SALT

FOR SOC, IR, AND DETECTION TEAMS

AI THREAT HUNTING

10 behaviors to investigate

Agent behavior

What the agent does right after it reads or is told something.

01 Execution after untrusted input

LOOK FOR Interpreter launches (node, python, shell) seconds after an agent reads an email, page, or ticket.

HUNT IN EDR or sandbox process logs joined to tool calls

WITH SALT AI-DR flags the injected prompt. Salt links it to the MCP and API calls that followed.

02 Alert fires after the action

LOOK FOR Injection or guardrail alerts stamped after the tool call they should have stopped.

HUNT IN Guardrail alerts joined to tool and API logs

WITH SALT Ties the late alert to the MCP and API calls it missed, in one view.

API call patterns and sequences

How agent identities call your APIs, and in what order.

03 Sequential object ID walking

LOOK FOR One agent or token stepping through IDs in order (R-4471, R-4472, R-4473) with no user request.

HUNT IN API gateway and WAF logs, grouped by identity

WITH SALT Detects record enumeration in live API traffic. AG-DR can block the agent in real time.

04 Bursts and endpoint fan-out

LOOK FOR One agent identity hitting many endpoints, retrying, or running far above its own normal request rate.

HUNT IN Gateway logs: calls per minute, endpoints per identity

WITH SALT Tells legitimate agents from malicious bots. AG-DR can block in real time.

05 Pivot from agent to direct API

LOOK FOR A source that probes an agent or MCP server, then calls the downstream API directly.

HUNT IN Per-source timelines across agent, MCP, API logs

WITH SALT Links the same source across agent, MCP, and API activity, even when it skips the agent.

These behaviors are general guidance. Adapt log sources to your telemetry.

Identity and access

Who the agent acts as, and what it can reach.

06 Agent-driven permission changes

LOOK FOR Role grants, policy edits, or config writes by an agent identity, often outside change windows.

HUNT IN IAM and cloud audit logs, non-human actors

WITH SALT AG-SPM flags over-scoped agent access before it's abused and maps what each agent can reach, so you know the blast radius.

07 Wrong credential type on agents

LOOK FOR Agent calls using human tokens, shared service accounts, or one embedded key from many origins.

HUNT IN IdP and API auth logs, token-to-source maps

WITH SALT AG-SPM finds hardcoded keys in code and missing auth in config, the root causes of credential misuse.

08 Exposed or shadow MCP servers

LOOK FOR Externally reachable MCP endpoints, or MCP servers with internal API edges missing from inventory.

HUNT IN Cloud and DNS inventory, MCP and gateway logs

WITH SALT Discovers every agent, MCP server, and tool, including shadow MCPs, and flags the ones exposed externally.

Data movement

What leaves, and in what form.

09 PII bulk read, encoded output

LOOK FOR Export-style requests returning PII, then Base64 or other encoded output that DLP may miss.

HUNT IN API response classification, DLP, agent output

WITH SALT Detects PII in API responses, including before it is encoded, and in model responses.

10 Outbound action or mass change

LOOK FOR External API calls, transactions, third-party contact, or bulk changes outside an agent's baseline.

HUNT IN Egress logs, payment and email APIs, audit trails

WITH SALT Links each outbound call back to the prompt and agent that triggered it, so you see why it happened, not just that it did.

THE SALT AGENTIC SECURITY PLATFORM

See every agent. Secure every layer.

Capabilities say when Salt acts. Layers say where it looks.

CAPABILITIES: WHEN SALT ACTS

AG-SPM Continuously

Agentic Security Posture Management

What do my agents have, and what can they reach?

- Discovers every agent, MCP server, and tool
- Flags exposed endpoints, hardcoded secrets, over-scoped access
- Maps posture to the EU AI Act, NIST, and SOC 2

AI-DR At the prompt

AI Detection and Response

What are my models being told, and what are they saying?

- Detects prompt injection and context manipulation
- Catches sensitive data leaking in model responses
- Links a model attack to the MCP and API attacks that follow

AG-DR At the action

Agentic Detection and Response

What are agents actually doing, and should they be?

- Detects record enumeration and PII exposure in live traffic
- Tells legitimate agents from malicious bots
- Blocks machine-speed abuse in real time

LAYERS: WHERE SALT LOOKS

AG-SPM continuously

AI-DR **AG-DR** in production

EXPOSURE
Salt Surface
Map external exposure

SOURCE CODE
Salt Code
Find hardcoded keys, enforce code policy

CONFIGURATION
Salt Connect
See APIs in your cloud

RUNTIME
Salt Collect and Protect
Analyze live traffic and block logic-based threats

One routine request, four findings the model check missed

Illustrative path: a billing agent exports refunds with customer contact details as a Base64 string.

Billing agent
- Model check passed

→

Refund MCP
- Internet-facing, externally exposed

EXPOSURE

→

Bulk-export tool
- Hardcoded backend API key

SOURCE CODE

→

Internal refund API
- No credential validation

CONFIGURATION

→

Customer PII
- ID enumeration, PII returned before encoding

RUNTIME

Your agents are already acting. See your own graph.

Get an Agentic Security assessment

Book a demo

Capabilities as described at salt.security/platform. 100+ enterprise deployments.