CYBER DEFENSE MAGAZINE

eMAGAZINE

JUNE 2026

CYBERSECURITY SOFTWARE IS ONLY AS GOOD AS THE PEOPLE CONFIGURING IT

Cyber Defense Magazine Editor's Choice Book Corner

The HIPAA Security Rule 2026 Problem Every Small Healthcare Practice Is Gambling on Right Now

How Not to Handle a Cyber Incident

From Compliance To Resilience: Why Modern CISOs Must Rethink Cybersecurity Strategy In 2026

...and much more...

MORE INSIDE

Agentic AI In Cyber Security: From Detection To Autonomous Response

By Tarun Wig, Founder & CEO, Innefu Labs Limited

Imagine a situation: your security team receives an alert at 3:57 AM. One of the user accounts is showing some abnormal behaviour such as accessing sensitive files at a speed no human could manage to, and that too from a location that is geographically impossible given their last login. By the time the on-call analysts wake up, read the alert, pull up the logs, and begin triaging, eleven minutes have passed. In a ransomware scenario, eleven minutes is the entire ballgame.

Though detection has improved dramatically. Decision-making frameworks have matured. And yet breaches keep happening. Not because teams fail to see the threat, but because the window between seeing that threat and actually stopping it, is simply too narrow for human-paced response. The attack surface has not just expanded; it has accelerated. And in that gap between decision and action, organisations are losing ground.

This is where agentic AI enters the picture. Not AI that alerts. Not AI that advises. AI that acts.

What is Agentic AI? Defining the Shift

The term "agentic" refers to AI systems that are goal-directed and capable of taking sequence of actions autonomously to achieve an objective. In a security context, this means AI that does not stop at surfacing a threat, it responds to it.

Where traditional AI-powered tools produce outputs for humans to act on, agentic AI systems execute actions directly, such as isolating a compromised endpoint from the network, revoking a user's access credentials, blocking malicious traffic at the firewall, quarantining a suspicious file, or triggering a rollback of a misconfigured cloud resource. These are not mere recommendations.

These are actual interventions.

This is a radical change in the way we consider the role of AI in security. Detection is passive. Decision support is assistive. Agentic AI is operational.

The Speed Problem: Why Human-in-the-Loop Models Are Hitting Their Limits

Cost of a Data Breach Report by IBM has always indicated that the longer an attacker stays undetected the more expensive it is to breach. But dwell time is only part of the problem. Even after detection, containment takes time. Time that's spent triaging, escalating, approving, and executing a response across multiple tools and teams.

In a typical security operations centre (SOC), the process looks something like - an alert fires up, an analyst investigates, a senior analyst confirms, then a playbook is consulted, a response action is approved, and then; finally, remediation begins. Each step introduces latency. And in high-velocity attacks, that latency is attacker's ally.

Human analysts cannot be replaced in terms of judgment, context and accountability. But the mechanical steps in that chain of response: correlation, confirmation, playbook selection, execution, are precisely where machine speed provides with an asymmetric advantage. Agentic AI compresses that chain from minutes to seconds, without removing humans from oversight.

The Trust Problem: Guardrails, Oversight and Human Control

Autonomy in security systems raises an uncomfortable question: what happens when the agent gets it wrong?

Automated response systems can cause real damage, blocking legitimate users, impacting business-critical processes, or triggering cascading failures in interconnected systems. The risk of a false positive causing operational harm is real, and it's the primary reason why many organisations have been slow to adopt autonomous response capabilities.

The answer is not to abandon autonomy but to architect it responsibly. Effective agentic systems operate with confidence thresholds. Autonomous action is taken only when detection confidence exceeds a defined threshold; below that, the agent escalates to a human.

They operate within set boundaries, defining clearly which actions the agent can take itself versus which require approval. And they maintain complete audit trails i.e. every action is logged, timestamped, and attributed, enabling post-incident review and continuous improvement.

The goal here is not to replace human judgment. It's to ensure that human judgment is applied where it matters most, such as, on complex, ambiguous, high-stakes decisions; while agents handle the high-confidence, time-critical responses that humans simply cannot execute fast enough.

AI vs. AI: The Arms Race Dimension

Adversaries are increasingly deploying AI-assisted attacks such as automated reconnaissance, generating convincing phishing content at scale, adapting malware to evade detection, and probing defences faster than any human red team could.

In this environment, a purely human-paced defence will always be in disadvantage. Machine-speed attacks require machine-speed responses. Agentic AI is not a competitive advantage in this context, it's a necessity. Organisations that rely solely on human response cycles against AI-assisted adversaries are, in effect, bringing a very deliberate process to a very fast fight.

What Security Leaders Need to Do Now

For CISOs and security leaders, agentic AI is not a distant future, it's a present-day procurement and governance question.

The first step would be an honest assessment of response latency. How long does it actually take your organisation to go from detection to containment? Where in that chain are the bottlenecks? Those bottlenecks are where agentic capabilities deliver the highest return.

The second would be building a governance framework before deploying such autonomous systems. Define the scope of autonomous action. Establish confidence thresholds. Map the escalation paths. Ensure audit trails are built into the architecture and not bolted on afterward.

Third is cultural. Security teams will have to trust these autonomous systems enough to let them act. Such trust is earned through transparency, explainability, and demonstrated accuracy over time. The organisational confidence that makes broader deployment possible is built by piloting agentic capabilities in lower-risk environments, reviewing agent decisions alongside human ones, and iterating on confidence thresholds.

The organisations that will be most resilient in the years ahead are not those with the most alerts or the most tools. They are the ones that have closed the gap between detection and action, and agentic AI is the most powerful mechanism available to do exactly that.

About the Author

Tarun Wig is a self-made entrepreneur and business executive with a decade long outstanding track record in building, growing and leading high-performing, startups in India.

He co-founded and is currently involved full-time at Innefu Labs, a research-oriented group using Artificial Intelligence and Machine Learning for National and Cyber Security. Innefu is a developer of niche and innovative security solutions such as the Intelligence Fusion Centre, Open-Source Intelligence, Forensics Analytical Toolkits, as well as Biometric Authentication. The solutions are running across multiple intelligence and critical infrastructure organizations.

Under Tarun's leadership, Innefu has more than 100+ installations across Indian Subcontinent, Middle East and Southeast Asia, the company is today a de-facto leader in developing and deploying AI for National and Cyber Security.

Tarun can be reached online at digital@innefu.com and at the company website https://innefu.com/