Courtesy of
RS Risk Solutions Ltd

How Agentic AI Is Changing the Cyber-threat Landscape

Artificial intelligence (AI) is undergoing a fundamental shift. Systems that once served as assistive tools (e.g., predictive models and generative AI) are now evolving into agentic systems capable of planning and executing complex, multistep tasks with minimal human input. As AI systems' ability to plan and execute tasks autonomously accelerates, so too does the potential speed and scale of cyber-attacks. Compounding this risk, traditional cyber-defences were designed to react to human-driven threats rather than automated, machine-speed attacks. As such, organisations may be increasingly vulnerable to losses arising from cyber incidents unless their security frameworks evolve to keep pace.

What Makes Agentic AI Different?

Agentic AI refers to systems that can plan, execute and adapt to tasks in real time with minimal human oversight. These systems consist of autonomous, interconnected components—often referred to as AI agents—that work together. Components may include a large language model (LLM) for reasoning and decision-making; integrated tools such as APIs, scanners and scripts to interact with external systems; and feedback loops to assess results, learn from outcomes and refine actions.

Unlike generative AI, which primarily produces outputs in response to human prompts, agentic AI is designed to act independently. In the context of cyber-security, AI's capabilities mean that it is moving beyond passive support and can directly enable threat actors to carry out cyber-attacks. Rather than waiting for instructions, agentic systems can autonomously scan networks, identify vulnerabilities, write and test malicious code, and access or exfiltrate data. They can also retain memory between sessions, allowing them to build on actions, observe outcomes and retry approaches until they meet their objectives. As a result, cyber-attacks may become more persistent and difficult to detect.

How Attackers Are Using Agentic AI

Cyber-criminals may use agentic AI to achieve several malicious objectives, including the following:

- **Autonomous network movement**—Agentic AI can independently map a network's architecture, identify critical assets (e.g., financial data) and escalate privileges (e.g., by exploiting system misconfigurations) to gain broader access. This enables cyber-criminals to move laterally across systems, significantly increasing the scope and impact of intrusions. Such movement may occur rapidly. According to CrowdStrike's 2026 Global Threat report, the average time between an attacker's initial access and their first instance of lateral movement fell to just 29 minutes in 2025, with the fastest observed time just 27 seconds.
- **Highly personalised social engineering at scale**—Agentic AI can autonomously harvest data from sources such as social media and public records to generate highly-targeted phishing and social engineering campaigns at scale. Unlike traditional phishing, which relies on generic messages sent to large audiences, these systems can refine their approach in real time based on the success or failure of prior attempts, potentially exposing organisations to more frequent and sophisticated threats, including email compromise and deepfake impersonation.
- **Memory poisoning**—Cyber-criminals can introduce false or malicious information into an agentic system's memory to alter its behaviour over time. For example, a procurement AI could be poisoned to route payments for a specific vendor to an external account, triggering fraudulent payments when a legitimate invoice is later processed.
- **Attacks at scale**—Agentic AI can be used to deploy multiple agents to launch cyber-attacks simultaneously. This approach increases the volume and speed of attacks, potentially

overwhelming impacted organisations' defences and making threats more difficult to detect and contain.

This shift toward more autonomous, agentic AI-driven cyber-attacks is already evident in real-world activity. In September 2025, a Chinese state-sponsored group used an AI coding tool to target approximately 30 organisations in a coordinated campaign, showing how rapidly these capabilities are being applied in practice.

What Organisations Should Do

Organisations can consider the following measures to reduce the potential impact of agentic AI:

- **Microsegment networks**. Organisations should divide networks into smaller isolated zones, each with its own security rules, to reduce the risk of lateral movement. This approach, known as microsegmentation, restricts communication between devices, applications and other network components. Alongside this, organisations should adopt a zero-trust approach, where all connections, including internal ones, are continuously verified, and access is limited to only what is necessary, following the principle of least privilege.
- **Implement behaviour-based endpoint detection**. Organisations should implement behaviour- and anomaly-based monitoring alongside antivirus tools. While antivirus software detects known threats such as malware signatures, agentic AI can change its signature with each execution to evade detection. Behaviour-based monitoring, in contrast, identifies deviations from normal system activity, using defined or learned baselines to detect and respond to previously unknown threats.
- **Audit permissions on internal AI tools.** Organisations should regularly review and audit the permissions granted to AI tools and agents to ensure they are appropriate for their defined tasks, thereby reducing the risk of exploitation. This includes limiting access to specific systems (e.g., allowing an agent that schedules meetings access to a calendar application, but not to corporate email systems) and, where possible, using just-in-time permissions so access is temporary and available only when needed, rather than permanent.
- **Update incident response plans for AI-driven threats**. Organisations should review and update their incident response plans to account for the speed and autonomy of AI-driven attacks, which can unfold far more quickly than traditional incidents. Regular tabletop exercises should be used to test how teams respond and to identify any gaps in readiness. Plans should include scenarios where AI tools and agents are compromised, such as through manipulation or data poisoning.
- **Speak with a broker**. Organisations should work with brokers to ensure their cyber-insurance policies adequately address evolving AI-related risks. Most cyber-policies today are silent on AI—coverage is neither expressly granted nor excluded—so organisations should seek clarity on how AI-enabled attacks are treated, seek affirmative endorsements where available and watch for emerging AI exclusions at renewal, including coverage for incidents involving misused or compromised AI tools and losses arising from autonomous or automated activity.

Conclusion

As agentic AI increases the speed and scale of cyber-attacks, organisations must adapt their security frameworks accordingly, strengthening governance, improving preparedness and evolving controls to detect, contain and respond to more autonomous attacks. Reviewing cyber-insurance policies can help ensure adequate financial coverage. Contact us today for further information.