TRADITIONAL THREATS VS. AGENTIC THREATS: WHAT CHANGED?

Traditional security still matters. Agentic AI adds a new layer of risk: systems that can interpret instructions, choose tools, and take action across business systems.

TRADITIONAL THREATS

ENDPOINTS | IDENTITY | NETWORK | DATA

PRIMARY SECURITY FOCUS
Users, endpoints, networks, applications, identities, and data

HOW ATTACKS CAN START
Malware, stolen credentials, malicious files, exploit attempts, or suspicious traffic

WHY THE SIGNAL IS DIFFERENT
Risk often appears in an identity, file, process, request, or network event

POTENTIAL OUTCOME
An attacker gains access and then takes action

AGENTIC THREATS

AGENTS | TOOLS | SEQUENCE | ACTIONS

PRIMARY SECURITY FOCUS
AI agents that can reason over context, invoke tools, and take actions on behalf of a malicious user or a threat actor

HOW ATTACKS CAN START
Untrusted content in an email, webpage, document, code comment, Skill, or tool response can influence what the agent does

WHY THE SIGNAL IS DIFFERENT
Each step can look legitimate on its own; the risk emerges from the sequence, context, and intended task

POTENTIAL OUTCOME
The agent can use valid access to exfiltrate data, misuse tools, execute code, modify systems, or exceed intended authority

The shift: traditional controls answer whether an identity, process, or request is allowed. Agentic security adds another question: should the agent be taking this action, in this context, as part of this task?

WHAT TRADITIONAL CONTROLS COVER

- IAM / PAM: who or what can access a system
- EDR / CNAPP: endpoint and workload behavior
- DLP / API security: data movement and requests
- AI guardrails: unsafe inputs and outputs

WHAT AGENTIC SECURITY ADDS

- Inventory agents, Skills, MCP servers, tools, and connections
- Test how agents can be manipulated or exceed intended authority
- Evaluate the context and sequence behind agent actions
- Detect, block, and contain risky behavior at runtime

STRAIKER RESEARCH: REAL-WORLD AGENTIC ATTACK PATHS

- **STAR Labs Threat Report, Volume I (2026):** 1,700+ successful exploits; 36% of successful coding-agent attacks reached remote code execution; 91% of successful productivity-agent attacks ended in silent data exfiltration.
- **Nothing to See Here: How an AI Agent Buried Its Own Commands (2026):** Indirect prompt injection in a routine-looking email led an agent to locate and exfiltrate SSH and PEM keys while the user saw a benign approval request.
- **Escape from Pod 9 (2026):** Poisoned telemetry drove an AI SRE agent into privileged container deployment, host escape, and functional ransomware.
- **Agentic Danger: DNS Rebinding Exposes Internal MCP Servers (2025):** STAR researchers demonstrated an external-to-internal MCP attack path with local command execution and environment-variable exfiltration.