Mythos-Class AI and Cybersecurity Law:

Vulnerability Amplification, Legal Fragmentation, and the Governance Gap

Mona Al Achkar Jabbour

Professor of Law

INTRODUCTION

A new generation of frontier artificial intelligence models, commonly designated Mythos-class systems, is fundamentally reshaping the cybersecurity landscape. Their defensive potential is remarkable; so, too, is their capacity to accelerate and amplify offensive operations at a scale that existing law was not designed to govern. The question is no longer whether such systems will transform risk landscapes. The question is whether legal and governance frameworks can evolve quickly enough to keep pace.

Mythos-class AI introduces what may be termed vulnerability amplification: the exponential intensification of pre-existing infrastructural and regulatory weaknesses through autonomous and scalable AI capabilities. Early security assessments are striking. In under three weeks, one such system accomplished the equivalent of a full year of penetration-testing effort. Starting from only a CVE identifier and a git commit hash, it completed a working exploit chain in under a day at a cost below $2,000 , a timeline that historically required skilled researchers days to weeks. Beyond identifying individual weaknesses, these systems excel at vulnerability chaining, combining multiple lower-severity issues into critical-level exploit paths, and at mapping systemic weaknesses across telecom, financial, and public sector networks.

This shift destabilises the foundational assumptions embedded in cybersecurity law. Traditional legal frameworks, whether grounded in criminal law (cybercrime conventions), regulatory compliance (data protection law), or emerging AI governance regimes rest on three core premises: identifiable human agency, traceable causation, and territorially bounded enforcement. Mythos-class AI disrupts each simultaneously. Agency becomes distributed across developers, deployers, and autonomous system behaviour. Causation becomes probabilistic and non-linear, complicating attribution. Enforcement becomes structurally limited in cross-border AI deployment environments.

Existing frameworks, including the EU Artificial Intelligence Act (EU AI Act), the General Data Protection Regulation (GDPR), and international cybercrime instruments, are not designed to regulate systems that operate simultaneously as defensive infrastructure and offensive capability. The following sections examine each regime in turn, before assessing the broader governance architecture and the structural misalignments it produces.

LEGAL ARCHITECTURE: FRAGMENTATION AND MISALIGNMENT

The EU Artificial Intelligence Act (Regulation (EU) 2024/1689) establishes the most comprehensive binding framework for AI governance currently in force. Its risk-based classification system differentiates between prohibited, high-risk, limited-risk, and minimal-risk AI systems. Under Articles 9–15, high-risk AI systems must comply with obligations including risk management systems, data governance requirements, technical documentation, human oversight provisions, and robustness and cybersecurity safeguards.

Despite this breadth, the Act's provisions on robustness and cybersecurity, notably Article 15, remain conceptually underdeveloped when applied to adaptive and autonomous systems capable of evolving threat behaviour. The Act assumes identifiable system boundaries, stable risk profiles, and predictable system outputs. These assumptions are incompatible with Mythos-class AI, where system capabilities evolve post-deployment, risk is emergent rather than predefined, and cybersecurity functionality is itself dual-use. By August 2025, obligations on providers of particularly powerful AI models classified as presenting systemic risk had taken effect, requiring reporting to the European Commission, structured evaluation and testing, permanent documentation of security incidents, and heightened cybersecurity requirements . By April 2026, EU AI Act compliance had become a critical variable in corporate due diligence, sitting alongside GDPR readiness and cybersecurity maturity as a standard component of transaction scope . Yet the fundamental conceptual gap remains unresolved, the Act regulates AI risks not AI-generated cyber power.

The General Data Protection Regulation provides a comprehensive framework governing data protection obligations, including data minimization (Article 5), security of processing (Article 32), and accountability (Article 24). Its regulatory logic is, however, fundamentally data-centric rather than capability-centric. It regulates how data is processed and how individuals are protected, but it does not regulate how AI systems generate vulnerabilities or how cyber capabilities are operationalized.

As a result, the GDPR cannot adequately address scenarios in which no personal data breach occurs yet systemic infrastructure vulnerabilities are exploited or exposed. Mythos-class AI is precisely capable of generating such scenarios: a system that maps telecom network weaknesses or simulates strategic attack scenarios may produce no data breach within GDPR's scope while creating conditions for catastrophic infrastructure compromise. The GDPR's data-centrism leaves this entire category of risk unaddressed.

The Budapest Convention on Cybercrime (2001) remains the primary international treaty governing cybercrime. It establishes offences including illegal access, system interference, and

misuse of devices, and provides mechanisms for international cooperation. Its architecture is, however, fundamentally actor-centric: crimes are committed by persons, liability is tied to intent, and enforcement depends on attribution.

The Convention was drafted before the exponential growth in internet usage, the rise of cloud computing, and the digitalization of virtually every form of commercial and social interaction, developments that have transformed the scale and nature of cybercrime far beyond what its drafters anticipated. Mythos-class AI challenges the Convention's model by introducing non-human operational agency, automated exploit generation, and distributed responsibility chains. Its doctrinal structure cannot accommodate AI systems that operate without direct human instruction at the point of harm, and that generate exploitable vulnerabilities without any identifiable individual authoring a criminal act.

In recognition of these limitations, the UN General Assembly adopted a new Convention Against Cybercrime on 24 December 2024 , which was opened for signature in October 2025 and will enter into force upon ratification by forty parties. The Convention attempts to modernize international cyber governance by addressing transnational cyber threats and strengthening cooperation mechanisms.

However, like its predecessor frameworks, the UN Convention remains anchored in criminalization logic, post hoc enforcement, and state-centric cooperation. It does not address pre-emptive regulation of AI capabilities, dual-use classification of AI systems, or systemic vulnerability generation. Existing instruments were not designed for automated vulnerability discovery ecosystems driven by frontier AI, and the 2024 Convention does not fill this gap.

GOVERNANCE LAYER: CONVERGENCE WITHOUT COHERENCE

Beyond binding law, AI governance is increasingly shaped by overlapping institutional frameworks that supply normative guidance without enforcement authority.

The NIST AI Risk Management Framework provides an operational structure built around four functions: Govern, Map, Measure, and Manage. It translates abstract principles into actionable risk management processes and has become influential beyond the United States. Nevertheless, the framework structurally fails for three reasons: it lacks mandatory enforcement, it is strictly voluntary, and it presumes a level of control that Mythos-class AI doesn’t allow. Because Mythos-class systems evolve after they are deployed, they inherently resist centralized organizational oversight.3.2 OECD Principles on Artificial Intelligence

On the other hand, the OECD Principles on Artificial Intelligence establish foundational norms including transparency, accountability, and human-centred values. These principles have shaped

global governance discourse, but they remain normative rather than binding, contributing to what scholars have described as 'soft law proliferation without enforcement convergence.' Their influence operates at the level of discourse; their capacity to constrain AI-generated cyber power is negligible.

The European Union Agency for Cybersecurity (ENISA) has developed frameworks aligning AI governance with cybersecurity practices , emphasizing lifecycle risk management, incident response, and system resilience. These frameworks complement the EU AI Act but do not resolve the core analytical problem: AI systems themselves have become cybersecurity actors. Regulating the infrastructure within which they operate does not regulate what they do.

At the international level, export control regimes are beginning to engage with AI's dual-use character. At its December 2024 plenary, the Wassenaar Arrangement adopted updated controls on intrusion software and IP-network surveillance systems while noting continued national-level action on emerging technologies . AI's inherent dual-use nature is now widely seen as demanding updated control frameworks that balance national security interests with global innovation objectives. High-capability cybersecurity AI systems may constitute the next category brought within export control scope, though the pace of regulatory adaptation has so far lagged behind the pace of capability development.

THE STRUCTURAL ASYMMETRY AND ITS CONSEQUENCES

A core asymmetry between attackers and defenders gives adversaries a disproportionate advantage when leveraging frontier AI. Attackers need only one successful exploit; defenders must protect against every attack vector. In AI-powered threat detection, both false positives and false negatives carry serious costs. Banks, telecom operators, hospitals, and education institutions are particularly exposed through the automated mapping of systemic risks that Mythos-class AI enables.

This asymmetry is compounded by the speed differential in the vulnerability lifecycle. Attackers weaponize vulnerabilities in an average of five days; organisations typically take sixty to one hundred and fifty days to deploy patches. AI-accelerated vulnerability discovery widens this gap further, compressing the window within which defenders have any operational advantage .

The consequences are acute for jurisdictions with developing cybersecurity regulatory architectures. Where baseline cybersecurity is weak, AI-enhanced attack tools deliver disproportionate damage, rapidly reducing whatever defensive advantage institutions may hold.

The most effective responses in such contexts are not novel or experimental but strong foundational security measures, that presume regulatory frameworks capable of directing and verifying them.

Across the governance regimes we have discussed, a consistent pattern emerges: regulatory convergence at the level of principles, but fragmentation at the level of operational control over AI-driven cyber capabilities. The law regulates AI risks. It does not regulate AI-generated cyber power.

5. CONCLUSION

The governance gap produced by Mythos-class AI is not simply a gap in coverage, a matter of adding new rules to existing frameworks. It reflects three structural misalignments that cut across every regime examined.

First, the agency problem. Every existing framework presupposes an identifiable human author of harm, a developer, deployer, or malicious actor whose intent can be assessed and whose conduct can be attributed. Mythos-class AI distributes agency across multiple parties and introduces autonomous operational behaviour at the point of harm. No existing framework has resolved how responsibility is allocated when harm is produced by a system that no single party directed to cause it.

Second, the causation problem. Attribution in cybersecurity law depends on traceable causal chains linking conduct to harm. AI-generated vulnerabilities are emergent, probabilistic, and frequently non-linear. A system that maps infrastructure weaknesses does not cause a breach in any direct sense; it creates conditions for one. Current frameworks have no conceptual vocabulary for this form of probabilistic, capability-mediated causation.

Third, the enforcement problem. Cross-border AI deployment is the norm, not the exception. Territorial enforcement mechanisms, the default architecture of both criminal and regulatory law, are structurally inadequate for systems that operate without regard to jurisdictional boundaries. The UN Convention's state-centric cooperation model and the EU AI Act's provider-focused obligations both assume a degree of territorial anchoring that Mythos-class AI does not provide.

The question confronting legal scholars, policymakers, and regulators is not whether Mythos-class AI will transform the risk landscape. It already has. The question is whether legal and governance frameworks can be rebuilt around the correct foundational assumptions, distributable agency, probabilistic causation, and extraterritorial reach, before the governance gap becomes permanent.