WHITE PAPER

Agentic AI Threats

A MITRE ATT&CK; Analysis

How autonomous systems change attack tempo, trust boundaries, and enterprise defense

---

Agentic AI does not need to become fully autonomous cyber warfare to materially change risk. The nearer-term challenge is more practical: semi-autonomous systems can sustain reconnaissance, adapt phishing and infrastructure, chain credentials, operate through browsers and SaaS, and continuously exploit the trust relationships modern enterprises depend on.

Patrick M. Hayes

Integrated Assurance®

Condensed from the 2nd Edition analysis • September 2026

EXECUTIVE SUMMARY

The threat is not intelligence. It is operational persistence.

Agentic AI changes cyber operations because it can combine reasoning, memory, browser interaction, API access, tool orchestration, and multi-agent coordination into sustained operational loops. Traditional automation accelerated individual tasks. Agentic systems can interpret feedback, choose what to do next, and continue pursuing an objective with limited human supervision.

That changes the economics of attack. A smaller adversary team can maintain continuous reconnaissance, personalize social engineering, test identity workflows, rotate infrastructure, correlate exposure across cloud and SaaS environments, and adapt when defenders intervene. The result is not necessarily a fully autonomous attacker. It is a persistent, semi-autonomous operating model that can act at a tempo many security programs were not designed to match.

MITRE ATT&CK; remains useful because the underlying adversary behaviors still matter. What changes is the operational context around those behaviors. Reconnaissance may continue throughout an intrusion. Credential access can focus on sessions, OAuth grants, API keys, and delegated permissions. Lateral movement can occur through SaaS relationships and cloud roles rather than host-to-host traversal. Malicious activity can increasingly look like legitimate work performed through trusted interfaces.

Five conclusions frame this white paper:

- **AI is both a tool and an attack surface.** Enterprise AI systems inherit data access, permissions, connectors, memory, and workflow authority that can themselves be manipulated.
- **Identity is becoming the operational perimeter.** Valid sessions, delegated access, OAuth grants, browser cookies, and federated trust can be more useful to an attacker than malware.
- **The kill chain is becoming an adaptive loop.** Agentic systems can overlap reconnaissance, discovery, credential access, persistence, and infrastructure adaptation continuously.
- **Visibility must become behavioral and cross-domain.** Endpoint, cloud, SaaS, identity, browser, API, and AI telemetry cannot remain isolated if attackers correlate across them.
- **Resilience becomes a strategic objective.** Organizations must be able to validate trust, contain abnormal behavior, recover operational confidence, and continue operating when prevention fails.

What makes Agentic AI different

Agentic AI refers to systems capable of autonomous goal pursuit, adaptive reasoning, memory retention, environmental awareness, and multi-step execution across connected systems and tools. The distinction from earlier automation is not simply that the models are more capable. It is that the operating loop can persist.

An agent can observe an environment, act, evaluate the result, retain context, revise its approach, and continue. When that loop is connected to browsers, APIs, cloud services, identity systems, SaaS platforms, or enterprise automation, the system gains the ability to work through the same operational pathways used by legitimate users.

The attack surface expands from systems to relationships

Modern enterprises are built from interconnected trust: federated identities, OAuth grants, SaaS connectors, browser sessions, cloud roles, APIs, copilots, automation platforms, and shared data stores. Agentic AI is well suited to mapping these relationships because the environment is increasingly machine-readable and continuously changing.

The practical consequence is that compromise may no longer begin with malicious code. An attacker can pursue trusted authentication, manipulate an AI workflow, poison retrieved context, exploit delegated permissions, or operate through an authenticated browser session. The security problem becomes less about whether a tool is approved and more about whether the behavior occurring through that tool still makes operational sense.

A compressed operating model

- Continuous low-volume reconnaissance rather than a discrete pre-attack phase.
- Adaptive social engineering that changes tone, timing, channel, and context based on target behavior.
- Browser operation that can navigate legitimate interfaces instead of relying only on malware or APIs.
- Infrastructure rotation and payload variation in response to detection.
- Persistent correlation of identities, permissions, SaaS relationships, APIs, and cloud configuration.
- Multi-agent specialization that divides reconnaissance, phishing, infrastructure, monitoring, and data movement across coordinated agents.

ATT&CK; still applies. The tempo and context change.

MITRE ATT&CK; remains a strong way to reason about adversary behavior because its value is behavioral rather than dependent on a particular malware family. Agentic AI does not invalidate the framework. It changes how tactics can be coordinated, how long they can persist, and how quickly an adversary can adapt.

The table below condenses the manuscript's stage-by-stage analysis into an executive view of how agentic capabilities alter each ATT&CK; tactic.

Tactic	Agentic shift	Defensive implication
Reconnaissance TA0043	Continuous monitoring of infrastructure, identities, public data, SaaS adoption, APIs, and organizational change.	Low-volume activity may persist indefinitely and feed every later phase.
Resource Development TA0042	Automated domains, cloud infrastructure, accounts, phishing assets, payload variation, and proxy rotation.	Disrupted infrastructure can be regenerated faster.
Initial Access TA0001	Adaptive phishing, trusted SaaS workflows, identity abuse, browser interaction, and exposed services.	Access attempts can personalize and change in real time.
Execution TA0002	Tool use, APIs, cloud functions, browser actions, automation workflows, and legitimate interpreters.	Execution may blend into approved business activity.
Persistence TA0003	Tokens, OAuth grants, cloud roles, SaaS access, browser sessions, automation hooks, and adaptive mechanisms.	Persistence can move with the environment rather than remain fixed.
Privilege Escalation TA0004	Permission chaining across cloud, SaaS, identity, workflows, and delegated trust.	Moderate weaknesses can be correlated into escalation paths.
Defense Evasion TA0005	Behavioral adaptation, infrastructure rotation, timing changes, legitimate tools, and session-based activity.	Static signatures and fixed indicators decay faster.
Credential Access TA0006	Passwords plus tokens, cookies, API keys, OAuth grants, temporary credentials, and AI tool access.	Identity artifacts become high-value operational assets.
Discovery TA0007	Continuous mapping of roles, services, workflows, cloud assets, data stores, and trust relationships.	Discovery becomes persistent environmental awareness.
Lateral Movement TA0008	Federation, SaaS relationships, cloud role inheritance, session reuse, integrations, and shared automation.	Movement may occur without traditional network traversal.
Collection TA0009	Automated identification, prioritization, and staging of data across SaaS, cloud, documents, and AI-connected stores.	AI can contextualize which data is operationally valuable.
Command & Control TA0011	Adaptive channels, cloud services, legitimate platforms, rotating infrastructure, and variable timing.	C2 can become more fluid and difficult to disrupt.
Exfiltration TA0010	API-driven transfer, cloud synchronization, SaaS sharing, browser workflows, and staged automation.	Data movement can resemble normal enterprise activity.
Impact TA0040	Disruption of workflows, data, identity, orchestration, synchronization, AI context, and business operations.	Operational trust and continuity become direct targets.

Identity, browsers, APIs, and AI workflows converge

The manuscript's analysis repeatedly points to the same structural change: enterprise boundaries are dissolving. Organizations increasingly operate through federated identity, cloud services, browser-centric applications, APIs, SaaS integrations, orchestration platforms, and AI assistants. Security boundaries no longer align neatly with organizational boundaries.

Identity-centric attacks

A valid identity can provide access to cloud administration, collaboration systems, financial applications, customer data, source repositories, automation platforms, and AI copilots. Agentic systems can continuously map role assignments, access changes, authentication patterns, API scopes, and SaaS relationships, looking for chains of permission that are easy to miss when reviewed in isolation.

Browser-centric operations

Browser-operating agents create a different detection problem. They can navigate portals, complete forms, read dashboards, trigger workflows, download documents, and act inside authenticated sessions. From a security tool's perspective, much of that activity may look legitimate. The distinguishing signal is increasingly behavioral context rather than the presence of malware.

AI becomes an aggregation layer

Enterprise assistants and copilots are often connected to email, documents, CRM data, cloud storage, collaboration platforms, development environments, and operational dashboards. Their usefulness comes from concentrated context and delegated access. Those same characteristics make the surrounding workflow a high-value target.

> **An attacker may not need to compromise the underlying model. Manipulating retrieved context, abusing a connector, hijacking an authenticated session, or exploiting delegated authority may be enough to influence the outcome.**

Threats that sit beyond a traditional kill chain

The source analysis extends beyond conventional ATT&CK; mapping because some AI-enabled risks target the decision and orchestration layer itself. These risks are important precisely because they may not resemble a conventional intrusion.

Prompt injection and indirect prompt manipulation

Instructions embedded in content, external data, websites, documents, or tool outputs can influence an AI system's behavior. The risk grows when the system can take action rather than only generate text.

Retrieval-augmented generation poisoning

When AI systems rely on connected repositories and retrieved context, corrupted or adversarial source material can distort decisions and downstream actions without requiring direct model compromise.

Browser-operating agents

Agents that interact visually with web applications can emulate human workflows, operate through approved interfaces, and cross application boundaries that were designed around human judgment.

Multi-agent coordination

Specialized agents can divide operational roles, share context, and coordinate continuously. This can increase persistence and scale while reducing the amount of human supervision required.

Delegated trust exploitation

AI and automation systems frequently inherit permissions from users, service accounts, connectors, and orchestration platforms. The chain of delegated authority can become more important than any single system.

AI-to-AI interaction

As defensive and operational agents interact with other automated systems, organizations must consider manipulation, feedback loops, conflicting objectives, and the integrity of machine-generated context.

Static controls cannot match a continuously adapting environment

Most enterprise security programs were designed around relatively stable infrastructure, recognizable boundaries, human-paced attacks, and discrete security events. Modern environments are distributed across cloud, SaaS, identity, browsers, APIs, automation, and AI. Agentic systems amplify that complexity by continuously observing and adapting to it.

The defensive model must shift in six areas

1. Continuous exposure awareness

Move beyond periodic reviews toward continuous understanding of configuration drift, new APIs, SaaS onboarding, identity changes, cloud exposure, and AI workflow connections.

2. Identity and session telemetry

Monitor not only authentication success but session behavior, device consistency, token lifecycle, access velocity, SaaS interaction patterns, conditional access context, and delegated permissions.

3. Cross-domain behavioral correlation

Correlate signals across endpoints, cloud, SaaS, browsers, identity, APIs, and AI systems. The attacker can connect these domains, so defenders must be able to do the same.

4. AI workflow governance

Inventory where AI systems retrieve data, what tools they can invoke, what permissions they inherit, how memory is retained, and how context is validated before actions occur.

5. Machine-speed response

Automate containment and response where the operational risk is well understood. Human approval remains important, but response models cannot assume every decision can wait for manual triage.

6. Operational recovery

Plan for restoration of workflow integrity, identity trust, orchestration behavior, AI context, and business confidence, not only infrastructure recovery.

> **The goal is not simply more visibility. It is operational awareness at a speed that is relevant to the environment being defended.**

06 | EXPOSURE CHAINING

The real risk often exists between individually acceptable controls

A recurring theme in the analysis is that modern compromise can emerge from combinations of moderate exposures rather than one catastrophic weakness. A SaaS integration, OAuth scope, browser session, cloud storage location, AI assistant, and automation workflow may each appear acceptable when assessed independently. Together they can form a high-impact operational path.

STEP	EXPOSURE	WHY IT MATTERS
1	SaaS integration	Introduces a trusted connection
2	OAuth / delegated permission	Extends authority across systems
3	Browser or session context	Provides authenticated operating access
4	Cloud / data connection	Expands reachable information and services
5	AI assistant or agent	Adds reasoning, context, and tool use
6	Automation / orchestration	Turns access into repeatable action

This is why isolated control validation is increasingly insufficient. The organization needs to understand how trust propagates across the complete operational chain.

Questions leaders should ask

- Which systems can act on behalf of people, services, or other systems?
- Where do AI tools inherit access to business data or operational functions?
- Which permissions propagate through integrations without being reviewed as a complete chain?
- Can we distinguish normal automated behavior from malicious intent operating through legitimate interfaces?
- How quickly can we revoke trust, contain abnormal automation, and restore reliable operations?

Cybersecurity becomes an operational governance discipline

The strategic implication is broader than a new class of technical threats. Cybersecurity increasingly has to govern how workflows operate, how trust propagates, how automation inherits authority, how AI influences decisions, and how the organization continues operating when those mechanisms become unreliable.

This brings security closer to enterprise architecture, business continuity, governance, financial risk, and operational resilience. The question for leadership is no longer limited to whether controls are deployed. It is whether the organization understands the operational relationships those controls are meant to protect and whether it can validate them continuously.

Board and executive discussion points

- **Authority:** Where have we delegated decision or execution authority to AI and automation?
- **Dependency:** Which business processes now depend on AI, SaaS, browser workflows, APIs, or orchestration platforms to continue operating?
- **Trust:** How do we validate that identities, sessions, data, context, and automated actions remain trustworthy over time?
- **Visibility:** Can we correlate abnormal behavior across organizational and technology silos?
- **Resilience:** If an AI-enabled workflow becomes unreliable or maliciously influenced, can the business continue operating while trust is restored?
- **Accountability:** Who owns the operational consequences when automated decisions cross technology, security, risk, and business boundaries?

Operational complexity is scaling faster than organizational understanding. That gap is becoming a material source of risk.

A practical agenda for the next 90 days

The source manuscript argues for continuous telemetry, identity-centric controls, behavioral analytics, AI-assisted detection, deception, continuous validation of trust, and operational recovery readiness. The following condensed agenda translates those themes into a practical starting point.

Map AI-enabled operations

Identify copilots, agents, RAG systems, browser automation, orchestration tools, AI-connected repositories, and autonomous workflows. Document what they can access and what actions they can take.

Map trust paths

Trace identity federation, OAuth grants, service accounts, API keys, browser sessions, connectors, cloud roles, and delegated permissions that allow authority to propagate.

Prioritize high-consequence workflows

Focus first on workflows tied to finance, customer data, production, privileged administration, code deployment, security operations, and business continuity.

Correlate telemetry

Determine whether identity, SaaS, cloud, browser, API, endpoint, and AI workflow telemetry can be analyzed together rather than as separate control domains.

Exercise abnormal automation

Test how teams identify and contain a valid identity, browser session, or AI workflow behaving in an operationally abnormal way.

Define recovery of trust

Document how the organization will revoke delegated authority, validate data and context, re-establish reliable workflows, and return automation to service after compromise or manipulation.

CONCLUSION

The attack surface is increasingly how the enterprise operates

Agentic AI is not important because it creates an entirely new set of cybersecurity fundamentals. It is important because it changes the speed, scale, persistence, and adaptability with which familiar adversary behaviors can be executed. It also gives attackers new ways to operate through the trusted systems and workflows organizations use every day.

The enterprise is becoming a continuously connected operational trust ecosystem. SaaS, cloud, browsers, APIs, AI assistants, orchestration platforms, synchronization services, and delegated automation increasingly blur the distinction between inside and outside, human and automated, application and workflow.

That makes operational trust a central security problem. Organizations need to know not only whether access is authorized, but whether the behavior, context, workflow, and resulting decisions remain trustworthy. They need the ability to detect change across domains, respond at relevant speed, and recover operational confidence when trust degrades.

> **The defining question is shifting from whether compromise can be prevented perfectly to whether the organization can continue operating safely when an adaptive, AI-enabled environment becomes unstable.**

About the source analysis

This white paper is a condensed executive treatment of Patrick M. Hayes's broader 2nd Edition analysis, *Agentic AI Threats - A MITRE ATT&CK; Analysis*. The full work examines Agentic AI across the ATT&CK; lifecycle, AI-native attack paths, defensive architecture, operational trust, enterprise governance, resilience, and reference architectures.

About Patrick M. Hayes

Patrick M. Hayes is an author, strategist, executive advisor, and keynote speaker whose work focuses on Business Survivability, Integrated Assurance, cybersecurity, AI, governance, and organizational risk. His work examines how technology, operations, trust, and decision-making interact under conditions of rapid change and uncertainty.

patrickmhayes.com | info@integratedassurance.co