Prepared for
Halden Robotics

Date
November 8, 2026

Sample report

Point-in-time snapshot

AI Agent Security Assessment

Point-in-time snapshot across 5240 endpoints: AI agents, MCP servers and add-ins, the credential stores they can reach, personal accounts, browser and IDE extensions, and Copilot Studio agents, with a proposed enforcement policy.

5,240
Endpoints scanned

43
Finding types

8
Critical finding types

Sample report. Halden Robotics is a fictional company and all data is invented. Risk-management guidance based on telemetry observed at scan time. Not legal advice, a compliance attestation, or a conformity assessment.

Introduction

5,240
Endpoints scanned

November 8, 2026
Scan date

4
Categories: Agent Inventory, Credentials, Capabilities, Network

Severity legend

Critical	Active exposure of credentials, code execution, or sensitive data. Immediate action required.
High	Material risk to credentials, code execution, or data exposure.
Medium	Configuration concern; not immediately exploitable but should be addressed.
Low	Informational; track for awareness.

Glossary

AI Agent

Software that uses an LLM to read files, run commands, fetch data, or call APIs on a user's behalf (Cursor, Claude Code, GitHub Copilot Chat, ChatGPT, etc.).

MCP Server

A Model Context Protocol server that an AI agent connects to for more tools and data (for example GitHub, Slack, Jira or Google Drive).

Add-in

A skill, plugin, subagent instruction or mod that an AI agent loads to do more (for example a code-review skill or a GitHub plugin).

Extension

A browser or IDE extension that adds AI features. It can read pages or code, so its permissions matter.

SaaS Agent

An AI agent that runs inside a company's SaaS platform, such as Microsoft Copilot Studio, and not on an endpoint.

Credential Store

Sensitive credential locations on the endpoint: SSH keys, AWS, GCP and Azure credentials, Kubernetes config, GnuPG keyring, OS keychains.

Sandboxed

Whether the AI agent's process is restricted by the OS from reading files outside its own container. Non-sandboxed agents can read whatever the user can read.

Finding

A single observation surfaced by Bay's scan. Each finding has a Bay-product type, severity (Critical / High / Medium / Low), and target (endpoint + agent).

Executive summary

Agent Inventory

Critical

31

Finding types

2,969 endpoints affected
4,730 distinct entities

Credentials

Critical

5

Finding types

1,365 endpoints affected
1,612 distinct entities

Capabilities

Critical

7

Finding types

1,015 endpoints affected
1,155 distinct entities

Network

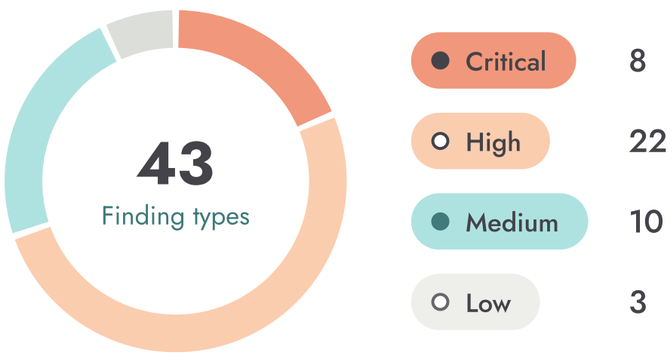
No findings

0

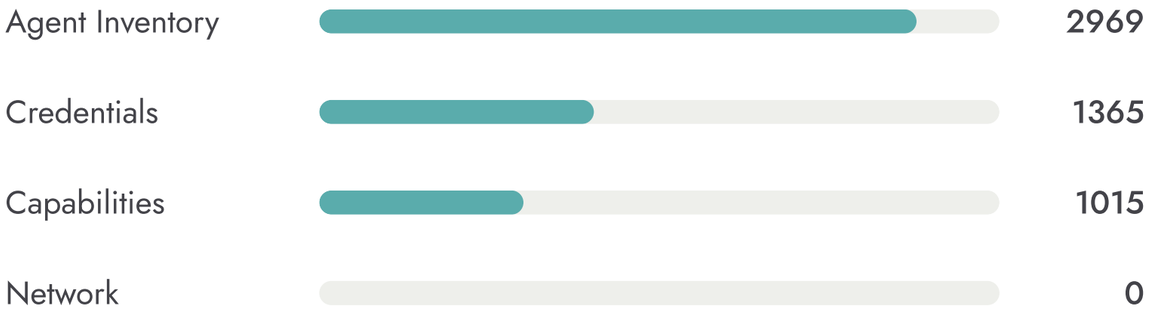
Finding types

0 endpoints affected
0 distinct entities

Finding types by severity



Endpoints affected by category



Top 5 findings by risk

#	ID	Finding	Category	Scope	Severity
1	F1	Claude Code Bypass-Permissions Mode Enabled	Agent Inventory	15 endpoints	Critical
2	F2	MCP Server Consent Bypassed	Agent Inventory	17 endpoints	Critical
3	F3	Malicious Extension Detected	Capabilities	21 endpoints	Critical
4	F4	Malicious Add-in Detected	Capabilities	14 endpoints	Critical
5	F5	Malicious Package Detected	Capabilities	13 endpoints	Critical

Top 5 endpoints by risk

#	Endpoint	Primary user	Total	Risk
1	HDN-WIN-20252	talvarez	23	Critical
2	HDN-WIN-35162	okristensen	15	Critical
3	Kofis-MacBook-Pro	knilsen	13	Critical
4	HDN-MBP-16682	kmenon	13	Critical
5	HDN-LNX-36281	byusupova	13	Critical

Top 5 AI agents by risk

#	Agent	Endpoints	Findings	Risk
1	Claude Code	1,460	5,112	High
2	Cursor	983	1,446	High
3	Murkquill	21	63	High
4	Codex CLI	421	670	High
5	Claude Desktop	1,312	1,806	High

AI agent inventory

Everything AI that Bay found across 5,240 endpoints: the agents, the MCP servers they connect to, and the skills, plugins and extensions that extend what they can do.

31	35	14	18	7
AI agents	MCP servers	Agent add-ins	Extensions	SaaS agents
8 agent types	8 approved, 27 to review	4 plugins, 3 skills, 7 other	7 browser, 11 IDE	Copilot Studio, 4 flagged

Skills, plugins and other add-ins

14 add-ins loaded by agents on Halden endpoints, from plugin marketplaces, shared repositories and individual installs.

Name	Type	Used by	Parent plugin	Endpoints	Risk
AGENTS.md	Instruction file	Codex CLI, Cursor		14	Critical
hub-shellwick	Plugin	Claude Code		43	High
copyfern-tools	Plugin	Claude Code		29	High
toolbrine-pack	Plugin	Claude Code		376	Medium
skillmoss	Plugin	Claude Code		296	Medium
sql-explorer	Skill	Claude Code	toolbrine-pack	201	Medium
commit-helper	Skill	Claude Code	skillmoss	244	Medium
frontend-design	Skill	Claude Code		938	Low
code-reviewer	Subagent	Claude Code	toolbrine-pack	335	Not rated
release-notes	Command	Claude Code	toolbrine-pack	281	Not rated
auto-format	Hook	Claude Code		693	Not rated
agent-session-log	Hook	Claude Code		358	Not rated
pre-commit-notify	Hook	Claude Code		211	Not rated
statusline	Mod	Claude Code		288	Not rated

hub-shellwick and copyfern-tools come from third-party plugin marketplaces and can run shell commands. See the example finding below.

AI agents by type

Coding agents	15
Browser agents	7
AI assistants	2
CLI agents	2
Autonomous agents	2
General agents	1
Automation bots	1
AI infrastructure	1

Most findings: Claude Code (5,112), Claude Desktop (1,806), Cursor (1,446)

MCP servers

Approved	8
To review	27
Local, no authentication	16

Set to remove by policy

pdfwhelk-mcp, webgrabber-mcp, sitewhittle-mcp

Database credentials in config

postgres-prod, mongodb

Approved

confluence, fetch, figma, gmail, google-drive, jira, linear, sentry

Riskiest extensions

Murkquill	Browser	21 endpoints	Critical
Claude in Chrome	Browser	352 endpoints	High
Pagethistle AI	Browser	231 endpoints	High
Pagegloam	Browser	174 endpoints	High
Commitwyrm	IDE	122 endpoints	High

SaaS agents (Copilot Studio)

Flagged	4
Pending	1
Approved	2

Run anonymously

Sales Quote Assistant, Vendor Inbox Triage

Run as a named employee

HR Helpdesk Agent, Finance Close Assistant, IT Onboarding Agent

Example finding

F12 Claude Code Third-Party Marketplace Registered

High

Open findings	Endpoints	Distinct entities
581	183	183

What's happening

Claude Code on this device trusts third-party plugin marketplaces. Plugins from these repos can install slash commands, skills, and MCP servers.

Recommendation

Remove third-party plugin marketplaces and review the plugins already installed from them, starting with hub-shellwick and copyfern-tools, which can run shell commands. Approve the marketplaces you keep under Trusted Sources. Bay can push managed settings that allow only approved marketplaces.