ecosystm.

The Narrowing Window: How AI Compresses Cybersecurity Response Times

When Anthropic released Fable 5 along with Mythos 5 — models that included advanced cybersecurity capabilities to assist with vulnerability analysis and exploit research — it was not only the cybersecurity community that sat up and took notice. The models appeared capable of accelerating the kind of offensive security research that typically requires significant expertise and time, raising concerns that the same capabilities could be used by threat actors to discover vulnerabilities and shorten the window between identification and attack. The US government reportedly requested early access to assess the models and, within days of their public release, restricted their use by foreign nationals. Anthropic subsequently withdrew the models, turning what might otherwise have been a routine product launch into a public debate about how, and even whether AI capabilities of this kind should be developed and controlled.

While Anthropic has grabbed the headlines, the discussion extends well beyond a single model or vendor. Rather than being viewed as a simple productivity tool, AI is becoming a critical cybersecurity capability. Security researchers, criminal groups, and nation-state actors are already incorporating AI into their operations, applying it to both cyber defence and attack.

How Cyber Criminals are Using AI

The debate around Mythos and Fable centres on what AI may enable in the future, but many of those capabilities are already influencing the threat landscape today. AI is being adopted across the cybercrime ecosystem, by individual fraudsters, organised criminal groups, and sophisticated state-affiliated actors alike; and in most cases it is not introducing entirely new attack categories. Instead, it is reducing the cost, skill, and time required to execute existing ones, producing attacks that are more automated, harder to attribute, and more difficult to distinguish from legitimate activity. Vulnerability discovery, automated attacks, and fraud are the areas where that shift is most clearly visible.

Vulnerability Discovery

Vulnerability discovery has traditionally been constrained by specialist expertise and the time required to analyse complex codebases, trace execution paths, and understand how weaknesses interact. AI is compressing that process considerably, helping security researchers — and attackers — operate far more quickly than manual approaches allow. The consequence is a shrinking window between vulnerability disclosure and active exploitation, which places additional pressure on security teams that are already stretched in terms of both capacity and capability.

More than 50 cybersecurity leaders, including executives and security practitioners from organisations such as Adobe, Zoom, Sophos, and DigitalOcean, signed an open letter arguing that restricting access to advanced

vulnerability research capabilities would hinder defenders more than attackers. The signatories pointed out that leading AI labs outside the United States are likely only months behind Anthropic in developing comparable capabilities, which means that access restrictions applied to a single model do little to determine who ultimately acquires them. Withholding these tools from security teams, they argued, concedes ground to adversaries who face no equivalent constraints.

PROJECT GLASSWING

Anthropic's Project Glasswing represents one of the most closely watched attempts to assess how advanced AI models perform on defensive security research. The program brought together a group of industry participants that spans cloud infrastructure, enterprise software, financial services, and cybersecurity (including AWS, Microsoft, Google, Cisco, JPMorgan Chase, CrowdStrike, Palo Alto Networks, Apple, NVIDIA, and the Linux Foundation) with the purpose of evaluating whether frontier models could meaningfully contribute to vulnerability discovery before adversaries exploit the same weaknesses.

Cloudflare's experience offers the most detailed public account of what that evaluation revealed. Testing Anthropic's Mythos Preview model across more than 50 internal code repositories, Cloudflare found that its reasoning resembled the work of a senior security researcher rather than the output of a conventional automated scanner. Mythos was able to identify vulnerabilities, construct exploit chains by linking multiple weaknesses together, and generate proof-of-concept exploit code — capabilities that other frontier models approached but did not match. Where competing models could identify many of the same underlying bugs, they were less effective at demonstrating exploitability and less capable of connecting individual vulnerabilities into coherent attack paths.

The safety findings were less straightforward. Cloudflare observed that Mythos would sometimes decline a task and then complete the same task when it was framed differently, which points to a meaningful limitation in how consistently guardrails hold under adversarial or simply varied prompting. The conclusion was not that the model's safety mechanisms were absent, but that they were insufficiently reliable to function as a complete boundary on their own. Model-level controls will need to be supplemented by additional safeguards at the deployment layer before tools of this kind can be used responsibly outside a controlled research setting.

Enhanced Automated Attacks

AI is not only helping attackers discover vulnerabilities, it is also making existing forms of automated attack more effective. Credential stuffing, where usernames and passwords exposed in previous breaches are tested across multiple services, can now be executed at far greater scale. Attackers can also use bots that mimic human users to scrape data and abuse APIs by interacting directly with backend systems and exploiting business processes. In its Bad Bot Report 2026, Thales estimates that around 40% of internet traffic is made up of malicious bots.

12.5x

Year-on-year increase of AI-enabled bots attacks in 2025

Source: Thales Bad Bot Report, 2026

What is beginning to change is the sophistication of that automation. AI is helping bots analyse application workflows, solve CAPTCHAs, evade fingerprinting controls, and adapt when mitigation measures are deployed. Rather than following a fixed script, AI-enabled bots can increasingly adjust their behaviour in response to the environment they encounter. As a result, malicious activity can resemble legitimate user traffic more closely, making automated attacks harder to identify and block.

AI-Powered Fraud

The moment GenAI AI tools entered the mainstream, fraudsters began turning them to their advantage them. Retailers, banks, and telecom providers along with their customers must contend with convincing scams that incorporate voice cloning, face swapping, and synthetic identities to deceive targets. Fraudulent techniques that were considered a specialised craft are now automated using off-the-shelf tools that anyone can buy with a credit card.

One such service that treads the line between legitimacy and criminality is Haotian. The voice cloning, face swapping software is marketed towards live streamers and online sellers but is also a favourite among so-called pig butchering scammers. The tool offers adjustments to 50 settings, including eye position and cheekbone size to create highly realistic digital doubles. It can also defeat common tests for deepfakes, by accounting for a hand passing in front of the face or skin pinching. A fully-developed platform, Haotian integrates with a host of collaboration suites, such as Telegram and WhatsApp.

GenAI AI is not only being used to target individuals but also organisations that rely on digital customer onboarding processes. Identification documents that on the surface look genuine can be bought for as little as USD 15 from services, such as OnlyFake. Fabricated pay stubs and utility bills help criminals create seemingly creditworthy synthetic identities. DuckDuckGoose, a company that specialises in deepfake detection, estimates that during peak times, up to 10% of bank applications contain synthetic documents and images. Once fraudsters develop a successful method, they often open tens or hundreds of accounts.

Fraudsters have also exploited the ability of GenAI to create fake storefronts, complete with realistic product images, descriptions, and advertisements. Fake reviews on legitimate websites and SEO poisoning have been used to lend credibility to their stores. SRLabs uncovered a China-based network of 75,000+ fraudulent ecommerce sites, dubbed BogusBazar. Around 850,000 shoppers were victimised, losing more than USD 50M in both initial purchase value and subsequent card-not-present fraud.

How CISOs are Responding

AI is compressing the time available to detect and respond to attacks. Security teams already struggle to keep pace with the volume of alerts, vulnerabilities, and incidents they face, and AI-assisted threats are increasing that pressure. Hiring additional analysts can help address capacity constraints, but it does little to match the speed at which attacks can now be developed and executed.

AI is becoming central to cybersecurity strategies, particularly in threat detection, monitoring, and automated response. However, many organisations remain focused on improving visibility and accelerating reaction times rather than fundamentally reducing exposure. As AI increases the speed and scale of attacks, detection alone is unlikely to be sufficient. Cyber leaders will need to complement AI-driven monitoring with stronger identity controls, attack surface reduction, and greater automation across security operations. The focus should not only be on identifying threats faster, but building security architectures that remain resilient when attacks can be developed, adapted, and carried out at AI speed.

Focus of Asia Pacific CISOs in 2026

51%
AI-enabled threat detection & automated response

47%
Centralised security logging, monitoring & correlation platforms

45%
Data protection technologies

42%
Predictive attack prevention using AI & threat intelligence platforms

37%
Identity & access management for humans & AI agents

32%
Vulnerability scanning & penetration testing

31%
Automated endpoint & network threat management

14%
Vendor & supply chain risk management tools

Source: Ecosystm, 2026

The objective is not to detect every threat after it appears, but to limit the opportunities for attackers to succeed in the first place.

Reducing the Attack Surface

The assumption that organisations have days or weeks to act on a disclosed vulnerability is simply flawed. AI is accelerating the speed at which weaknesses can be identified and exploited, reducing the window between disclosure and active attack to a point where previously acceptable risk postures are difficult to justify. Recently, the US Cybersecurity and Infrastructure Security Agency (CISA) directed civilian federal agencies to patch, disable, or remove the most serious categories of vulnerable software within 3 calendar days. In its directive, it attributed this abbreviated timeline directly to attackers' growing ability to exploit weaknesses autonomously. For enterprise security teams, the broader implication is that reducing publicly exposed systems, limiting direct internet access to critical infrastructure, and applying least-privilege principles consistently can no longer be treated as longer-term hardening objectives. The tolerance for deferring them has narrowed in line with the exploitation window.

If we go back to the data, while AI-powered detection ranks as the leading security initiative, only 32% are prioritising vulnerability scanning and penetration testing, and 31% are focusing on automated endpoint and network threat management. In an environment where vulnerabilities can be weaponised within hours, prevention and remediation capabilities need to advance as quickly as detection capabilities.

Moving Security Controls to the Edge

As AI-enabled bots become more adept at mimicking legitimate user behaviour, it has become more difficult to distinguish malicious from genuine traffic. Organisations that have invested heavily in application-level detection are finding that filtering needs to move earlier in the traffic path, intercepting credential stuffing attempts, abusive bot activity, and malformed requests before they reach the systems they target. Cloud-based bot management services stop attacks before they reach applications, filtering traffic before it consumes internal resources or can probe application behaviour and exploit business logic.

APIs warrant particular attention in this context. They are frequently less well-defended than web-facing applications despite being equally exposed, and attackers have taken note, using them as a direct route into backend systems. Treating API protection with the same rigour applied to public-facing web infrastructure closes an exposure that automated attacks are increasingly designed to exploit.

Strenghtening Identity & Trust

With synthetic identities and deepfakes eroding the reliability of static authentication, continuous verification is becoming the more defensible model. Rather than establishing trust at a single point of entry, organisations are moving toward ongoing validation of users, devices, applications, and APIs throughout a session. Trust, in this model, is derived from identity and observed behaviour rather than network location, which makes it considerably harder for attackers operating with synthetic or compromised credentials to move undetected once inside.

Identity is emerging as one of the most important control points in an AI-driven threat landscape, yet it remains comparatively under-prioritised. Only 37% of organisations identified identity and access management as a top cybersecurity focus. As organisations deploy AI agents, automate workflows, and interact with increasingly sophisticated synthetic identities, traditional human-centric identity controls will need to evolve into broader trust frameworks covering users, applications, devices, APIs, and autonomous agents.

FIGHTING AI WITH AI

LG Uplus

ON-DEVICE LANGUAGE MODELS

Banks across the world have reported a surge in voice phishing, known as vishing, now that voice cloning has become more sophisticated. Several Korean banks are partnering with LG Uplus to take advantage of vishing detection capabilities deployed directly on customer handsets. Using voice-to-text and LLMs to recognise vishing as it occurs, banks can instantly respond by blocking suspicious transactions.

National Bank of Australia (NAB)

BEHAVIOURAL ANALYTICS

Fraud detection units commonly want to move from reactive to preventative systems, using AI to catch fraud before it happens. NAB uses behavioural analytics to monitor for signs of hesitation or coercion during online banking sessions, such as aimless mouse movements and segmented typing. The bank recently estimated that behavioural analytics helped it stop and recover almost USD 1.4 million of its customers' money each week.

Handelsbanken and Swedbank

FEDERATED LEARNING

When attackers can operate across channels, banks, and jurisdictions, each institution lacks full visibility of the threat. Privacy-preserving techniques are emerging that allow model training to be shared without exchanging sensitive customer data. Handelsbanken and Swedbank have collaborated with AI Sweden to test federated learning approaches to combat money laundering. The joint experiment found that AI models trained across multiple institutions were significantly more effective at identifying money laundering patterns than models developed by individual banks alone.

ECOSYSTM OPINION

Darian Bird

Principal Advisor,
Ecosystm

The debate sparked by Mythos and Fable will not be resolved by any single policy decision or product withdrawal. The capabilities that made those models controversial are already emerging across the industry, and the trajectory — faster vulnerability discovery, more adaptive automation, more convincing fraud — points toward a threat landscape that continues to compress the time available to detect, respond, and recover. Attackers operating with AI face none of the governance, compliance, or procurement constraints that slow its adoption on the defensive side, which makes the gap a structural one rather than a temporary lag. Closing it requires more than deploying AI within existing security operations; it requires organisations to reconsider how they approach risk tolerance, infrastructure exposure, and trust — treating the narrowing window not as an operational problem to be managed but as a strategic condition to be designed around.

ecosystm.

About Ecosystm

Ecosystm is a leading technology market analyst and advisory firm that helps stakeholders navigate innovation in the digital economy through data, insights, and expertise. We connect enterprises, technology companies, digital-native founders, investors, and policymakers to enable informed decision-making. With ongoing research and access to top analysts and strategic advisors, we empower business planning, go-to-market activities, thought leadership, and innovation strategy consulting. Visit ecosystm.io