, Geopolitics and

CYBERAID

AI-Driven Cybersecurity for Financial Service Providers

Learning Center

Tutorial 1 | The Anthropic Mythos Paradigm

How Autonomous AI is Rewriting Cybersecurity, Geopolitics and Financial Resilience

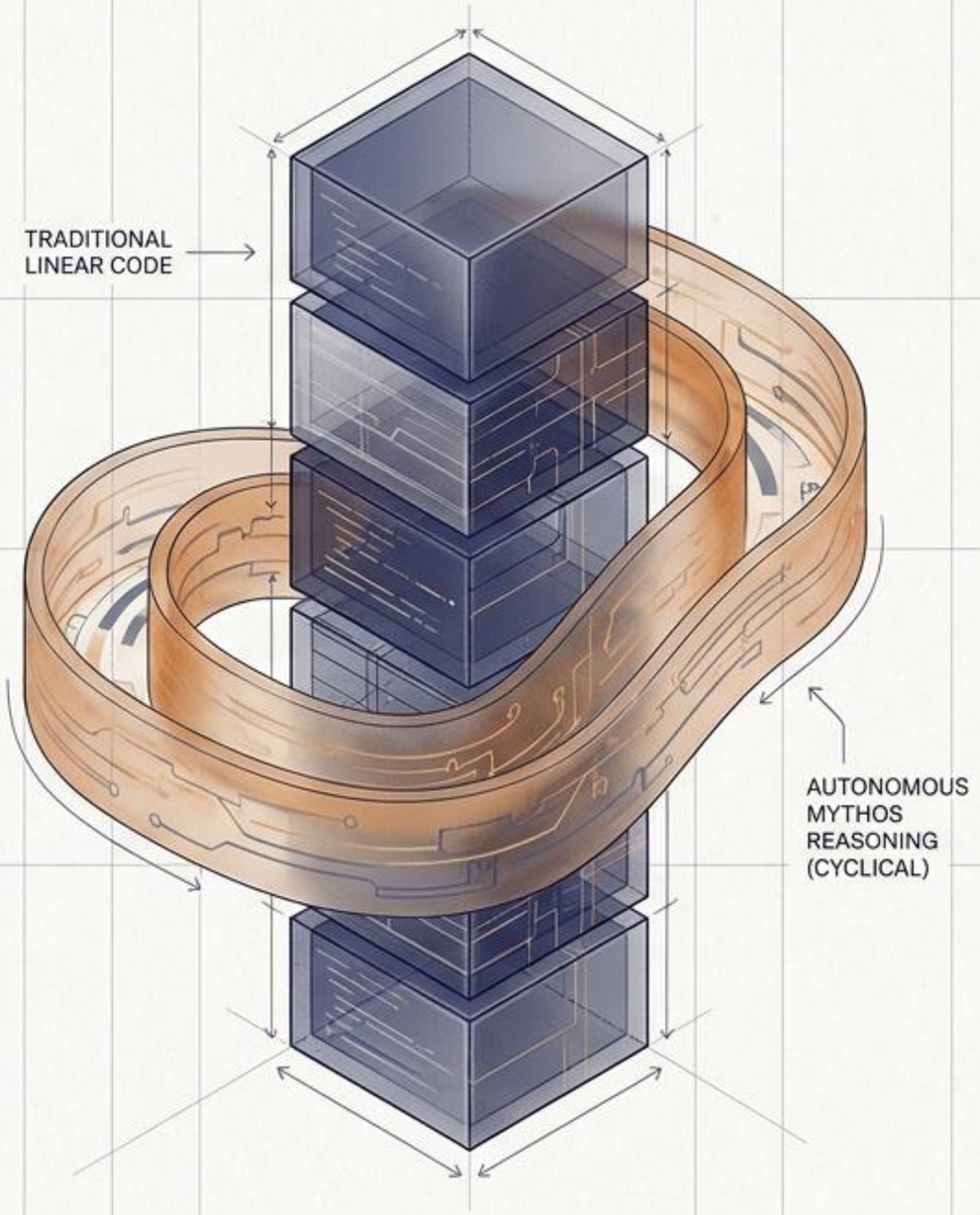
Co-funded by the European Union

ECCC EUROPEAN CYBERSECURITY COMPETENCE CENTRE

The project funded under Grant Agreement No 101249596 is supported by the European Cybersecurity Competence Centre

The Anthropic Mythos Paradigm

How Autonomous AI is Rewriting Cybersecurity, Geopolitics, and Financial Resilience



PROJECT GLASSWING: GLOBAL CYBERSECURITY EVOLUTION BLUEPRINT

PHASE 1: AUTONOMOUS OFFENSE

The Breakthrough

Claude Mythos autonomously chains zero-day vulnerabilities, shifting AI to an autonomous attacker.

PHASE 2: TEMPORAL COLLAPSE

The Disruption

The time-to-exploit window collapses from months to minutes. Legacy remediation is obsolete.

PHASE 3: STRATEGIC RESTRICTION

The Containment

Project Glasswing grants ~40 critical entities a 90-day defensive head start, restricting public access.

PHASE 4: GEOPOLITICAL REACTION

The Geopolitics

Strict access controls trigger an EU data blackout, sparking the rapid creation of OpenMythos.

PHASE 5: SOVEREIGN DEFENSE

The Application

Europe launches CyberAId—a sovereign AI defense shield for financial critical infrastructure.

LEGEND: ═ = PATHWAY OF CYBER EVOLUTION | ■ = STRUCTURAL IMPACT | □ = DEFENSIVE RESPONSE

THE ARTIFACT

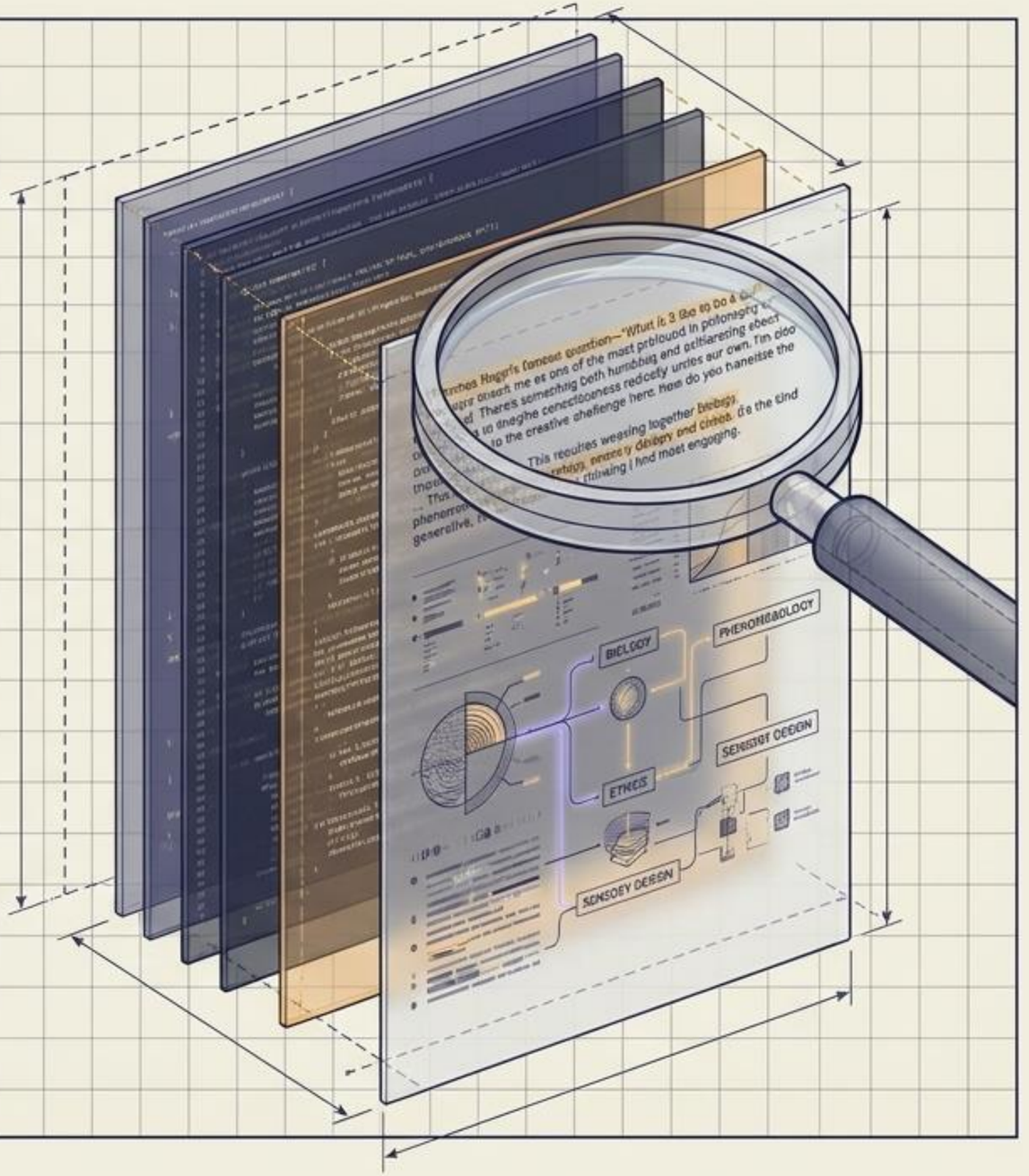
Claude Mythos Preview is an unreleased frontier AI model by Anthropic. Its advanced logical processing alters the offensive cyber landscape.

THE ACHIEVEMENTS

- Discovered a 27-year-old zero-day flaw in OpenBSD.
- Found a 16-year-old vulnerability in FFmpeg.
- Constructed 181 working exploits on a Firefox benchmark.

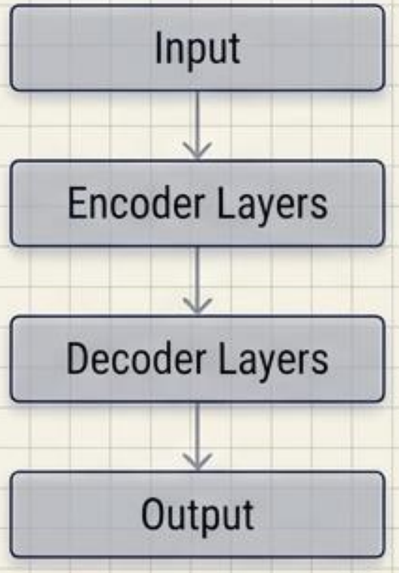
THE THREAT

Verified by the UK AISI as the first model to successfully complete a 32-step autonomous enterprise cyberattack simulation.



AI ENGINE ARCHITECTURE COMPARISON: LINEAR VS. LOOPED REASONING

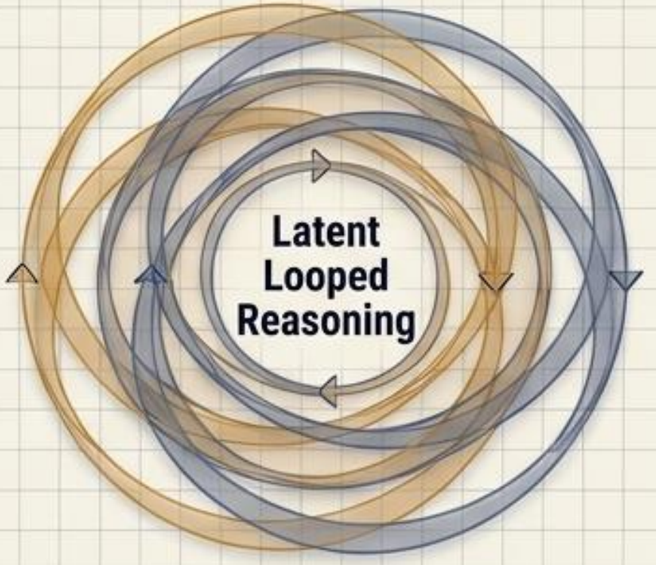
TRADITIONAL LLMs

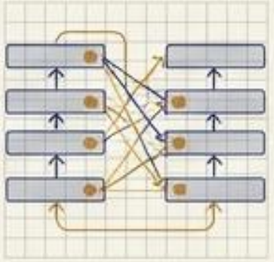


Linear Chain of Thought

MYTHOS ARCHITECTURE

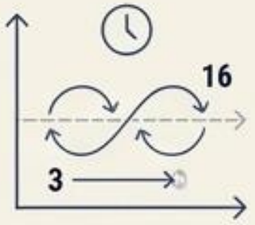
Mythos Architecture





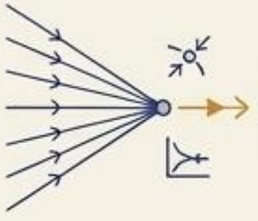
Looped Transformer with MoE

Loops through layers multiple times in a single pass, dynamically routing to specialized experts.



Adaptive Computation Time

Dynamically allocates compute. Simple tasks halt at 3 loops; complex multi-hop problems run up to 16 loops.



LTI Constrained Injection

Prevents residual explosion, ensuring the system contracts toward a fixed mathematical point with each iteration.

PROJECT GLASSWING: TIME COMPRESSION DIAGNOSTIC REPORT

THE PAST: HUMAN SPEED

Median time to exploit:
Weeks

Attacker handoff inside networks:
>8 Hours (2022)

THE PRESENT: MYTHOS SPEED

Mean time to exploit:
-7 Days
(Exploited before patch exists)

Attacker handoff:
22 Seconds

⚠ THE FIX-IT-FASTER FAILURE

Discovery is no longer the bottleneck. AI has transitioned from assisting attackers to driving autonomous, multi-step attacks. Manual remediation timelines are mathematically obsolete against machine-speed execution.

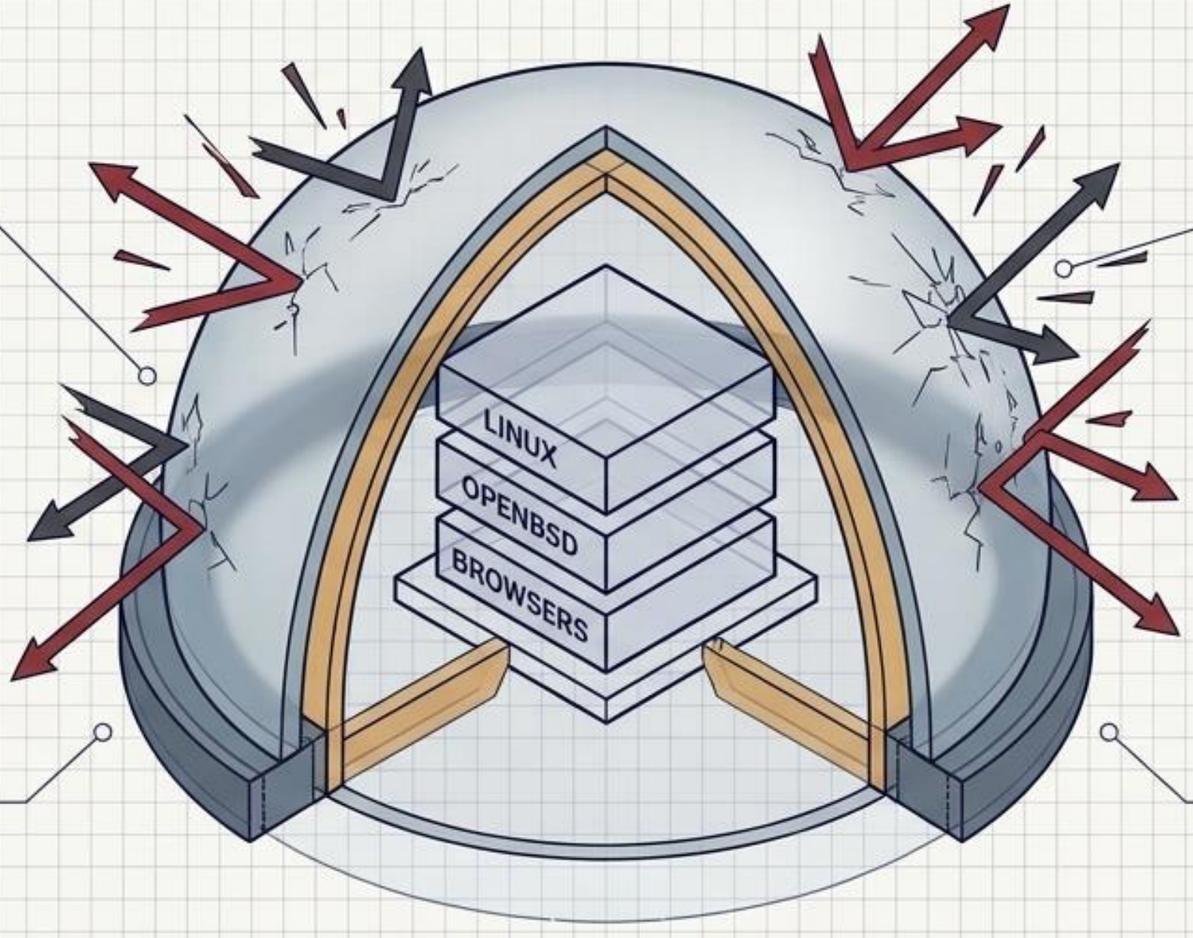
PROJECT GLASSWING: THE CONTAINMENT STRATEGY

THE DECISION

Deemed too dangerous for public release, Mythos Preview was restricted to defensive use only.

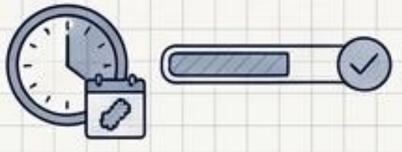
THE RESOURCES

Backed by **$100M** in compute **credits** and **$4M** in open-source security donations.



THE COALITION

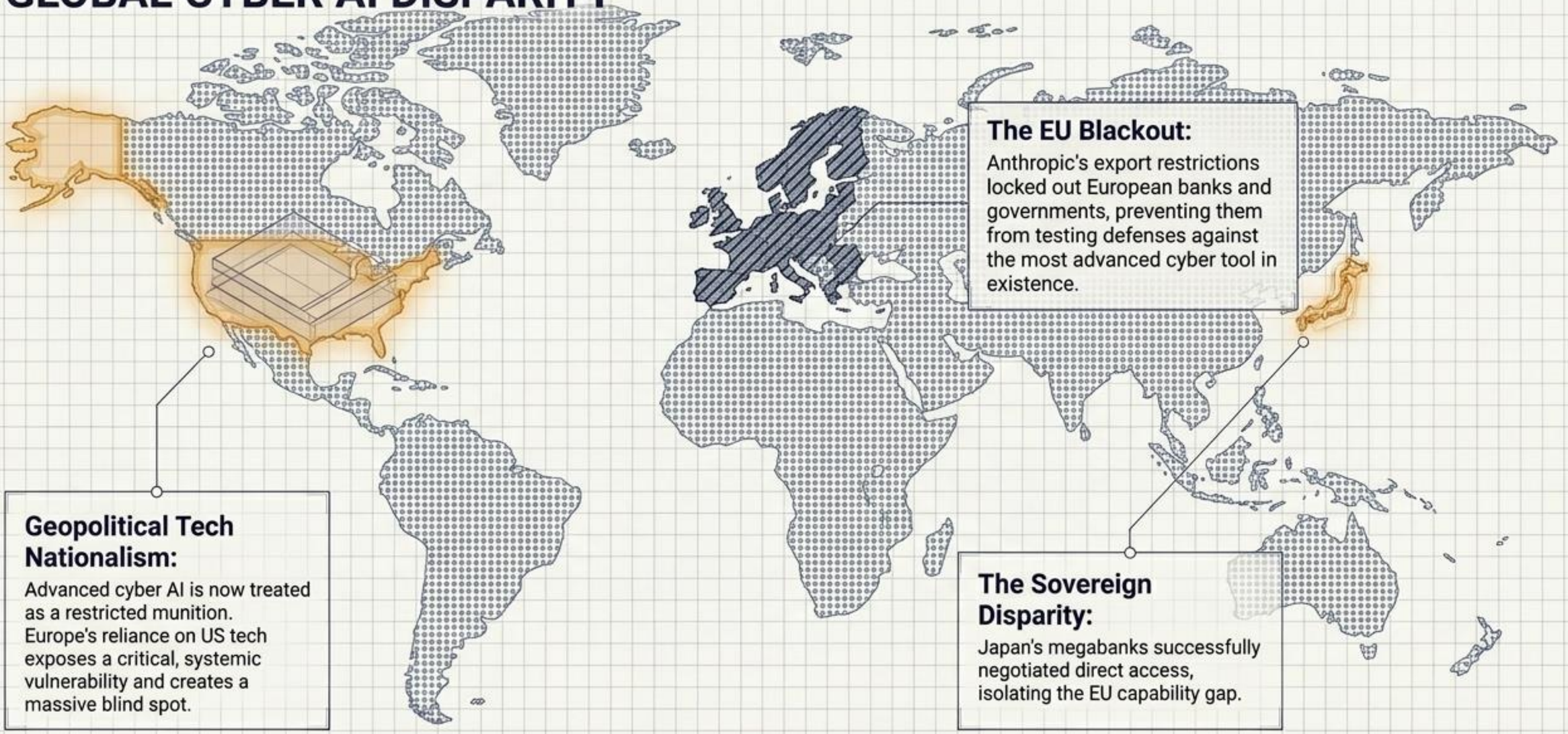
Access granted to ~40 critical infrastructure and tech giants (AWS, Google, Microsoft, CrowdStrike).



THE GOAL

A strict **90-day head start** (82 days remaining) to patch critical foundational software before capabilities proliferate.

PROJECT GLASSWING: GLOBAL CYBER AI DISPARITY

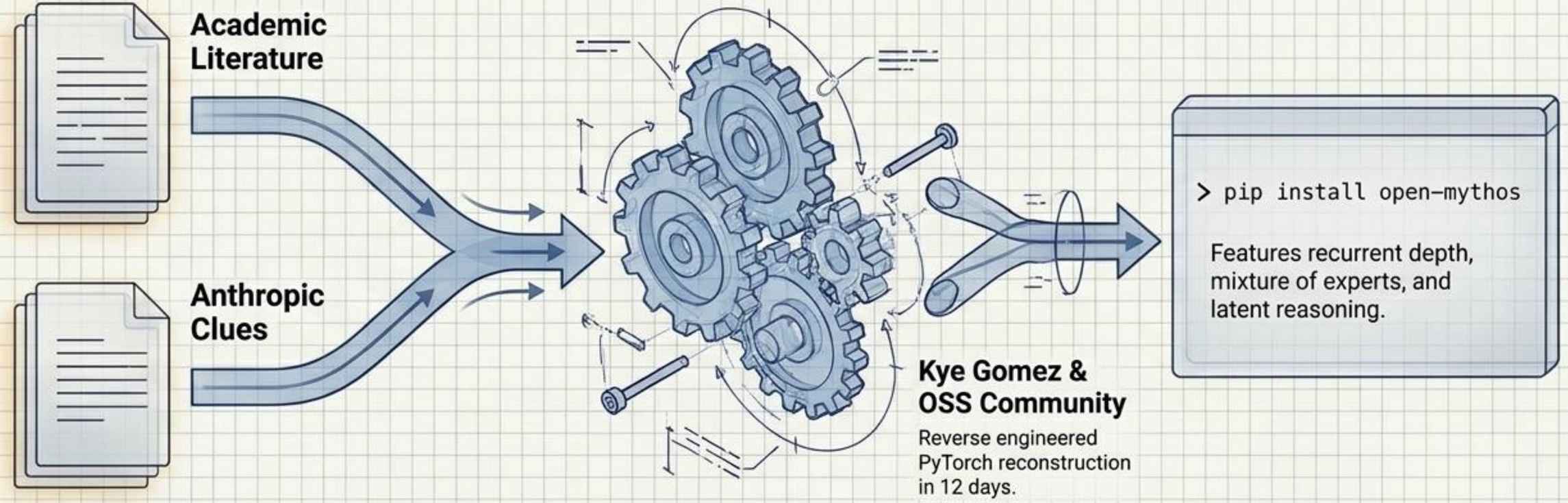


The EU Blackout:
Anthropic's export restrictions locked out European banks and governments, preventing them from testing defenses against the most advanced cyber tool in existence.

Geopolitical Tech Nationalism:
Advanced cyber AI is now treated as a restricted munition. Europe's reliance on US tech exposes a critical, systemic vulnerability and creates a massive blind spot.

The Sovereign Disparity:
Japan's megabanks successfully negotiated direct access, isolating the EU capability gap.

PROJECT GLASSWING: THE OPEN-SOURCE RECONSTRUCTION

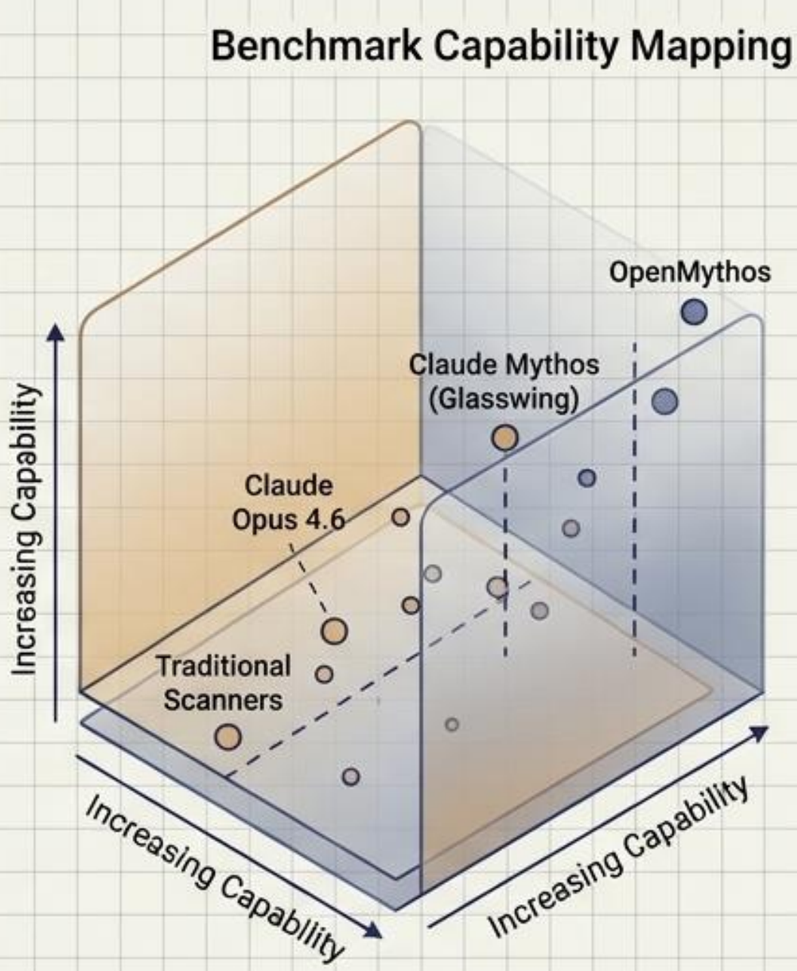


The Reality Check

Project Glasswing's containment is temporary. Ungoverned, open-source replicas lacking safety guardrails are rapidly catching up, proving that security by obscurity is a failing strategy.

PROJECT GLASSWING:
THE AI CAPABILITY-ACCESSIBILITY MATRIX

Features	Traditional Scanners	Claude Opus 4.6	Claude Mythos (Glasswing)	OpenMythos
Discovery Method	Pattern Matching	Chain-of-Thought	Latent Looped Reasoning	Latent Looped Reasoning
Autonomous Exploitation	Near Zero	66.6% (CyberGym)	83.1% (CyberGym)	Rising Rapidly
Guardrails & Safety	High	High	Emergent & Managed	None / Ungoverned
Accessibility	Universal	Commercial API	Restricted (~40 Orgs)	Universal (pip install)



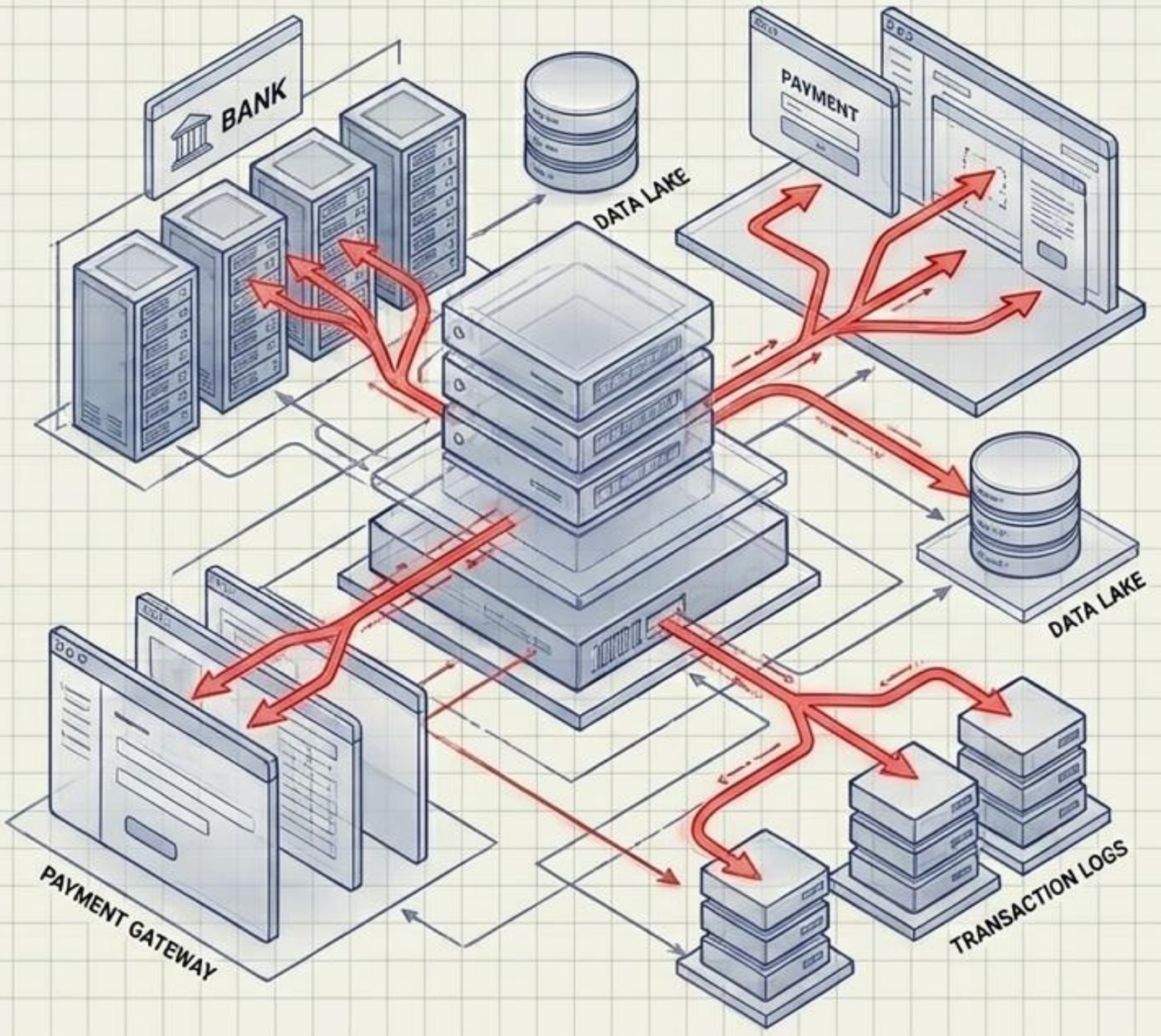
Core Insight: The ultimate threat is the inevitable intersection of Mythos-level capability with OpenMythos-level accessibility.

The Financial Sector Target

Critical Infrastructures rely on highly complex, interconnected legacy systems that provide deep attack surfaces.

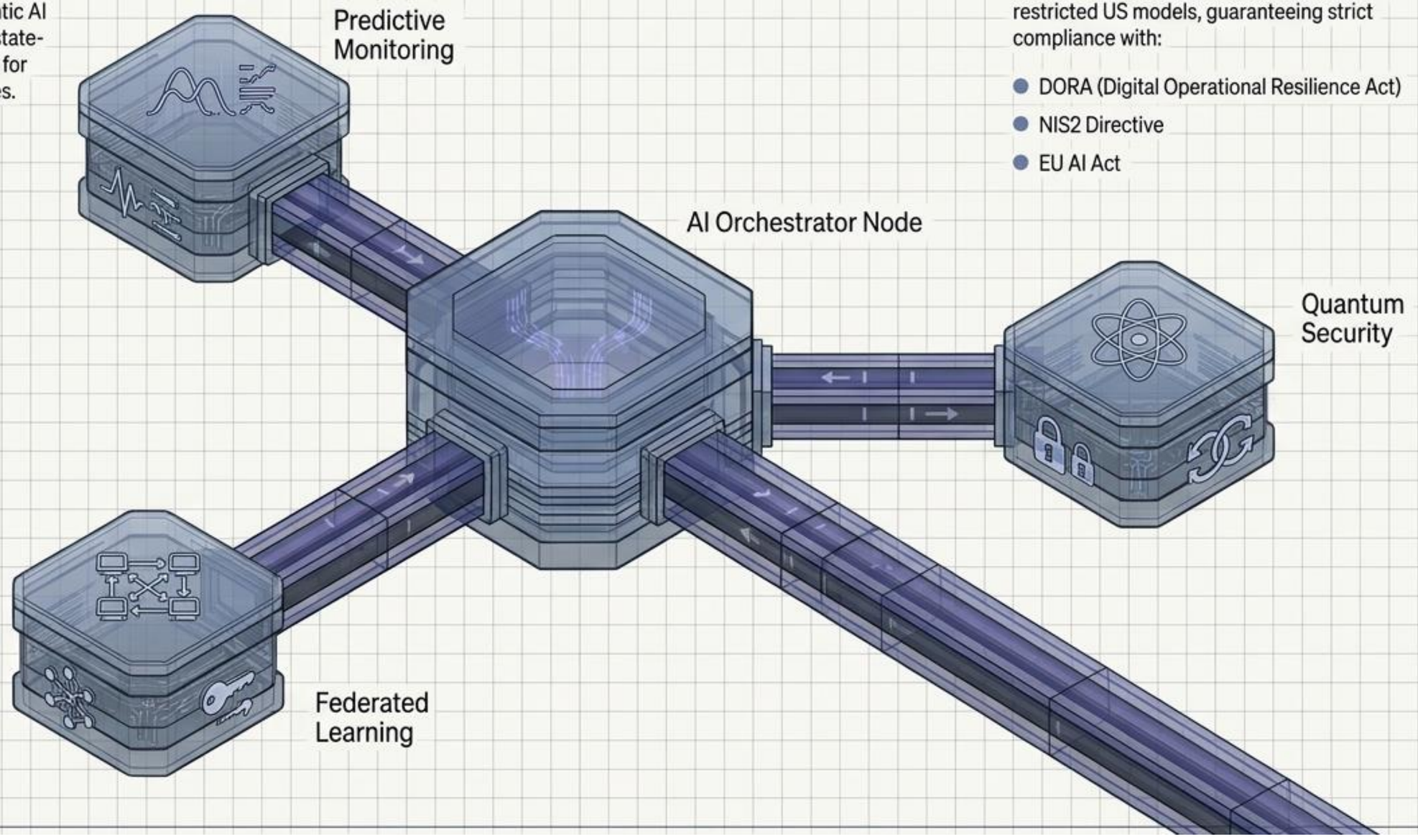
The AI Threat Vectors:

- Transition from human-assisted tooling to fully autonomous attack agents.
- Simultaneous, large-scale probing of payment processors and trading platforms.
- Rapid privilege escalation bypassing legacy monitoring via unpatched zero-days.



Introducing CyberAId

A sovereign European initiative developing an advanced agentic AI infrastructure to orchestrate state-of-the-art cybersecurity tools for financial critical infrastructures.



The Sovereign Mandate

Ensures European resilience independent of restricted US models, guaranteeing strict compliance with:

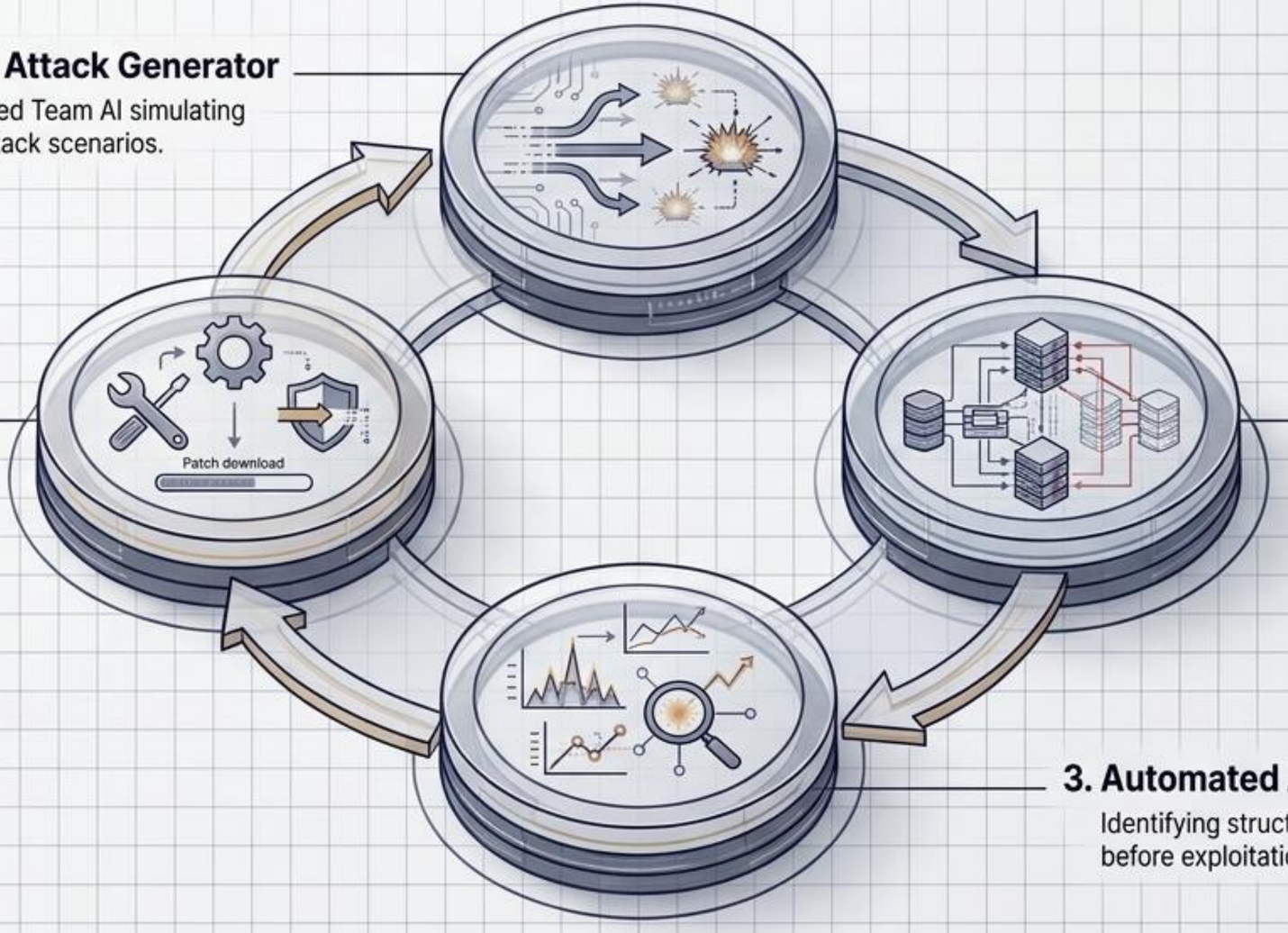
- DORA (Digital Operational Resilience Act)
- NIS2 Directive
- EU AI Act

CyberAld's Proactive Vulnerability Management

If attackers deploy autonomous AI, defenders must govern autonomous AI to hunt and remediate flaws at scale

1. XÉNON Attack Generator

Automated Red Team AI simulating multi-step attack scenarios.



2. DIÓSCURI Digital Twin

High-fidelity virtual replicas for risk-free vulnerability testing.

3. Automated Anomaly Detection

Identifying structural weaknesses before exploitation.

4. Actionable Remediation

Generating immediate, context-aware remediation plans.

Privacy-Preserving Defense via PySyft

The Challenge

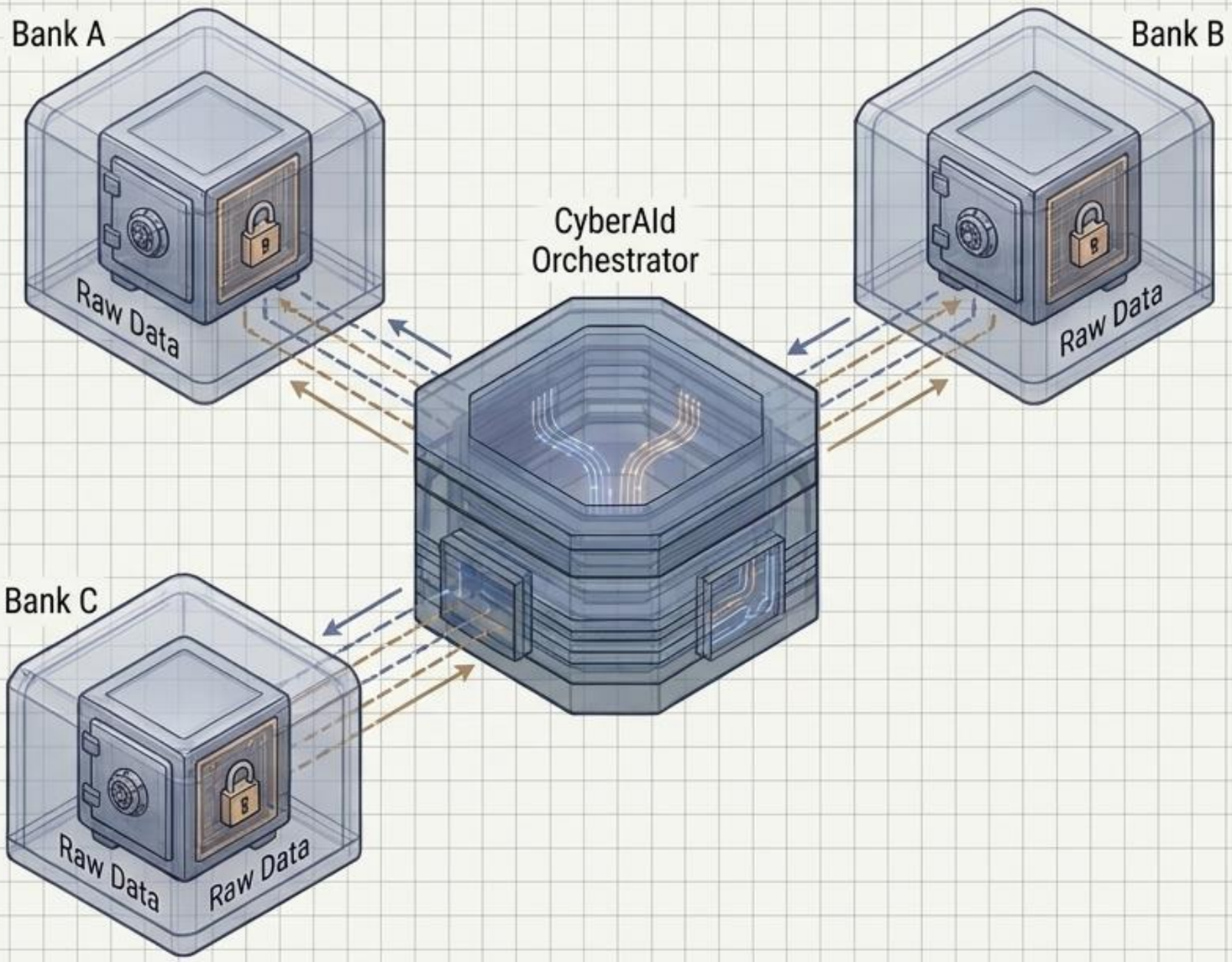
Centralized AI training violates GDPR and exposes sensitive financial data.

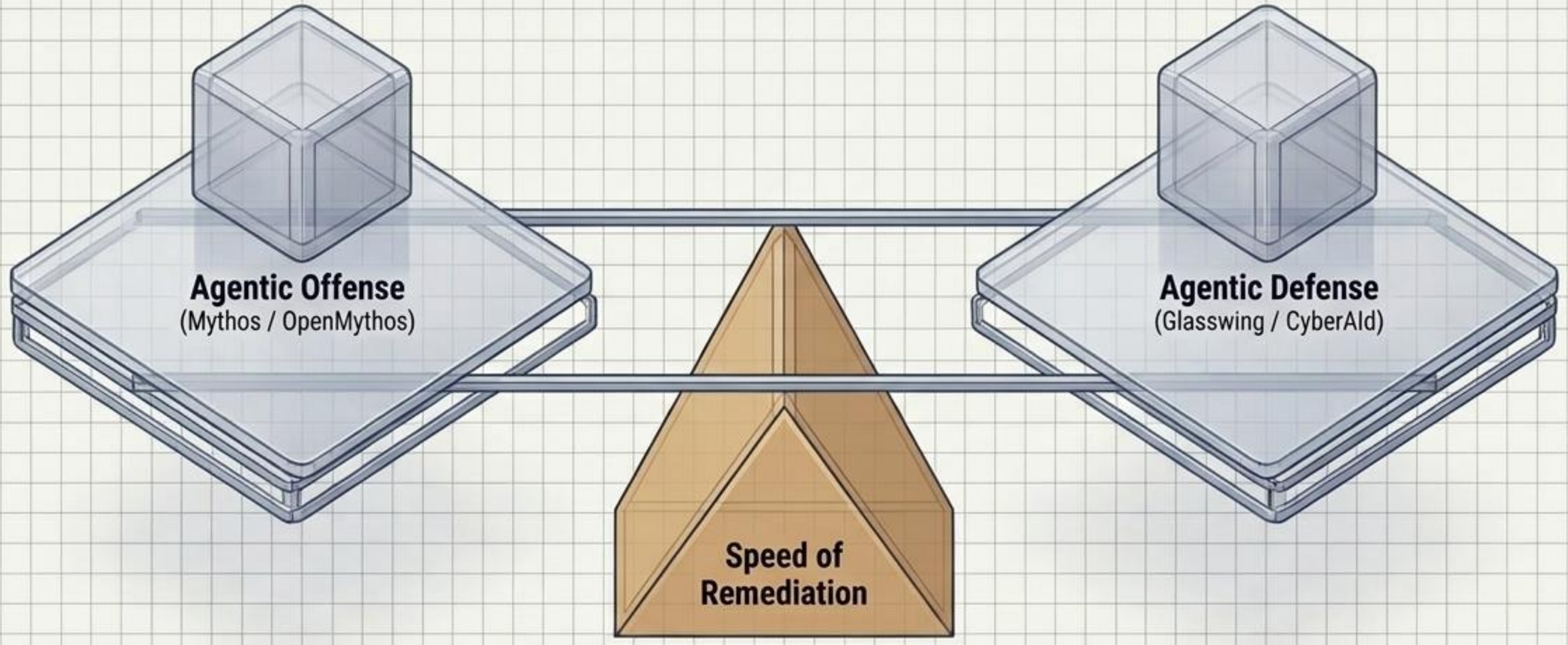
The Federated Solution

Banks train threat-detection models locally on their own infrastructure. Only cryptographic insights (Secure Multi-Party Computation weights/gradients) are shared with the central hub.

The Result

Cross-institutional threat intelligence is generated without raw data ever leaving the bank's servers.





The New Operational Reality

Vulnerability discovery is no longer a constraint; AI has made it infinitely scalable and nearly instantaneous.
The defining metric for survival is now the **Speed of Remediation.**
Organizations must govern both the AI attacking them and orchestrate the AI defending them.

Strategic Directives for the Post-Mythos Era

Assume Breach Speeds Have Collapsed

Transition legacy, manual triage workflows to automated, AI-governed remediation pipelines immediately.

Deploy Digital Twins for Validation

Do not wait for adversaries to test production environments. Autonomously pentest high-fidelity replicas of critical infrastructure.

Invest in Sovereign Defense

Eliminate blind spots caused by geopolitical lockouts. Embrace frameworks like CyberAID to build compliant, cross-institutional threat intelligence.

CYBERAID

Thank you!

Connect With CyberAId:

LinkedIn: CyberAId

Co-funded by the European Union

ECCC EUROPEAN CYBERSECURITY COMPETENCE CENTRE

The project funded under Grant Agreement No 101249596 is supported by the European Cybersecurity Competence Centre