Cybersecurity Growth

AI SECURITY GUIDANCE SERIES

PUBLIC GUIDANCE · JUNE 2026

Securing the Enterprise in the Age of Frontier AI

A CISO Field Guide

Practical guidance for controlling AI data loss, securing AI-assisted development, and defending against AI-powered threats.

Audience	Chief Information Security Officers and security leadership
Version	V1.0 | Published 23 June 2026
Classification	Public / Point-in-time guidance

This document reflects the AI threat and tooling landscape as of June 2026. Capabilities, products, regulations, and the status of frontier models referenced herein are evolving rapidly. Treat all specifics as point-in-time and re-validate before acting.

Executive Snapshot

Data loss to AI tools

Employees are pasting sensitive information into third-party AI tools faster than legacy DLP can see.

AI-assisted development

AI-generated code should be treated as untrusted code and routed through normal SDLC gates.

AI-powered threats

Frontier models change attacker speed more than attacker shape, raising pressure on mitigation throughput.

BOARD-LEVEL TAKEAWAY

The binding constraint is no longer finding vulnerabilities. It is the human capacity to triage, mitigate, patch, and contain them.

Contents

1. Executive Summary

AI has crossed from productivity accelerant to security-defining force. As of mid-2026, three shifts demand a CISO's direct attention. First, **employees are exfiltrating sensitive data into third-party AI tools** at scale, often without intent or awareness. Second, **AI-assisted software development** is injecting vulnerabilities and untrusted code into the SDLC faster than traditional review can catch. Third, **frontier models with strong cyber capabilities** appear to compress the attacker's timeline from vulnerability discovery to working exploit, particularly where vulnerability chaining and proof-of-concept generation can be automated.

This is not a responsible-AI policy paper. It is a security operating guide for controlling AI-related data exposure, software risk, and attacker acceleration.

The third shift is no longer theoretical. In April 2026, Anthropic launched **Project Glasswing**, giving roughly 50 partners early access to an unreleased frontier model, **Claude Mythos Preview**. Within about one month, partners and Anthropic collectively surfaced more than **10,000 high- or critical-severity vulnerabilities** across systemically important software. Cloudflare reported that the model could chain low-severity primitives into working exploit chains and generate proof-of-concept code in a scratch environment.

This guide maps a practical, control-oriented response to the public frameworks your auditors, board, and regulators already recognize: NIST CSF 2.0, MITRE ATT&CK, MITRE ATLAS, the OWASP Top 10 for LLM Applications, ISO/IEC 42001, and the EU AI Act. The core argument is blunt:

- **Data loss to AI tools is a data-governance and egress problem.** Solve it with AI-aware DLP, identity, and acceptable-use enforcement, not bans.
- **AI-generated code is untrusted code.** Treat it as supply chain and apply provenance, SAST/SCA, and human accountability.
- **Frontier-model attackers change speed, not shape.** The broad intrusion lifecycle is largely unchanged, so resilient architecture beats patch-velocity heroics.

> **BOTTOM LINE FOR THE BOARD**
>
> The binding constraint is no longer finding vulnerabilities. It is the human capacity to triage, patch, and contain them. Invest in remediation throughput and defensible architecture now, because the offensive capability demonstrated in controlled settings will proliferate. For critical exposed vulnerabilities, the first SLA should no longer be "patch within X days." It should be "mitigate within hours, remediate as fast as feasible."

2. How to Use This Guide

This guide is written for CISOs and security leaders building a practical AI-security program. Use Sections 5 through 7 as the control architecture, Section 8 as the governance model, and Section 9 as the 90-day action plan. Technical teams should treat the pseudo-policy examples as control intent, not production-ready implementation. Every figure and product reference is point-in-time; verify against current sources before acting.

CISO FIELD GUIDE

3. The 2026 AI Threat Landscape

10,000+
High/critical findings
Reported across controlled Project Glasswing access in roughly one month.

90.6%
Validated as true positives
Of 1,752 high/critical findings independently assessed.

Speed, not shape
Lifecycle stays detectable
Attack stages remain visible; only the timelines compress.

3.1 What Project Glasswing demonstrated

Anthropic's Project Glasswing initial update and Cloudflare's coverage report the following point-in-time figures and observations:

- **10,000+** high- or critical-severity vulnerabilities found by approximately 50 partners in the first month.
- **23,019** total issues identified across 1,000+ open-source projects scanned with Mythos Preview; 6,202 classified high or critical.
- Of 1,752 high/critical findings assessed by six independent firms and Anthropic, **90.6%** (1,587) validated as true positives.
- Cloudflare found roughly 2,000 bugs, including 400 high/critical, with a false-positive rate it considered better than human testers.
- Mozilla found and fixed 271 vulnerabilities in Firefox 150 while testing the model, about ten times the prior version.

Cloudflare's technical write-up makes the most operationally important point for defenders: the model's distinguishing capability was **chaining**. It could take low-severity bugs that would normally sit invisible in a backlog and stitch them into a single severe, demonstrably exploitable chain, then write and run proof-of-concept code to prove it.

3.2 A concrete example: CVE-2026-5194 (wolfSSL)

The clearest illustration of impact is **CVE-2026-5194**, an improper certificate validation flaw (CWE-295) in the wolfSSL TLS/SSL library, discovered by a researcher at Anthropic. wolfSSL is embedded in an estimated billions of devices worldwide: IoT, routers, industrial control systems, automotive, and aerospace.

- **Root cause:** missing hash/digest-size and OID validation during ECDSA signature verification, allowing smaller-than-required digests to be accepted under affected conditions.
- **Impact:** certificate verification can accept improperly constrained signatures, enabling TLS man-in-the-middle and trust-forgery scenarios. The risk is magnified because wolfSSL is embedded in downstream products, firmware, and statically linked builds where remediation lags.
- **Fix:** wolfSSL 5.9.1, released 8 April 2026. The hard part is downstream: statically linked builds and vendor firmware require rebuilds, so real-world remediation can lag by weeks to months.

The CISO lesson is not the bug. It is the asymmetry. A frontier model found a critical flaw in a library most organizations do not know they depend on, then demonstrated a working forgery attack. Your exposure lives in your software bill of materials, not just your first-party code.

3.3 Speed, not shape

Cloudflare's framing for defenders is precise: a model like Mythos Preview does not fundamentally change the shape of an intrusion. Reconnaissance, initial access, lateral movement, persistence, and exfiltration all still have to happen. What changes is speed and scale. This is good news for defenders, because every stage of MITRE ATT&CK remains a place to detect, slow, and contain. The architecture around a vulnerability matters more than the speed of the patch.

> **STATUS NOTE (POINT-IN-TIME)**
>
> As of June 2026, Mythos-class access remained controlled rather than generally available, though Anthropic had begun expanding Project Glasswing to roughly 150 additional vetted organizations across more than 15 countries. The defensive planning assumption should be that similar capabilities will proliferate, whether through commercial frontier models, controlled partner programs, or open-source alternatives.

CISO FIELD GUIDE

4. Anchoring to Public Frameworks

Do not invent a bespoke AI control taxonomy. Map AI risk onto frameworks your auditors, board, and regulators already recognize. The six below are the load-bearing references for 2026.

Framework	What it covers	How to use it for AI
NIST CSF 2.0	Govern, Identify, Protect, Detect, Respond, Recover	Use the new GOVERN function to own AI risk at the executive level; map every AI control to a CSF category.
MITRE ATT&CK	Adversary tactics and techniques against enterprise systems	Threat-model AI-accelerated attacks; the broad lifecycle is unchanged, so existing detections still apply.
MITRE ATLAS	Adversarial threat landscape specific to AI systems	Threat-model attacks on your AI: prompt injection, model evasion, data and model poisoning, extraction.
OWASP Top 10 for LLM Apps (2025)	Top risks in LLM-based applications	The build-time checklist for any team shipping LLM features or agents.
ISO/IEC 42001:2023	Certifiable AI Management System (AIMS)	The governance operating system; certify it to prove responsible-AI practice to customers and regulators.
EU AI Act	Risk-tiered legal obligations for AI systems	Governance, documentation, and risk-classification baseline; applies extraterritorially if you serve the EU.

4.1 OWASP Top 10 for LLM Applications (2025)

The current edition is the 2025 list from the OWASP GenAI Security Project. Security leaders should treat these as mandatory review gates for any LLM-enabled product.

ID	Risk	CISO priority
LLM01	Prompt Injection (direct and indirect)	Highest. No fool-proof fix; layer mitigations.
LLM02	Sensitive Information Disclosure	High. Ties directly to your DLP program.

ID	Risk	CISO priority
LLM03	Supply Chain	High. Models, datasets, adapters are dependencies.
LLM04	Data and Model Poisoning	Medium-High. Backdoors via training/fine-tune data.
LLM05	Improper Output Handling	High. Treat model output as untrusted input.
LLM06	Excessive Agency	High. Constrain tool and agent permissions tightly.
LLM07	System Prompt Leakage	Medium. System prompts are not security controls.
LLM08	Vector and Embedding Weaknesses	Medium. RAG is an attack surface, not a shield.
LLM09	Misinformation	Medium. Overreliance on unverified output.
LLM10	Unbounded Consumption	Medium. Cost and DoS via uncontrolled inference.

Two non-obvious truths from the 2025 guidance: system prompts are not firewalls, so if a secret is in the prompt, treat it as already disclosed; and RAG and fine-tuning ground a model but do not solve prompt injection, because the boundary between data and instruction is permanently blurred.

DOMAIN 1

5. Protecting Against Data Loss to AI Tools

The problem: employees paste source code, customer PII, contracts, financials, and credentials into consumer AI tools to get work done faster. Traditional DLP, built for email and USB, is often blind to this. Maps to OWASP LLM02 and NIST CSF PROTECT.

5.1 Why a blanket ban usually fails

A permanent ban drives usage to personal devices and accounts, creating shadow AI and eliminating all visibility. Temporary blocking may be necessary while controls are deployed, but the defensible long-term posture is to provide a sanctioned path and govern it. Three things make this work: an approved enterprise tool, identity-bound access, and inline inspection of what leaves.

5.2 Control architecture (layered)

1. **Govern first.** Publish an AI Acceptable Use Policy naming approved tools, prohibited data classes, and consequences. Without this, technical controls have no policy to enforce.
2. **Sanction an enterprise tier.** Provide an enterprise AI offering with contractual no-training-on-data terms, SSO, audit logging, and admin controls. This removes the incentive to use ungoverned consumer tools.
3. **Identity-bind access.** Force AI tools behind SSO/IdP with conditional access. Block personal-account logins to AI domains on managed devices via your secure web gateway.
4. **Inspect the egress.** Deploy AI-aware DLP that inspects prompts, pasted text, and file uploads in real time, and acts before data reaches the provider.
5. **Monitor and coach.** Log AI usage, alert on policy hits, and use in-the-moment user coaching rather than silent blocks that train avoidance.

5.3 Tooling landscape (by category)

As of mid-2026, AI-aware data protection clusters into four deployment models. The right choice depends on where you can intercept traffic and how mature your existing stack is.

Category	How it works	Representative tools	Trade-off
SSE / SASE inline	Cloud proxy inspects web/ SaaS traffic including AI domains	Zscaler, Netskope, Palo Alto Prisma, Forcepoint	Strong coverage; depends on a mature SSE rollout, adds backhaul.
Browser-layer	Extension or managed browser governs AI use at the tab	LayerX, Island, dope.security, Nightfall plugin	Fast to deploy; covers browser-accessible AI only.
Endpoint DLP	Agent inspects clipboard, uploads, local-app and local-LLM use	Microsoft Purview, Teramind, Symantec, Cyberhaven	Covers desktop AI apps and local models; agent overhead.
AI-native / prompt	LLM-based classification of prompts and responses	Prompt Security, Nightfall, Cyberhaven	Higher accuracy on intent; newer category, validate at scale.

> **SELECTION GUIDANCE**
>
> If you already run Zscaler or Netskope, extend it to AI before buying a point tool. If you need coverage in weeks, a managed browser or browser extension is the lightest lift. The differentiator that actually determines how much loss you prevent is classification quality and inspection location. Pattern/regex DLP and browser-only coverage both leave gaps that on-

device or LLM-based inspection closes. Validate that any vendor specifically supports the providers your staff use, such as ChatGPT, Claude, Gemini, and Copilot.

High-level technical controls

- **Secure web gateway rules:** force AI SaaS through inspection; block consumer AI domains for unmanaged identities; allow only the sanctioned enterprise tenant.
- **DLP classifiers:** tune detectors for crown-jewel data rather than relying only on generic templates.
- **Tenant controls:** disable training-on-data, enforce SSO, set retention limits, and export admin audit logs to your SIEM.
- **Detection content:** alert on first-seen AI domains, anomalous upload volume, and personal-account auth attempts.

```
# Conceptual SWG/CASB policy intent
IF destination IN ai_genai_category
   AND identity NOT IN corp_sso_group        -> BLOCK
IF upload.contains(PII | source_code | secrets)
   AND destination IN ai_genai_category      -> BLOCK + COACH + LOG
```

DOMAIN 2

6. Securing AI-Assisted Software Development

The problem: AI coding assistants and agents now write a material share of enterprise code. They accelerate delivery but also introduce insecure patterns, hallucinated or malicious dependencies, and leaked secrets. Maps to OWASP LLM03/LLM05, MITRE ATLAS, and NIST CSF PROTECT/IDENTIFY.

6.1 Core principle: AI-generated code is untrusted code

Treat output from a coding assistant exactly as you would code from an unvetted external contributor: it must pass the same provenance, review, scanning, and accountability gates. The productivity gain does not buy a trust exemption.

6.2 Specific risks

- **Insecure patterns at scale:** models reproduce common-but-wrong idioms, such as string-concatenated SQL, weak crypto defaults, and missing authorization checks.

- **Dependency hallucination / slopsquatting:** assistants suggest package names that do not exist; attackers pre-register those names with malicious payloads.
- **Secret leakage:** secrets pasted into prompts for help debugging, or generated code that hard-codes credentials.
- **Memory-unsafe code:** in C/C++, models can reproduce buffer-overflow and out-of-bounds classes that memory-safe languages eliminate at compile time. CVE-2026-5194 lived in exactly this kind of low-level code.
- **Prompt injection into agents:** coding agents that read issues, docs, or web content can be steered by injected instructions in that content.
- **Repository and workspace overreach:** assistants may gain broad read access to private repos, issues, pull requests, terminals, or cloud environments without a documented business need.

6.3 Control architecture across the SDLC

SDLC stage	Control	Implementation
Author	Govern the assistant	Enterprise tier with SSO, no-training terms, org policy on allowed repos/contexts; block secret-pasting via DLP.
Author	Scope repo/workspace access	Restrict coding assistants to approved repositories and contexts; require documented need for sensitive repo access.
Commit	Provenance and secrets	Signed commits; pre-commit secret scanning; flag AI-assisted commits for review weighting.
Build	SCA and dependency pinning	Verify every dependency exists and is pinned/hash-locked; block install of newly registered or low-reputation packages.
Test	SAST and DAST	Run static analysis on all merged code; gate on high-severity findings; add AI-specific test cases for agent inputs.
Review	Human accountability	A named human owns every merge; AI-heavy changes get heightened review, not lighter review.
Release	SBOM and signing	Generate an SBOM (CycloneDX/SPDX); sign artifacts; track third-party libs for fast CVE response.

High-level technical controls

```
# Conceptual CI policy
step: verify_dependencies
  for pkg in manifest:
    assert registry.exists(pkg) and pkg.age > 90d and pkg.reputation_ok
step: sast_scan      -> fail_if(severity >= HIGH)
```

```
step: secret_scan    -> fail_if(any_secret_found)
step: sbom_generate  -> publish(cyclonedx.json)
```

- **Constrain coding agents:** least-privilege tokens, no standing production credentials, and allowlisted tools/commands the agent may run.
- **Treat untrusted content as hostile:** when agents ingest issues, pull requests, or web pages, sanitize and sandbox. Never let fetched content carry execution authority.
- **Turn the capability around:** use AI-assisted security review and scanning defensively. The same models that find bugs for attackers find them for you. Prefer memory-safe languages such as Rust for new low-level code where practical.

DOMAIN 3

7. Defending Against AI-Powered Cyber Threats

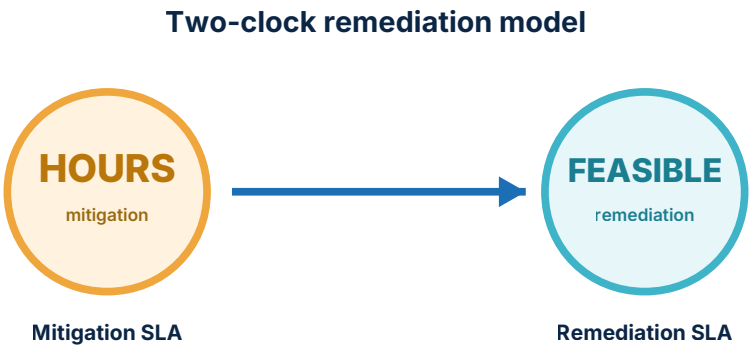
The problem: frontier models compress the attacker's path from discovery to working exploit. Patch-velocity alone cannot win a race against machine-speed exploit generation. The answer is architecture that assumes some vulnerabilities will always be unpatched. Maps to MITRE ATT&CK and NIST CSF DETECT/RESPOND/RECOVER.

7.1 The strategic reframe

As Cloudflare argued, the architecture around a vulnerability matters more than the speed of the patch. Because frontier models change speed but not the broad shape of an intrusion, every existing layer of defense-in-depth still works. It just has to be in place before the attacker arrives, not deployed in reaction.

7.2 Rewrite your remediation policy: mitigate in hours, not days

This is the single most urgent policy change for most organizations. Most remediation policies still encode a human-speed assumption: criticals patched within 15 or 30 days, highs within 60 or 90. That assumption is now dangerous. The gap between a vulnerability being discovered and a working exploit existing can collapse to hours, with no specialist skill required, because frontier models chain low-severity bugs and generate proof-of-concept exploits autonomously. The exposure window your day-based SLAs were calibrated for no longer exists, and the old triage heuristic, "hard to find equals hard to exploit," is invalid.



But be precise: "patch every critical in hours" is usually not achievable, and writing it as a literal SLA only produces paper non-compliance. Fixing a critical can take days to weeks of engineering; downstream rebuilds, as wolfSSL showed, take longer; and the Glasswing bottleneck is now fixing, not finding. The fix is to separate two clocks in written policy:

- **Mitigation SLA, hours.** For critical findings on exposed or internet-facing assets, deploy a compensating control within hours: a virtual patch or WAF rule, disabling the vulnerable feature, isolating or segmenting the asset, or taking it offline. This is the clock that must move from days to hours, and it is achievable.
- **Remediation SLA, compressed but feasibility-bound.** Apply the actual code or vendor patch as fast as the fix allows. Shorten patch testing and deployment windows, but do not commit to a timeline engineering cannot meet.

7.3 Defensive priorities

1. **Operationalize the two-clock model.** Codify the mitigate-in-hours / patch-as-feasible split in policy, with an emergency change path and pre-approved compensating controls.
2. **Know your SBOM.** You cannot patch wolfSSL-class dependencies you do not know you have. Maintain a current software bill of materials and map it to live CVE feeds for same-day triage.
3. **Architect for containment.** Assume breach. Microsegmentation, zero-trust access, and blast-radius limits mean a single exploited bug does not become lateral movement and exfiltration.
4. **Instrument the full ATT&CK chain.** Discovery and initial access may be faster, but persistence, lateral movement, command and control, and exfiltration still occur. Detection at these stages is where AI-accelerated attacks get caught.
5. **Prioritize remediation throughput.** The Glasswing bottleneck is human triage and patching, not bug-finding. Invest in evidence-driven triage, verified remediation, and automation that closes the loop.
6. **Harden cryptographic identity.** CVE-2026-5194 was a trust-forgery flaw. Inventory crypto libraries, enforce certificate validation rigor, and watch the post-quantum / crypto-agility transition.

Anchor on architecture: Anthropic noted the NIST and UK NCSC critical controls now matter more, precisely because they improve security without depending on any single patch landing in time. Mitigation speed plus containment is the foundation; patch velocity alone cannot win the race.

7.4 Defending your own AI systems (MITRE ATLAS)

Do not forget the inward-facing risk: attacks against the AI you deploy. Use MITRE ATLAS to threat-model these and apply OWASP LLM mitigations.

- **Prompt injection:** filter and isolate untrusted content; never let model output trigger privileged actions without a deterministic check.
- **Data/model poisoning:** vet training and fine-tuning data provenance; version and integrity-check models and adapters.
- **Model extraction and evasion:** rate-limit, monitor for systematic probing, and apply output controls.
- **Excessive agency:** constrain agent tools, scopes, and standing permissions; require human-in-the-loop for irreversible actions.

CISO FIELD GUIDE

8. Governance and Regulatory Alignment

One governance program should satisfy multiple masters at once. The map below shows how the load-bearing frameworks line up.

NIST CSF 2.0 GOVERN function	**OWASP LLM** Build-time gates	**MITRE ATT&CK** Attack chain
MITRE ATLAS AI-system attacks	**ISO/IEC 42001** AI Management System	**EU AI Act** Risk tiers

8.1 ISO/IEC 42001: the AI Management System

ISO/IEC 42001:2023 is the first certifiable standard for an AI Management System (AIMS). Think of it as the AI-specific counterpart to ISO 27001: where the EU AI Act is law and NIST CSF is a framework, ISO 42001 is the operating system for AI governance. It gives you a structured, auditable management system you can certify against to demonstrate responsible-AI practice to customers, partners, and regulators.

It follows the standard ISO management-system structure (clauses 4 through 10: context, leadership, planning, support, operation, performance evaluation, improvement) and adds **Annex A: 38 AI-specific controls across 9 control objectives (A.2 through A.10)**. As with ISO 27001, you document applicability and may exclude controls only with a justified, written rationale. Undocumented exclusions are a major nonconformity.

Annex A objective	Focus	CISO action
A.2 AI policies	Management direction for AI	Publish AI policy; align it with existing security, privacy, and ethics policies.
A.3 Internal organization	Roles and accountability	Assign AI roles and responsibilities; define reporting and escalation lines.
A.4 Resources for AI systems	Data, tooling, compute, people	Document the resources behind each AI system: data, models, compute, skills.

Annex A objective	Focus	CISO action
A.5 Impact assessment	Effects on individuals and society	Run AI impact assessments; map to EU AI Act risk tiers and fundamental-rights impact.
A.6 AI system life cycle	Responsible design to retirement	Embed security and responsible-AI gates across the full SDLC.
A.7 Data for AI systems	Data quality, provenance, governance	Govern training/fine-tuning data provenance, quality, and handling.
A.8 Information for interested parties	Transparency to users and stakeholders	Document system purpose, limitations, and disclosures for affected parties.
A.9 Responsible use	Use of AI within the organization	Enforce acceptable-use; this is where your AI DLP program lives.
A.10 Third-party relationships	Suppliers and customers	Flow AIMS obligations down to vendors; manage third-party and embedded-AI risk.

WHY CERTIFY

Certification is not box-ticking if you do it for the right reason. It forces a single, auditable governance artifact that satisfies multiple masters at once: one AIMS maps to EU AI Act documentation duties, NIST CSF GOVERN, and customer due-diligence questionnaires. For organizations selling to enterprise, regulated, or public-sector buyers, ISO/IEC 42001 is a strong candidate for the governance structure to build around.

8.2 EU AI Act

The EU AI Act is the first comprehensive, risk-tiered AI law and applies extraterritorially when organizations place AI systems on the EU market or affect EU persons. It classifies systems by risk, with escalating obligations. Even if you are not EU-based, it is becoming a baseline for AI governance documentation, transparency, and risk management. Treat compliance work as reusable governance, not a regional checkbox.

- **Action:** inventory AI systems, classify by EU AI Act risk tier, and document data sources, intended use, and human oversight for anything high-risk.
- **Action:** align this documentation with NIST CSF GOVERN so one governance artifact serves multiple frameworks.

8.3 Building the AI governance program

1. Stand up an AI governance body with security, legal, privacy, and engineering representation.
2. Maintain an AI system inventory covering first-party, embedded, and third-party AI features in SaaS.
3. Publish acceptable-use, secure-development, and incident-response policies that explicitly cover AI.
4. Adopt ISO/IEC 42001 as the AIMS backbone; produce a Statement of Applicability and target certification if you sell to enterprise or public sector.
5. Map every control to NIST CSF 2.0 and report posture to the board on a regular cadence.
6. Re-validate this entire program quarterly. The landscape in this guide will shift.

CISO FIELD GUIDE

9. 90-Day CISO Action Plan

A prioritized starting point. Sequence by your current gaps, not strictly top-to-bottom.

0-30 DAYS Govern

- Publish an AI Acceptable Use Policy naming approved tools and prohibited data classes.
- Stand up an AI governance body; begin an AI system inventory.
- Adopt ISO/IEC 42001 as the AIMS backbone; draft a Statement of Applicability against its Annex A controls.
- Brief the board on the speed-not-shape reframe and the remediation-throughput bottleneck.

30-60 DAYS Protect data

- Sanction an enterprise AI tier with no-training terms, SSO, and audit logging.
- Extend existing SSE/SASE or deploy browser/endpoint DLP with AI-aware classification.
- Block personal-account AI logins on managed devices; route AI traffic through inspection.

60-90 DAYS

Secure development

- Mandate that AI-generated code passes the same review, SAST/SCA, and provenance gates as any code.
- Add dependency-existence and reputation checks to CI; pin and hash-lock dependencies.
- Scope coding-assistant access to approved repositories and workspaces.
- Generate and maintain SBOMs; sign release artifacts; enforce secret scanning.

ONGOING

Defend

- Rewrite remediation SLAs to a two-clock model: mitigate critical, exposed findings within hours; patch as fast as feasible. Stand up an emergency change process.
- Automate SBOM-to-CVE mapping for same-day exposure triage, including statically linked and firmware libraries.
- Shorten patch testing and deployment timelines; ensure the pipeline absorbs larger patch volumes.
- Enforce microsegmentation and zero-trust; instrument detection across the full ATT&CK chain.
- Threat-model your own AI systems with MITRE ATLAS and apply OWASP LLM mitigations.
- Where eligible, turn frontier-model capability inward for defensive vulnerability discovery.

CISO FIELD GUIDE

10. References and Disclaimer

All URLs and figures verified as of 23 June 2026. Frontier-model status, product capabilities, and figures are point-in-time and subject to change.

1. NIST Cybersecurity Framework (CSF) 2.0. nist.gov/cyberframework

2. MITRE ATT&CK. attack.mitre.org

3. MITRE ATLAS. atlas.mitre.org

4. OWASP Top 10 for Large Language Model Applications (2025). owasp.org/www-project-top-10-for-large-language-model-applications

5. ISO/IEC 42001:2023, Information technology, Artificial intelligence, Management system. iso.org/standard/81230.html

6. EU Artificial Intelligence Act. artificialintelligenceact.eu

7. Anthropic, Project Glasswing program overview. anthropic.com/glasswing

8. Anthropic, Project Glasswing: An initial update, 22 May 2026. anthropic.com/research/glasswing-initial-update

9. Cloudflare, Project Glasswing: What Mythos showed us, 18 May 2026. blog.cloudflare.com/cyber-frontier-models

10. Anthropic, Expanding Project Glasswing, 2 June 2026. anthropic.com/news/expanding-project-glasswing

11. Cloudflare, Defend against frontier cyber models. blog.cloudflare.com/frontier-model-defense

12. NVD, CVE-2026-5194, wolfSSL improper certificate validation. nvd.nist.gov/vuln/detail/CVE-2026-5194

Cybersecurity Growth

Executive cybersecurity leadership for companies that cannot afford failure.

WORK WITH CSG

Need executive security leadership for AI, compliance, or security program growth?

Cybersecurity Growth helps security leaders, CEOs, boards, and growth-stage companies translate risk into practical action.

Fractional CISO & Interim Leadership

Executive security leadership without hiring full-time.

AI, Risk & Compliance Advisory

Practical guidance across AI security, SOC 2, ISO, CMMC, and cloud risk.

Board, Customer & Audit Readiness

Clear security narratives for boards, customers, auditors, and investors.

> From AI security strategy to audit readiness, CSG helps turn executive cybersecurity expectations into operating reality.

Book a Call

cybersecuritygrowth.com