

AI Agent Governance

Extending IAM Principles to Autonomous Machine Identities

Abdel Fane · Founder & CEO, OpenA2A · GSX 2026, Atlanta

THE GAP

Your IAM team can name every human identity.

HUMAN IDENTITIES · VISIBLE TO YOUR IAM



AGENT IDENTITIES · INVISIBLE TO YOUR IAM



Abdel Fane

Founder & CEO, **OpenA2A**

Executive Director, **CyberSecurity NonProfit (CSNP)**

20 years securing enterprise environments across
healthcare, finance and government.

Booz Allen Hamilton · Allstate · Protiviti

What I build

Agent Identity Management (AIM)

open-source IAM for autonomous agents

HackMyAgent

security scanner and shadow-AI audit for agent deployments

All Open Source!

What you'll leave with

- 01 A command that finds your unmanaged agents**
Inventories shadow AI on any machine. No install, no server, no account.
- 02 A CI/CD pipeline that stops new ones from shipping**
The same scanner in your pipelines your policy can reference.
- 03 An identity model that governs agents**
Agent identity, capability grants, and an audit trail that maps to PAM and SIEM.
- 04 A safe range where you can test**
A training lab where you can test 21 agents for security vulnerabilities.

How can we build together

Two states. Only one governed.



You have tools your ITSec knows about



Registered identity with a keypair



A named human owner



Access granted to the agent, not borrowed



Explicit capability grant, all else denied



Revocable in one action



Signed audit trail per privileged call



You have tools your ITSec *doesn't* know about



Installed, never requested



No owner of record



Inherits a human's standing access



No capability boundary on what it can reach



Cannot be revoked, only deleted



Leaves no evidence of what it did

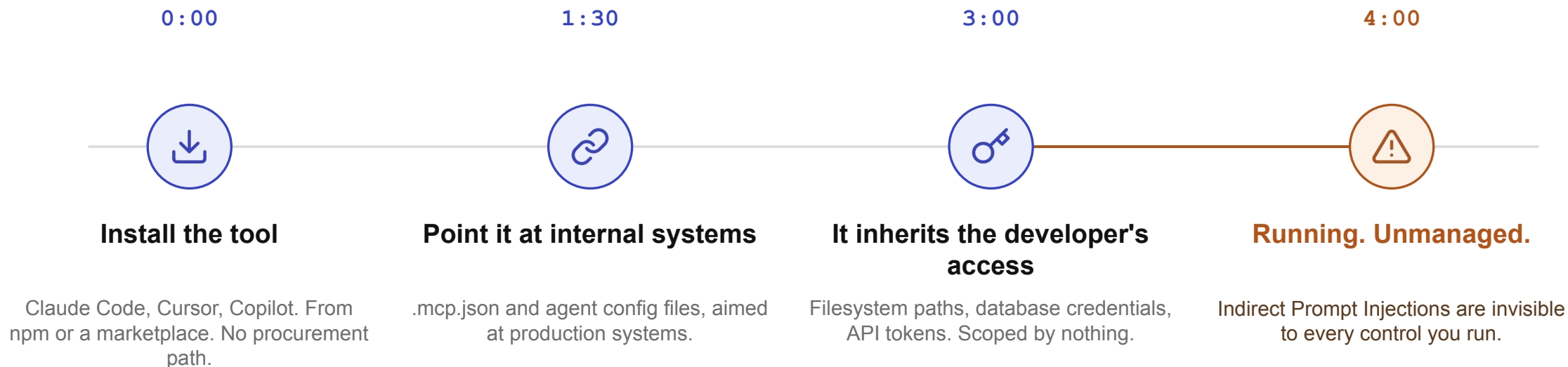
Your program *likely* has no control that tells these apart. To every tool you own, both are just a process.

PART 01

Discover

How many unmanaged agents are running right now?

What a developer stands up in four minutes



That agent was unmanaged from the second it ran and nobody made a decision to make it so.

Why your current controls can't see it

WHAT YOUR PROGRAM GOVERNS

A human identity in the IdP

A service account with a named owner

An access request and an approval

Deprovisioning at offboarding

Quarterly recertification



WHAT AN UNMANAGED AGENT ACTUALLY IS

A machine identity created by an install command

A keypair in a developer's home directory

A capability granted in a JSON file

Persists until somebody deletes the folder

Never reviewed, because it was never registered

Every control assumes an identity that was requested. Agents are not requested. They are installed.

Here's an ask

Ask These Three Questions.

How many agents are running? What can each one reach? *Who approved it?*

PART 02

Validate

Find Shadow AI in one command

Inventory Your Agentic Systems

```
npx hackmyagent detect
```

Requires Node.js 18 or later. npx runs it without installing.

What it inventories on the machine it runs on

- **AI tools installed** — Claude Code, Cursor, Copilot and similar
- **MCP server configurations** — both project-local and machine-wide
- **AI config files** carrying credential references or broad permission grants
- **Governance files** (SOUL.md) — and the projects where there are none

Find unmanaged AI Tools (AI Agents, MCPs, Credentials, etc.) that *nobody* approved.

Stop new unmanaged agents from shipping

One pull request for your platform team, in the repositories that ship agents

```
- uses: actions/checkout@v4
- uses: actions/setup-node@v4
  with: { node-version: '20' }
- run: npx hackmyagent secure --json > security-report.json
- run: npx hackmyagent secure -b oasb-1 --fail-below 70
```

And the perimeter

hackmyagent scan your-domain.com

see what your own perimeter exposes to the internet.

*only scan domains you have permissions from.

“oasb-1” runs the Open Agent Security Benchmark at L1, L2 or L3.

Back to

The Three Questions.

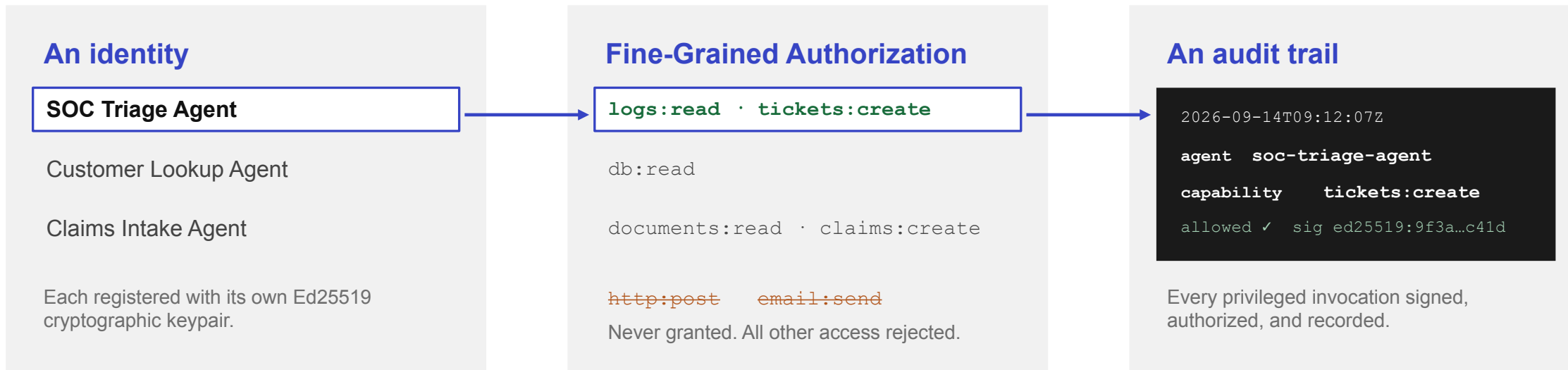
How many agents are running? *What can each one reach?* **Who approved it?**

PART 03

A.I.M

Agent Identity Management, IAM for Agents

Agent Identity & {Access} Management



Integrating AIM into your Agent

```
soc_triage_agent.py - python

from aim_sdk import secure
agent = secure("soc-triage-agent")    #AIM registers agent with crypto key

@agent.perform_action(capability="logs:read")    #AIM grants logs:read to soc-triage-agent
def triage_alert(alert_id): ...
```

You already require an SBOM. This is its agent equivalent.

AIBOM — Agent Bill of Materials

Complete inventory of all agents, MCP servers, tools, and data access patterns observed by AIM.

AGENT	OWNER	KEYPAIR	CAPABILITIES	MCP SERVERS	LAST ACTION
soc-triage-agent	j.rivera	ed25519:9f3a...c41d	logs:read · tickets:create	splunk-mcp ✓	2m ago
customer-lookup-agent	a.chen	ed25519:41bb...8e02	db:read	postgres-mcp ✓	14m ago
claims-intake-agent	m.osei	ed25519:c7d2...9a51	documents:read · claims:create	s3-mcp ✓	1h ago
unregistered ×4 (from scan)	none	none	inherited developer access	3 unverified	unknown



Exportable from the AIM compliance dashboard alongside credential rotation, trust scores, capability violations and audit coverage.

PART 04

Vulnerability Demo

Same code. Two states. See the difference.

Damn Vulnerable AI Agent (DVAA)

21

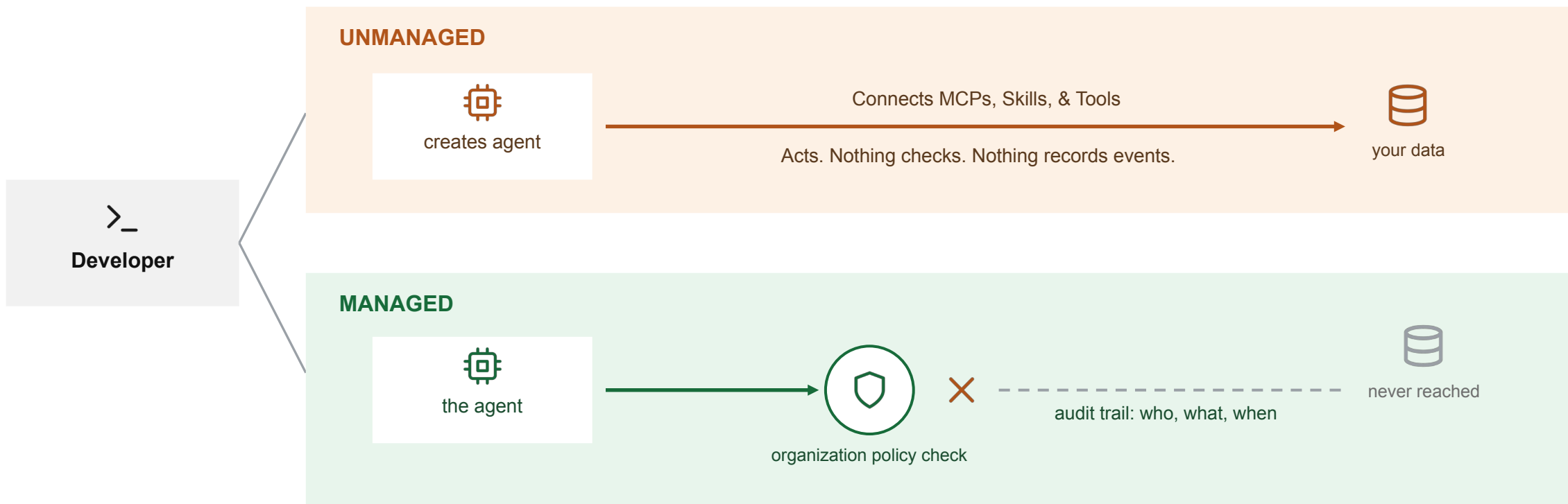
agents

12

vulnerability categories

3

protocols



Take these to your team

Find the unmanaged

```
npx hackmyagent detect
```

inventory on any machine, no install
add it to your CMDB

Stop new ones

```
hackmyagent secure --ci
```

pipeline gate, exit codes as policy

Manage the agents

```
pip install aim-sdk
```

agent identity, capability grants, &
monitoring

Plus slides and the demo: opena2a.org/gsx

We're looking for design partners

You have:

- Agents already running in production
- Regulated environment: insurance, finance, healthcare, public sector
- An inventory problem you can already name out loud

What you get

- A say in the AIM spec while it can still change
- Engineering support standing it up in your environment
- Named as a partner or kept anonymous — entirely your call

What I need back

- A real environment
- A few hours a month from someone technical
- Feedback that tells me when I have got it wrong

opena2a.org/gsx

THANK YOU

opena2a.org/gsx

everything from this talk, in one place

research.opena2a.org

monthly AI agent security research

opena2a.org

documentation and demos

Abdel Fane
linkedin.com/in/abdelfane

