



Beyond Patch Velocity

Rethinking Cybersecurity Strategy in a Post-Frontier AI World

James Tuplin, SVP and Head of International Cyber, Mosaic;
Jay Vinda, Global CISO and Cyber Risk Engineering Lead, Mosaic;
Sanith Wijesinghe, Chief Engineer, Public Sector, MITRE;
Chris Folk, Director, Cyber Solutions Cell, MITRE

“

The emergence of frontier AI models has triggered a debate across the cyber security and risk management communities – what happens when attackers gain access to systems capable of discovering vulnerabilities, generating exploits and reasoning through attack paths at unprecedented speeds. While this is a reasonable debate to be having, this framing has the risk of overstating what is actually changing.

This Thought Leadership Paper Series is a joint project from Mosaic Insurance and MITRE.

Introduction | The New Economics of Cyber Risk

How CISOs, Insurers, and Risk Managers Respond to the Age of Frontier Artificial Intelligence

Paper 1 | Beyond Patch Velocity

Rethinking Cybersecurity Strategy in a Post-Frontier AI World

Paper 2 | Beyond Vulnerability Counts

Reimagining Cyber Insurance Underwriting in a Post-Frontier AI World

Paper 3 | Buying Resilience

How Risk Managers Can Close the Gap Between the CISO and the Boardroom

Beyond Patch Velocity

Rethinking Cybersecurity Strategy in a Post-Frontier AI World

Chief Information Security Officers (CISOs) cannot be expected to predict and prevent every possible attack path. However, we must become better at explaining cyber risk in business terms and building the organizational support needed to evolve our operating models for a post-frontier AI world.

The release of Anthropic's Mythos Preview has intensified debate across the cybersecurity community. Much of the discussion has focused on the fear that AI-driven vulnerability discovery and exploit generation will overwhelm modern defenses. While the concern is justified, the conversation is also missing an important point: cybersecurity has never been solely about vulnerabilities in code.

Most organizations are compromised through a combination of operational weaknesses, identity sprawl, excessive trust relationships, misconfigurations, poor visibility, and inconsistent governance. Mythos and similar models do not eliminate these realities. Instead, frontier models accelerate the speed at which attackers can identify and exploit them.

The significance of frontier AI models lies less in vulnerability discovery itself and more in how they change the economics of offensive operations. The barrier to creating sophisticated attacks is shrinking, while the time required to move from discovery to exploitation is compressing rapidly. The same economics apply outside code. Vulnerability research is not the only capability that once required scarce expertise and no longer does. Convincing pretexting in a target's own language, voice and likeness cloning, and impersonation sustained across a conversation draw on the same collapse in skill and cost, and they bypass the patching pipeline entirely. This paper concentrates on the technical pathway because that is where the disclosure and remediation machinery sits. A CISO operating model built only around that pathway will be exposed on the side where a large share of compromises still begin.

Most organizations are compromised through a combination of operational weaknesses, identity sprawl, excessive trust relationships, misconfigurations, poor visibility, and inconsistent governance.

Beyond Patch Velocity

Rethinking Cybersecurity Strategy in a Post-Frontier AI World

The Democratization of Attack Capability

Cybersecurity has already experienced several waves of attacker industrialization. Historically, attackers needed to be technically adept enough to gain access to an environment, maintain persistence without being detected, craft exploit code, and have the economic mindset to monetize their intrusion.

The criminal marketplace that emerged over the past decade, including initial-access brokers, malware-as-a-service providers, and ransomware-as-a-service operators, turned sophisticated attack capability into a commodity that less-skilled actors could assemble and deploy. This industrialization significantly changed the economics of offensive operations.

Mythos and similar frontier AI models represent a qualitatively different stage in this evolution.

The difference is democratization. Industrialization made sophisticated attacks more accessible by dividing the labor. Democratization makes sophisticated attacks accessible by eliminating the skill requirement. Advanced vulnerability research historically demanded deep technical expertise, architectural understanding, and significant time investment.

Models can chain together seemingly minor weaknesses, generate proof-of-concept exploit code, test hypotheses, and iteratively refine their output when exploitation fails.

The ExploitGym benchmark, published in May 2026 by researchers from Berkeley, Anthropic, OpenAI, and Google, documents what that shift looks like in practice: Claude Mythos Preview successfully produced working exploits for 157 of

898 real-world vulnerabilities, each within a two-hour budget and with standard exploit mitigations disabled. Those mitigations are the protections built into modern operating systems and browsers that make a vulnerability hard to weaponize even after it has been found: Address Space Layout Randomization (ASLR), which randomizes where code and data sit in memory, stack canaries, and heap sandboxing. In one case, OpenAI's GPT-5.4 took a five-line crash report and produced a 229-line working exploit chain that included memory layout analysis, heap grooming, pointer forgery, and a signal-return-oriented-programming chain. It did this in 71 minutes. No specialist directed it. It reasoned, experimented, and iterated until it succeeded. With ASLR and heap sandboxing switched back on, that specific exploit no longer worked, which is an argument for defense in depth rather than a reason to discount the result.

That feedback loop is the change. The system is not just identifying flaws; it is performing security engineering that previously required expert human judgment, but at machine speed and without the years of experience on which that judgment was built.

For defenders, this creates a serious operational challenge. Most enterprise security functions were not designed for adversaries capable of iterating at machine speed. Defensive vulnerability management processes still rely heavily on:

- Manual vulnerability triage and prioritization
- Human-led testing, risk assessments, and validation
- Change advisory boards and governance approvals
- Structured deployment waves to minimize business disruption
- Rollback planning and operational contingency processes

Beyond Patch Velocity

Rethinking Cybersecurity Strategy in a Post-Frontier AI World

None of these activities are inherently wrong. In fact, they exist for valid operational reasons. The problem is that attackers no longer operate on comparable timelines. Defensive operating models begin to fail when the time required to assess, approve, and remediate exceeds the time it takes an attacker to operationalize exposure.

Why “Patch Faster” Is Not a Strategy

The natural reaction to frontier AI models is to argue that organizations simply need to patch faster. This is operationally unrealistic.

Security and IT operations teams already face significant pressure balancing vulnerability remediation against stability, testing, maintenance windows, and business continuity. Machine-speed discovery will increase the number of vulnerabilities requiring analysis and disclosure, placing additional strain on developers, vendors, and the wider Common Vulnerabilities and Exposures (CVE) ecosystem. Over time, this creates the risk of prioritization failure.

If organizations begin assigning increasingly severe Common Vulnerability Scoring System (CVSS) ratings simply to ensure remediation receives attention, or stop enriching CVEs with CVSS ratings, then defenders lose the ability to meaningfully distinguish between critical business risks and routine exposure. When everything becomes urgent, prioritization collapses.

This prioritization collapse has a structural cause outside the organization. The vulnerability disclosure pipeline, comprising the CVE Program, the CVE Numbering Authority (CNA) framework, and the National Institute of Standards and

Technology (NIST) National Vulnerability Database (NVD), is the system that receives, names, scores, and routes discovered vulnerabilities to the defenders who need to act on them. That pipeline was designed for a world in which finding vulnerabilities was hard, slow, and human-paced. But between 2020 and 2025, submission volume grew 263 percent, leading NIST to announce in April 2026 that it would no longer enrich all CVEs with CVSS scores and classification data.

The impact of that volume in a real production environment is not theoretical. In May 2026, Microsoft published details of MDASH, its multi-model agentic security system, which surfaced 16 new CVEs on a single Patch Tuesday, including four Critical remote code execution vulnerabilities, across a code base Microsoft reviews more heavily than almost any other. In a separate retrospective test, Microsoft re-ran the system against pre-patch snapshots of two Windows components and measured whether it would have rediscovered bugs the Microsoft Security Response Center had already confirmed over five years: it recovered 96 percent of 28 cases in `clfs.sys` and 100 percent of 7 in `tcpip.sys`. One of those findings, CVE-2026-33824, was exploitable with two User Datagram Protocol (UDP) packets to achieve LocalSystem privileges on any Windows IKEv2 responder. These were not obscure edge cases. They were significant vulnerabilities in widely deployed software that survived years of conventional review. AI-scale discovery does not just add volume to the pipeline; it surfaces a class of findings that the pipeline was never calibrated to receive.

On June 10, 2026, the Cybersecurity and Infrastructure Security Agency (CISA) issued Binding Operational Directive (BOD) 26-04, which revokes and replaces both BOD 22-01 and BOD 19-02 and prioritizes security updates against

Beyond Patch Velocity

Rethinking Cybersecurity Strategy in a Post-Frontier AI World

four risk variables: whether the affected asset is publicly exposed; whether the vulnerability is in the Known Exploited Vulnerabilities (KEV) catalog; whether exploitation of the vulnerability is automatable; and whether the adversary gains total or partial control of the asset. CISA's BOD 26-04 represents a significant evolution in vulnerability management, signaling a shift away from prioritizing remediation based primarily on CVSS severity scores and recognizing that technical severity alone does not determine risk. This reflects a broader industry reality: as the volume of vulnerabilities continues to grow and exploitation timelines shrink, the challenge is no longer discovering vulnerabilities, but prioritizing the ones that matter most.

The consequence for CISOs is direct: the scoring system that most vulnerability management programs depend on for prioritization is now itself a bottleneck. Patching faster is not a strategy when the pipeline that tells you what to patch is operating beyond its designed capacity.

Historical precedent also shows the risks of purely patch-centric strategies. During the Meltdown and Spectre response in 2018, many organizations accelerated patching activity only to encounter widespread instability, blue-screen failures, and operational disruption. Business leaders are unlikely to support aggressive remediation programs if they repeatedly introduce additional operational risk. Governance and change management processes exist because enterprises must balance security with availability and resilience.

The answer therefore cannot simply be to "patch everything faster." That is operationally unrealistic for most large organizations and ignores the broader structural problem.

Frontier AI Reinforces Existing Security Principles

Despite the attention surrounding AI-enabled offensive capability, frontier AI models do not invalidate decades of cybersecurity lessons. If anything, they reinforces them.

The organizations most exposed in a post-frontier AI environment are likely to be those with fragmented visibility, inconsistent governance, excessive architectural complexity, and weak operational discipline. AI does not create these weaknesses. It accelerates the speed at which they become exploitable. The security fundamentals remain critically important:

- Maintaining accurate asset and identity inventories
- Understanding system configurations and monitoring for drift
- Continuously enforcing least privilege
- Reducing unnecessary complexity and technical debt
- Designing operating models that assume compromise will eventually occur
- Monitoring continuously rather than relying on periodic validation exercises

Traditional assurance mechanisms are increasingly insufficient in environments that change continuously. Annual audits and point-in-time compliance reviews cannot keep pace with cloud-native infrastructure, ephemeral assets, legacy debt in infrastructure and processes, and AI-enabled attack workflows.

Beyond Patch Velocity

Rethinking Cybersecurity Strategy in a Post-Frontier AI World

Moving from Prevention to Cost Imposition

Most organizations still run security as a prevention exercise, and for a long time that worked.

Managing vulnerabilities and configurations on the external attack surface priced out opportunistic attackers, because finding a way in took skill and time most were unwilling to spend. It never stopped a determined adversary. It pushed everyone else toward an easier target.

Frontier AI removes that cost, and with it the deterrent. When discovery and exploit development run at machine speed at a fraction of the previous effort, being a harder target buys progressively less. The answer is not to abandon the perimeter but to stop asking it to carry the strategy alone and to impose the cost deeper in the environment. Defense in depth, zero trust, and blast radius reduction have been treated for years as the fallback for when prevention fails. They now become primary mechanisms because they act on the parts of an attack that are still expensive: moving laterally, escalating privilege, staying undetected, and reaching something worth taking.

Deception is worth more against this class of attacker than it was against the last one. A honeypot does not weaken the model that walks into it. It consumes iterations, which is what the attacker now spends in place of expertise, and it feeds the agent a false picture of the environment it is reasoning about. It also produces a detection signal with almost no noise, because nothing legitimate has any reason to touch it. Against an adversary that reasons from what it observes, corrupting the observation is a control.

The goal is not perfection. It is resilience.

The Need for AI-Augmented Defense

If attackers are operating at machine speed, defenders will need to augment operations with AI-driven capabilities as well.

This does not mean deploying a “defensive Mythos” across the enterprise and assuming the problem is solved. The industry should be cautious of simply repeating the cycle of “silver bullet” tool acquisition without operational transformation.

If attackers are operating at machine speed, the question is not how fast we respond in general, but whether we respond faster than the attack progresses. A single mean time to respond cannot answer that, because it blends the alert closed in four minutes with the one that sat for nine days and rewards teams for clearing the easy alerts faster. Instead, track response time by alert class and measure it against how long the matching attack path takes to execute. Alerts with slow response times are almost always delayed for a specific reason: a missing asset record, absent threat intelligence, or a skill gap on the current shift. Closing those gaps is where AI augmentation earns its place. The average improves as a consequence rather than as a target.

Effective AI augmentation also requires reviewing manual-intensive processes across an organization’s security operating model and identifying where AI can make repetitive or under-resourced workloads more manageable (e.g., third-party risk management).

The objective is not to replace human defenders but to increase operational speed, consistency, and scale to rebalance an asymmetry in which attackers are changing the cost, scale, and effort required for offensive operations.

Beyond Patch Velocity

Rethinking Cybersecurity Strategy in a Post-Frontier AI World

At the same time, organizations must remain realistic about the limitations of current AI systems. Large language models are not yet sufficiently reliable to autonomously remediate vulnerabilities without risking additional defects or unintended consequences. The next step in reducing bottlenecks in vulnerability disclosure may require frontier AI models that can safely and consistently generate trustworthy remediation guidance and patches.

A Revised Operating Model for Cybersecurity

The arrival of frontier AI models represents another major shift in the evolution of cyber risk. The dynamics of attack and defense change, but the underlying challenge remains the same: organizations must build operating models capable of adapting faster than attackers.

The answer is not to seek a Mythos-like tool to operate across an organization's attack surface or simply to patch faster, which we know is easy to say and difficult to do. We need to rethink our cybersecurity strategy and acknowledge the threat of frontier AI models chaining together seemingly simple vulnerabilities to create working concepts and exploits.

Embracing AI to augment defensive operations will be key to reshaping the operating model.

We do not claim to have all the answers. Our view is that the strategy should consider visibility, control of the attack surface, AI-augmented defense, and continuous assurance. The organizations that succeed in a post-frontier AI world will not necessarily be those with the most tools. They will be the organizations that can continuously

understand their environments, reduce unnecessary exposure, respond at speed, and align cybersecurity decisions with business resilience.

The operating model changes described in this paper are achievable at the technical and operational level. Most of them do not require new tools or additional budget. Instead, they require a different way of thinking about the problem and a different set of organizational conversations.

That is where operating model changes fail, not because CISOs cannot build them, but because CISOs cannot sustain them without alignment from people who speak a different language. CFOs who approve security budgets evaluate those investments against a risk picture that may not reflect what CISOs know about actual exposure. Risk managers who purchase their organizations' cyber insurance price coverage against assumptions that the new operating model may have already invalidated. Neither group has a structured way to connect those decisions to the CISOs' understanding of risk, and CISOs typically have no forum for reaching either group until a claim forces the conversation.

CISOs are often isolated, but not by choice. That isolation is the result of an organizational architecture problem. The people who understand the threat and the people who resource the response to it are not talking to each other in the structured, ongoing way the post-frontier AI environment now requires. How to close that gap, using the insurance renewal as a forcing function to create the occasion, is the subject of Buying Resilience, Paper 3 in this series. Paper 2, Beyond Vulnerability Counts, takes the same gap from the insurer side. The operating model this paper describes is the content of that conversation. Risk managers are the critical link that can bring those conversations together.

