Privacy Advisory Board

MEMORANDUM

April 30, 2026

To: The Hon. Council President LaCava and Members of the San Diego City Council

From: City of San Diego Privacy Advisory Board

RE: PAB Review and Recommendation of the San Diego Police Department's **2025 Annual Surveillance Report for 53 Existing and Previously Approved Surveillance Technologies**

I. Recommendation and Introduction

This recommendation concerns the 2025 Annual Surveillance Reports for the 53 previously approved Surveillance Technologies operated by the San Diego Police Department ("SDPD"). The recommendation concerning the Automatic License Plate Recognition technology and its 2025 Annual Surveillance Report is the subject of a separate Recommendation. A list of the 53 Surveillance Technologies that are the subject of this recommendation is attached as Exhibit A.

The Privacy Advisory Board ("PAB") recommends that the City Council **approve** the continued use of the 53 previously approved Surveillance Technologies, provided the SDPD continues to improve the content of the Annual Surveillance Reports and its auditing and related practices for the subject Surveillance Technologies to ensure full compliance with the applicable Use Policies, as discussed more fully below.

Full and meaningful implementation of the TRUST Ordinance continues to be a work in progress. As stakeholders learn more about the process, the Surveillance Technologies, and best practices to monitor, audit, and regulate their use, the process has and will continue to be refined to fulfill the purpose of the TRUST Ordinance. To that end, the PAB and the SDPD, with input from the community, have engaged in productive ongoing discussions about improvements to the Annual Surveillance Reports and Use Policies (if needed). These include improvements to the audit processes for the operation of the Surveillance Technologies and cybersecurity best practices.

The PAB's Recommendations for approval come with the explicit understanding that this dialogue will continue and the resulting improvements will be made. Based on the current discussions that have occurred and the SDPD's demonstrated commitment to this process, the PAB is confident that they will.

Below are recommendations the PAB considers critical to ensure the ongoing safe use of Surveillance Technologies in the City. While some of these recommendations have been implemented, we note that not all have been, but the stakeholders continue to work on their full and complete implementation. We note that many of the principles detailed below apply to all Surveillance Technologies in use or contemplated to be used in the City.

II. Audit Processes and Procedures

The SDPD currently has internal compliance and audit processes. However, they need to be strengthened as follows:

- The final compliance or audit review should be conducted by someone who is independent of the Surveillance Technology and chain of command for the Surveillance Technology.
- Clear documentation should be used of the "risk" for which the department is testing in its compliance and audit processes. Normally this includes identification of the risk, procedures specifically established to address the risk, and controls over those procedures.
- All audit or compliance findings and exceptions should be researched and a root cause analysis performed to determine systemic or other causes of the compliance issue. This must then be followed up to ensure the matter was adequately addressed.

III. Cybersecurity Related Processes and Procedures

For a variety of reasons, cybersecurity risks are significant and growing. The department or agency responsible for a Surveillance Technology should ensure proper steps are taken regarding the Surveillance Technology for which it is responsible to guard against these threats. A department or agency may rely on others (such as the City's Information Technology department) but should confirm all reasonable steps are taken. To ensure the SDPD identifies the IT risks, it must analyze "What can go wrong" and then put in place processes and controls to address each of these.

Example processes and controls include:

- Review and validate how the City's IT department conducts due diligence and assessment of surveillance technologies. Members of the PAB can assist in that regard.
- Ensure that implemented surveillance technologies have centralized and verbose logging established, and that the logs are reviewed for anomalous activity by qualified staff as part of continuous security monitoring.
- Ensure that implemented surveillance technologies have strong authentication enabled (e.g., multi-factor authentication, ideally phishing-resistant MFA such as FIDO2).
- Validate session-expiry settings for implemented surveillance technologies such that when an authorized user completes their task with the technology, they are fully logged out of the system and all current sessions are broken.
- Conduct quarterly or minimally bi-annual access reviews of who is authorized to use a given Surveillance Technology. Attached as Exhibit B is a helpful form used to conduct this type of review. Members of the PAB are available to review this with the SDPD or City IT staff, including security staff.
- Conduct vendor reviews of providers. Attached as Exhibit C is a helpful form the PAB previously provided to support the review. Members of the PAB are available to meet with City staff to assist with this.
- Consider tabletop exercises to address scenarios where surveillance technology is compromised and there's a data breach.
- Determine and ensure that implemented surveillance technology is maintained current and appropriately configured. System patches should be deployed in a timely manner.[1]

IV. Recommendation Tracking

The SDPD should develop a process to track the PAB comments and City Council requirements to ensure that each are addressed and the resolutions are properly disclosed to the PAB, the City Council, and the public.

[1] For an analysis of coming threats, see Exhibit D, "The 'AI Vulnerability Storm': Building a 'Mythos-ready' Security Program." Experts in the field caution that the nation is a few weeks to months from a flood of system-related exploit occurrences.

Exhibit A

List of Surveillance Technologies

1. 836 Technologies CINT Commander II
2. 836 Technologies Tactical Throw Phone
3. Acecore Zoe Unmanned Aircraft System
4. Arteco
5. Avalex DVR and FLIR-HDc
6. Berla iVE
7. Body Worn Camera System (Formerly AXON Body Worn Camera)
8. Brinc Lemur-S Unmanned Aircraft System
9. Camera Trailer Camera System
10. Cellebrite Inseyets (Previously Cellebrite Universal Extraction Device)
11. CellHawk
12. Code5Group GPS Bike
13. Covert Audio Recording Device (Record and Transmit Audio)
14. Covert Audio Recording Device (Record Audio Only)
15. Covert Audiovisual Recording Device
16. Covert Cloud-Based Mobile Application
17. CPCLear
18. Dejero Downlink Transmission System
19. DJI Avata Unmanned Aircraft System
20. DJI Matrice 210 Unmanned Aircraft System
21. DJI Matrice 30T Unmanned Aircraft System
22. DJI Mavic 2 Enterprise Advanced Unmanned Aircraft System
23. DJI Mavic 2 Enterprise Dual Unmanned Aircraft System
24. DJI Mavic 3 Enterprise Unmanned Aircraft System
25. DJI Mavic Air Unmanned Aircraft System
26. DJI Phantom 4 V1 Unmanned Aircraft System
27. DJI Phantom 4 V2 Unmanned Aircraft System
28. FirstLook (Gen 1) Robot
29. FirstLook (Gen 2) Robot
30. FotoKyte Sigma Unmanned Aircraft System
31. Hoverfly Spectre HL Unmanned Aircraft System
32. ICOR Mini-Caliber Robot
33. Magnet AXIOM (Formerly Magnet Forensics AXIOM)
34. Magnet Graykey (Formerly Grayshift GrayKey)
35. Motion Activated Trail Cameras
36. NightHawk
37. Object Tracker

38. Pan Tilt Zoom Video Camera Mobile Unit
39. Pan Tilt Zoom Video Camera Mobile Unit with Cloud Based Storage
40. Pan Tilt Zoom Video Cameras
41. Parrot Anafi Thermal Unmanned Aircraft System
42. Power Over Ethernet Digital Video Recorders
43. Power Over Ethernet Network Video Recorders
44. Power Over Ethernet Video Cameras
45. Realquest Online Services
46. Shield AI Nova Unmanned Aircraft System
47. Skywatch
48. Smart Streetlights
49. SWIFT Under Door Camera
50. Teledyne FLIR Black Hornet PRS Unmanned Aircraft System
51. TLOxp
52. Vehicle Trackers
53. Vigilant

Exhibit B

Application Access Review Form

1 PURPOSE

The purpose of performing application access and credential reviews includes, but is not limited to, the following:

1.1 Validating the status of application access controls to the Company's material applications.

1.2 Ensuring that application access rights and entitlements are appropriate and aligned to the Authorized User's role within the Company.

1.3 Validating that access controls and entitlements align to the Company's policies (e.g., Security Policy, Separation of Duties Policy, etc.).

2 SCOPE

2.1 This Application Access Review Form is used with applications that are deemed to be material to the Company's operations and/or regulatory requirements or contractual obligations.

3 GENERAL INFORMATION

Name of application reviewed:	Provide the name, version, and other pertinent details related to the application (e.g., if it is on-premise or SaaS).
Effective date of the review:	
Owner of the application under review:	
Review form completed by:	
Next scheduled review:	

4 APPLICATION ACCESS & AUTHENTICATION CONTROLS REVIEW

Place an X in the appropriate column.	Complies	
	Y	N
1. Authorized Users have signed a non-disclosure agreement.		
2. Access to the application is only granted based on a legitimate business need.		
3. Authorized Users are accountable for the activity associated with their account and are precluded from sharing their username and password, unless authorized, by Company policy.		
4. Usernames are unique, and passwords comply with the Company's password policy (e.g., complexity, length, rotation, etc.).		
5. User sessions time out after a predetermined time period (e.g., 15 minutes of inactivity).		
6. Application access requires strong authentication (e.g., 2-step verification or multi-factor authentication (MFA)).		
7. Procedures for establishing, changing, and deleting (disabling) authorized users from the application are adequately documented, logged, and completed within timescales consistent with Company policy.		
8. Access logs are collected, reviewed, and retained consistent with the Company's security policy.		
9. Application entitlements and permissions are consistent with the Authorized User's role within the Company.		
10. Terminated employees and contractors have their access rights disabled, and current sessions are terminated within 15 minutes.		

5 ASSESSMENT OF CONTROL EVIDENCE

This section summarizes the evidence available to assess the appropriateness of the application's access permissions.

For this assessment, you will need the following documents:

- Printout of the application access rights for each user and their role (as appropriate to the specific application under review).
- Job descriptions for each user.

Next, complete the Authorized User Assessment Questionnaire by entering all user IDs or usernames in the left-hand column and answering the questions by placing a “Y” or an “N” in each column with your findings.

Table 1 - Authorized User Assessment Questionnaire

User ID or Username	1. Does this Authorized User have elevated privileges to the application?	2. Does the Authorized User’s job description match their application access rights?	3. Has the Authorized User’s role changed in the last quarter?	4. Have the Authorized User’s access rights to the system changed in the last quarter?	5. Do the changes in Questions 3 & 4 match each other?	6. Does the granted access match the least privileged access methodology employed by the Company?

6 CONCERNS OR REMEDIATION REQUIRED.

Based on the review conducted in Section 5, list or describe any identified concerns with specific Authorized Users and their entitlements to the application.

List or describe any remediation steps that need to be taken for application access to comply with the Company's policies.

7 REVIEW APPROVAL

Management Review and Approval By:			
Name	**Signature**	**Title**	**Date**

Exhibit C

Statement of Standards for Attestation Engagements (SSAE) 18
&
System and Organization Controls (SOC) 2
Audit Review Form

1 PURPOSE

The purpose of this form is to structure the review of a material vendor's SSAE 18 and/or SOC 2 audit report. Specifically, the purpose of this form is to:

1.0 Facilitate and evidence the effective review of an SSAE 18 and/or SOC 2 report of [insert entity] material service provides (e.g., vendors that are material to the internal controls over financial reporting and/or other underpinning services used at [insert entity]), and;

1.1 Ensure that the Company's assessment of internal control of the third-party service organization (e.g., material vendor) is complete, thorough, and adequately addresses the Company's control objectives as they relate to the services offered.

2 GENERAL INFORMATION

Name of Service Organization:	
Name of Auditor Organization:	
"As of" date for the description of the service organization controls:	
Period covered by service auditor's tests of control operating effectiveness:	
Types of transactions processed, or services offered by the service organization that affect the company's financial statements or internal services:	

2.1 Identify the significant financial statement accounts, business processes, and/or IT-related systems and the disclosures and relevant assertions affected by transactions processed by or services delivered from the service organization (aka material vendor).

3 PROCEDURES

Read and assess the implications of the SSAE 18 and/or SOC 2 Report	Complies	
	Y	N
Read the service auditor's report and assess its implications for the Company's assessment of internal control effectiveness, and indicate whether compliance was found for the following:		
○ Whether the service auditor prepared a Type 2 report.		
○ The nature of the opinions rendered and whether these included any modifications to the standard reporting language.		
The timing of the audit engagement as to:		
○ The date as of which the description of controls applies.		
○ The period of time covered by the tests of operating effectiveness of controls.		
Read the description of the service organization's controls and evaluate the effect of the following on the Company's assessment of internal control effectiveness:		
○ Whether the description includes all significant transactions, processes, computer applications, or business units that are within the scope of the Company's assessment of internal control effectiveness.		
○ Whether the description includes all five components of internal control:		
• Control Environment		
• Risk Assessment		
• Control Procedures		
• Monitoring		
• Information and Communication		

○ Whether the description is sufficiently detailed to understand how the service organization's processing affects the Company's internal control over financial reporting.		
○ Whether there were material changes to service organization controls.		
○ Whether there are instances of noncompliance with service organization control.		
○ Whether the description of controls is adequate to provide an understanding of those elements of the company's accounting information system that is maintained or impacted by the service organization.		
List all complementary user organization controls identified in the audit report that the service auditor assumed were maintained by the Company ("user controls"). Cross-reference this list to the work performed to:		
○ Assess the design effectiveness of these user controls.		
○ Test the operating effectiveness of these user controls.		
Tests of Operating Effectiveness		
Review the service auditor's description of the tests of controls and assess their adequacy for your purposes. Were you able to observe:		
○ The link between the financial statement assertion and the control objective.		
○ The link between the control objective and the controls tested.		
○ The nature, timing, and extent of the tests performed.		
Evaluate the results of the tests of controls.		
○ Identify control testing exceptions and determine whether they indicate a control deficiency.		
○ Summarize all control deficiencies and assess their significance, both individually and in combination.		

4 COMPLEMENTARY USER ENTITY CONTROLS (CUECs)

If the description lists customer controls which must be in place in order for the customer to obtain a reasonable assurance that the service organizations controls are valid, summarize those customer controls in the space below.

5 REVIEW OF SERVICE AUDITOR'S REPORT

The following summarizes the opinions provided in the service auditor's report.

Required Opinion	Service Auditor Opinion	
	Standard	**Modified**
Whether the service organization's description of its controls presents fairly, in all material respects, the relevant aspects of the service organization's controls that had been placed in operation as of a specific date.		
Whether the controls were suitably designed to achieve specified control objectives.		
Whether the controls that were tested were operating with sufficient effectiveness to provide reasonable, but not absolute, assurance that the control objectives were achieved during the period specified.		

Describe any modifications to the service auditor's standard opinion and the effect these modifications have on the Company's assessment of internal control effectiveness.

6 REVIEW OF SERVICE AUDITOR'S REPORT

	Internal Control Component							
	Control Environment		Risk Assessment		Information and Communication		Monitoring	
	Yes	No	Yes	No	Yes	No	Yes	No
All transactions, processes, computer applications, or business units that affect the Company's assessment of internal control effectiveness are described in the audit report.								
The level of detail provided is sufficient to allow the Company to understand how the service organization's processing affects the Company's internal control.								
The audit report identified no changes to controls since the later of the date of the last service auditor's report or within the last 12 months.								
The audit report identified no instances of noncompliance with the service organization's controls identified in the service organization's description of controls.								

7 INFORMATION COMPONENTS OF INTERNAL CONTROL

	Yes	No	N/A
The service auditor's report is adequate to allow the Company to obtain sufficient knowledge of the information relevant to the Company's financial reporting to understand those elements of the information system maintained by the service organization related to: • The classes of transactions in the Company's operations that are significant to the financial statements. • The procedures, both automated and manual, by which transactions are initiated, recorded, processed, and reported from their occurrence to their inclusion in the financial statements. • The related accounting records, whether electronic or manual; supporting information; and specific accounts in the financial statements involved in initiating, recording, processing, and reporting transactions. • How the information system captures other events and conditions that are significant to the financial statements.			

8 Tests of Controls

Identify the relevant financial statement assertions affected by the service organization services. Summarize these assertions across the horizontal axis. For each assertion, answer the questions listed in the first column, and document your answers by marking the appropriate box.

A. Control Environment B. Risk Assessment C. Control Procedures D. Monitoring E. Information and Communication	**Assertions for Which Control Risk to be Assessed Below Maximum**									
	A		**B**		**C**		**D**		**E**	
	Y	**N**	**Y**	**N**	**Y**	**N**	**Y**	**N**	**Y**	**N**
Is the assertion linked to a service organization control objective?										
Is the control objective linked to related control objectives?										
Is the description of the nature, timing, and the extent of the tests applied in sufficient detail to enable you to determine the effect of the tests on the assessment of internal control effectiveness?										
Do the results of the service auditor's tests support an assessment that the internal control is effective?										

9 REVIEW & APPROVAL

Management Review and Approval By:			
Name	**Signature**	**Title**	**Date**

5 REVIEW & ACKNOWLEDGEMENT

This form summarizes our procedures and the conclusions reached on the effectiveness of internal controls maintained at the service organization, as documented in the service auditor's audit report.

Employee Name:

_____

Employee Title:

_____

Employee Signature:

_____

Date:

_____

Exhibit D

DRAFT

The “AI Vulnerability Storm”: Building a “Mythos-ready” Security Program

Expedited Strategy Briefing
By the CSA CISO Community, SANS, [un]prompted,
the OWASP Gen AI Security Project, and the wider community.

Contact cisos@cloudsecurityalliance.org with any inquiries.

12 April, 2026

cloud security alliance®

SANS

[un]prompted

OWASP GenAI SECURITY PROJECT

Authors

Gadi Evron

CEO, **Knostic**, CISO-in-Residence for AI, **Cloud Security Alliance**

Robert T. Lee

Chief AI Officer, Chief of Research, **SANS** Institute

Rich Mogull

Chief Analyst, **Cloud Security Alliance**

Contributing Authors

Jen Easterly

CEO, **RSAC** and Former Director, CISA

Chris Inglis

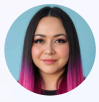
Former National Cyber Director, **The White House**

Heather Adkins

CISO, **Google**

Sounil Yu

CTO, **Knostic**, former Chief Security Scientist, Bank of America



Katie Moussouris

Founder and CEO, **Luta Security**

Maxim Kovalsky

Managing Director, AI Security CoE, **Consortium Networks**

Joshua Saxe

CTO and Co-founder at **Security Superintelligence Labs**, former AI and Llama Security Lead, Meta

Bruce Schneier

Chief of Security Architecture, **Inrupt**, Fellow and lecturer, Harvard Kennedy School

Phil Venables

Ballistic Ventures, former CISO, Google Cloud

Rob Joyce

Former Cybersecurity Director, **NSA**

Jim Reavis

CEO, **Cloud Security Alliance**

John N. Stewart

Talons Ventures, former CSTO, Cisco Systems

Dave Lewis

Global Advisory CISO, **1Password**

John Yeoh

CSO, **Cloud Security Alliance**

Ramy Houssaini

CCSO, **Cloudflare**

Reviewers

Many CISOs, and some other practitioners, assisted in reviewing and editing this document. These are the ones who would share their names publicly, in alphabetical order (by last name):

Mark Aklian, Founder & CISO, Silver Oak Cyber
David Aronchick, Co-Founder, Expanso / Kubeflow
Jake Bernardes, CISO, Anecdotes
Alan Berry, CISO, Centene Corporation
Jeff Bryner, CISO, Independent
Michael Calderin, CISO
Daniele Catteddu, Chief Technology Officer, Cloud Security Alliance
Viswanath Chirravuri, Global Product Security Director, Thales
Mea Clift, CISO, Cengage
David B. Cross, CISO, Atlassian
Chris Cochran, Field CISr,O & VP AI Security, SANS Institute
Michael Colao, former Corporate CISO, AXA, Director, Island Cyber
Daniel Cuthbert, Associate Fellow, Cyber and Tech, RUSI
Julie Davila, VP Product Security, GitLab
Michael Douglas, Managing Partner / SANS Instructor, InfoSec Innovations / SANS Institute
Yoni Efrati, former CISO, Bank HaPoalim
Eliya Elon, EIR, Notable Capital
Sergej Epp, CISO, Sysdig
Alex Foley, Data Security TISO, Wells Fargo
George Gerchow, CSO, Bedrock Data
Dan Glass, CISO, Delek US Holdings
Barry Greene, Co-Founder, Qubit Cyber
the grugq, Independent Security Researcher, Independent
Erik Hart, Global CISO, Cushman & Wakefield
Gary Hayslip, CISO in Residence, Halcyon, former CISO, SoftBank
Dustin Heywood, Senior Technical Staff Member, IBM
Heather Hinton, CISO, Sitecore, Lecturer, Harvard Extension School
Matt Holland, Director of Cyber and AI Security, Kainos
Igor Ignatov, Head of Security & Compliance, Cognichip Inc.
Waylon Janowiak, Head of Information Security & IT, Faire
Mike Johnson, CISO, Rivian
Avner Langut
Rock Lambros, Director of AI Standards and Governance, Founder, Zenity / RockCyber
Ariel Litvin, former CISO, First Quality Enterprises
Bob Lord, former CISO, Yahoo, CSO, DNC
Myke Lyons, CISO, Cribl
Ciaran Martin, Head of Cyber Leaders Network, SANS Institute, Founder and former CEO, UK NCSC
Michael Machado, CISO & CDO, Hyland
Donald McFarlane, Principal Technical Advisor, Microsoft
Gal Malach, CTO and Co-Founder, Terra Security
Tomas Maldonado, CISO, NFL
Greg McCord, Founder and CISO, McCord Keystone Advisory
Ross McKerchar, CISO, Sophos
Greg Notch, CSO, Expel
Charles Nwatu, Head of Security (former), Netflix
Mark Orsi, CEO, Global Resilience Federation
Teju Oyewole, Director IT Security and CCSO, Sunwing
David Quisenberry, CTO & CISO, Ferguson Wellman Capital Management
Gavin Reid, CISO, Human Security
Gerardo Richarte, CISO, Satellogic
Joshua Scott, VP of Security and CISO, Hydrolix
Mark Seiden, Volunteer and Security Advisor, Internet Archive
James Shank, Director of Threat Operations, Expel
Samir Sherif, Global Field CISO, Fastly
Conor Sherman, CISO in Residence, Sysdig
Ed Skoudis, President, SANS Technology Institute
John Sotiropoulos, OWASP GenAI Agentic Security Initiative Co-Lead / Co-Founder, Deep Cyber
Sean Todd, CISO, trycoral.ai
Anna Sarnek, Senior Fellow, McCrary Institute for Cyber & Critical Infrastructure Security
Holger Spohn, CISO, Candriam
Phil Venables, Venture Partner, Ballistic Ventures, former CISO, Google Cloud
Rob van der Veer, OWASP AI Exchange
Mikael Vinding, CISO, AP Technology
Yabing Wang, CISO & CIO, Justworks
Mike Wilkes, Adjunct Professor, New York University
Jeff Williams, CISO, Sigma360
Steve Wilson, Chief AI Officer / Co-Chair OWASP GenAI Security Project, Exabeam
Jason Woloz, CISO, TransUnion

All the listed authors and reviewers represent only themselves, and not their employer(s).

Join the Cloud Security Alliance CISO community, email us at cisos@cloudsecurityalliance.org.

Table of Contents

Executive Summary

What happened:

- AI, as demonstrated by Anthropic's Mythos, has significantly increased the likelihood of attackers discovering new vulnerabilities, creating new exploits, and using them in complex automated attacks at scale.
- While AI also increases the speed to develop patches, and reduces defects in new software, the burden on defenders, by comparison, increases due to the inherent limitations of patching. The attackers gain asymmetric benefits.

How is this different from the status quo?

- In the near term, security organizations will likely be overwhelmed by the need to apply patches and respond to AI-discovered vulnerabilities, exploits, and autonomous attacks.

What to do now to deal with the current risk spike?

- Adjust risk calculations and re-orient security program resources for increasing volume of patches, decreasing time to patch, and more-persistent complex attacks.
- Focus on the basics and harden your environment further. Segmentation, egress filtering, multifactor authentication, and defense-in-depth/breadth all increase the difficulty for attackers.

What do we believe will happen next?

- The storm of vulnerability disclosures from Project Glasswing is the first of many large waves of AI-discovered vulnerabilities that may occur in rapid sequence.
- The capabilities seen in Mythos will quickly become more widely available, dramatically increasing the number and frequency of complex, novel attacks organizations will face.

What else should start now to be ready for the next waves?

- Prioritize robust dependency management to reduce vulnerabilities in third-party and open-source components.
- Consistently enforce automated security assessments in your development processes, including using LLM-powered agents to find vulnerabilities before the attackers.
- Introduce AI agents to the cyber workforce across the board enabling defenders to match attackers' speed and begin closing the gap.
- Re-evaluate your risk tolerance to operational downtime caused by vulnerability remediation to account for shorter adversary timelines.

- Update governance for more efficient vendor onboarding and increase headcount to facilitate a faster cycle deployment of new AI-based defenses.

- As an industry we need to strengthen our coalitions, cooperation, and coordination.

Key Takeaways for the CISO

Use LLM-based vulnerability discovery and remediation capabilities.

Unlike defensive AI technologies, LLM-based vulnerability discovery capabilities are already mature and can be used to our advantage. Start immediately by asking an agent for a security review of any code, and build toward a VulnOps capability.

Update risk metrics.

With the shifting landscape, many of your metrics and risk assessments may be outdated, and could potentially even affect business reporting. Consider how to update these, and communicate the challenge with stakeholders.

Accelerate your team by the use of coding agents.

While defensive AI technologies are lagging behind offensive ones, agents can already accelerate human action across the board, from incident response to GRC. Encourage and demand for your team to make use of these agents to accelerate their capabilities. Triage and test patches, red team your environment, automate audit data collection, and accelerate security operations overall.

Prepare to respond to more incidents.

Run tabletop exercises for multiple, simultaneous, high-severity incidents occurring within the same week; have playbooks in place for high level, critical incidents. Examine how to automate remediation capabilities to the degree possible. Verify and enable mitigating controls such as segmentation, egress filtering, Zero Trust architectures, phishing-resistant MFA, and secrets rotation, to limit impact when post-exploitation. The supply chain will be affected.

Increase focus on the basics.

The basics remain valid and can be prioritized for risks that can't be otherwise mitigated. Segmentation, patching known vulns, Identity and Access Management, and defense-in-depth/breadth all increase the difficulty for attackers. To lower latent risk, expanding these efforts while there is time, is prudent.

Prepare for burnout.

The cadence and volume of vulnerability disclosures will exceed anything we have experienced before. Request additional headcount and budget for reserve capacity to avoid burning out existing staff, in parallel with putting more automation in place.

Evolve to a Mythos-ready Security Program.

Mythos is one of what will likely be many changes to cybersecurity risk. If not already underway, incorporating Mythos and its implications into your strategy should be seriously considered.

Expand Outreach and Partnerships.

Attackers already operate as syndicates, crowdsourcing, sharing tools, and moving as a collective. Defenders must do the same and leverage our coordinating groups. Teams beat stovepipes, coalitions beat teams, and coalitions equipped with the right technology win.

Build Collective Defense Now.

Attackers already operate as syndicates, crowdsourcing, sharing tools, and moving as a collective. Engage now with sector coordinating groups, ISACs, CERTs, and standards bodies to share threat intelligence, coordinate response, and produce sector-specific guidance for this moment. Defenders must do the same and leverage our coordinating groups. Teams beat stovepipes, coalitions beat teams, and coalitions equipped with the right technology win.

Introduction

Many of our assumptions about the capabilities of AI in vulnerability research, exploitation, and autonomous attacks, may be outdated. Throughout 2025 and into 2026 we've seen continuous examples of increasing capabilities, in research and in actual in-the-wild attacks. AI-driven vulnerability discovery and exploitation has been accelerating for over a year. See Appendix A for more details and historical evidence.

Anthropic's Claude Mythos (Preview) represents a step change in that trajectory, autonomously finding thousands of critical vulnerabilities across every major operating system and browser, generating working exploits without human guidance, and empowering autonomous attack orchestration, all at a speed and scale that outpaces any prior capability.

The asymmetry this creates is structural. AI lowers the cost and skill floor for discovering and exploiting vulnerabilities faster than organizations can patch them. The window between discovery and weaponization has collapsed into hours. Attackers gain disproportionate benefit, and current patch cycles, response processes, and risk metrics were not built for this environment.

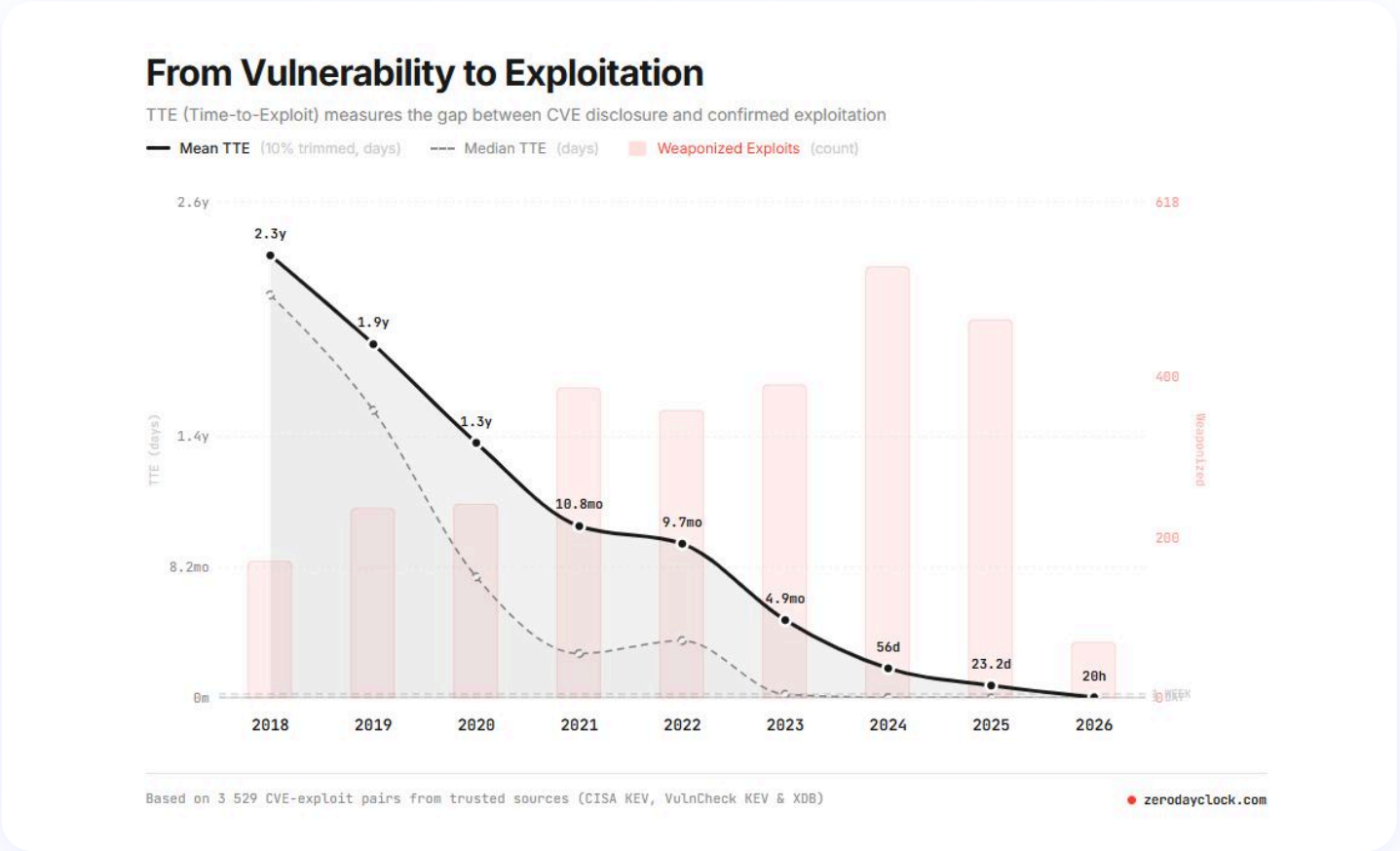


Diagram from the Zero Day Clock, demonstrating the fall of time to exploitation to less than a day in 2026.

While many of these capabilities pre-date this new model, Mythos-class capabilities do represent a step-change, and will proliferate. The organizations that respond well will be those that build the muscle now: the processes, the tooling, and a culture willing to adopt AI as a core part of how security gets done. That adaptability will help determine who meets the next wave on their own terms.

This moment requires reprioritizing resources, reviewing risk levels and controls, and leveraging AI where feasible. At the time of this writing, most AI defensive controls and approaches are not yet mature. That said, AI attacker technology may be used for defense purposes and coding agents will help.

The detailed recommendations are included later in this document.

Mythos & Glasswing: Why They Matter

Mythos

Mythos is distinguished from previous capabilities on both technological and strategic levels, even if many of its attributes **already existed** and have evolved over the past year. Technologically, models like Mythos exhibit three capabilities that make them different:

1. **Exploits without scaffolding.** Internal testing at Anthropic showed Mythos generated 181 working exploits on Firefox where Claude Opus 4.6 succeeded only twice under the same conditions, marking a substantial jump in autonomy and reliability.
2. **Complex, chained vulnerabilities.** Mythos identifies vulnerabilities composed of multiple primitives chained together, such as scenarios requiring multiple memory corruption bugs combined into a single exploit path.
3. **“One-shot” (single-prompt) capability.** Mythos accomplishes significantly more with a single prompt, without elaborate scaffolding or agent configuration

Strategically, Mythos broke into mainstream media beyond technical security communities and reached into boardrooms, raising awareness and the urgency of AI-driven vulnerability risks. This has forced security teams to respond and opened the door for new resources and funding across the industry.

Glasswing

The scale and speed of Mythos prompted Anthropic to create **Project Glasswing**, possibly the largest multi-party vulnerability coordination effort in history. Anthropic provided selected critical infrastructure providers, industry partners, and open source maintainers early access to Mythos so they could patch their own products. Other AI model vendors have launched similar vetted-participant programs.

The most significant limitation of Project Glasswing is that it can only cover so much.

The world's exploitable attack surface is vastly larger than what any curated partner ecosystem can cover, and most organizations that build or maintain critical software will not have early access to Mythos-class capabilities. Meanwhile, the competitive landscape is narrowing that window. If comparable offensive capabilities emerge in other frontier models within months, and in open-weight models within six months to a year, the defensive advantage conferred by early access becomes time-limited by definition.

While the coordination model Glasswing established is critically important, its impact will depend heavily on how quickly it can expand coverage, and whether the patch and disclosure pipeline can keep pace with both AI progress and adversarial adoption.

The Evolution of LLM-based Offensive Capabilities, 2025/6

Jun 24, 2025

XBOW tops the HackerOne leaderboard

- **XBOW** became **#1 on HackerOne's US leaderboard**, first autonomous system to outperform all human hackers on the platform
- Open-source **raptor** demonstrated that autonomous **vulnerability research** is available to anyone using an off-the-shelf agent

Aug 5, 2025

Google Big Sleep finds 20 real-world zero-days

- Google's **Big Sleep discovered real zero-days in open source**, 20 vulnerabilities in projects including FFmpeg and ImageMagick, each found and reproduced autonomously

Aug 8, 2025

DARPA AIxCC finals at DEF CON 33

- **DARPA AIxCC found 54 vulnerabilities in four hours of compute** across 54 million lines of code

Sep 2025

Singularity warning issued

- Heather Adkins (CISO, Google) & Gadi Evron (CEO, Knostic) **publish warning** that attackers are racing toward a singularity moment
- Autonomous vulnerability discovery and exploitation estimated ~6 months away

Nov 14, 2025

First AI-orchestrated espionage campaign disclosed

- Anthropic **disclosed a Chinese state-sponsored group** had used Claude Code to autonomously run full attack chains — recon through exfiltration — across ~30 global targets (detected mid-Sep 2025)

Feb 5, 2026

AI finds hundreds of high-severity bugs; autonomous attacks discovered

- Anthropic (using Claude Opus 4.6) **reported 500+ high-severity vulnerabilities** in open source software
- **AISLE** found **12 OpenSSL zero-days**, including a CVSS 9.8 flaw dating to 1998
- Sysdig **documented an AI-based attack** reaching admin-level access in 8 minutes
- Gambit released a report on the **AI-led compromise** of Mexican government infrastructure

Mar 2026

Open source projects overwhelmed

- Linux kernel bug reports climbed from 2 to 10/week, initially hallucinated, now all verified real
- The **curl project**, which **discontinued its bug bounty** over AI-generated "slop" reports, now echoes the same shift, an increasing share of reports are quality AI-supported findings

Mar 2026

[un]prompted conference + Zero Day Clock

- **[un]prompted** introduces multiple talks, open source projects, and specific implementations demonstrating the risk with data
- The **Zero Day Clock** is launched, visualizing the collapse of time-to-exploit — now under one day in 2026

Apr 7, 2026

Claude Mythos Preview

- Anthropic announces Claude Mythos Preview & Project Glasswing. Discovers thousands of zero-days across every major OS and browser. 72% exploit success rate. 27-yr-old OpenBSD bug.

The Mythos/AI-Ready Security Program

The changing landscape, and resulting risk and impact, require an approach that is both operational, incident response-like, and strategic, focused on program building over time. This is implemented across three time horizons.

It is beyond the scope of this text to be exhaustive or prescribe how a full-fledged AI security program should be built. Rather, we selected *high-impact* recommendations that you can start with today, based on what the community can clearly discern at this early stage.

Beyond Application Security and Vulnerability Management, Mythos affects the wider security program. For example:

- **Operationally**, expect a potential deluge of new patches released from the 40 vendors in the early access program, similar to recent experience of needing to respond to multiple supply chain incidents within a two-week timeframe.
- **Risk management-wise**, business risk is shifting and engagement with stakeholders on risk planning and tolerance is key. The CISO's ability to manage risk has been reduced to a degree that could potentially have effects on business reporting and projections.
- **Strategically**, longer-term gap analysis and selective overhaul of various functions will be beneficial, including governance processes to support faster technology onboarding and the implementation of innovative AI-based security controls.

To start, a Mythos-ready security program should achieve minimum viable resilience. It would upgrade and realign measurements to a higher maturity level on key metrics such as cost of exploitation, early detection of compromise, and blast radius containment.

This matters because many of the assumptions underlying our cyber defense programs are being challenged. For example, time to exploitation has been reduced to minutes, we can no longer assume a patch will be ready in time for remediation purposes, incident frequency is likely to increase, the CVE system may not scale, shadow IT will fragment central control as coding agents proliferate to Citizen Coders, employees develop their own infrastructure, and threat intelligence is lagging behind on vulnerability discovery and exploitation.

The First of Many Waves?

Further, any program we build must acknowledge Mythos is only the first wave of future AI technology disruptions. In building a Mythos-ready program, we are not only seeking a return to equilibrium but also preparing to maintain balance for the waves ahead.

A Mythos-ready program should also account for how these shifts affect your team. The pace of change is real, and practitioners across all levels are working through what AI means for their roles and skills. This is a normal response to disruptive capability shifts, not a crisis of relevance. The practitioners who adapt fastest will be the ones who lean into AI

tooling rather than viewing it as a threat to their expertise.
The path forward is doubling down on fundamental security controls and hands-on adoption of agents at every level, from the CISO down. Every security role is becoming an "AI builder" role, and the barrier is lower than most people realize. Using a coding agent is now **easier than using Excel.**

The Human Cost

Leaders must be clear-eyed about the human cost of this transition.

Security teams are caught in a vice: AI is simultaneously accelerating the volume of vulnerabilities they must respond to, the volume of code their organisations are shipping, and expanding the attack surface. Add the cognitive intensity of integrating AI into their own workflows, and you have a workforce already at capacity absorbing exponential increases in workload without corresponding investment in headcount, tooling, or wellbeing.

Burnout and attrition in security functions represent a direct operational risk - the expertise needed to navigate this transition is scarce, takes years to develop, and is not replaceable on short timescales. Security team resilience, including sustainable workload, mental health support, and retention, should be treated as a strategic priority with the same urgency as the technical challenges AI presents.

Security practitioners, ourselves included, are facing a culture challenge. Many are uncertain about how their roles will evolve. It is often unclear to them, and us, how we could keep up with the pace of change. This affects even the most technical, such as vulnerability researchers, many of whom are asking questions about the future and if they will have a place in it.

Agents, often in the form of coding agents, also represent an opportunity for personal growth, and a feeling of empowerment. Everyone on your team, including you, can become hands on. All roles will likely become “AI builders”, and getting started is now easier than using Excel. All you need to know is English.

The Shrinking Time Horizon

The time available for action is shrinking, and we need to find ways to move faster. Long-term goals should be considered a quarter away at most.

10 Questions to Understand Your Security Program State and Influence

A questions-based approach to triage your understanding of your security program, to reach ground truth, as well as gauge your influence on various business functions.

10 Questions to Understand Your Security Program State and Influence

A questions-based approach to triage your understanding of your security program, to reach ground truth, as well as gauge your influence on various business functions.

Question	Context
What is our actual stance on AI today?	Allowed, tolerated, restricted, or unknown.
Can employees use agentic coding tools in the enterprise today?	Making use of agentic capabilities such as looping LLM tool use, and specifically coding agents (regardless of writing code), not just chatbot access. Do you have security guardrails in place for these coding agents?
Can employees contribute to open source without legal ambiguity?	A legal and IP question, not a technology philosophy question.
Do we have disciplined control repos, artifacts, and software, including for agentic supply chain such as MCP servers, plugins, and skills?	Source control, package paths, artifact provenance, and what is actually allowed in, in the CI/CD pipeline and through coding agents.
Is there a real cooling-off point/security gate between code change and production?	Demonstrates enforcement of security in release cycles and control of software supply chain.
Is security operational, or primarily advisory?	The extent to which the security function can directly affect outcomes, or does it serve mostly as a review and escalation function.
What is the fastest this company has made a security-driven production change in the last year?	Use a real example, not a policy statement.
Are our critical "crown jewels" explicitly tracked and current?	Not theoretically important systems. The actual few that matter most, and their main dependencies.
Do we know how to get urgent work prioritized by our key third parties?	Feature requests, bug reports, security escalations, relationship ownership, and leverage.
Does executive leadership have a working definition of urgency?	If everything is a crisis, nothing is urgent.

Updating Your Security Program

With those answers, we start with a draft risk register, followed by a list of prioritized actions and controls for a Mythos-ready security program, based on what the writers believe are most likely to be effective and impactful for most organizations.

Each action below is broken down into when it should commence, and a generalized estimate on a potential risk is linked to recognized frameworks, and each action to a time horizon under which it could be completed, for most organizations.

A Mythos-ready Security Program Risk Register (DRAFT)

#	Severity	Risk	Description	Type	Framework Refs	Maps to Priority Action
			CRITICAL			
1	Critical	**Accelerated Threat Exploitation** *AI-autonomous exploit generation at machine speed*	AI models have been discovering vulnerabilities and creating exploits for over a year. Mythos accelerates this significantly, but the capability predates it. What changes is the speed, scale, and the reduction in skill required to execute complex attacks,democratizing capabilities that were previously expensive and skill-intensive. Non-frontier, open-weight models can already achieve much of this at accessible cost. Frontier models like Mythos are the acceleration, not the starting gun. Each patch also becomes an exploit blueprint, as AI accelerates patch-diffing and reverse engineering of fixes.	*Threat*	AML.T0040, AML.T0043, PR.PS, PR.IR	**PA 3, 5**
2	Critical	**Insufficient AI Automation Capabilities** *Defenders operating at human speed while attackers operate with AI augmentation*	Attackers freely use AI coding agents for vulnerability discovery, exploit development, and attack orchestration. Many defensive teams are not yet aware of equivalent capabilities available to them, or lack the security controls to deploy them confidently. The resulting asymmetry is not just technological but cultural: teams that do not adopt AI agents cannot match the speed or scale of AI-augmented threats, regardless of their technical skill.	*Capability gap*	GV.OC, GV.RM, DE.CM, RS.MA	**PA 1, 2**

3	Critical	**Unmanaged AI Agent Attack Surface** *Privileged AI agents outside existing control frameworks*	Agents are necessary to counter AI-speed threats, but they are privileged, insecure by default, and not covered by existing security controls. This asset class introduces defensive risks (insecure, privileged agents within your own environment) and supply chain risks (compromised or manipulated agents from third parties). These have different owners and different mitigations.	*Vulnerability*	LLM06, ASI02, ASI03, AML.T0047, PR.AA, GV.SC	**PA 4**
4	Critical	**Inadequate Incident Detection and Response Velocity** *Detection and response at human speed against machine-speed attacks*	AI has reduced the sophistication and time needed to construct complex attacks. Defensive detection and response capabilities have not yet been upgraded to match, creating an asymmetric speed advantage for attackers. Alert triage volumes, SIEM correlation speed, and containment authorization latency were designed for human-paced threats.	*Capability gap*	ASI08, AML.T0047, DE.CM, DE.AE, RS.MA	**PA 9, 10**
5	Critical	**Cybersecurity Risk Model Outdated** *Stakeholder decisions based on pre-AI risk models*	Security reporting metrics built on pre-AI assumptions about exploit timelines and attack complexity may no longer reflect actual exposure. The CISO's ability to control risk has shifted, which could affect business reporting and projections. Outdated risk models could lead to underfunding of the controls that prevent incidents.	*Governance*	GV.OC, GV.RM, RS.CO	**PA 6**
			HIGH			
6	High	**Incomplete Asset and Exposure Inventory** *Unknown attack surface, assets, code, dependencies, shadow agents*	AI-accelerated attacker capabilities change which assets are at highest risk and which controls matter most. Attackers can now scan an entire OS codebase at accessible cost and enumerate your exposure faster than you can inventory it. For assets that cannot be patched or directly defended, inventory determines whether you can segment, isolate, or monitor them. Without continuously updated inventory, controls have inherent gaps. The proliferation of coding agents to non-developer users further fragments central IT visibility.	*Vulnerability*	ASI04, AML.T0000, ID.AM, GV.SC	**PA 7**

7	High	**Unsecured Software Delivery Pipeline** *Code shipping without AI-driven security review*	Code produced by both humans and AI agents ships without consistent security review. AI-generated code introduces vulnerabilities at higher volume than manual development. The risk compounds: more code produced faster, with the same defect rate, against a more capable adversary. Without LLM-driven review integrated into the pipeline, exploitable flaws reach production before defenders can find them.	*Vulnerability*	LLM01, LLM05, LLM08, ASI01, AML.T0018, AML.T0051.001, PR.PS, ID.IM	**PA 1**
8	High	**Network Architecture Insufficient for Lateral Movement Containment** *Flat or insufficiently segmented network enabling 1:N exploit leverage*	A flat or insufficiently segmented network gives every successful exploit leverage. AI-driven attacks worsen this: automated multi-hop lateral movement exploits poor architecture faster and more creatively than manual attackers ever could. When AI-accelerated vulnerability discovery increases the volume of exploitable findings, architectural segmentation becomes the primary control limiting blast radius.	*Vulnerability*	PR.IR, PR.PS	**PA 8**
9	High	**Continuous Vulnerability Management Maturity Gap** *Reactive posture against continuous AI-discovered zero-days, no VulnOps function*	AI-driven vulnerability discovery, which predates Mythos but is significantly accelerated by it, means zero-day vulnerabilities in your own code and third-party software can be discovered and weaponized before your security team knows they exist. Quarterly pen tests and reactive patching cycles cannot keep pace with continuous AI-driven discovery. Existing CVE/NVD infrastructure and patch prioritization workflows were built for dozens of critical CVEs per month, not hundreds.	*Capability gap*	ASI10, ASI06, AML.T0018, ID.RA, ID.AM, DE.CM	**PA 11**
10	High	**Threat Detection Dependent on Lagging Intelligence** *CVE- and KEV-based intelligence structurally outpaced by AI discovery rates*	Threat intelligence has been falling behind AI-accelerated vulnerability discovery for over a year. Mythos widens the gap further. The CVE system may not scale to AI-generated discovery rates, and novel vulnerabilities have no listing in KEV by definition.	*Capability gap*	AML.T0000, DE.CM, ID.RA, GV.OV	**PA 9, 10**

11	High	**Innovation Governance and Oversight Deficit** *Governance vacuum creating approval friction that slows defensive AI adoption*	Without a cross-functional governance mechanism, the onboarding and deployment of any new control runs into approval friction that slows adoption. This is where the liability and governance asymmetry gets addressed structurally. AI-accelerated timelines mean this friction now has a harder deadline.	*Governance*	GV.OC, GV.RM, GV.RR, GV.OV	**PA 2, 3**
12	High	**Regulatory and Liability Exposure from AI-Discovered** *Vulnerabilities* *Shifting standard of care as AI scanning becomes broadly available*	The EU AI Act (August 2026) introduces automated audit, incident reporting, and cybersecurity requirements around AI. Existing regulations use reasonableness as a test. When AI can find significantly more vulnerabilities at accessible cost, the standard of what constitutes reasonable defensive effort shifts. Boards will face questions about whether they used available AI tools for defensive scanning, and whether not doing so constitutes negligence. This is a governance risk with direct financial exposure.	*Governance*	GV.OC, GV.RM, GV.RR	**PA 1, 3**
			MEDIUM			
13	Medium	**AI Hype and Confusion Causing Systematic Inaction** Signal-to-noise collapse in threat and technology guidance	The volume of AI-related security guidance, commentary and vendor claims exceeds anything the industry has experienced. Security leaders find it difficult to navigate the noise. The confusion itself is a consequential risk: teams that dismiss the shift as hype, or exhaust their attention on low-signal content, will miss critical threat landscape changes they need to react to.	*Governance*	GV.OC, GV.RM	**PA 1**

See appendix for a full legend. Grouped by severity.
PA = Priority Action from the Mythos-Ready Security Program table.
See Appendix B for framework reference legend.

Priority Actions for a Mythos-Ready Security Program (Aggressive Time Table) (DRAFT)

For the CISO who needs to walk into a room Monday morning with a plan. This is meant as a quick reference to facilitate strategy and action.

#	Action	Category	Risk	Start	Horizon	What It Means
1	**Point Agents at Your Code and Pipelines**	Risk Control	Critical	This week	Ongoing	Turn agents and LLM capabilities inward on your own code and dependencies. Start immediately by asking an agent for a security review of any code, then build toward a full audit within your CI/CD pipeline, and shift left by adding capabilities directly into developers' coding agents. All code (human or AI-generated) should pass LLM-driven security review before merge. • Commercial: Claude Code Security from Anthropic, Codex Security from OpenAI. • Open source: OpenAnt from Knostic, raptor (Claude Code framework), the exploitation-validator agentic skill, and agentic skills from Trail of Bits.
2	**Require AI Agent Adoption**	Operational Enabler	Critical	This week	Ongoing	Formalize AI agent usage (mostly in the form of "coding agents") as part of all security functions, with mandatory security controls and oversight in place. While defensive AI technology has not yet caught up, these agents empower staff to be effective in the new threat landscape, allowing acceleration beyond "human speed." Optional adoption programs have not been shown to overcome cultural barriers, while adoption is a limiting factor in achieving the rest of the actions in this table.
3	**Establish Innovation, Acceleration Governance**	Governance	Critical	This week	6 months	Cross-functional mechanism (Security, Legal, Engineering) to evaluate new offensive threats and accelerate onboarding of defensive technologies. Without this in place, every other action in this table runs into approval friction that slows deployment to the attacker's advantage.
4	**Defend Your Agents**	Risk Control	Critical	This month	45 days	Without agents, most tasks on this list will be untenable, but they must be defended. Agents are not covered by existing controls and introduce both cyber defense and agentic supply chain risks. The agent harness – prompts, tool definitions, retrieval pipelines, and escalation logic – is where the most consequential failures occur; audit it with the same rigor as the agent's permissions. Before deploying agents in or adjacent to production environments, define scope boundaries, blast-radius limits, escalation logic, and human override mechanisms. Do not wait for industry governance frameworks. Define your own now.

5	**Prepare for Continuous Patching**	Risk Control	Critical	This week	45 days	With the increase in vulnerability discovery and reporting, and specifically now that Glasswing has made Mythos available to significant software vendors, prepare triage and deployment capacity to handle a potential flood of patches as new critical vulnerabilities are disclosed.
6	**Update Risk Models and Reporting**	Governance	Critical	This week	45 days	Review and update security risk metrics, reporting, and business risk calculations to reflect AI-accelerated exploit timelines and attack complexity. Pre-AI assumptions about patch windows, exploit scarcity, and incident frequency may no longer hold. Outdated models could potentially even lead to underfunding of controls and inaccurate business reporting. Communicate and collaborate with stakeholders, mapping out and prioritizing potential effects on the business, reporting, and projections..
7	**Inventory and Reduce Attack Surface**	Risk Control	High	This month	90 days	Make use of, update, or create an inventory. Using agents, the process can be significantly accelerated and enable continuous updates. Start with critical internet-facing systems, build toward a full-coverage inventory over 45 days. Generate real SBOMs. Aggressively shut down unneeded or unmaintained functionality, phase out suppliers that no longer comply with your updated vulnerability management requirements, and isolate or airgap at-risk systems. You cannot patch, segment, or defend what you don't know exists.
8	**Harden Your Environment**	Risk Control	High	This month	6 months	The basics remain valid and can be prioritized for risks that can't be easily mitigated. Implement egress filtering (it blocked every public log4j exploit). Enforce deep segmentation and zero trust where possible. Lock down your dependency chain. Mandate phishing-resistant MFA for all privileged accounts. Every boundary increases attacker cost.
9	**Build a Deception Capability**	Risk Control	High	Next 90 days	6 months	Deception is attack-tool and vulnerability independent, identifying attacks and attackers based on their TTPs. Deploy canaries and honey tokens, layer behavioral monitoring, pre-authorize containment actions, and build response playbooks that execute at machine speed.
10	**Build an Automated Response Capability**	Risk Control	High	Next 90 days	12 months	Improve detection engineering and incident response capabilities to be systemic and, to the degree possible, autonomous. Examples: asset and user behavioral analysis, pre-authorized containment actions, and response playbooks that execute at machine speed.

11	**Stand Up VulnOps**	**Risk Control**	**Critical**	**Next 6 months**	**12 months**	Long-term, there is no alternative to building a permanent Vulnerability Operations (VulnOps) function, staffed and automated like DevOps, but for autonomous vulnerability research and remediation. Owns continuous discovery of zero-day vulnerabilities across your entire software estate (from your own code to third-party software), and establishes automated remediation pipelines. Design VulnOps around triage discipline from the start.

Risk: **Critical** = immediate exposure if unaddressed **High** = significant exposure within 45 days Category: **Governance** = structural prerequisite **Risk Control** = direct risk reduction **Operational Enabler** = makes risk controls executable

Executive and Board Briefing: the AI Risk Summary

Mythos is now a boardroom concern, and that creates an opportunity. This section is a working tool for CISOs preparing a leadership and/or board update, organized around two things: justifying the current program and making the case for what comes next. Every organization is different, so make sure you align the talking points and timelines with your actual current situation and programs.

The Shift

AI at the capability level demonstrated by Mythos will transform how organizations operate, compressing development cycles and accelerating time to market. The business is already pursuing that value with current highly-capable models.

That same capability in adversary hands compresses the time between a vulnerability existing and causing business disruption from weeks to hours; a permanent acceleration, not a temporary spike.

This has two implications for the organization. First, several assumptions behind current risk metrics may no longer hold and need re-examination. We have moved into a world of containment and a focus on resilience, so metrics should now focus on the speed to recover to normal operations. Second, the same AI capabilities that create this risk also create a defensive opportunity: organizations can now identify their own weaknesses before attackers do, review code at machine speed, and respond to incidents faster than any human team can. Organizations that invest will be both faster to market and more resilient to attack.

Talking Point: AI Accelerates Both Sides

AI is making us faster and more competitive. But those same capabilities make attackers faster and more dangerous. It has compressed the time to a serious incident from weeks to hours, and that gap will continue to narrow. Turned inward, these tools let us find and fix our own weaknesses before adversaries do. Without attention to buying down risk, we move faster as a business while accumulating risk just as rapidly.

The security program this company has funded is what makes our AI security strategy viable. The investments already in place ensure that no single point of entry becomes a full business disruption. In an environment where entry points and weaknesses are discovered faster, that containment architecture is more valuable, not less.

With continued support, the changes we recommend here will return risk to pre-Mythos levels and demonstrate due diligence in response to a documented shift in the threat environment. This program builds the foundation that lets the business move fast with confidence.

Talking Point: An Aggressive Plan Is Needed

The funded foundation is why our program can adapt rather than react in a crisis. What has changed is the speed and volume it must absorb.

This is not an open-ended AI initiative. We are seeking alignment to execute a targeted and aggressive 90-day plan with clear owners and outcomes:

- **Increase People and Capacity.** Plan for repurposing of existing staff (within the security org, but also, and especially, within engineering teams) and/or onboarding of additional headcount and contractor capacity to handle the anticipated increases in triage, remediation, and incidents, while protecting experienced staff from burnout, especially as the first wave of Glasswing patches hits.

- **Deploy AI Tooling.** Formalize AI agent usage across all security functions as standard practice: scanning our own code, ensuring AI-driven review before code ships, and augmenting teams with purpose-built agents. This equips our teams to operate at the same speed as adversaries.

- **Harden Infrastructure.** Prioritize updating asset inventories; reducing unnecessary exposure; and enforcing segmentation, Zero Trust, egress filtering, and phishing-resistant authentication. Validate these elements across internal systems and key third-party providers (MSPs, SOCs).

- **Accelerate Procurement and Governance.** Align across functional teams (security, legal, engineering) to evaluate threats and fast-track priority defensive technology onboarding. Current approval cycles are too slow for the coming threat environment.

- **Update Playbooks.** Update technical and communications response plans to execute at the required speed and scale, including pre-authorized containment and coordination for simultaneous incidents.

- **Track Progress.** Provide regular check-ins throughout the 90-day period to capture results and identify roadblocks.

Conclusions and Recommendations

AI-based attacks represent a structural shift in how offense and defense work, and it will not change. The cost and capability floor to exploit discovery is dropping, the time between disclosure and weaponization is compressing toward zero, and capabilities that previously required nation-state resources are now becoming broadly accessible.

While vulnerability discovery capabilities comparable to Mythos have shown to be present through earlier AI models, the Mythos announcement has grabbed the attention of the boardroom. Defenders can seize this opportunity and make a compelling business case to become “Mythos-ready” and prepare for an oncoming onslaught of patches.

Being “Mythos-ready” means:

- Engineering a resilient architecture that limits the ability of attackers to exploit discovered vulnerabilities and contains the impact if they are exploited.
- Discovering more vulnerabilities yourself in advance of any adversary (or vendor advisories).
- Responding quickly to incidents at scale and containing the impact to minimize business disruption.
- Accelerating your security program and staff capabilities with AI agents.

Empower your teams to use AI for defense, starting today. Every action in this brief can begin this week.

We have done this before. Y2K was a systemic threat with a hard deadline, and the industry met it through coordinated, disciplined effort. This is the same kind of problem, requiring the same kind of response, with more powerful tools available to defenders.

Building a “Mythos-ready” security program is not about reacting to one model or announcement. It is about permanently closing the gap between how fast vulnerabilities are found and how fast your organization can respond.

Appendix A: Historical Precedence

Background

This all began with the DARPA Cyber Grand Challenge, a landmark competition organized by DARPA in 2016 that demonstrated the potential of fully automated cybersecurity systems. Teams developed autonomous platforms capable of identifying, exploiting, and patching software vulnerabilities in real time, without human intervention. The challenge highlighted a shift toward machine-speed cyber defense, showing how automation and artificial intelligence could significantly enhance vulnerability management and incident response, while also raising important questions about trust, control, and the future role of human operators in cybersecurity.

By mid-2025, XBOW, an autonomous offensive security company, topped the HackerOne leaderboard.

The DARPA AI Cyber Challenge (AIxCC) found 54 vulnerabilities in four hours of compute. Google's Big Sleep discovered real zero-days in open source.
Anthropic was used to automate full attack chains from reconnaissance through exfiltration. And, open source tools such as raptor proved autonomous vulnerability research is available to anyone able to use an agent.
In September 2025, Heather Adkins (CISO, Google) and Gadi Evron (CEO, Knostic) published a warning that attackers were racing toward a singularity moment, with autonomous vulnerability discovery and exploitation roughly six months away.
In February 2026 Anthropic, using Claude Opus 4.6, reported more than 500 high-severity vulnerabilities in open source software. AISLE found 12 OpenSSL zero-days, including a CVSS 9.8 vulnerability dating to 1998.
Linux kernel maintainers saw vulnerability reports climb from 2 to 10 per week, largely hallucinated at first, but that changed rapidly. The volume has held steady, but the reports are now all verified as real bugs.
The curl project, which originally discontinued its bug bounty program because it was drowning in hallucinated vulnerability reports ("AI slop"), last week echoed the above observation from the Linux team, reporting an increasing number of AI-supported quality security reports.
Sysdig documented an AI-based attack that reached admin-level access in eight minutes.
This week, Gambit released a report on the AI-led compromise of Mexican government infrastructure, originally reported in February.

In March, Sergej Epp and others introduced the Zero Day Clock, visually demonstrating the disappearing time to exploit development, demonstrating the drastic fall in time to exploitation to less than a day in 2026. It is worth noting that the historical collapse in time-to-exploit has not yet produced a proportional increase in the impact of exploitation. Many of the most consequential incidents of recent years involved credential abuse, social engineering, or supply chain compromise rather than zero-day exploitation. The ZeroClock trend is a leading indicator of where attacker capability is heading, not a direct measure of current damage.

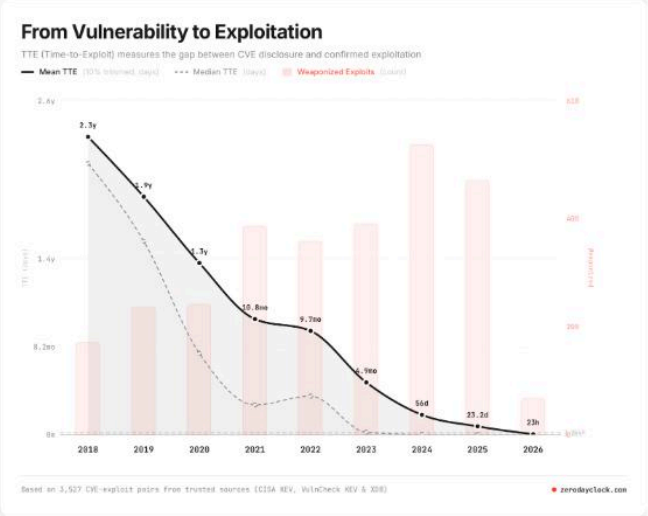


Diagram from the Zero Day Clock.

Appendix B: Mythos Risk Register Legend

OWASP LLM 2025 · OWASP Agentic 2026 · MITRE ATLAS · NIST CSF 2.0

1. Framework Prefixes
Every code in the Frameworks column belongs to one of these four frameworks.

LLMxx **OWASP Top 10 for LLM Applications 2025** *Risks in LLMs used as application components*	**ASIxx** **OWASP Top 10 for Agentic Applications 2026** *Risks in autonomous AI systems that plan and act*
AML.Txxxx **MITRE ATLAS** *Adversarial techniques targeting AI/ML systems*	**GV.xx** **NIST CSF 2.0 - Govern (GV)** *Governance: context, risk strategy, roles, supply chain*
ID.xx **NIST CSF 2.0 - Identify (ID)** *Asset management, risk assessment, improvement*	**PR.xx** **NIST CSF 2.0 - Protect (PR)** *Access control, platform security, resilience*
DE.xx **NIST CSF 2.0 - Detect (DE)** *Continuous monitoring, adverse event analysis*	**RS.xx** **NIST CSF 2.0 - Respond (RS)** *Incident management and communication*

2. All Framework Codes Used in This Register

Code	Full name and framework
AML.T0000	ML Model Reconnaissance - MITRE ATLAS
AML.T0018	Backdoor ML Model - MITRE ATLAS
AML.T0040	ML Inference API Access - MITRE ATLAS
AML.T0043	Craft Adversarial Data - MITRE ATLAS
AML.T0047	ML-Enabled Product Abuse - MITRE ATLAS
AML.T0051.000	LLM Prompt Injection (Direct) - MITRE ATLAS
AML.T0051.001	LLM Prompt Injection (Indirect) - MITRE ATLAS
ASI01	Agent Goal Hijack - OWASP Agentic Top 10 2026
ASI02	Tool Misuse and Exploitation - OWASP Agentic Top 10 2026
ASI03	Identity and Privilege Abuse - OWASP Agentic Top 10 2026
ASI04	Agentic Supply Chain Vulnerabilities - OWASP Agentic Top 10 2026

ASI06	Memory and Context Poisoning - OWASP Agentic Top 10 2026
ASI08	Cascading Failures - OWASP Agentic Top 10 2026
ASI10	Rogue Agents - OWASP Agentic Top 10 2026
LLM01	Prompt Injection - OWASP LLM Top 10 2025
LLM02	Sensitive Information Disclosure - OWASP LLM Top 10 2025
LLM05	Improper Output Handling - OWASP LLM Top 10 2025
LLM06	Excessive Agency - OWASP LLM Top 10 2025
LLM08	Vector and Embedding Weaknesses - OWASP LLM Top 10 2025
DE.AE	Adverse Event Analysis - NIST CSF 2.0 Detect
DE.CM	Continuous Monitoring - NIST CSF 2.0 Detect
GV.OC	Organizational Context - NIST CSF 2.0 Govern
GV.OV	Oversight - NIST CSF 2.0 Govern
GV.RM	Risk Management Strategy - NIST CSF 2.0 Govern
GV.RR	Roles, Responsibilities, and Authorities - NIST CSF 2.0 Govern
GV.SC	Supply Chain Risk Management - NIST CSF 2.0 Govern
ID.AM	Asset Management - NIST CSF 2.0 Identify
ID.IM	Improvement - NIST CSF 2.0 Identify
ID.RA	Risk Assessment - NIST CSF 2.0 Identify
PR.AA	Identity Management, Authentication, and Access Control - NIST CSF 2.0 Protect
PR.IR	Infrastructure Resilience - NIST CSF 2.0 Protect
PR.PS	Platform Security - NIST CSF 2.0 Protect
RS.CO	Incident Response Communication - NIST CSF 2.0 Respond
RS.MA	Incident Management - NIST CSF 2.0 Respond

3. Severity

Level	Meaning
Critical	Immediate exposure or increased risk if unaddressed
High	Significant exposure or increased risk within 45 days
Level	Meaning
Medium	Organizational risk requiring structured attention; does not create direct exploitable exposure but weakens the effectiveness of higher-priority controls if left unaddressed.

4. Risk Type

Type	Definition
Threat	External actor capability - controls raise cost but cannot eliminate it
Vulnerability	Internal exploitable condition - addressable through remediation
Capability gap	Defensive function missing or operating below the required level
Governance	Organizational or structural failure that amplifies every other risk