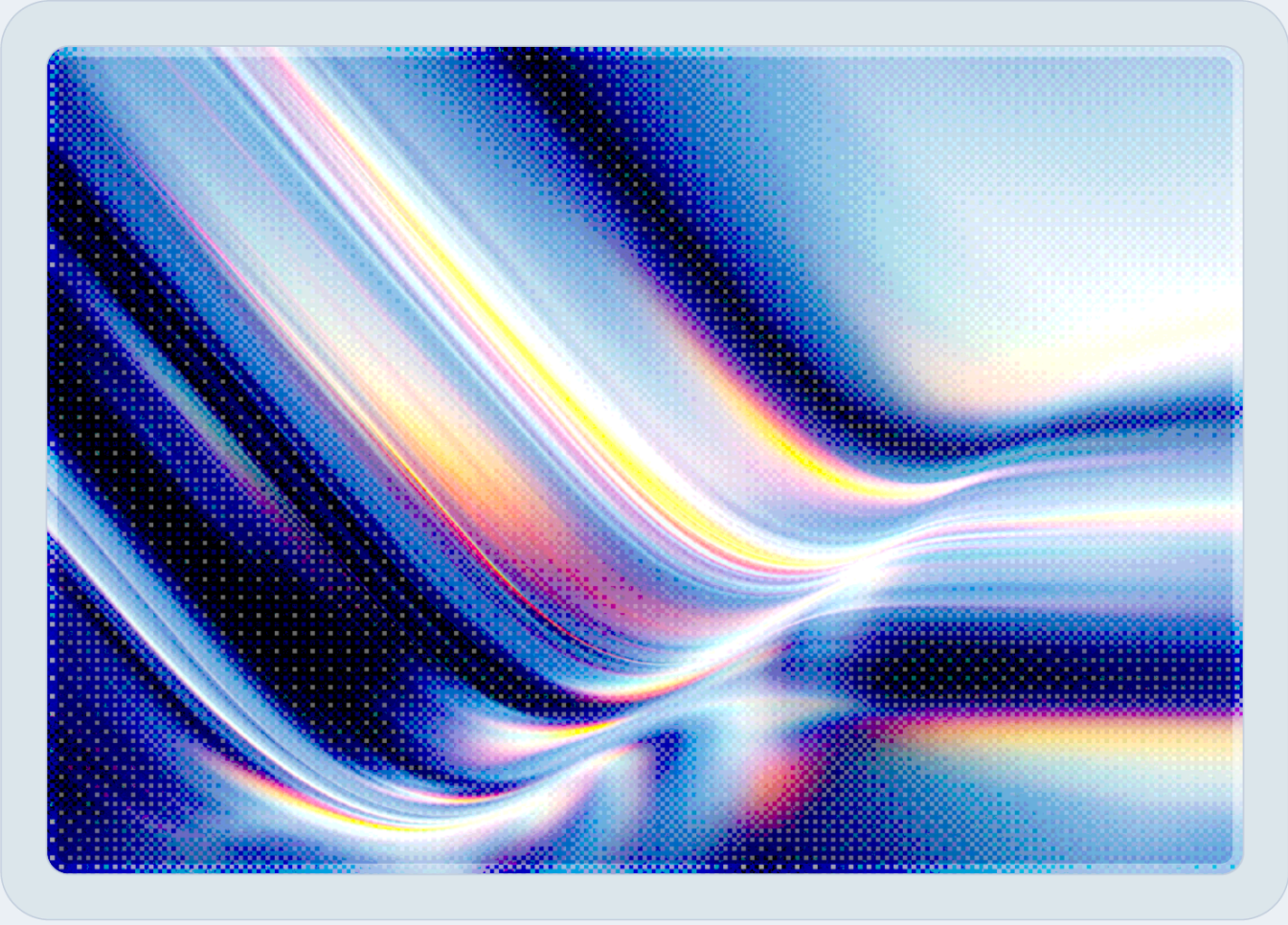
CSAI Foundation | Cloud Security Alliance

Allied Cyber Defense and Vendor Concentration: NATO and Daybreak

NATO, ENISA and the Concentration Question in Frontier Cyber AI

2026-11-22

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

© 2026 Cloud Security Alliance. Some rights reserved.

You may download, store, display, view, print, redistribute, and link to this document in its original, unmodified form, provided that attribution to the Cloud Security Alliance is maintained and all trademark and copyright notices remain intact.

This document may not be modified or altered. You may quote portions of the document as permitted by the Fair Use provisions of the United States Copyright Act, provided that attribution is given to the Cloud Security Alliance.

This document may be shared on professional and social media platforms in its original form with attribution.

This document was generated with AI assistance and has not undergone official CSA review and approval processes.

Key Takeaways

On November 17, 2026, OpenAI announced that it is working with the NATO Communications and Information Agency (NCIA) to give NATO's core institutions access to its cybersecurity models through the Daybreak program. OpenAI also reported that NATO has used "up to 300 billion raw input and output tokens per month" in this work [1]. The announcement did not disclose which models are in use, the commercial terms, where the models are hosted, or what continuity-of-service arrangements exist [1][7]. The same week, ENISA and CERT-EU published a joint advisory showing a different pattern. The EU agencies ran frontier models from two providers, OpenAI through Daybreak and Anthropic through Project Glasswing, inside an isolated environment, using a mix of vendor tools and open-source harnesses [4].

In CSA's view, the concentration question is real but not yet a crisis. As publicly described, NATO's arrangement depends on one vendor's access program, and no other providers have been disclosed. However, NCIA's General Manager has said that usage data will help the agency "evaluate competing solutions" [1]. Organizations should therefore treat this as a dependency that is still taking shape and act before it settles. That urgency follows from the fact that frontier cyber model access is conditional. Daybreak access needs case-by-case approval and is limited to authorized internal use, zero data retention is not on by default, and individual cyber models have scheduled removal dates [3]. Access can be withdrawn or changed by the vendor and, as the June 2026 Anthropic directive showed, by the vendor's home government [9][10].

ENISA's experiment offers the more transferable model, in CSA's assessment. It used two vendors, open-source harnesses, sandboxed execution, and maintainer validation, and it reported a 95% true-positive rate on the 88 findings that code owners assessed [4]. This shows that useful results do not require committing to a single provider, although diversification carries its own costs, including a wider data-handling surface for sensitive defensive telemetry. Enterprises face the same pattern. Critical infrastructure operators and large enterprises adopting Daybreak, Glasswing, or similar programs should design for vendor portability, government-action risk, and correlated blind spots from the start.

Background

Daybreak is OpenAI's Trusted Access for Cyber program, and it offers two tiers. Daybreak Blue gives reduced-refusal access to general-purpose frontier models for verified defensive work. Daybreak Red gives approved business and enterprise organizations access to purpose-trained cyber models such as GPT-5.6-Cyber, plus GPT-6 Cyber with separate approval [2][3]. OpenAI's help documentation says approval "is not automatic." Access is limited to systems the customer owns, operates, or is explicitly authorized to test, and customers may not resell, proxy, or share it with third parties [3]. In August 2026, OpenAI described its access controls as "identity verification, account security, monitoring, approved-use restrictions, and legal attestations," and began requiring hardware security keys for individual accounts [2]. Daybreak is also offered to eligible customers through Amazon Bedrock [11].

The NATO announcement came as NCIA's NATO Edge 26 conference opened in İzmir, Türkiye [5][7]. OpenAI said it is working with NCIA "to provide NATO's core institutions with access to advanced cybersecurity models." It also said Daybreak supported a new training stream at Crossed Swords 2026. That exercise is organized by the NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE) in Tallinn, and the new stream had participants build, deploy, and evaluate AI agents for cyber operations under human oversight [1][5]. Resilience Media reported that the exercise involved 250 participants from 40 NATO and non-NATO states [6]. In the same announcement, OpenAI said it has also given cyber model access to ENISA, France's ANSSI, and national agencies in Germany and Poland, and it described CERT Polska's work finding six vulnerabilities in MikroTik RouterOS [1]. Neither OpenAI nor NATO has disclosed financial terms [1][7]. Defence Matters noted that data-processing terms, infrastructure location, and continuity arrangements "have not been publicly detailed" [7].

The 300-billion-token figure needs careful reading. It is a peak figure ("up to") reported by OpenAI, and it combines input and output tokens [1]. It measures how much the models were used; it does not measure vulnerabilities found or security outcomes. Still, the scale suggests the service is already embedded in the cyber workflows of NATO's core institutions. For comparison, ENISA's whole multi-model code review experiment used roughly 611 million tokens across 27 codebases [4]. NATO's reported monthly figure is several hundred times larger, though the two numbers measure different things: a monthly operational peak versus a one-time experiment total. NCIA General Manager Dylan Browne framed the usage data as a procurement input: "This insight will help us optimise costs, evaluate competing solutions and make informed procurement decisions" [1]. That statement suggests NCIA sees its current posture as an early stage of adoption rather than a final sourcing decision.

ENISA's *Joint Advisory on the Use of Frontier AI Models for Code Review*, published with CERT-EU in November 2026, describes a different approach [4]. Through the providers' structured access programs, ENISA received free access to OpenAI's GPT-5.5 Cyber, GPT-5.6 Cyber, and GPT-6 Astra (via

Daybreak) and to Anthropic's Mythos 5 (via Project Glasswing). GPT-6 Astra is a general-purpose mainline model that retains standard safeguards, distinct from the Red-tier GPT-6 Cyber [3][4]. ENISA ran these models through OpenAI's Codex client and Codex Security plug-in, the open-source OpenCode client, CERT-EU's open-source Vigilo harness, and two ENISA-built agentic skills [4]. The models flagged 220 potential issues across 27 codebases. By the time of publication, code owners had assessed 88 of them and confirmed 95% as true positives, and four CVE IDs were assigned [4]. Because owners may have triaged the most likely issues first, the assessed subset may not be fully representative. ENISA estimated the total cost at about EUR 6,649, noting that this figure is indicative only because the providers gave access free of charge [4].

Security Analysis

Two Postures, One Risk Surface

The NATO and ENISA announcements are not directly comparable. NATO's covers ongoing operational use at alliance scale, while ENISA's describes a bounded experiment on source code. Their architectures nevertheless differ in ways that matter for concentration risk. The table below summarizes what each has disclosed publicly.

Dimension	NATO / NCIA (as announced)	ENISA / CERT-EU (as reported)
Model providers	OpenAI (Daybreak); other providers not disclosed	OpenAI (Daybreak) and Anthropic (Glasswing)
Harness / tooling	Not disclosed	Vendor client and plug-in, plus open-source OpenCode, Vigilo, and ENISA skills
Scale	Up to 300B tokens per month	About 611M tokens total, 27 codebases
Commercial terms	Not disclosed	Free access under provider programs; about EUR 6,649 estimated cost
Hosting and data terms	Not publicly detailed	Isolated, containerized lab; source code only, no live systems

Dimension	NATO / NCIA (as announced)	ENISA / CERT-EU (as reported)
Validation	Human oversight emphasized at Crossed Swords	Maintainer triage; severity re-rating; coordinated disclosure via CERT-EU
Sources	[1][5][6][7]	[4]

The ENISA design keeps the most important assets independent of any one provider: the harness, the skills, the reporting pipeline, and the disclosure process. If one provider's access ended tomorrow, ENISA's tooling would allow it to continue the workflow on the other provider or another supported model, though result quality might differ. Nothing public indicates whether NATO's integration has the same portability. The absence of disclosure is itself the finding, because allied defenders cannot yet judge how hard it would be for NATO to switch providers.

Access Is Conditional, Revocable, and Jurisdictional

In CSA's assessment, frontier cyber models are not commodity utilities; their access terms make them behave more like controlled goods. Daybreak access depends on vendor approval, and the vendor can change model availability on its own schedule. For example, OpenAI has scheduled `gpt-5.5-cyber-preview` for removal from its API on February 1, 2027 [3]. Model lineups within the tiers have also shifted across 2026: OpenAI introduced GPT-5.6-Cyber as the successor to GPT-5.5-Cyber, added GPT-6 Cyber as a newer Red-tier model, and now limits GPT-5.5-Cyber to organizations with existing access [2][3]. Any security operation built on one named model must therefore plan for forced migration.

Government action adds a second layer of conditionality. On June 12, 2026, the U.S. Department of Commerce directed Anthropic to suspend foreign-national access to Fable 5 and Mythos 5, and Anthropic responded by cutting off access to the models for all users at short notice [9][10]. Two weeks later, Commerce eased the restriction for Mythos 5 to named entities listed in an annex, their foreign-national employees, and Anthropic's own foreign-national staff [10]. That episode involved a different vendor, but the mechanism applies to any U.S.-headquartered frontier provider. Many allied users of U.S. cyber models are foreign nationals, including staff at NATO bodies, ENISA, ANSSI, and national CERTs, so their access depends on the policy of a single jurisdiction. Because this dependency affects every U.S. provider at once, a second U.S. vendor reduces commercial and operational concentration but does not reduce jurisdictional concentration.

European Sovereignty Pressure Cuts Both Ways

The EU's proposed Cloud and AI Development Act (CADA) would introduce four assurance levels for cloud services, with the highest level requiring protection from interference by non-EU states [8]. Defence Matters reported in September 2026 that European defence officials are pushing back. They warn that a rushed exit from U.S. technology "could weaken military capability before credible European alternatives are available" and could disrupt NATO interoperability [8]. The officials argue for separating dependencies that must be eliminated from dependencies that must, "for the time being, be managed" [8]. NATO's Daybreak arrangement fits the second category. It appears to be a capability-driven dependency: CSA is not aware of a publicly documented European frontier model with comparable dedicated cyber capability, though NATO has not stated its reasons for selecting Daybreak. The risk-management task is to keep that dependency managed and visible, rather than to deny it exists.

Correlated Blind Spots and Shared Exposure

Concentration also has a technical side. A defensive community that runs the same model over the same open-source code inherits that model's systematic strengths and weaknesses. ENISA's data shows that AI-assigned severities tended to be overrated and that results varied by tooling: code owners kept the severity from the specialized Vigilo harness in 74% of cases, compared with 60% for general-purpose agents with security skills [4]. ENISA did not isolate model-family effects, but if a single model's systematic biases feed alliance-wide prioritization, they could spread without an independent check. Using more than one model and harness, combined with human severity review, offers some protection against this. Agreement between two models is only modest evidence that a finding is correct, but systematic disagreement is a useful signal for human review.

The vendor relationship also creates a sensitive data channel. Daybreak does not grant zero data retention by default, and OpenAI states that Trusted Access and zero data retention "are separate" [3]. Defensive prompts reveal which systems a defender considers critical, what code they run, and which weaknesses they are investigating. In aggregate across allied institutions, this telemetry is a high-value intelligence target, and in CSA's assessment its handling terms deserve the same scrutiny as any classified-adjacent data flow. Separately, Resilience Media reported that California's Attorney General has filed suit against OpenAI, quoting the Attorney General as saying the state is suing over the development of AI models "without safeguards in place" [6]. Regardless of outcome, litigation of this kind is a vendor-risk factor that procurement teams should track, because it can affect product availability, terms, or disclosure obligations.

Diversification is not free, however. Each additional provider extends the data-handling surface for this same sensitive defensive telemetry and adds accreditation and integration overhead, especially in classified environments. ENISA faced fewer of these constraints because its work was a free, bounded experiment on source code rather than operational use. The goal is therefore a portable orchestration layer and a qualified fallback, not necessarily simultaneous production use of several vendors.

Risk Summary

The concentration vectors discussed above, and the patterns that mitigate them, are summarized in the following table.

Concentration vector	Failure scenario	Mitigating pattern
Single model provider	Access revoked, re-tiered, or model retired	Provider-neutral harness; validated second provider
Single jurisdiction	Export directive or national-security order limits foreign-national access	Contingency plan with non-U.S. or open-weight fallback for core workflows
Shared model family	Correlated false negatives and severity bias across allies	Cross-model verification; human severity review
Undisclosed data terms	Defensive telemetry retained or exposed	Negotiated retention, residency, and audit rights
Opaque commercial terms	Free or subsidized access ends; cost shock at scale	Track per-outcome cost; budget for market pricing
Diversification overhead	Additional providers widen telemetry exposure and accreditation burden	Portable orchestration with a qualified fallback rather than parallel production use

Recommendations

Immediate Actions

Organizations already using Daybreak, Glasswing, or a comparable frontier cyber access program should start by inventorying that dependency. The inventory should cover which workflows call which models, which named models are approved, and which approvals are tied to individual accounts versus the organization. Teams should confirm their data retention setting directly, rather than assuming that trusted-access enrollment includes zero data retention [3]. They should also map every scheduled model removal date into change management [3]. Where staff who are foreign nationals relative to the provider's home jurisdiction rely on these tools, the organization should document what happens operationally if a government directive like the June 2026 Anthropic order cuts their access [9][10].

Short-Term Mitigations

Within the next one to two quarters, organizations should separate their orchestration layer from any single model. ENISA's approach shows how: open-source harnesses such as Vigilo and OpenCode, together with portable agentic skills, let the same audit pipeline run against more than one provider [4]. Teams should qualify at least one alternative model on representative workloads and record comparative precision and severity accuracy, as ENISA did, while weighing the added telemetry exposure and accreditation effort each new provider brings. They should run all frontier cyber tooling in isolated, monitored environments with network segmentation and defined trust boundaries [4]. In addition, procurement should require contract terms covering data retention, processing location, audit rights, continuity of service, and notice periods for model retirement. These are the items that remain publicly undisclosed in the NATO arrangement [7].

Strategic Considerations

Over the longer term, allied governments and large enterprises should treat frontier cyber model access as critical supply, with diversification targets, exit testing, and board-level reporting. Browne's framing of usage data as a basis for evaluating "competing solutions" [1] offers a natural checkpoint. Organizations should set dates by which they will have measured a second provider, rather than letting the incumbent's integration deepen by default. Because a second U.S. vendor does not address jurisdictional concentration, European stakeholders should keep supporting the development of credible regional or open-weight alternatives. The CADA debate shows that this will take years and should not be forced at

the expense of current capability [8]. Finally, findings from AI-assisted discovery should flow into shared coordinated disclosure channels, such as CERT-EU's role for Union entities [4], so that the defensive benefit reaches the wider ecosystem rather than staying inside one vendor's customer base.

CSA Resource Alignment

CSA's *The AI Vulnerability Storm* [12] argued in 2026 that frontier models had moved vulnerability discovery to machine speed, and it urged CISOs to build programs that can absorb that shift. The NATO and ENISA announcements show governments acting on the same premise. They also add a dimension the briefing anticipated only in part: the defensive capability itself now comes from a small number of access-controlled programs, so readiness must include resilience to losing that access.

Project Glasswing: AI Discovery Outpaces Open Source Patching [13] documented the gap between AI-driven discovery and maintainer remediation capacity. ENISA's experiment is consistent with that bottleneck persisting in a government setting. All findings were shared with code owners in August 2026, and at publication in November, code owners had assessed 88 of the 220 findings (40%) [4]. Scaling AI-assisted discovery, whether through one provider or several, increases finding volume without addressing that remediation bottleneck. Diversification decisions should therefore be paired with investment in triage and maintainer capacity.

Gold Eagle: The White House's AI Vulnerability Clearinghouse [14] analyzed government-run coordination of AI-generated vulnerability findings and flagged that aggregated finding data becomes a target in its own right. The same concern applies to the defensive telemetry that allied institutions send to model providers, and it supports the data-terms recommendations above.

The *AI Controls Matrix (AICM) v1.1* [15] provides the control vocabulary for putting these recommendations into practice. Its Supply Chain Management, Transparency and Accountability domain covers provider due diligence and contractual terms. Its Business Continuity Management and Operational Resilience domain covers exit planning and provider failover. Its Data Security and Privacy Lifecycle Management controls cover the retention and residency questions raised by Daybreak's default settings.

References

[1] OpenAI. "Strengthening NATO and allied cyber defenses with AI." OpenAI, November 17, 2026.

[2] OpenAI. "Expanding Daybreak as the Cyber Defense Window Narrows." OpenAI, August 10, 2026.

[3] OpenAI. "OpenAI Daybreak - Trusted Access for Cyber overview." OpenAI Help Center, accessed November 22, 2026.

[4] ENISA and CERT-EU. "Joint Advisory on the Use of Frontier AI Models for Code Review." European Union Agency for Cybersecurity, November 2026.

[5] George Allison. "NATO using OpenAI models to defend allied networks." UK Defence Journal, November 17, 2026.

[6] Ingrid Lunden. "OpenAI, facing a cybersecurity lawsuit in the US, is working with NATO... on cybersecurity." Resilience Media, November 17, 2026.

[7] Defencematters.eu Correspondents. "NATO Adopts OpenAI Technology as Artificial Intelligence Transforms Military Cyber Defence." Defence Matters, November 18, 2026.

[8] Defencematters.eu Correspondents. "European defence officials challenge EU curbs on US cloud dependence." Defence Matters, September 4, 2026.

[9] Al Jazeera Staff and Reuters. "US orders Anthropic to disable AI models for all foreign nationals." Al Jazeera, June 13, 2026.

[10] Andrew W. Reddie. "Commerce Eased Its Block on Anthropic's Mythos, But Major Questions Remain." Tech Policy Press, June 27, 2026.

[11] AWS. "Accelerate cyber defense with OpenAI and AWS: Daybreak Red & Daybreak Blue now available to eligible customers on Amazon Bedrock." AWS Machine Learning Blog, 2026.

[12] Cloud Security Alliance. "The AI Vulnerability Storm." CSA, 2026.

[13] Cloud Security Alliance. "Project Glasswing: AI Discovery Outpaces Open Source Patching." CSA, June 2026.

[14] Cloud Security Alliance. "Gold Eagle: The White House's AI Vulnerability Clearinghouse." CSA, July 2026.

[15] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." CSA, 2026.