Above

FIELD GUIDE · INSIDER RISK IN THE AGENTIC ERA

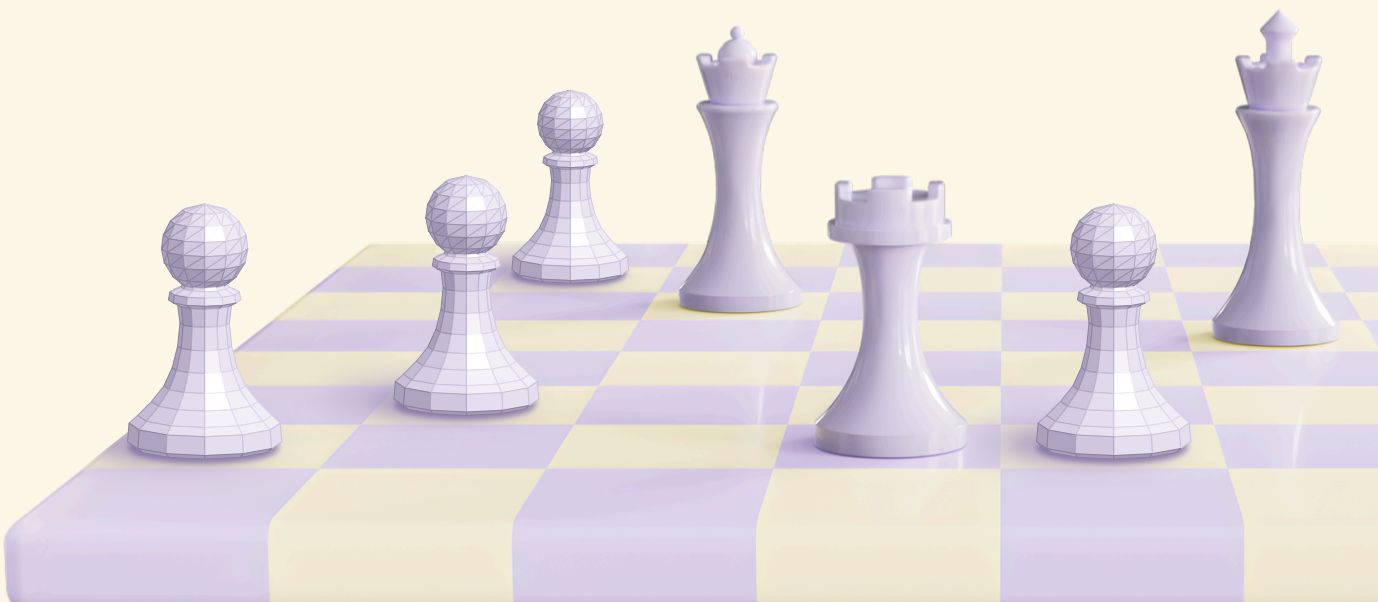
# The Ultimate Guide to the Agentic Insider

A practitioner's guide to the non-human actors now operating inside your trust boundary, on your employees' credentials, at machine speed.

Your newest insider has no badge, no manager, and full use of a real employee's permissions. Every control you own was built on the assumption that an identity maps to a person you can ask.

**Inside:** the five forms of the agentic insider · the five assumptions agentic activity breaks · the telemetry to collect before the first incident · five questions that resolve one · how attribution works when nobody typed.

**Above Security**



DEFINITION

# The insider who was never hired

For twenty years, insider risk meant a person. The person is still there, but they have started delegating. An employee grants an agent access to their mailbox, their repository and their drive, then asks it to get something done. The agent works at machine speed, under the employee's identity, with the employee's permissions, and your logs record the employee.

**109:1**
machine identities per human identity. 79 of the 109 are AI agents.
PALO ALTO NETWORKS IDENTITY SECURITY LANDSCAPE, 2026

**45%**
of employees now use AI regularly on corporate devices, up from 15%
VERIZON DBIR, 2026

**67%**
of that AI use is signed in with a non-corporate account
VERIZON DBIR, 2026

**30%**
of organizations have immutable audit logging of what agents do
IDENTITY SECURITY LANDSCAPE, 2026

**An agentic insider is a non-human actor operating inside your trust boundary on human authority.** It is not a person and it is not malware. It is a delegation, and delegation is the one thing your identity model, your data controls and your case management were never asked to represent.

WHERE THE INDUSTRY LANDED

The 2026 Verizon DBIR advises that **"we should pay special attention to service and machine accounts"** as the ones most likely to be leveraged for modern cyber threats. Shadow AI is now the third most common non-malicious insider action in DLP datasets, a fourfold rise in a single year.

## Five forms of AI agents

| FORM | HOW IT GETS ITS ACCESS | WHAT IT LOOKS LIKE IN YOUR TELEMETRY |
|---|---|---|
| **The delegated agent** SANCTIONED | An OAuth consent grant from the employee. Enterprise assistants and copilots with connectors into mail, files, tickets and code. | A known user, an unknown tempo. Correct scopes, exercised at a rate no human works at. |
| **The shadow agent** UNSANCTIONED | A personal AI account, a local desktop client on a managed laptop, or a connector the employee wired up themselves. | Frequently nothing. If it runs on a personal identity, the only trace is on the endpoint or the egress. |
| **The agentic browser** SESSION-BORNE | The employee's live authenticated session, cookies and all. No new credential is ever issued. | A logged-in human, clicking. Indistinguishable from work by design. |
| **The promoted service identity** NON-HUMAN | A long-lived token, API key or service account, quietly repurposed to drive an agent instead of a script. | An automation that starts making choices. Same identity, new and unbounded behavior. |
| **The hijacked agent** INJECTED | An attacker's instruction, embedded in a document, ticket, email or web page the agent was asked to read. | A trusted employee exfiltrating data. The insider is real. The intent belongs to someone outside. |

The last row is the one most programs are least ready for. An agentic insider can be created without the employee ever intending it.

WHY THE STACK MISSES IT

# Five assumptions that agentic activity breaks

Nothing in the agentic insider story requires a new exploit. That is the difficulty. Every layer of the detection stack rests on a premise about how work happens, and delegation invalidates the premise without triggering a single control.

| CONTROL | THE ASSUMPTION IT WAS BUILT ON | WHAT AGENTIC ACTIVITY DOES TO IT |
|---|---|---|
| **DLP** | Sensitive data is recognizable at the moment it moves. | The agent reads, reasons and rewrites. The summary carries the secret; the fingerprint, the classifier and the regex do not follow it. |
| **UEBA** | Deviation from an established baseline is a usable proxy for risk. | Every agentic action is a deviation. Baseline it in and you go blind. Alert on it and you drown. Deviation was never the same thing as risk. |
| **IAM and PAM** | An identity maps to a person, and a person can be asked what they did. | One identity now covers a human and a delegate. The audit trail names the human either way, so attribution stops at the account. |
| **EDR** | Harmful action is accompanied by harmful code. | Nothing malicious executes. Sanctioned APIs, granted scopes, approved applications, ordinary TLS. |
| **SIEM correlation** | Sessions and thresholds follow human tempo. | Four hundred reads in ninety seconds is either an outage or an incident, and the rule was written for neither. |

THE STRUCTURAL CHANGE

**Intent moved out of the action and into the instruction.** For a human insider, the action carried the intent: the download, the upload, the forward. For an agentic insider, the action is neutral and the intent lives in a prompt, a system message, or a paragraph buried in a file the agent was told to summarize. Almost no enterprise records that layer, which is why agentic cases so often end with a timeline nobody can interpret.

**Attribution collapses before triage begins.** When an alert fires on an agentic action, the analyst inherits three unanswered questions at once: whose credentials, whose instruction, and whose accountability. Existing tooling answers the first and is silent on the other two, so the case stalls in the queue rather than resolving.

None of these controls is wrong. They are aimed at a layer where the interesting decision no longer happens. A rule can still tell you that ten thousand records were read. It cannot tell you that an employee asked an assistant to prepare a handover document, that the assistant read the entire customer table to write it, and that the resulting file was shared to a personal address forty minutes later. That is one behavioral narrative with three participants, and it is the unit of work that agentic risk actually produces.

FOR DETECTION ENGINEERS

# You cannot investigate what you never recorded

Only three in ten organizations keep immutable logs of what their agents do, and roughly a third can revoke an agent's credentials at all. Agentic coverage is not a new sensor. It is three layers of existing telemetry that most teams collect separately and have never joined.

### AUTHORIZATION LAYER

- OAuth consent grants, with scope, grantor and grant date
- Third-party and internal app registrations per tenant
- Connector and MCP server inventory, including local clients
- Token age, last rotation, last use
- Browser extension inventory on managed devices

### ACTION LAYER

- Requests per session and per minute, by identity
- API to interactive ratio for each account
- Client id and user agent on every API call
- Bulk read followed by a single narrow write
- Access to resources the human owner has never opened

### OUTPUT LAYER

- Destination of every write, share and send
- New external sharing and link creation
- Messages and files authored by agent clients
- Downloads and clipboard events on the endpoint
- Pushes from non-interactive sessions

The joins are what matter. An OAuth grant on its own is governance paperwork. A bulk read on its own is a backup job. A grant issued on Tuesday, exercised at a thousand reads an hour on Wednesday, resolving into an external share on Thursday, is a case. Build the correlation on identity and time, not on a signature.

### SIX QUESTIONS YOUR TELEMETRY SHOULD ALREADY ANSWER

- [ ] Can you list every agent holding a live grant in your tenant today, sanctioned or not?
- [ ] Within one session, can you separate agent-initiated actions from human-initiated ones?
- [ ] Can you retrieve the instruction that produced a given action, and where that instruction came from?
- [ ] Can you revoke one agent's access without disabling the employee who owns it?
- [ ] Do you know which agents can currently reach your most sensitive repositories, drives and records?
- [ ] Can you reconstruct an agent's full activity ninety days after the fact?

A no to question three is the common case, and it is the one that turns a two-hour investigation into a two-week one.

### IF YOU ONLY DO TWO THINGS THIS QUARTER

Inventory every live consent grant and connector in the tenant, recording who granted it, at what scope, and when it was last exercised. Then start retaining client id and user agent on API calls, so agent activity is separable from human activity inside a single session. Neither move needs new tooling, and without them no agentic case can be closed.

FOR IR AND THE SOC

# Five questions that resolve an agentic case

An agentic incident arrives looking like an employee doing something inexplicable. Run it as an investigation, not as an alert triage, and it resolves in a predictable order. Skip a step and you end up disciplining someone for an instruction they never wrote.

**Establish the actor**
MINUTES 0 TO 30

**Recover the instruction**
HOUR 1

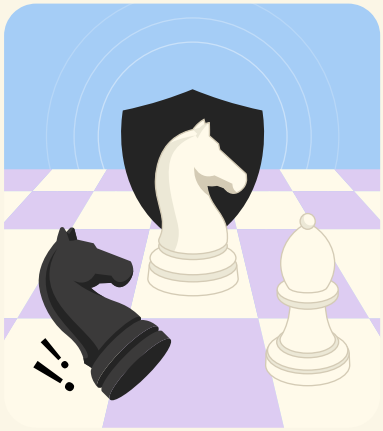
**Trace the data**
HOURS 1 TO 4

**Assign accountability**
DAY 1 TO 2

1. **Who acted?** Which identity, which client id, which token, and was a human present at the keyboard when it happened. Interactive and non-interactive activity on one account is the first fork in the case.
2. **Who authorized it?** Which consent grant, at what scope, issued by whom and when, and whether it is still live. This is the difference between a governance gap and a compromise.
3. **What instructed it?** A prompt, a scheduled task, a tool call, or text inside content the agent was asked to process. If the instruction came from a document, a ticket or a web page, you are handling a prompt injection wearing an insider's identity, and the response is containment, not HR.
4. **What moved?** Separate the read set from the write set. What the agent could see matters as much as what it sent, because the read set is your disclosure scope whether data left or not.
5. **Who is accountable?** The delegating employee owns the outcome. Culpability depends entirely on question three, and that distinction is the whole reason the instruction layer has to be logged before you need it.

### CONTAINMENT MOVES SPECIFIC TO AGENTS

- Revoke the grant, not the human. Suspending the employee stops the work and leaves the delegation intact.
- Rotate the token the agent authenticated with, then check where else that secret is in use.
- Preserve the session and instruction history first. Conversation retention is often shorter than your case timeline.
- Re-scope before you re-enable. An agent restored at its original permissions reproduces the incident.



**The first agentic insider most teams meet is not malicious.** It is a competent employee who delegated more access than they realized to a tool that worked faster than anyone expected. Treating that case as an investigation rather than a violation is what keeps the next hundred employees telling you what they have connected.

WHERE THIS GOES

# Insider risk stops being a policy problem

Policy-based insider risk depended on being able to enumerate what bad looks like in advance. That was already strained by human behavior, which drifts, and it does not survive contact with delegated non-human actors whose capabilities change with every release note. The agentic insider does not need a better rule. It needs the question answered: what happened here, and does it indicate risk.

### The policy posture

WHAT MOST PROGRAMS RUN TODAY

Enumerate the bad patterns. Write rules. Tune thresholds. Wait for a match, then hand an analyst an alert and a starting point. Every new agent, connector and model release adds rules faster than anyone retires them, and the tuning debt compounds until the signal is unusable.

### The investigation posture

WHAT AGENTIC INSIDER RISK DEMANDS

Observe behavior continuously across identities, applications, endpoints and data. Reason about who acted, what changed, on whose authority, and why. Deliver a finished narrative with a timeline and a recommended action instead of a fragment that still needs a human to assemble it.

That shift is what Above Security was built for. Above is a proactive insider risk platform that replaces static rules with continuous behavioral investigation, run by a fleet of AI investigative agents. It observes activity across identities, SaaS, endpoints and collaboration tools, reasons about how risk is forming, and produces investigations rather than alerts: a behavioral timeline, the context, the reasoning, and what to do next. Where an action warrants it, real-time behavioral nudging can intervene in the moment with a contextual warning or a safer path, before an incident exists to investigate. Agentic AI, custom GPTs, personal AI and credential exposure are handled inside one behavioral framework, alongside the human scenarios they now overlap with.

This guide is published ungated and free to share, quote and adapt with attribution. If it is useful to your team, the same material works as an internal briefing: **above.security**

SOURCES

**Verizon.** 2026 Data Breach Investigations Report. Figures cited: regular AI use on corporate devices, non-corporate account sign-in, shadow AI as a non-malicious insider action, and guidance on service and machine accounts.

**Palo Alto Networks.** 2026 Identity Security Landscape, survey of 2,930 cybersecurity decision-makers. Figures cited: machine to human identity ratio, AI agent share, agent credential revocation, immutable audit logging.

**Above Security.** Product and positioning claims in the closing section describe the Above platform and are the publisher's own.

CHECKMATE, INSIDER THREAT.