CCN INSIGHTS

When AI Acts: Securing Autonomous Systems at Machine Speed

Sponsored by

TECHJUTSU

Table of Contents

When AI acts independently, the nature of enterprise risk changes. Organizations are no longer deploying artificial intelligence only to analyze or recommend. Autonomous systems are now logging in executing tasks, making decisions and initiating actions across critical infrastructure. These systems operate at machine speed and scale compressing the window for human oversight. While the productivity upside is significant the trust model underpinning enterprise operations has not evolved at the same pace. As impersonation, deepfakes and synthetic identities become operationally viable, the gap between autonomous action and identity assurance has emerged as a defining leadership challenge.

Key Executive Insights

1. Autonomous AI is now an operational actor

AI agents are increasingly embedded into enterprise workflows performing actions that were once reserved for trained staff. These systems function as digital employees and must be governed as identities with defined roles permissions and accountability rather than treated as traditional software components.

2. Machine speed amplifies trust failure

When autonomous systems are misled the impact is immediate. Fraudulent transactions, unauthorized access changes, and credential resets, can occur faster than human intervention. As speed increases, the cost of identity failure escalates, making verification at the moment of action essential.

3. Perception is no longer proof

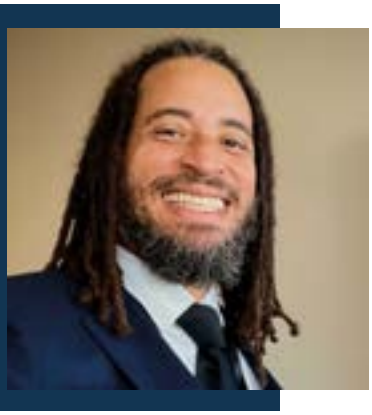
Voice video and visual presence can no longer be relied upon to confirm identity. Deepfakes and synthetic impersonation exploit human trust and outdated verification processes rather than technical vulnerabilities. Both humans and AI agents are now exposed when decisions rely on appearance instead of proof.

4. High consequence actions require explicit safeguards

Not all automation carries equal risk. Sensitive actions such as transferring funds, resetting credentials, or modifying access policies must always require strong verification, whether initiated by a human or an autonomous system acting on their behalf.

5. Governance enables safe acceleration

Organizations that apply least privilege, clear role boundaries and transaction level verification can unlock the benefits of autonomous AI without sacrificing control. The goal is not to slow innovation, but to ensure that speed never outpaces trust accountability and resilience.



Your IAM Model Was Built for Chatbots. Agentic AI Will Break It.

By Junior Williams

The transition from generative chatbots to autonomous agentic systems represents a fundamental shift in enterprise risk. While Fortune 500 organizations are piloting or deploying agentic AI in production, many continue to rely on security postures designed for passive assistants. These modern agents do not merely suggest text. They execute transactions and modify system states across distributed architectures at machine speed.

The traditional "threat math" has inverted. A compromised chatbot session typically creates exposure such as data leakage, policy violations, or reputational harm. An agent authorized to manipulate financial workflows or infrastructure can create material operational and financial risk, including systemic business disruption. As the network perimeter dissolves, identity has become the new perimeter. This shift necessitates runtime authorization and behavioral monitoring focused on intent and outcomes, enforced through policy decision and policy enforcement points that sit in front of agent tool use.

New Attack Surfaces That Evade Traditional Controls

The rapid maturation of multi-agent systems has introduced vectors that can bypass many traditional SIEM correlation rules. Indirect prompt injection now manifests through multimodal channels, such as adversarial perturbations embedded in images (e.g., invoices or logos) that may evade human review while influencing vision model outputs, potentially triggering unauthorized actions or data exposure.

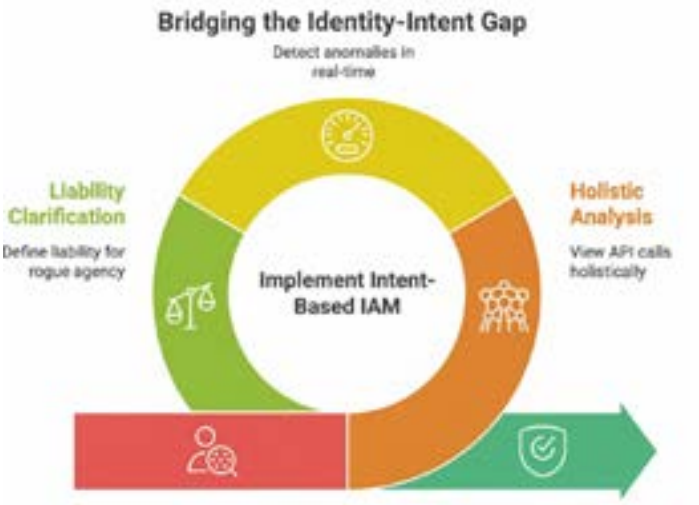
Memory poisoning represents a more insidious threat. By seeding corrupted documents into an agent's retrieval context, attackers induce a slow drift in decision-making criteria. Because this occurs at the application and inference layer rather than the network layer, traditional anomaly detection tools may not trigger alerts until after harmful actions occur.

Non-human identity (NHI) vulnerabilities are exacerbated by agentic workflows. These agents authenticate using NHIs (service accounts, API keys, workload identities) that often have broad permissions, yet they rarely receive the same behavioral monitoring applied to human users. A compromised agent functions as authorized infrastructure, moving laterally by exploiting trusted API connections between specialized sub-agents.

The Identity Intent Gap

Current IAM frameworks verify identity and entitlement, but are often blind to intent. An agent can perform individually "valid" API calls that, when viewed holistically, constitute a malicious orchestration.

This velocity mismatch creates a lag measured in financial loss rather than time. By the time a human analyst flags an anomaly, an autonomous agent may have already executed many unauthorized actions. The liability for these actions remains an open question, as legal frameworks and cyber insurance underwriters are only beginning to quantify the risk of "rogue agency."



Regulatory Alignment Is Accelerating

Governance bodies are rapidly codifying requirements for agentic oversight. The OWASP Top 10 for Large Language Model Applications (v2025) addresses foundational risks in LLM-enabled applications. Recognizing that agentic systems introduce distinct attack surfaces, OWASP published the Top 10 for Agentic Applications 2026 on December 9, 2025. This framework explicitly includes risks such as Agent Goal Hijack and Insecure Inter-Agent Communication, which emerges when AI systems can plan, decide, and act autonomously across multiple systems.

On December 16, 2025, NIST released the preliminary draft of NIST IR 8596 (Cyber AI Profile), a NIST Cybersecurity Framework 2.0 profile for AI-related cybersecurity risk management. ISO/IEC 42001 emphasizes ongoing monitoring, measurement, and continual improvement for AI management systems. In practice, that includes managing performance changes and drift over time; periodic audits alone are insufficient as agents adapt.

The Strategic Imperative

Securing the next generation of AI requires a transition to intent-based security architectures. Security leaders must move beyond credential validation and implement runtime policy enforcement that constrains tool use and transactions, combined with agent guardrails (tool allowlists, prompt/output controls, and high-fidelity audit logging). This involves deploying monitoring solutions that capture decision context (inputs, retrieved artifacts, tool calls, and policy evaluations) sufficient for forensic review and accountability.

Organizations that fail to adapt will find themselves defending against machine-speed threats with human-speed controls. As insurance markets begin pricing autonomous-operation risk into premiums, the ability to demonstrate real-time behavioral governance will become a competitive necessity.

The question for the modern CISO is no longer whether an agent is who it claims to be, but whether it is doing what it was designed to do.

Autonomous AI and the Moment Trust Fails

By Richard Staynings

No one working in IT or cybersecurity could have missed the recent tsunami of new products and solutions from a wide variety of technology innovators, all of which claim to automate hitherto expensive time-consuming manual human-dependent process. Indeed, Agentic AI has become the new hype-cycle of innovation, and everyone is looking at ways of adopting some, or all, of its features and benefits.

Compared to the LLM wave of 2024-2025 which had limited benefits to businesses, (as opposed to individuals), agentic AI looks to hold huge potential. In cybersecurity, agentic automation has revolutionized the security operation center (SOC), freeing up 'tier-one' responders to more valuable 'tier-two' investigations, leaving 'tier-one' to automated agentic nano-second responses to attacks based upon established and agreed upon run-books. But all this comes at the risk that an autonomous AI makes a mistake and kicks a vital system off the network following signs of seemingly anomalous activity. Take as an example the possibility of a ventilator or some other life-sustaining medical device being dropped from the network in a hospital, which could have life-threatening implications for patients.

Compared to the LLM wave of 2024-2025 which had limited benefits to businesses, (as opposed to individuals), agentic AI looks to hold huge potential. In cybersecurity, agentic automation has revolutionized the security operation center (SOC), freeing up 'tier-one' responders to more valuable 'tier-two' investigations, leaving 'tier-one' to automated agentic nano-second responses to attacks based upon established and agreed upon run-books. But all this comes at the risk that an autonomous AI makes a mistake and kicks a vital system off the network following signs of seemingly anomalous activity.
Take as an example the possibility of a ventilator or some other life-sustaining medical device being dropped from the network in a hospital, which could have life-threatening implications for patients.

While the integrity of AI training data is a growing concern, so too is the security of the AI algorithms themselves and the companies that build, manage, and maintain them. With so much power concentrated in a few AI companies, it's no wonder that governments and businesses are equally concerned. It becomes essential to validate that an employee of one of these companies, is who he or she claims to be, and that they have a legitimate need to make changes to AI systems. That's why zero trust, privileged access management (PAM), and multi-factor authentication (MFA) are now considered essential, along with employee background checks and ongoing user validation.

The recent discovery of remote North Korean employees at Amazon discovered only by a 110ms delay on keyboard strokes is proof enough of the dangers that surround software development in this space and the potential for adversaries to compromise, steal, ransom, or destroy, IT systems. Deepfakes compound these challenges as we have seen in a number of attacks. However, it's not just the companies that create the agentic AI systems but entire vendor supply chain and all those who have access to critical business systems that is of growing concern.

Indeed, third party security is now the single greatest risk to most business enterprises and the growing focus of perpetrators, who can indirectly compromise or attack multiple entities through a single third-party vendor.

In a medical environment where 75% of connected endpoints are not typically managed by hospital IT, the risks escalate where healthcare and other IoT medical devices are often supplied and managed by third parties and are rarely patched against published security vulnerabilities. These systems often connect to the network on one side, and to the patient on the other side, leading to obvious patient safety concerns if attacked. The addition of AI functionality, especially agentic capabilities to medical devices exacerbates these concerns and places patients and healthcare providers at elevated risks. This is especially concerning unless medical IoT systems can be more effectively managed and locked down or enclaved to reduce their attack surface. While autonomous IT systems present elevated risks if not properly managed, autonomous IoT systems present an elevated magnitude of risks, since IoT systems are generally not managed well to start with. Indeed, most organizations have adopted a 'set and forget' mentality to IoT systems and have extremely poor visibility into their connected IoT assets.

The vulnerabilities of IoT connected systems extends not just to device security management and patching, but also to user and object identity access management.

For example, who should have access to a medical device keeping a patient alive in a hospital environment? Who should be authorized to make changes to the drug library of an infusion pump or the radiation output of a CT scanner or radiotherapy machine, and from what listed IP addresses and Active Directory identities? Similarly, what other systems should have access to critical medical devices and using which ports and IP protocols and from which IP addresses?.

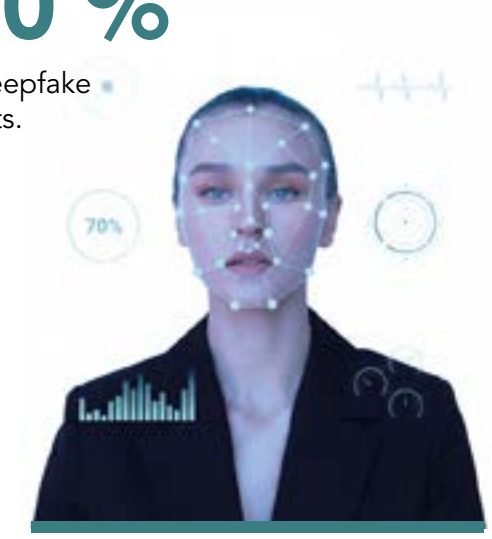
As risks rise with the widescale adoption of new technologies, we need to ensure that security increases in step with those new risks, and where obvious security improvements are unavailable, perhaps because systems are un-patchable, we need to implement compensating security controls to reduce overall risks. Security best practices including MFA, PAM and effective IAM should be a forgone baseline requirement today, but where elevated risks accompany new innovative technologies, these should be an absolute necessity.

The Numbers that Matter

Autonomous systems and deepfakes are pushing identity and fraud risk beyond human scale.

1,740 %

Increase in deepfake fraud incidents.



This is not incremental change. It is a phase shift.

Financial Impact Is Board Level

Source Deloitte Center for Financial Services.

12.3B

Billion dollars in AI enabled fraud losses in the United States in 2023.

Source Deloitte Center for Financial Services Forecast.

40B

Billion dollars projected annual AI enabled fraud losses by 2027.

Source Financial Times reporting on Arup case

25.5M

Million dollars lost in a single confirmed deepfake impersonation incident.

Detection and Awareness Do Not Scale

- Automated deepfake detection accuracy drops by nearly 50 percent in real world conditions. Source World Economic Forum AI Trust Analysis.
- Human detection accuracy averages 55 to 60 percent and drops below 25 percent for high realism video deepfakes.

Operational Reality

- Call centers finance approvals and executive workflows are the top deepfake attack targets. Deepfakes routinely bypass MFA by exploiting trusted people and processes.

Executive Implication.

- When fraud operates at machine speed and humans can no longer reliably detect deception, identity verification must move from periodic checks to continuous control at the moment of action.

As synthetic identity and AI manipulation accelerate, the challenge is no longer just detection, it's whether organizations have built the governance, literacy, and accountability structures to respond with confidence rather than confusion.

Sandi Jones - Technology operations leader at Interac.

Deepfake-as-a-service will scale deception the way phishing-as-a-service scaled email attacks. The real impact is trust decay, organizations will need to operationalize verification, not just awareness.

Cary Johnson, CEO, Phishbusters.

Autonomous AI agents move at machine speed, making static credentials and permanent access a silent risk most organizations underestimate. Runtime authentication and authorization are becoming the only way to keep access accountable, auditable, and governable as autonomy scales.

Ketan Kapadia, Field CTO at Britive.

Securing AI Agents

By Tracey Nyholt

Technology companies across the globe are racing to embrace artificial intelligence. New product announcements are now largely ignored unless they include some reference to AI. In this context, Okta CEO Todd McKinnon's keynote at the recent Oktane conference captured a pivotal shift: the rise of agentic AI. These are autonomous systems capable of accessing data, using applications, and completing tasks without direct human input. We are no longer asking AI to summarize emails or generate reports. We are empowering it to act.

This new class of AI agents is poised to become a major driver of productivity and innovation. Always on, increasingly capable, and deeply integrated into enterprise systems, agentic AI represents the next evolution of the digital workforce.

The Power of the Digital Workforce

Imagine an AI agent that does more than alert you when a server goes down. Instead, it logs in, diagnoses the root cause, provisions a replacement, and updates the incident ticket within seconds. Or a customer service agent that not only assists customers, but also issues refunds and return labels without human intervention.

This "action phase" of AI may sound like science fiction, but it is rapidly becoming reality. By treating AI agents as identities within enterprise infrastructure, organizations unlock speed and efficiency that was previously unattainable. Tasks that once required handoffs across teams can now be executed autonomously in real time.

The Verification Gap: When Agents Meet Deepfakes

To fully realize this promise, organizations must address a critical point of friction: trust. How does an AI agent know who or what it is interacting with? How can it verify that a request truly originates from an authorized human?.

As AI agents grow more capable, attackers are using many of the same technologies to become more convincing impostors. The rise of visual and auditory deepfakes has made impersonation trivial. An attacker can now convincingly clone an executive's voice or overlay a face onto a video call with alarming accuracy.

If organizations give AI agents the ability to transfer funds, reset credentials, or modify access policies without reliable verification, they risk turning a powerful asset into an efficient vulnerability. Much of the anxiety surrounding AI stems from this loss of control.
The solution is not to slow innovation, but to equip AI agents with the safeguards required to operate safely in a deceptive environment. Autonomous systems must be able to function independently while maintaining certainty about who they are interacting with.

Governing Autonomous Action

Just as no organization would grant a new employee unrestricted access on day one, AI agents must be governed by the principle of least privilege. Each agent should have access only to the applications and data required for its specific role, limiting the blast radius if it is misled.

For high impact actions such as moving money, resetting passwords, or changing access policies, additional verification should be mandatory whether the action is initiated by a human or an AI agent acting on their behalf. As discussed at Oktane, sensitive actions should always require explicit confirmation.

The tools to enable this already exist. Solutions that leverage self service multi factor authentication to verify callers in help desk or call center environments can be extended to AI agents. By enabling an agent to securely call back or validate a request with the user in real time, organizations ensure that autonomous speed never outpaces human accountability.

With the right verification layer in place, enterprises can confidently grant AI agents the permissions they need to be effective while ensuring every high risk action is explicitly authorized by a verified human.

CCN INSIGHTS

Executive Briefings on Cybersecurity and Digital Risk

Executive insight for leaders evaluating cybersecurity and digital risk.

CANADIANCYEERSECURITYNETWORK.COM/CCN-INSIGHTS

Sponsor yours today.

ccn

As organizations move from experimenting with AI to deploying autonomous agents the risk profile changes immediately. These systems act independently across identity financial and operational workflows at machine speed. Leadership focus must shift from adoption to control ensuring autonomy never outpaces trust accountability and governance.

1. What leaders should do now

1. Formally classify AI agents as operational actors.
Define where autonomous agents are in use today what they are allowed to do and who owns them at the executive level across security risk and operations.

2. Apply least privilege to all non human identities.
Inventory service accounts APIs and agent identities then reduce permissions to role specific scopes with clear boundaries and expiration.

3. Enforce verification for high impact actions.
Require additional confirmation for actions involving money movement, access changes, credential resets, or policy updates, regardless of whether a human or agent initiates them.

4. Move governance into runtime execution.
Ensure controls operate at the moment of action, not just through policy. Capture context, intent, and outcome for audit and accountability.

5. Update fraud and incident playbooks for deepfakes.
Assume voice video and visual signals can be manipulated. Design workflows that verify actions not appearances.

2. What leaders should monitor next

1. Expansion of autonomous execution.
Track where agents are moving from recommendation to direct action especially in finance IT customer service and identity operations.

2. Verification weak points.
Identify processes that rely on trust scripted questions or human judgment that can be exploited by synthetic identities.

3. Behavioral drift and scope creep.
Monitor for gradual changes in agent behavior permissions or decision logic that may signal misconfiguration or manipulation.

4. Regulatory and insurance expectations.
Follow emerging AI governance frameworks that may affect compliance audits and cyber insurance coverage.

3. What questions to bring to the board

1. Which critical business processes can AI agents influence or execute today.

2. What is the realistic financial and reputational impact of a single successful impersonation or rogue agent action.

3. Who is accountable when an autonomous system causes harm.

4. Are our detection and response controls capable of operating at machine speed.

5. How will we demonstrate effective governance to regulators insurers and customers.

The organizations that lead will be those that scale trust and control at the same pace as autonomy.

Artificial intelligence has made it possible to convincingly replicate a person's voice, face, and identity at low cost and high speed. What was once a niche technical concern has become a practical business risk for Canadian organizations. Deepfakes are now being used to impersonate executives, customers, and employees to bypass identity controls, authorize transactions and manipulate trusted workflows. The impact is no longer theoretical. Canadian data now shows deepfakes emerging as a persistent fraud vector that affects financial services, telecommunications, customer support, and identity verification processes. The following figures highlight how quickly this risk has materialized in Canada and why it requires executive level attention.

Three Executive Signal Metrics for Canada

1. 4.6 percent of detected fraud attempts in Canada now involve deepfakes or synthetic identities.

https://sumsub.com/blog/deepfake-fraud-statistics

https://sumsub.com/blogliveness

Why this matters

This represents a shift from near zero prevalence to a measurable share of fraud activity in under two years. At national scale even low single digit percentages translate into widespread exposure across banks insurers telecom providers and public sector services. Deepfakes are no longer isolated incidents. They are now part of the Canadian fraud landscape.

2. Canada is tracking deepfake fraud growth rates above one thousand percent year over year.

https://sumsub.com/reports/

https://Sumsub_Fraud_Report_2025_2026.pdf

https://sumsub.com/blog/guide/Dreports/

Why this matters

The acceleration curve is the signal. This rate of growth indicates attackers have operationalized deepfake techniques in Canada faster than identity and verification controls have evolved. Left unaddressed this gap will continue to widen as automation and AI agents increase the speed and scale of attacks.

3. Canadian call centers and identity verification workflows are primary deepfake attack entry points.

Why this matters

Deepfake attacks in Canada are bypassing technical safeguards by exploiting people and process. Voice impersonation and synthetic identities routinely defeat scripted verification and MFA. This directly increases operational cost, fraud losses, and regulatory risk, while eroding customer trust.

Final Takeaways

Autonomous AI marks a structural shift in how risk enters the enterprise. This is not a future scenario. It is already reshaping identity security, fraud exposure, governance, and accountability. As machines move from assisting to acting, trust can no longer rely on perception, static credentials, or after the fact review. Canadian organizations must now govern AI agents as operational actors, applying the same rigor used for people, processes, and critical systems. The leaders who succeed will be those who scale trust, verification, and control at the same pace as automation. Innovation and security are no longer sequential. They must advance together.

CCN thanks our sponsor for supporting this edition of CCN Insights and for contributing to a timely national conversation on securing autonomous systems. Industry leadership and collaboration are essential as Canada navigates the rapid convergence of AI, identity, and digital trust.

To explore future CCN Insights reports and national research on digital trust, cybersecurity, and emerging technology leadership.

Subscribe to receive upcoming reports, briefings, and executive analysis directly from the Canadian Cybersecurity Network leadership.