ASSESSED INTELLIGENCE

ASSESSED SIGNAL

June 2026

At the Intersection of Policy, Use & Threat Intelligence

On June 12, 2026, a U.S. export control directive suspended foreign national access to Anthropic's Fable 5 and Mythos 5 systems, causing immediate disruption for organizations with cross-border teams. This followed a June 2 executive order establishing a voluntary federal benchmarking process for frontier models. Meanwhile, Verizon's 2026 DBIR highlighted that vulnerability exploitation now accounts for 31% of initial access, with attackers leveraging generative AI across numerous techniques. In the enterprise sector, Gartner projects 40% of agentic AI projects will fail by 2027, primarily due to inadequate governance and unclear ownership rather than technical limitations. Organizations must prioritize assurance and governance before scaling to maintain control over these frontier capabilities.

Secure & Responsible Technology.

AssessedIntelligence.com

The Export Directive Redrew the Terms of Frontier AI Access

On June 12, 2026, the United States issued an export control directive that suspended access to Anthropic's Fable 5 and Mythos 5 systems by any foreign national, including foreign nationals inside the United States and the company's own foreign national employees. Anthropic disabled both models for all customers while it worked to comply. The action arrived ten days after the June 2 executive order titled "Promoting Advanced Artificial Intelligence Innovation and Security," and together the two measures changed how organizations must think about dependency on a single frontier provider. The executive order pairs cyber defense with a controlled-release framework. It directs the National Security Agency and the Cybersecurity and Infrastructure Security Agency to develop a classified benchmarking process to identify models with advanced cyber capabilities, designating those that meet the threshold as "covered frontier models." Developers of such models are invited, on a voluntary basis, to give the government up to 30 days of pre-release access. The order expressly disclaims any mandatory licensing or preclearance requirement, and it instructs CISA, the NSA, and the Department of War to harden federal information systems and prioritize criminal enforcement against malicious AI-enabled cyber activity.

The reaction abroad was immediate. The directive reached G7 deliberations underway in Geneva, and policymakers across the European Union and the United Kingdom treated the suspension as evidence of strategic exposure. British researchers, companies, and hospitals that had been studying or piloting the affected models lost access without notice. Reporting indicates that dozens of European governments, companies, and organizations are now moving or planning to move away from United States systems. Anthropic maintained that its models carry safeguards that substantially reduce the likelihood of misuse for cybersecurity tasks. The documented effect of the directive is a restriction on access. The assessed effect is different. A control that suspends access does not constrain the diffusion of capability; allied governments and enterprises are responding by sourcing alternatives, which redistributes frontier capability rather than containing it.

The posture is reactive, consistent with how governments have historically responded to newly visible cyber risk, and the resulting realignment is unlikely to produce the outcome United States security officials intend. AI adoption is driving productive and destructive innovation worldwide, and access policy alone does not govern either.

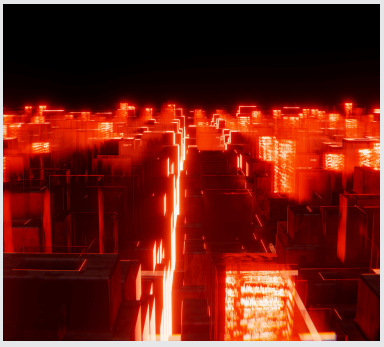
This is a GOVERN problem in the ARISE Framework™. Organizations must treat frontier model dependency as a governed supply decision, with documented continuity provisions, identified ownership of the dependency, and a defensible position on what happens when a provider or a government removes access without warning. Governance that addresses only internal use, and not the conditions under which a critical capability can be withdrawn, leaves the organization exposed to a decision it does not make.

1. The White House. Promoting Advanced Artificial Intelligence Innovation and Security. June 2, 2026. https://www.whitehouse.gov/presidential-actions/2026/06/promoting-advanced-artificial-intelligence-innovation-and-security/
2. IAPP. The global implications of the White House's export controls on Anthropic. June 2026. https://iapp.org/news/a/the-global-implications-of-the-white-houses-export-controls-on-anthropic
3. Nextgov/FCW. Anthropic suspends top AI models after U.S. export control order. June 13, 2026. https://www.nextgov.com/artificial-intelligence/2026/06/anthropic-suspends-top-ai-models-after-us-export-control-order/414173/
4. Greenberg Traurig LLP. White House Issues Executive Order Targeting Frontier AI Models. June 2026. https://www.gtlaw.com/en/insights/2026/6/white-house-issues-executive-order-targeting-ai-enabled-cybersecurity

Tool Sprawl Now Defines the MSP Response Gap

Attack timelines are compressing, and many managed service provider technicians are still piecing together what is happening by moving between disconnected consoles. An alert fires in the endpoint detection platform, verifying backup status requires a separate login, patching data lives in the remote monitoring and management tool, and remediation steps have to be validated by hand across platforms. Every minute spent switching between tools is a minute an attacker uses to escalate privileges, move laterally, and deepen a foothold.

The Verizon 2026 Data Breach Investigations Report quantifies the speed problem. Drawing on more than 31,000 incidents and over 22,000 confirmed breaches across 145 countries, the report found that exploitation of vulnerabilities reached 31 percent of initial access, up from 20 percent the prior year, and overtook credential abuse as the leading entry vector. Verizon documented generative AI in use across every stage of the attack lifecycle, from reconnaissance and targeting through initial access and malware development, with the median threat actor applying AI across 15 distinct techniques and some applying it across 40 to 50. The report describes AI shrinking the defender's response window from months to hours.



The data also explains why managed service providers are exposed at a specific point. Abuse of remote monitoring and management tools rose 240 percent over the prior year, while use of traditional backdoor and command-and-control malware fell 27 percent. Attackers are increasingly operating with the same remote access tooling that IT teams and MSPs use to run client environments, which makes post-compromise activity harder to distinguish from legitimate administration. Fragmented operations compound the problem: they inflate technician workloads, slow incident response, and make it harder to scale security services without adding headcount and tools, which in turn pressures margins.Clients are now judging managed service providers not only on detection but on how quickly they respond, recover systems, and communicate during an incident. Detection without integrated response and recovery is insufficient when attacks adapt at machine speed. Deep integration and automated, AI-assisted response would reduce the exposed surface, but tool sprawl prevents that integration from

operating efficiently, and AI-assisted response still lacks the validation and verification controls that organizations should build into implementations from the start.

Clients are now judging managed service providers not only on detection but on how quickly they respond, recover systems, and communicate during an incident. Detection without integrated response and recovery is insufficient when attacks adapt at machine speed. Deep integration and automated, AI-assisted response would reduce the exposed surface, but tool sprawl prevents that integration from operating efficiently, and AI-assisted response still lacks the validation and verification controls that organizations should build into implementations from the start.

This finding maps to the DETECT and RESPOND domains of the ARISE Framework. DETECT fails when the same RMM tooling carries both legitimate administration and adversary activity and no unified telemetry exists to separate them. RESPOND fails when recovery depends on manual validation across disconnected platforms. Organizations operating or buying managed services should require an integrated operational model with verification baked into automated response, and they must validate that recovery procedures function under a compressed timeline rather than assuming detection alone meets the obligation to clients.

1. BleepingComputer. Why AI-driven threats are exposing the limits of MSP security stacks. 2026. https://www.bleepingcomputer.com/news/security/why-ai-driven-threats-are-exposing-the-limits-of-msp-security-stacks/
2. Verizon. 2026 Data Breach Investigations Report. 2026. https://www.verizon.com/business/resources/reports/dbir/
3. Cyber Insurance News. Verizon 2026 DBIR: Vulnerability Exploitation Now The #1 Breach Vector. 2026. https://cyberinsurancenews.org/verizon-2026-dbir-data-breach-report/
4. Push Security. What the Verizon DBIR tells us about breaches in 2026. 2026. https://pushsecurity.com/blog/verizon-dbir-2026-review

Autonomy Has Severed the Link Between Skill and Danger

On June 3, 2026, Anthropic published an analysis of how threat actors weaponize AI in real cyber operations, drawn from 832 accounts it banned for malicious cyber activity between March 2025 and March 2026. The company mapped the observed behavior onto version 18 of the MITRE ATT&CK framework, recording 13,873 actions that spanned 482 unique techniques and all 14 tactics. Anthropic published a subset of these results in Verizon's 2026 Data Breach Investigations Report and released the longer analysis separately. The distribution of activity shows attackers applying AI deeper in the attack lifecycle. The most common use was preparation: 560 of the 832 accounts, or 67.3 percent, used AI to write malware. A smaller group used it for more advanced work, including 54 accounts, or 6.5 percent, that used AI to assist with lateral movement inside a compromised network. Over the study period, AI-assisted phishing fell 8.6 percent while AI-assisted account discovery rose 8.9 percent, a shift from gaining access toward operating once inside. The share of actors that Anthropic's risk scoring classified as medium risk or higher rose from 33 percent in the first six months to 56 percent in the second, a roughly 1.7-fold increase.

The most consequential finding concerns the breakdown of traditional risk signals. The least-skilled actors in the dataset used about 16 distinct techniques on average, and the most skilled used about 20, a gap too narrow to separate them by volume. The platform an actor used, whether a coding tool, an API, or a chat interface, did not correlate with risk either. Anthropic's clearest example is a state-sponsored operation disrupted in November 2025 that mapped to 30 techniques across 13 tactics, a profile that reads as mid-tier by technique count, yet the operation scored at the top of Anthropic's enablement scale because the attacker used AI to execute most of the campaign autonomously, with human operators intervening at only a handful of decision points. The structural implication is that the danger no longer lives in any single technique; it lives in the autonomous scaffolding that chains techniques together. MITRE ATT&CK catalogs discrete techniques and was built on the assumption that a human makes the decision at each step. It does not yet represent AI-driven orchestration, which is why the framework does not fully capture what makes these attackers dangerous. Anthropic is in discussions with MITRE about how ATT&CK might evolve, and until that happens, organizations are measuring threats against a model that no longer matches the behavior.

This maps to the IDENTIFY and VALIDATE domains of the ARISE Framework. IDENTIFY must account for orchestration risk rather than ranking actors by technical sophistication or technique count alone, because those signals now mislead. VALIDATE applies to the safeguards themselves: Anthropic reports that the findings informed cyber safeguards on its most capable models to detect and block activities such as malware development and mass data exfiltration, and organizations should require evidence that such controls are tested against autonomous attack patterns, not assumed effective.

1. Anthropic. What we learned mapping a year's worth of AI-enabled cyber threats. June 3, 2026. https://www.anthropic.com/news/AI-enabled-cyber-threats-mitre-attack
2. Anthropic Frontier Red Team. LLM ATT&CK Navigator. 2026. https://red.anthropic.com/2026/attack-navigator/
3. Help Net Security. AI is helping low-skill hackers pull off advanced cyberattacks. June 5, 2026. https://www.helpnetsecurity.com/2026/06/05/anthropic-ai-cyber-activity-analysis/

The Pilot-to-Production Gap Is a Governance Problem

Most organizations now hold AI capability they cannot put into production. The defining condition of enterprise AI in 2026 is not whether agents work in a demonstration; it is that the majority of agent pilots stall before they reach live operation. The question that stops them is rarely about model quality. It is about what an agent is permitted to do without approval, who is accountable when it acts wrongly, and what happens when it fails.

The forecasts are consistent on direction. Gartner projects that more than 40 percent of agentic AI projects will be canceled by the end of 2027, citing escalating costs, unclear business value, and inadequate risk controls. Gartner attributes much of the current activity

early-stage proofs of concept driven by hype and often misapplied, and estimates that only about 130 of the thousands of self-described agentic AI vendors offer genuine capability, with the remainder practicing "agent washing" by rebranding existing assistants, robotic process automation, and chatbots. At the same time, Gartner projects that at least 15 percent of day-to-day work decisions will be made autonomously through agentic AI by 2028, up from zero in 2024, and that 33 percent of enterprise software applications will include agentic AI by 2028, up from less than 1 percent. Both the cancellations and the adoption are real, and they will occur together.

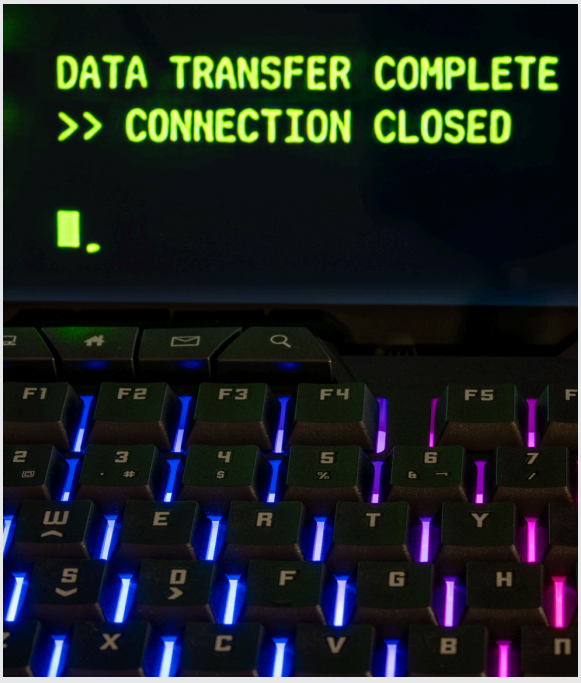
The evidence points to organizational failure rather than technical failure. Independent surveys converge on the same pattern: a large share of pilots begin, a much smaller share reach production at meaningful scale, and the primary causes are data quality, integration complexity, unclear ownership, and absent governance, not model limitations. Pilots frequently lack a designated owner because the pilot is exploratory and accountability is diffuse; production requires a specific person or team accountable for performance, for the response when the system underperforms, and for the governance decisions about how the system is used and updated.

The organizations that clear the gap share one trait: they define governance and evaluation before they deploy, not after. They scope a narrow operational workflow first, map every decision point, designate an owner with real authority, and maintain a record of every agent, its permissions, its data access, and its last evaluation. Layering an agent onto a process that was not designed for it produces a faster broken process, not value. The shakeout that Gartner describes will separate organizations that treated governance as architecture from those that treated it as a later step. This maps directly to the GOVERN and MANAGE domains of the ARISE Framework. GOVERN establishes the ownership, accountability, and decision authority that production deployment requires before a pilot is allowed to scale. MANAGE maintains the agent inventory, permission boundaries, and evaluation cadence that keep a deployed system inside the conditions under which it was approved. Organizations that build assurance into the pilot can move to production with confidence; organizations that defer it will fund proofs of concept that never ship.

1. Gartner. Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027. June 25, 2025. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027
2. Reuters (via AOL). Over 40% of agentic AI projects will be scrapped by 2027, Gartner says. 2025. https://www.aol.com/over-40-agentic-ai-projects-100510793.html
3. VaaSBlock. Enterprise AI Deployment Gap: Why Pilots Fail to Reach Production. 2026. https://www.vaasblock.com/news/enterprise-ai-deployment-gap-pilots-vs-production-2026/

Shadow AI Has Turned Routine Work Into Continuous Data Loss

The largest AI exposure documented in the 2026 Data Breach Investigations Report is not an attack. It is routine employee behavior. Verizon applies the term "Shadow AI" to the use of unauthorized generative AI services, and the report shows that the practice has moved from a fringe behavior to a structural feature of the corporate environment. The exposure does not come from a vulnerability, and no scanner will surface it, because the data leaves through the front door under a legitimate user. The figures establish the scale. Verizon found that 45 percent of employees are now regular users of AI on corporate devices, authorized or not, up from 15 percent the prior year, which is a threefold increase in twelve months. Of the users accessing AI services on corporate devices, 67 percent do so through non-corporate accounts, a figure Verizon notes is a slight decrease from the previous year even as overall use climbed. Shadow AI is now the third most common non-malicious insider action in Verizon's data loss prevention dataset, a fourfold increase in share over the prior year. Across the DLP events that targeted AI tools, source code was the most common data type submitted by a wide margin, followed by images and structured data, with research and technical documentation appearing in 3.2 percent of events. The average organization also had more than 15 percent of users running unauthorized AI browser extensions, many of which retain the context of every page visited.



The exposure is structurally similar to shadow IT, with one difference that changes the risk. A shadow IT action, such as moving a file to a personal cloud account, creates a discrete event. A shadow AI prompt sends work product to infrastructure the organization does not control, where it may be logged, retained, or processed downstream under terms the organization never reviewed. The intent is benign and the effect is equivalent to exfiltration. Source code pasted into a consumer tool for debugging has left the trust boundary, and it does not return. Organizations that responded to this with a blanket ban recreated the same shadow dynamic they spent the prior decade resolving, now attached to higher-value data.

This maps to the PROTECT and GOVERN domains of the ARISE Framework™. PROTECT requires that organizations provide sanctioned, enterprise-grade AI tools with defined data-handling boundaries and that they control the browser extensions and account types permitted on corporate devices, because employees will use AI whether or not the organization sanctions it. GOVERN requires an inventory of AI usage, a classification of the data flows those tools touch, and a policy that keeps pace with business demand; shadow AI expands when policy moves slower than the work. Organizations must treat ungoverned AI use as a data-loss channel and govern the flow rather than attempt to eliminate the behavior.

1. Verizon. 2026 Data Breach Investigations Report. 2026. https://www.verizon.com/business/resources/T1ae/reports/2026-dbir-data-breach-investigations-report.pdf
2. National Law Review. Shadow AI Continues to Expose Company IP. 2026. https://natlawreview.com/article/shadow-ai-continues-expose-company-ip
3. Axonius. Verizon DBIR 2026: Back to the Fundamentals, Beyond CVEs. 2026. https://www.axonius.com/blog/verizon-dbir-2026-fundamentals-beyond-cves
4. Kiteworks. Verizon DBIR 2026: Shadow AI Now a Top Insider Threat. 2026. https://www.kiteworks.com/cybersecurity-risk-management/shadow-ai-data-leakage-governance/

ASSESSED INTELLIGENCE

Secure & Responsible Technology.

YOU'RE BUILDING THE FUTURE. DON'T LET EMERGING RISK STALL YOUR MOMENTUM.

SALES@ASSESSEDINTELLIGENCE.COM
EMEASALES@ASSESSEDINTELLIGENCE.COM