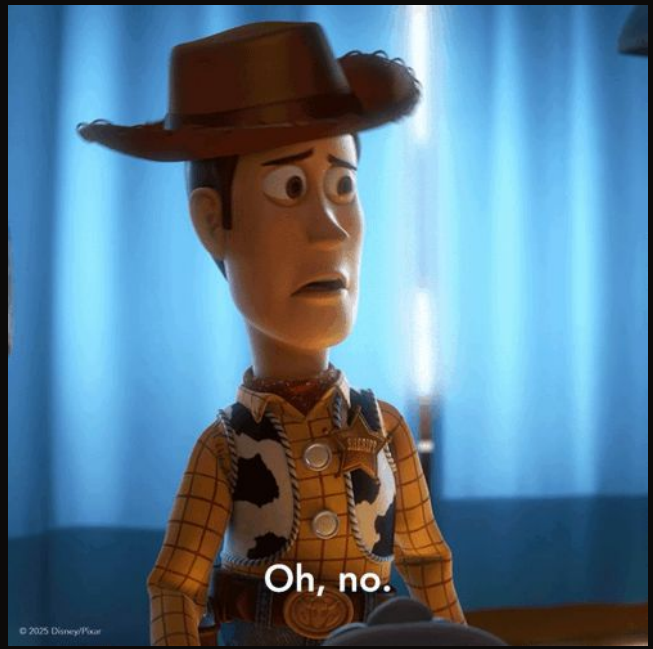
- Your endpoints have containers
- Windows endpoints also have WSL VMs
- Those WSL VMs have containers
- MCP configs, skills, rules, and instructions can live in repos, profiles, and devices
- These get loaded automatically by most coding agents
- There are containers elsewhere (k8s, k3s, etc)
- And it keeps going...

Problem

- Invisible with existing tooling
  - Network requires TLS intercept (non-trivial)
  - Endpoint doesn’t surface text file events and doesn’t have file content
  - Runtime only analysis is a last line of defense
  - And so on…

None of them tell the complete story:

- When did the artifact appear?
- Who brought it?
- Where else is it in your organization?
- What is it (the content)?
- Why is it malicious/bad?



Runlayer

Find it, find it all

Visibility

Every endpoint/container has:

- User folders for AI client apps
  - MCPs, skills, plugins
- “Enterprise” locations
  - Restricted access config for enterprise management
- Project folders
  - Arbitrary places based on where users store projects
- Container volumes
  - Expensive to get visibility when not running
- Running processes
  - MCPs listening on TCP ports
  - Custom agents
- Web access (perplexity, etc)

Visibility

More challenges:

- This has to run on every endpoint (deployment at scale)
- The landscape is ever evolving (updates)
- Open source options are partial solutions (completeness)
- Performance impacts on endpoints (not another agent)

Visibility

- It’s a full time job
- I would know
- it’s my full time job
- And my agents
- And their agents

And we haven’t even talked about control yet...

P.S. I’m hiring 🤣

Wrap Up

TL;DR

- Things you can’t see are influencing your agents
- You can’t control what you can’t see
- Complete, scalable visibility is non-trivial
- EDR doesn’t have the granularity for this

If you do one thing with all of this:

- Get visibility on your Shadow AI on your devices (and ideally your organization)
  - It will surprise you how much is there right now

Runlayer

Questions?

Contact Me