sysdig

Your visual guide to securing AI agents

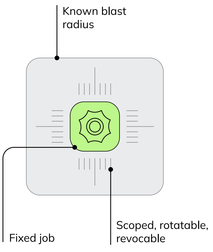
AI agents are powerful tools because they operate independently, retrieving data, executing commands, and interacting with tools, all with real credentials. That also makes them a target.

To run agentic AI systems at scale and at their full potential, you first need a clear understanding of how these systems can be compromised, and what you can do to keep them safe.

AI agents vs. static accounts: an expanding blast radius

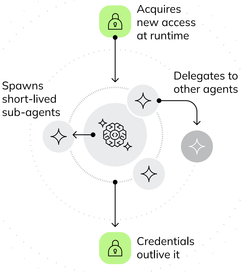
Service Account

Blast radius is fixed at deployment



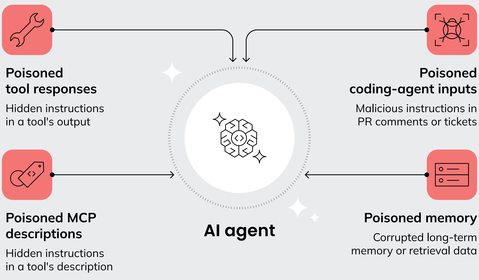
AI Agent

Reach expands as it runs, blast radius unknown



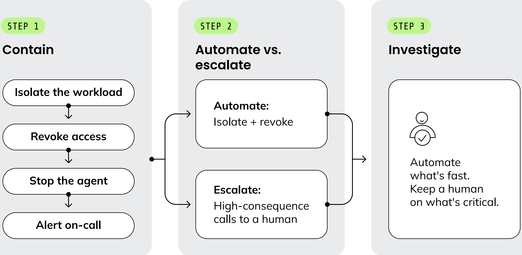
How AI agents get compromised

Every path arrives through an input the agent already trusts.



How to respond to a compromised AI agent

Agents move at machine speed. Respond in minutes, not hours.



How to build an agentic AI security program

STEP 1

Map your attack surface

Know your defense

Inventory where agents run, the tools and MCP servers they reach, and the credentials they hold

Threat-model with a method like MAESTRO

STEP 2

Layer your defenses

No single control is enough

Prevent: scoped identity, least privilege, sandboxing

Detect: watch what the agent does at runtime

Respond: contain fast, keep humans on the big calls

STEP 3

Put tooling in place

Detection needs somewhere to run

Open-source runtime detection (like Falco) as the foundation

A managed layer to scale rules, alerts, and response

sysdig

INFOGRAPHIC

COPYRIGHT © 2026 SYSDIG, INC
ALL RIGHTS RESERVED.
ING-813 REV. A 01/26

Want to learn more?

Get the full blueprint on securing AI workloads the right way.

GET THE BLUEPRINT →