VIEW POINT

THE ROLE OF SECURITY-FOCUSED AI AGENTS: CAPABILITIES, ADVANTAGES, AND APPLICATIONS

Abstract

As cyber threats grow in scale, speed, and sophistication, traditional alert-driven security models are struggling to keep pace. Security-focused AI agents are emerging as a critical evolution in cyber defence, enabling organisations to move from reactive response to proactive, decision-oriented security operations. This article explores the role of AI agents for cybersecurity, examining their core capabilities, advantages, and real-world use cases, from agentic AI for threat hunting to autonomous SOC operations and risk-based remediation. It also addresses how these agents integrate with existing security stacks, the governance and readiness considerations leaders must evaluate, and the measurable business impact of adoption. For organisations navigating complex, regulated environments, the article provides a pragmatic framework for adopting autonomous security capabilities with control, accountability, and resilience.

Infosys®
Navigate your next

The pace of the modern attack surface expansion outstrips the capacity of security teams to monitor, investigate, and contain threats. As adversaries adopt automation and AI to operate at machine speed, organisations are reassessing whether traditional security automation remains sufficient.

Security-focused AI agents, autonomous, goal-oriented systems that observe, reason, and act across complex environments, are emerging as a meaningful evolution in cyber defence. Unlike conventional tools that primarily detect and alert, these agents can actively hunt threats, coordinate response actions, and adapt to changing attacker behaviour. We will examine how AI agents for cybersecurity are reshaping security operations, their advantages, practical AI security agent use cases, and the governance considerations organisations must address before adoption.

What AI agents for cybersecurity are and why they matter now

Security-focused AI agents are autonomous software entities designed to achieve defined security outcomes rather than execute predefined scripts. Instead of responding to individual alerts in isolation, they operate against broader objectives such as identifying active threats, containing incidents, or validating security posture. They maintain contextual awareness across systems and time, plan and execute multi-step actions, invoke tools and workflows independently, and learn from outcomes to refine their behaviour. This represents a shift from task automation to decision-oriented security intelligence.

Their growing relevance reflects sustained operational strain within security organisations. As environments become more distributed and dynamic, security teams face increasing decision latency, fragmented visibility, and escalating response complexity. Skills shortages and alert fatigue further constrain the ability to investigate and act at the speed required, even as the financial and operational consequences of delayed containment remain significant.

Together, these pressures are accelerating the move toward autonomous security operations that can scale decision-making while preserving control and accountability.

Core capabilities of security-focused AI agents

Security-focused AI agents move security operations from alert-driven response to continuous, outcome-oriented defence by combining contextual intelligence, autonomous decision-making, and adaptive execution.

Continuous telemetry fusion and contextual reasoning

AI agents ingest and correlate telemetry across endpoints, networks, cloud workloads, identity systems, and threat intelligence feeds to maintain persistent context, enabling more accurate differentiation between genuine threats and benign anomalies.

Agentic AI for threat hunting

Agentic AI enables hypothesis-driven threat hunting by proactively identifying patterns such as lateral movement, privilege escalation, and low-and-slow exfiltration that often evade static, rule-based detection.

Orchestration and autonomous response

Within defined policy boundaries, AI agents coordinate containment and remediation actions, such as endpoint isolation or credential revocation, significantly reducing mean time to respond, particularly during large-scale or off-hours incidents.

Adaptive learning and playbook optimisation

By learning from outcomes and attacker behaviour over time, agents continuously refine detection logic and response playbooks, reducing false positives and improving operational efficiency.

Advantages of autonomous security operations

The adoption of autonomous security operations delivers measurable benefits across people, process, and technology, particularly in complex and distributed environments.

Speed and scale

AI agents can execute thousands of investigations and correlations in parallel across hybrid estates, a critical capability as attackers compress the time between initial access and impact. Faster detection and response materially limit breach severity.

Cost efficiency

By accelerating detection and containment while reducing manual investigation effort, autonomous security operations lower the overall cost of incidents and improve the efficiency of security teams.

Reduced analyst fatigue

By filtering noise and automating repetitive investigative steps, AI agents allow analysts to focus on complex cases, strategic analysis, and threat modelling, improving both productivity and retention.

Improved resilience and consistency

Unlike human-driven processes, AI agents apply security policies consistently and without fatigue, improving compliance outcomes and reducing variability in incident handling.

How security-focused AI agents integrate with existing security stacks

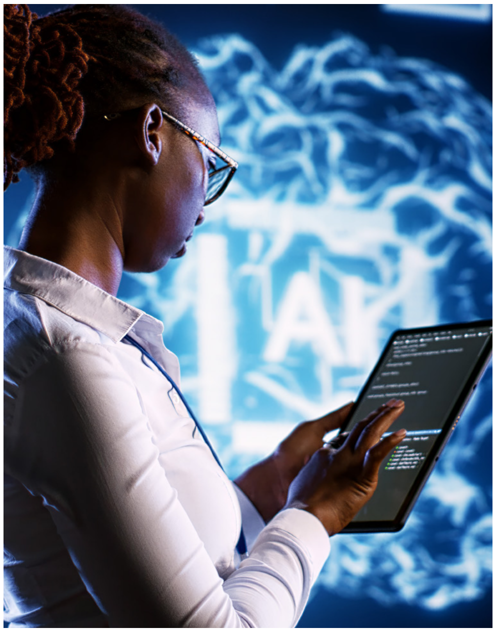
For most organisations, adopting security-focused AI agents does not require replacing existing security investments. Instead, value is realised when agents are layered across established tools such as Security Information and Event Management (SIEM), SECURITY Orchestration, Automation, and Response (SOAR), Endpoint Detection and Response (EDR), Extended Detection and Response (XDR), cloud security platforms, identity systems, and vulnerability scanners.

AI agents function as an orchestration and intelligence layer, connecting disparate security controls into a cohesive operational fabric. By ingesting telemetry from endpoints, cloud workloads, identity providers, and logs, agents reason across a unified context and invoke actions through SOAR workflows or native security APIs. Research on the convergence of SIEM, SOAR, and AI highlights that this integrated approach improves incident detection, investigation, and response efficiency while maximising the return on existing security platforms.

API maturity is a critical enabler. Well-documented, real-time APIs allow agents to investigate and respond reliably without brittle scripting. This architecture also supports phased adoption, with many organisations starting in assistive mode before progressing to conditional automation for well-defined scenarios.

Evaluating readiness: key questions leaders should ask before adoption

Before deploying security-focused AI agents, leaders should assess readiness across technology, operating model, and governance. Agentic capabilities amplify existing strengths, but they also expose underlying weaknesses. Deploying them on fragile foundations risks automating inconsistency rather than improving security outcomes.



A small set of strategic questions can clarify readiness:

- Is security telemetry complete, timely, and reliable across endpoints, cloud environments, and identities?
- Are detection, escalation, and response processes clearly defined, with unambiguous ownership?
- Which decisions can be safely automated, and where must human approval remain, particularly for actions affecting access, availability, or customer experience?
- How will success be measured? Through reduced dwell time, faster containment, and improved analyst productivity, rather than alert volume?
- Finally, is AI governance aligned with existing risk, compliance, and audit frameworks to ensure accountability and traceability?

Organisations aligning AI adoption with enterprise risk management are better positioned to scale autonomous security operations safely and sustainably.

AI security agent use cases in practice

AI security agent use cases are emerging where scale, speed, and consistency matter most, enabling organisations to operationalise proactive defence, risk-based prioritisation, and disciplined response across complex security environments.

Proactive threat hunting at scale

Security agents continuously hunt for unknown threats by analysing behavioural deviations rather than relying on known signatures. Organisations piloting agentic threat hunting report faster identification of stealthy attacks and more complete investigation trails.

Autonomous security operations centre (SOC) assistants

In modern SOCs, agents triage alerts, enrich incidents with historical and threat intelligence context, prioritise based on business impact, and recommend or execute response actions. The integration of multiple AI security agents into enterprise SOC platforms reflects a broader industry move toward agent-driven security operations.

Vulnerability prioritisation and remediation

Agents correlate vulnerability data with asset criticality, exploit intelligence, and exposure context to prioritise remediation efforts, moving organisations beyond CVSS scores toward risk-based vulnerability management.

Automated incident forensics and reporting

After containment, agents can automatically assemble forensic artefacts, map attacker behaviour to MITRE ATT&CK, and generate audit-ready reports. This significantly reduces post-incident workload and supports regulatory compliance.

Measurable business impact of security-focused AI agents

Recent industry analysis indicates that AI and automation are reshaping the economics of data breaches in favour of defenders. Organisations that integrate AI-driven security capabilities are significantly shortening breach detection and containment timelines, in some cases by around 80 days, while reducing average breach costs by approximately USD 1.9 million compared with organisations without such capabilities. This improvement is reflected in a broader decline in average breach costs globally, driven in part by faster containment enabled by AI-augmented security operations.

Speed remains a decisive factor in limiting the impact of security incidents. The longer a threat goes undetected or unresolved, the greater the financial and operational damage, particularly in insider-driven scenarios where dwell time is often extended.

Governing autonomy: risk, accountability, and standards alignment

While the benefits of security-focused AI agents are clear, sustained value depends on disciplined governance. Autonomous systems must operate within clearly defined decision boundaries, particularly where actions affect system availability, access rights, or customer experience. Human-in-the-loop thresholds, escalation paths, and exception handling should be defined early to prevent unintended outcomes.

Guidance from ENISA and NIST stresses that AI-driven security systems require controls at least as rigorous as those applied to human-operated processes. This includes end-to-end auditability of agent decisions, protection against prompt injection and model manipulation, and safeguards against data poisoning and supply-chain risk. Clear accountability for agent behaviour, supported by a strong model and data security, is essential to maintaining operational trust.

As policy and standards continue to evolve, frameworks such as the NIST AI Risk Management Framework provide a structured way to align agentic AI for threat hunting and broader AI security agent use cases with enterprise risk and compliance expectations, enabling responsible scale.

The future of security-focused AI agents: from assistance to autonomous defence

Over the next three to five years, security-focused AI agents will transition from assistive tools to foundational elements of enterprise cybersecurity. The emphasis will move from reactive automation to predictive and preventive security, with agents continuously assessing risk, anticipating attack paths, and strengthening controls ahead of exploitation.

Multi-agent architectures will enable specialised agents across identity, cloud, endpoint, data, and application security to collaborate in real time, supporting coordinated defence decisions that reduce risk without disrupting business operations. At the same time, alignment with enterprise AI governance frameworks will become non-negotiable. Explainability, auditability, and policy compliance will be embedded by design, enabling adoption in regulated industries. As adversaries increasingly use AI-driven techniques, agentic defence will favour organisations that invest early in governed, autonomous security capabilities.

Pragmatic adoption: turning autonomy into advantage

Security-focused AI agents represent a meaningful shift in how organisations defend digital environments. By enabling agentic AI for threat hunting, accelerating response, and reducing operational strain, they offer a powerful advantage in an increasingly hostile threat landscape. However, success depends on disciplined implementation, such as strong telemetry foundations, clear governance, and incremental automation. Organisations that treat AI agents as trusted collaborators rather than unchecked replacements will be best positioned to realise the benefits of autonomous security operations while managing their risks.

For more information, contact infosysbpm@infosys.com

Infosys®
Navigate your next