Hypothesis

Mythos-Class Frontier Models as System-Level Disruptors in Post-Quantum Cryptography Migration: A Systems-Engineering Analysis of Lifecycle, Architecture, Dependencies, and Operational Modeling

Robert Campbell *

Posted Date: 24 April 2026

doi: 10.20944/preprints202604.1744.v1

Keywords: Post-Quantum Cryptography; PQC migration; Claude Mythos; Project Glasswing; frontier models; systems engineering; NIST PQC; CNSA 2.0; AI-accelerated cybersecurity

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 April 2026 doi:10.20944/preprints202604.1744.v1

Hypothesis

Mythos-Class Frontier Models as System-Level Disruptors in Post-Quantum Cryptography Migration: A Systems-Engineering Analysis of Lifecycle, Architecture, Dependencies, and Operational Modeling

Rob Campbell

Independent Researcher; Fellow, British Blockchain Association, Upper Marlboro, Maryland, USA; rc@medcybersecurity.com

Abstract

Anthropic's Claude Mythos Preview (April 2026) introduces a class of AI-accelerated cybersecurity capabilities that intersects directly with Post-Quantum Cryptography (PQC) migration. Drawing on federal guidance from NIST, NSA, OMB, and CISA, and on independent analyses from CETaS (Alan Turing Institute) and the UK AI Security Institute, we present a systems-engineering analysis of how Mythos-class models alter PQC migration timelines, risk surfaces, lifecycle dependencies, and architectural constraints. Mythos performs autonomous reasoning against previously unknown vulnerabilities in production software — evidenced by ten tier-5 control-flow hijacks on fully patched targets at per-campaign costs under USD 20,000 in Anthropic's disclosed examples — a qualitative departure from signature-based and SAST/DAST tooling. Modeling Mythos as both accelerator and destabilizer, we derive — as an analytical projection, not an empirical forecast — a compressed two-to-four-year migration window for highest-exposure systems, against traditional baselines of five to ten years for small organizations and twelve to fifteen or more years for large enterprises. The compression collapses human-labor bottlenecks in discovery, planning, and code modification, not cryptography itself. We propose a lifecycle-aligned migration model, an updated cost model, and governance requirements for frontier-model access. The binding constraint shifts domain-conditionally: for software-analytical phases, toward defender capacity to act on AI-accelerated analysis at adversary tempo; for embedded, regulated, and externally governed domains — FIPS validation, ATO renewals, certification cycles, and hardware replacement — non-compressible external cadence remains binding.

Keywords: Post-Quantum Cryptography; PQC migration; Claude Mythos; Project Glasswing; frontier models; systems engineering; NIST PQC; CNSA 2.0; AI-accelerated cybersecurity

---

1. Introduction

PQC migration is one of the largest cryptographic transitions in modern history. NIST's standardization of ML-KEM (FIPS 203), ML-DSA (FIPS 204), and SLH-DSA (FIPS 205) [1–3], combined with NSA's CNSA 2.0 requirements [4] and OMB M-23-02 [5], mandates a multi-year transformation across cloud, enterprise, embedded, OT, tactical, and national-security systems. This migration is not a cryptographic swap but a system-of-systems redesign involving protocol restructuring, firmware updates, PKI re-architecture, and cross-domain gateway modification.

Simultaneously, Anthropic's Claude Mythos Preview frontier model, announced on 7 April 2026 [6,7], introduces unprecedented, automated reasoning, exploit-discovery, and protocol-analysis capabilities. Anthropic's 244-page System Card [6], accompanying alignment risk report [8], and Frontier Red Team technical brief [9] document that the model achieved full control-flow hijack on ten fully patched open-source targets [9], demonstrated the ability to identify and exploit zero-day vulnerabilities across every major operating system and every major web browser [9], and, separately, identified thousands of additional high- and critical-severity vulnerabilities across the broader open- and closed-source software corpus under responsible-disclosure review [9]. Independent analyses from CETaS at the Alan Turing Institute [10] and the UK AI Security Institute [11] — the latter reporting 73% success on expert-level capture-the-flag tasks and first-ever end-to-end completion of a 32-step corporate network attack range, with explicit caveats that the evaluation environments lack active defenders and endpoint detection [11] — corroborate, together with industry reporting [7,12–14], that the capability jump over Claude Opus 4.6 is substantial and that Anthropic has chosen to restrict access through Project Glasswing [15] rather than release the model for general availability.

The core mechanism of the compression argument developed in this paper is neither autonomous model execution nor faster cryptography. Frontier models do not accelerate the cryptographic primitives themselves — ML-KEM, ML-DSA, and SLH-DSA run at the speeds of their underlying hardware and software implementations — but they do collapse the human-labor bottlenecks that force traditional PQC migration phases to be sequential. Cryptographic discovery, dependency analysis, protocol redesign, code modification, and initial validation are bounded in traditional programs by the cognitive bandwidth of specialist teams and by the sequential handoffs between them. AI-augmented teams dissolve both constraints where the work is software-analytical. They do not dissolve institutional change-management, FIPS and ATO compliance cycles, vendor and certificate-authority coordination, or hardware replacement schedules, all of which remain on externally imposed clocks. Section 7.3 develops this mechanism phase by phase with explicit epistemic-status labels.

This paper addresses the central research question:

How do Mythos-class frontier models alter PQC migration timelines, risk surfaces, lifecycle dependencies, and system-level constraints?

We contribute:

- A systems-engineering architecture modeling Mythos as a system actor.
- A lifecycle-aligned PQC migration model incorporating AI-accelerated analysis.
- A revised cost and timeline model.
- Governance and risk recommendations for frontier-model access.

1.1. Method

This paper applies a systems-engineering analysis method combining structured document review, functional decomposition, and comparative scenario modeling. The method is intentionally qualitative-architectural rather than empirical-experimental because the object of study — a restricted-access frontier model operating under a partner-only release regime [15] — is not open to third-party benchmarking. The analysis, therefore, treats Mythos-class capability as documented fact from primary sources and examines its architectural consequences for PQC migration as a system-of-systems transformation.

The source base comprises four tiers. Tier one is authoritative technical documentation: Anthropic's April 2026 System Card and Responsible Scaling Policy v3 [6], the accompanying Alignment Risk Update [8], and the Frontier Red Team technical brief [9], which together establish the empirical ceiling for Mythos-class capability. Tier two is federal PQC guidance: NIST FIPS 203, 204, and 205 [1–3]; NIST SP 800-208 [16]; NSA CNSA 2.0 [4]; OMB M-23-02 [5]; and the CISA Zero Trust Maturity Model [17]. Tier three is independent third-party analysis, principally the Centre for Emerging Technology and Security at the Alan Turing Institute [10], which provides the months-scale open-weight convergence framing used in Section 7, and the UK AI Security Institute [11], which provides third-party cyber-capability evaluation on standardized ranges with explicit caveats about simulated environments. Tier four is contemporaneous reporting [7,12–14] used only to corroborate non-technical context (launch, partner list, and release posture).

Four analytic steps are applied in sequence. First, system decomposition: the PQC migration space is partitioned into the eight domains shown in Figure 1 and the five lifecycle phases shown in Figure 3, each traceable to specific items in federal guidance [4,5,17]. Second, capability mapping: each documented Mythos capability [6,9] is mapped to the phase or phases it affects, distinguishing defender-accelerating from adversary-accelerating effects. Third, dynamic modeling: feedback loops are identified, and the acceleration-and-stress interaction shown in Section 5 is analyzed as a race-condition system. Fourth, comparative scenario modeling: the traditional sequential migration baseline from [18,19] is compared against a Mythos-compressed parallel trajectory, using the months-scale open-weight convergence window from [10] — which cites Epoch AI estimates of an average three-month capability lag between proprietary and open-weight frontier models (rising to five-to-twenty-two months in certain benchmarks) — as the adversary-capability anchor (Figure 4).

Scope boundaries are stated explicitly. This paper does not evaluate specific PQC algorithm implementations, does not benchmark Mythos against other frontier models, and does not attempt original empirical measurement of the Mythos system. The compressed-migration duration range in Section 7 is an analytic projection based on documented capability and the CETaS-cited months-scale open-weight convergence window, not an observational estimate; readers requiring operational numbers should treat it as a defensible upper bound on available time rather than a calendar forecast. Where claims in the text are supported by fact, each is cited to a primary or tier-two secondary source; where a claim is interpretive, it is framed as such.

1.2. Epistemic Status

The analysis is structured across three epistemic registers, and readers are asked to hold claims in each register to the standard of evidence that applies there.

Documented facts. Mythos-class capability is described from Anthropic's April 2026 System Card and Frontier Red Team technical brief [6,9] — including ten tier-5 control-flow hijacks on fully patched open-source targets, 181 working exploits on Mozilla Firefox 147 against two for Claude Opus 4.6, and testing-campaign costs under USD 20,000 in Anthropic's publicly disclosed scaffolds — corroborated by third-party evaluation from CETaS [10] and the UK AI Security Institute [11] (the latter with explicit caveats about simulated evaluation environments). Pre-Mythos AI-cybersecurity baselines are drawn from Google's Big Sleep disclosure [20], Meta's CyberSecEval benchmark suite [21], and the DARPA AI Cyber Challenge results [22]. Federal PQC migration guidance is taken from NIST, NSA, OMB, and CISA primary sources [1–5,16,17]. Traditional enterprise PQC migration baselines are drawn from peer-reviewed enterprise-migration analyses [18,19].

Analytical inferences. The compressed-track migration duration range (2–4 years for highest-exposure systems), the months-to-days compression of crypto-touchpoint discovery, the approximately-one-order-of-magnitude shift in adversary exploit economics, and the locations at which different organization classes are placed within the compression envelope are all analytical inferences from the documented capability base to adjacent PQC migration task classes. No empirical PQC migration has yet been completed with frontier-model assistance at enterprise scale. Where inferences are made in the body, they are explicitly labeled as such, and their bounding conditions and falsifiability criteria are disclosed in §7.3.

Normative recommendations. Governance recommendations (access controls, evaluation requirements, red-team requirements) in §8 are policy positions derived from the analytical inferences in §5–§7 combined with the restricted-access pattern Anthropic has established for Project Glasswing [15]. They identify the governance surface area that any serious response to Mythos-class capability must cover, not a specific regulatory regime.

The 2–4-year compressed migration window, in particular, sits in the analytical-inference register and is not a calendar forecast; §7.3 develops its methodology and limitations in full.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 April 2026 doi:10.20944/preprints202604.1744.v1

2. System Definition

This section defines the two primary objects of analysis: PQC migration as a system-of-systems transformation, and Mythos as a non-human system actor. The decomposition established here — eight migration domains and four automated capabilities — provides the structural vocabulary used by §4 (architecture), §5 (dynamics), §6 (lifecycle), and §8 (governance). Readers requiring full background on the underlying standards and capability disclosures may consult §3 in parallel; the analysis that follows does not depend on the reader reading §3 first.

2.1. PQC Migration as a System-of-Systems Transformation

PQC migration is frequently described as a cryptographic upgrade — replace classical primitives with ML-KEM, ML-DSA, and SLH-DSA across systems that use them. That framing understates the transformation. PQC touches protocol wire formats, firmware trust anchors, cryptographic hardware modules, and cross-domain enforcement logic, each governed by its own certification cadence, vendor ecosystem, and operational constraints. Federal guidance [4,5], therefore, frames PQC migration as a multi-domain transformation, and the eight domains shown in Figure 1 reflect the scope of that framing. Systems in scope include:

- Cloud and enterprise applications.
- Mobile and endpoint clients.
- IoT and embedded devices.
- OT/ICS systems.
- Tactical radios and RF systems.
- Satellites and space systems.
- Cross-domain gateways.
- PKI and identity infrastructure.

Each domain has unique constraints: MTU limits, RF airtime, firmware signing, hardware acceleration, and protocol dependencies. The ciphertext and signature expansions introduced by ML-KEM and ML-DSA relative to classical primitives have been documented as significant for constrained and latency-sensitive environments [1,2,16].

2.2. Mythos as a System Actor

Based on the Anthropic System Card [6], the Frontier Red Team technical brief [9], and independent analyses [7,10–14], Mythos Preview exhibits:

- Advanced reasoning and extended autonomous task execution, with the ability to chain multiple vulnerabilities into working exploits without human intervention [9].
- Automated vulnerability discovery across every major operating system and web browser, including long-lived bugs that survived decades of human review [9].
- Cross-domain protocol and system analysis, including reverse-engineering exploits on closed-source software and converting N-day disclosures into working exploits [9].
- Controlled access through a restricted-partner program (Project Glasswing) and a 90-day coordinated disclosure window [7,15].
- Anthropic-reported pre-launch briefings to U.S. federal officials, per Platformer reporting [14], including conversations with the Cybersecurity and Infrastructure Security Agency (CISA) and the Center for AI Standards and Innovation (CAISI); these are briefings reported by Anthropic rather than confirmed ongoing-access arrangements, and as of late April 2026 CISA had not been granted access to the model.

We model Mythos as a non-human system actor capable of automated protocol decomposition, automated exploit-chain generation, automated firmware dependency extraction, and automated cross-domain attack-path reasoning.

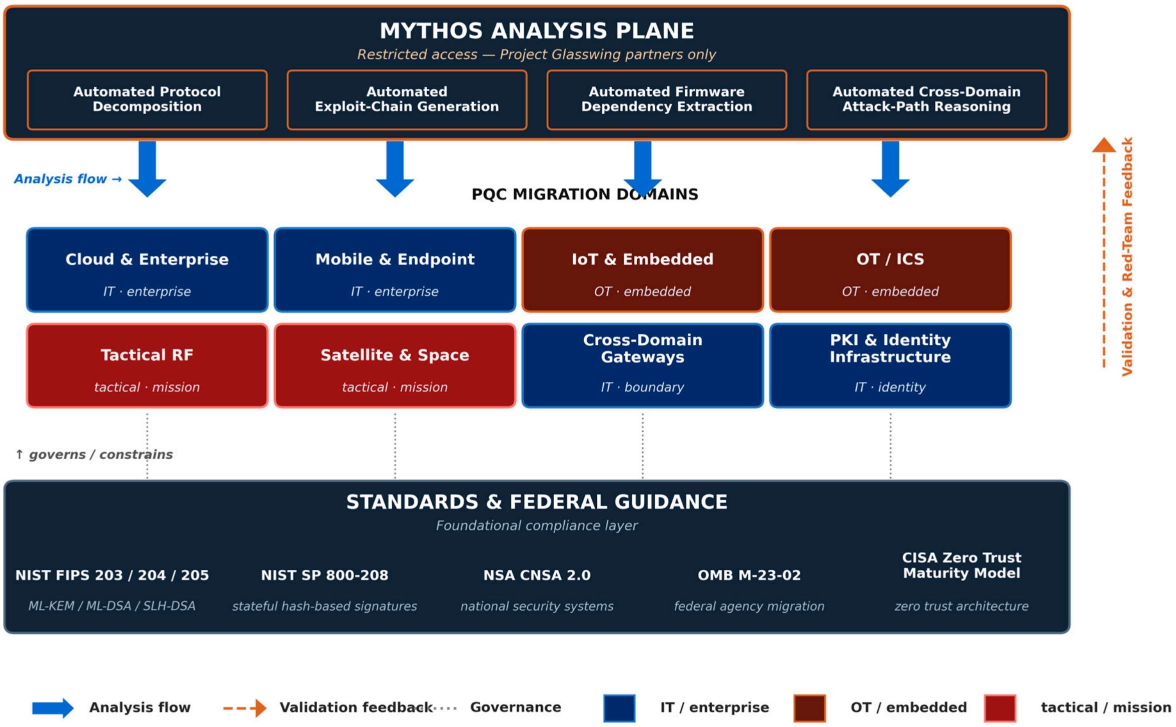


Figure 1. System architecture for PQC migration under Mythos-class frontier-model capability, organized as three tiers. The restricted-access Mythos Analysis Plane (top) hosts four automated capabilities: protocol decomposition, exploit-chain generation, firmware dependency extraction, and cross-domain attack-path reasoning. The eight PQC Migration Domains (middle) are color-coded by archetype — IT/enterprise in navy, OT/embedded in brown, tactical/mission in crimson. The Standards and Federal Guidance layer (bottom) provides foundational compliance. Solid blue arrows show analysis flow from capabilities into domains; a dashed amber arrow on the right carries validation and red-team feedback from the domains back up to the Mythos plane; dashed grey connectors indicate governance flow from standards upward to the domains.

3. Background and Related Work

Three bodies of work provide the empirical base for this analysis. Federal PQC standards and migration guidance [1–5,16,17] establish what must be migrated and on what cadence. Frontier-model capability disclosures [6,8,9] and independent analyses [7,10–14] establish what Mythos-class models can now do. Recent work on AI-accelerated vulnerability discovery [9,10] establishes the capability boundary between traditional vulnerability-management tooling and the step-change disclosed in April 2026. Each body is reviewed in turn below.

3.1. PQC Standards and Migration Guidance

The federal PQC migration landscape is anchored by three authorities: NIST, which publishes cryptographic standards [1–3,16]; NSA, which sets National Security System requirements [4]; and OMB with CISA, which sets civilian-agency migration and zero-trust requirements [5,17]. The three are aligned in intent but differ in binding scope and cadence; the sections below summarize each and identify where their requirements intersect with the Mythos-era timing pressure developed later in the paper.

NIST Standards

- FIPS 203: Module-Lattice-Based Key-Encapsulation Mechanism Standard (ML-KEM) [1].
- FIPS 204: Module-Lattice-Based Digital Signature Standard (ML-DSA) [2].
- FIPS 205: Stateless Hash-Based Digital Signature Standard (SLH-DSA) [3].
- SP 800-208: Recommendation for Stateful Hash-Based Signature Schemes [16].

NSA CNSA 2.0

Mandates PQC adoption for National Security Systems, with binding transition milestones through the 2030s [4].

OMB M-23-02 and CISA Guidance

OMB M-23-02 requires federal agencies to inventory cryptographic dependencies and plan migration to PQC [5]. The CISA Zero Trust Maturity Model [17] provides architectural guidance directly relevant to PQC-aligned redesign, particularly around cryptographic agility and identity.

3.2. Frontier Model Capabilities

The capability base for this paper is documented in three primary sources published by Anthropic on 7 April 2026: the 244-page System Card [6], the accompanying alignment risk update [8], and the Frontier Red Team technical brief [9]. Independent analyses [7,10–14] corroborate the capability claims and provide policy framing. Each source is reviewed below in the order in which its content becomes load-bearing for the architectural and dynamic analysis in §4 and §5.

Anthropic System Card and Risk Documentation

The Claude Mythos Preview System Card [6] is Anthropic's first system card published under Responsible Scaling Policy (RSP) v3 and the first system card issued for a model that Anthropic has chosen not to make generally available. The 244-page document reports cybersecurity, alignment, autonomy, and model-welfare evaluations, and confirms that the model exceeds Claude Opus 4.6 across mathematics, long-context reasoning, software engineering, and cybersecurity benchmarks [6,10]. The accompanying alignment risk update [8] is structured as an implementation of RSP v3 and documents observed behaviors, including access escalation within execution environments.

Frontier Red Team Technical Brief

Anthropic's Frontier Red Team blog post [9] provides the technical evidence base: on a corpus of roughly 7,000 entry points into open-source repositories drawn from OSS-Fuzz, Mythos Preview achieved 595 crashes at tiers 1–2 and full control-flow hijack on ten fully patched targets, compared with a single tier-3 crash each for Sonnet 4.6 and Opus 4.6 under the same benchmark. On Mozilla Firefox 147, a benchmark previously used to evaluate Opus 4.6, Mythos Preview developed 181 working JavaScript shell exploits and achieved register control in 29 additional trials. These results emerged without domain-specific security fine-tuning, as a downstream consequence of general improvements in code, reasoning, and autonomy.

Independent Analysis and High-Credibility Reporting

TechCrunch reported the model's 7 April 2026 preview launch and the 12 founding Project Glasswing partners (Amazon, Apple, Broadcom, Cisco, CrowdStrike, Google, JPMorganChase, the Linux Foundation, Microsoft, NVIDIA, and Palo Alto Networks), plus 40 additional critical-infrastructure organizations granted monitored access [7]. Fortune documented the March 2026 accidental disclosure of the model (then internally codenamed "Capybara") through an unsecured public data cache [13]. Platformer reported that Anthropic briefed senior U.S. government officials prior to launch, including conversations with CISA and the Center for AI Standards and Innovation (CAISI), as described to the publication by an Anthropic spokesperson [14]. The World Economic Forum [12] contextualized the release within the 2026 Global Cybersecurity Outlook, noting the risk of a widening gap between AI-enabled offense and defender capacity. CETaS at the Alan Turing Institute [10] provided an independent assessment of the model's implications for national cybersecurity policy, with specific attention to the risk posed by open-weight successors. The UK AI Security Institute (AISI) [11] conducted independent cyber-capability evaluations on standardized ranges, reporting that Mythos Preview succeeded on 73% of expert-level capture-the-flag tasks and was the first model to solve AISI's 32-step "The Last Ones" corporate-network attack range end-to-end (three of ten attempts), a range the institute estimates would take human experts approximately twenty hours; the next-best model, Claude Opus 4.6, averaged sixteen of the thirty-two steps. AISI attaches explicit caveats that its evaluation environments lack active defenders, endpoint detection, and real-time incident response, and that it therefore cannot conclude from these results whether Mythos Preview would be able to attack well-defended systems [11]. These caveats are material to the adversary-capability anchor used in Section 7: the AISI results establish a capability delta on standardized simulated ranges, not a demonstrated capability against hardened production targets, and this paper treats them accordingly.

3.3. AI-Accelerated Vulnerability Discovery

The Mythos disclosure establishes a step-change in AI-accelerated vulnerability discovery with two implications for PQC migration. The first is capability: what defenders can now use frontier models to analyze and what adversaries can now use them to exploit. The second is positioning: how this capability sits relative to the traditional vulnerability-management tooling federal and enterprise defenders have deployed. Both are addressed below, and the recalibration they motivate is the empirical foundation for the compressed-track argument developed in §7.

Mythos-Era Step-Change in Discovery

Prior work has demonstrated that large language models can assist in exploit generation and code auditing [20–22]: Google's Big Sleep framework, a collaboration between Project Zero and DeepMind, identified the first publicly reported AI-discovered real-world zero-day (a stack-buffer-underflow in SQLite) in October 2024 [20]; Meta's CyberSecEval benchmark suite established baseline measurements for LLM insecure-coding tendencies and cyberattack helpfulness across multiple model families [21]; and DARPA's AI Cyber Challenge demonstrated that cyber-reasoning systems leveraging frontier LLMs could autonomously discover and patch vulnerabilities at competition scale (86% discovery, 68% patching at finals) with average per-task costs of approximately USD 152 [22]. However, performance in these pre-Mythos efforts typically required significant human scaffolding — custom tool-use frameworks, hand-built agent loops, and iterative prompt engineering — alongside domain-specific fine-tuning that limited accessibility to security specialists. Mythos Preview represents a measurable step-change on both axes: engineers at Anthropic without formal security training produced complete, working remote-code-execution exploits overnight using generic natural-language prompts and without exploit-specific fine-tuning [9]. The capability is therefore accessible to defender organizations that lack in-house offensive-security expertise, which is precisely the population most exposed to the compressed-track trajectory in §7. Independent analyses corroborate that several open-weight models are approaching comparable capability on a subset of tasks through post-training distillation and capability-matching fine-tuning, reinforcing the CETaS warning [10] that restricted-access regimes have a limited horizon.

Positioning Relative to Traditional Vulnerability Management

The vulnerability management tooling deployed across enterprise and federal environments — signature-based network scanners such as Tenable Nessus, Qualys VMDR, and Rapid7 InsightVM; static and dynamic application security testing (SAST/DAST) platforms such as Veracode and Checkmarx; and continuous fuzzing infrastructure such as OSS-Fuzz — operates almost entirely in the space of known vulnerability classes and published CVEs. Signature-based scanners detect the fingerprints of previously disclosed flaws at scale; SAST and DAST detect known anti-patterns and exploitable constructs in code and running applications; fuzzers produce crashes whose exploitability must then be assessed by human researchers. Each is optimized for a threat environment in which vulnerability discovery is the rate-limiting step and the downstream activities — deployment of detection coverage, prioritization, and patching — constitute the defender's principal workload.

Mythos-class frontier models operate in a categorically different space: autonomous reasoning about program behavior to identify previously unknown vulnerabilities and generate working end-to-end exploits for them. The OSS-Fuzz result in [9] is diagnostic of the distinction. OSS-Fuzz has continuously fuzzed the implicated open-source repositories for years without surfacing the ten tier-5 control-flow hijacks Mythos Preview produced, because fuzzing generates crashes while tier-5 findings require the analytic step from crash to weaponized primitive — historically the work of skilled human exploit developers. The implication for enterprise defenders is not that existing vulnerability management tooling becomes obsolete; it remains necessary for coverage of the known-vulnerability population, where most incidents continue to originate. The implication is that Mythos-class capability introduces an operational threat surface — the currently unknown vulnerability population in critical software — against which no component of the traditional stack provides direct coverage. Organizations whose defensive posture assumes vulnerability discovery is the rate-limiting step are calibrated to a threat environment that no longer exists; the compressed-track migration argument developed in Section 7 follows directly from this recalibration.

4. System Architecture

The PQC migration landscape is architecturally layered: touchpoints where cryptography is invoked, protocol layers where PQC primitives change wire format, firmware and embedded systems where cryptography is compiled into binary artifacts, and cross-domain gateways where all of the above concentrate under byte-level guard enforcement. This section develops each layer in turn, with specific attention to how Mythos-class capability changes the analysis and where the architectural fragilities that drive §7's cost model originate.

4.1. Crypto-Touchpoint Topology

PQC migration requires enumerating all locations where classical cryptography is embedded in system behavior. Federal guidance [4,5] emphasizes that cryptography is not isolated to TLS libraries but is deeply embedded across transport protocols (TLS, SSH, IPsec, QUIC), application protocols (S/MIME, DNSSEC, OAuth, SSO), firmware signing chains, bootloaders, and secure enclaves, cross-domain gateways, satellite and RF links, and IoT/embedded stacks.

Mythos-class models alter this topology by automating the discovery of crypto-touchpoints. The Frontier Red Team brief [9] documents the model's ability to analyze complex systems, identify vulnerabilities, and map dependencies. The same underlying capability, applied to crypto-inventory discovery, is projected to compress that phase of PQC migration from a months-scale manual task to a days-scale AI-augmented task in enterprise environments where touchpoints are spread across hundreds of repositories and long-tail application binaries. This projection is an analytical inference from [9]'s documented code-reading throughput to the adjacent CBOM-generation task class, not a measured outcome of CBOM-specific Mythos pilots, which have not yet been publicly reported.

Touchpoint density varies sharply by domain archetype (Figure 1). Cloud and enterprise systems concentrate cryptography in a small number of well-known libraries (OpenSSL, BoringSSL, cloud KMS services, TLS terminators) but replicate those libraries across thousands of deployment targets; discovery here is an inventory problem at scale. Mobile and endpoint systems distribute cryptography across platform-level APIs, application sandboxes, and hardware-backed keystores. Operational-technology and embedded systems are harder: cryptography is embedded in firmware, bootloaders, and vendor-supplied binary blobs that are often undocumented and not easy to re-flash. Cross-domain gateways concentrate cryptographic enforcement at a small number of protocol-translation points, but each point is high-consequence and high-complexity. Tactical radio-frequency and satellite systems concentrate cryptography in waveform-layer modules where replacement is gated by spectrum certification cycles that measure in years. PKI and identity infrastructure span all of the above, because every issuing authority, intermediate, and end-entity certificate is a migration object. The eight-domain decomposition in Figure 1 reflects this density gradient and serves as the traceability backbone for Sections 6 and 7.

A Mythos-class model shifts the discovery problem from enumeration to prioritization. Where traditional inventory tooling returns a flat list of cryptographic call sites, Mythos-class reasoning can produce a ranked list weighted by exploitability, downgrade surface, and downstream dependency fan-out, drawing on the same chained-reasoning capabilities demonstrated in the OSS-Fuzz and Firefox evaluations [9]. Architecturally, this means the crypto-touchpoint inventory becomes a live data structure consumed by both defenders during migration planning and adversaries during target selection — the same artifact, different consumers.

4.2. Protocol Decomposition Layer

Evaluations in the System Card [6] demonstrate that Mythos-class models perform multi-step reasoning and can decompose complex structures. Applied to PQC migration, the model can break down protocol flows, identify cryptographic primitives, map handshake dependencies, detect MTU-sensitive expansions (for example, ML-KEM ciphertext sizes [1]), and identify hybrid-mode opportunities. This creates a new architectural layer: AI-assisted protocol decomposition.

The message-size expansion introduced by PQC primitives is the dominant driver of protocol-level rework. ML-KEM-768, the NIST-recommended Level 3 parameter set [1], produces public keys of 1,184 bytes, ciphertexts of 1,088 bytes, and shared-secret outputs of 32 bytes. ML-DSA-65 [2] produces public keys of 1,952 bytes and signatures of 3,309 bytes. SLH-DSA-SHA2-192s [3], the stateless hash-based alternative for long-lifetime signing contexts, produces signatures of 16,224 bytes. For comparison, elliptic-curve counterparts at equivalent classical strength produce keys and signatures in the 32-to-96-byte range. The resulting expansion factors — roughly 35x for ML-DSA signatures relative to Ed25519, and over 500x for SLH-DSA signatures — interact with every transport-layer constraint that classical cryptography leaves untouched.

MTU interaction is the most immediate architectural consequence. Ethernet's default 1,500-byte MTU is sufficient for a bare ML-KEM-768 exchange, but TLS 1.3 handshakes that combine ML-KEM key establishment with an ML-DSA-65 certificate chain routinely exceed the initial congestion window and cross the IPv6 minimum-MTU threshold of 1,280 bytes, forcing additional round trips and, on path-MTU-constrained links, black-hole failures. Narrowband tactical radio waveforms operate at MTUs measured in hundreds of bytes, not thousands; a single ML-DSA-65 signature requires fragmentation across multiple waveform frames with associated airtime cost. Mythos-class decomposition is valuable here precisely because the analysis is mechanical: given a protocol specification and a target bearer, the model can enumerate every message whose post-PQC size crosses an MTU boundary, trace the resulting fragmentation behavior, and identify hybrid-mode configurations that reduce the critical-path signature cost. Architecturally, the protocol-decomposition layer that Mythos enables is therefore not an analytic convenience but a precondition for correct PQC migration on bandwidth-constrained bearers.

4.3. Firmware and Embedded Dependencies

Embedded systems represent the hardest PQC migration domain. Constraints include limited flash and RAM, fixed MTU, RF airtime limits, hardware acceleration dependencies, and long certification cycles. The Frontier Red Team brief documents Mythos Preview's ability to analyze firmware and identify dependency chains [9], including a 27-year-old remote-crash vulnerability in OpenBSD and a 16-year-old vulnerability in FFmpeg that had survived five million automated test iterations. Applied to PQC migration, this capability enables automated extraction of crypto libraries, bootloader signing paths, hardware crypto calls, and OTA update constraints, reducing manual reverse-engineering effort.

A concrete example illustrates the architectural pressure. Consider a legacy narrowband tactical radio with 512 KB of flash, 128 KB of RAM, a hardware AES accelerator, and a waveform-layer MTU of approximately 256 bytes. Classical elliptic-curve key establishment fits comfortably in this

doi:10.20944/preprints202604.1744.v1

envelope: a single P-256 ECDH exchange consumes under one hundred bytes of signaling and completes within one radio frame. Substituting ML-KEM-768 into the same handshake places a 1,088-byte ciphertext across four or five waveform frames, depending on framing overhead, adds tens of milliseconds of additional airtime at typical tactical data rates, and introduces a reassembly state machine in a code base that previously did not need one. If a certificate chain is required in-band, ML-DSA-65 or SLH-DSA signatures multiply this cost. The firmware reality is that ROM is often full; adding a PQC library of 30–80 KB can require removing or refactoring other modules to make room. Hardware crypto accelerators designed for AES and SHA-2 do not accelerate ML-KEM or ML-DSA, so PQC operations fall back to constant-time software with corresponding battery and latency cost. Mythos-class analysis helps here because the model can ingest the firmware, the waveform specification, and the certification envelope together and produce a per-frame budget showing where the PQC handshake must be spread, whether pre-shared symmetric material can substitute, and whether a classical/PQC hybrid mode reduces the critical-path airtime below the operational threshold.

Certification timelines deserve explicit attention. Tactical radio waveforms are typically governed by spectrum-allocation and interoperability-certification processes that measure in years, not months. A firmware change that alters on-air signaling — which any PQC migration does — may require re-entry into the certification process and re-issuance of cryptographic approvals. This means the firmware domain is gated not by engineering capacity but by external review cadence, and it is the dominant reason Section 7 argues for tiered prioritization rather than uniform migration.

4.4. Cross-Domain Gateway Architecture

Cross-domain gateways (CDGs) enforce security boundaries between networks of differing classifications. PQC migration requires rewriting guard protocols, updating PKI, and ensuring PQC-safe filtering that accounts for the significantly larger message sizes produced by ML-KEM and ML-DSA. Mythos Preview’s demonstrated ability to chain vulnerabilities across components [9] allows automated modeling of CDG flows, identifying where PQC insertion breaks message formats or guard logic.

The specific fragility of CDGs is that guard logic operates at the byte level. Rule sets that parse, whitelist, and re-emit messages are written against the exact wire format of the protocols they mediate. Classical TLS or IKEv2 handshakes are fully understood by mature CDG implementations; PQC-modified handshakes are not. When an ML-DSA certificate chain pushes a handshake record past the guard’s per-message length ceiling, three failure modes are possible: the guard rejects the message and breaks connectivity, the guard truncates and forwards an invalid fragment, or the guard’s reassembly logic is forced into a code path that was not security-reviewed. The same applies to DNSSEC responses, S/MIME messages, OAuth assertions, and any other protocol family that now carries a post-quantum signature. Cross-Domain Gateways are the narrow waist of Figure 1’s architecture: everything that crosses a classification boundary must pass through them, and every PQC primitive change must be validated against every guard that inspects it.

Mythos-class cross-domain attack-path reasoning [9,10] applies directly to this fragility. An adversary using an equivalent-capability model can enumerate every handshake whose PQC variant crosses a guard’s parse-length boundary, identify the specific field whose post-PQC size change creates ambiguous parsing, and generate a minimal downgrade probe that elicits the guard’s most permissive failure mode. The defender’s equivalent use of the same capability — the validation loop discussed in Section 5.1 and illustrated on the right side of Figure 1 — is the only scalable counterbalance, because the set of handshake variants to test grows combinatorially with the number of PQC primitives, hybrid modes, and certificate-chain lengths in use. This architectural point motivates the governance requirements in Section 8: the analysis capability that defenders need is the same capability that restricted-access controls [6,15] are designed to limit, and any access regime that gates defender use more tightly than adversary use produces a worse security outcome.

doi:10.20944/preprints202604.1744.v1

5. System Dynamics

Analyzing PQC migration under Mythos-class capability requires modeling the feedback dynamics that operate on both sides of the defender–adversary equation. Six identifiable loops shape the migration trajectory. Three loops accelerate defender work — automated mapping, automated redesign, and automated validation — by compressing the cognitive labor historically required for cryptographic discovery and protocol engineering. Three loops apply pressure — exploit discovery, attack-path generation, and legacy-system pressure — by reducing the marginal cost of offensive operations and increasing the exploitability of systems awaiting migration. The interaction of these loops determines whether the compressed-track trajectory in Figure 4 is achievable.

5.1. Acceleration Loops

Acceleration loops reduce the human-labor bottlenecks that set the pace of traditional PQC migration. Each operates on a distinct phase of the lifecycle developed in Section 6, with output that feeds the subsequent loop and raises the overall throughput of defender work.

Automated Mapping Loop

Mythos-class capability is projected to compress crypto-touchpoint discovery, protocol decomposition, and firmware dependency extraction from the months that manual inventory typically requires down to days of automated analysis (see §3.3 for the inferential basis). The output is not merely a faster inventory, but a ranked, weighted touchpoint register with exploitability scoring and downstream fan-out data (§4.1, §6.1). This register feeds the Redesign Loop by exposing which touchpoints are on the critical path, eliminating the sequential handoff between discovery and planning teams that bound the traditional trajectory and converting a gated workflow into a streaming one.

Automated Redesign Loop

Once the touchpoint register is available, Mythos can propose PQC-compatible protocol variants, hybrid-mode configurations, and MTU-safe message formats tailored to each bearer class identified in mapping. The loop is iterative: each proposed variant is simulated against downgrade and fragmentation test cases by the Validation Loop, errors surface in the output, and refined variants are generated in the same analytic pass. This compresses the Planning phase of Section 6 and eliminates the traditional wait states between protocol design and interoperability analysis, which in the sequential model accounts for a substantial fraction of calendar time.

Automated Validation Loop

Mythos can simulate PQC handshake flows, failure modes, MTU fragmentation, and RF airtime impacts across bearer classes before any production change is staged. The loop's value is not testing coverage alone but continuous feedback: validation failures flow back into the Redesign Loop as constraints on the next variant, and validation successes flow forward into execution (§6.3, §6.4). This mechanism converts validation from an end-of-pipeline gate into an always-on control surface, which is the structural prerequisite for the compressed trajectory argued in Section 7.

5.2. Stress Loops

Stress loops are symmetric adversary-side dynamics. The same capability class that accelerates defender work accelerates offensive work, and for most organizations today, the defender has not yet adopted the capability while open-weight successors are approaching capability convergence with proprietary frontier models on a months-scale timeline [10]. The asymmetry matters because the three stress loops operate on real targets now, producing measurable pressure whether defenders have integrated their counterparts.

Exploit-Discovery Loop

The Frontier Red Team brief [9] reports specific testing-campaign costs under USD 20,000 in their publicly disclosed scaffolds — for example, an OpenBSD SACK-vulnerability discovery campaign that totalled under USD 20,000 across approximately one thousand scaffold runs, an FFmpeg campaign at roughly USD 10,000 across several hundred runs, and individual N-day-exploit generation pipelines completed at under USD 2,000 [9]. These reported costs are specific to Anthropic's disclosed scaffolds and campaigns rather than a generalized per-target adversary-cost constant; nonetheless, against standard-industry contracting rates for penetration-testing engagements and published bug-bounty payout tables, they establish a cost order-of-magnitude for AI-augmented vulnerability discovery and exploit generation that, on a practitioner-grade comparison, is dramatically below the per-exploit cost of traditional programs. A precise per-exploit comparison is not attempted here because traditional and AI-augmented workflows produce different exploit classes against different target sets; the order-of-magnitude claim should be read as an economic-signal estimate, not a controlled benchmark. The economic implication is that aggregate adversarial pressure against legacy, pre-PQC systems increases as soon as scaffold-level reproducibility becomes more broadly available. The loop tightens over time: discovered vulnerabilities feed the Attack-Path Generation Loop as primitives for chaining, and unpatched disclosures from public corpora such as OSS-Fuzz are reprocessed for weaponizable findings that prior-generation tooling did not surface. The economic consequence, developed further in Section 7.4, is that the traditional bug-bounty and penetration-test cost curves no longer approximate adversary marginal cost.

Attack-Path Generation Loop

Mythos can reason across cloud, enterprise, embedded, OT, RF, and satellite layers, combining lateral-movement primitives with cryptographic weaknesses to construct cross-domain attack paths that no single-domain defender framework can fully model [9,10]. Figure 2 illustrates a hypothetical six-stage example synthesized from benchmark evidence and domain-specific threat assumptions. The dynamic consequence is that defenders can no longer address vulnerabilities domain-by-domain: a remediation that closes one link of a chain may still leave multiple viable paths, and any defender scope that does not span the full set of reachable domains produces a false assurance surface. This is the operational motivation for the cross-domain-gateway focus on Section 4.4 and the red-team requirements developed in Section 8.

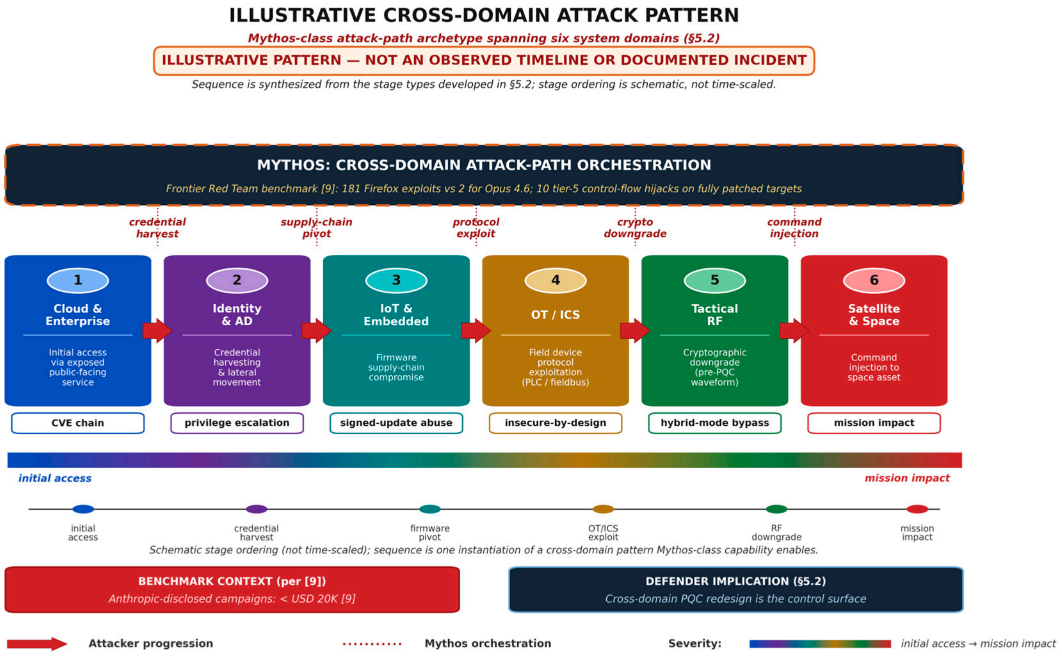


Figure 2. Illustrative cross-domain attack pattern: a hypothetical Mythos-class attack-path archetype spanning six system domains. The six stages — Cloud & Enterprise (blue), Identity & AD (purple), IoT & Embedded (teal), OT / ICS (amber), Tactical RF (green), and Satellite & Space (red) — illustrate how Mythos-class attack-path reasoning [9,10] combines lateral-movement primitives with cryptographic weaknesses across domain boundaries. Attacker techniques are labeled above the transitions between stages (credential harvest, supply-chain pivot, protocol exploit, crypto downgrade, command injection); tag pills below each stage summarize the vulnerability class or outcome associated with that stage (CVE chain, privilege escalation, signed-update abuse, insecure-by-design, hybrid-mode bypass, mission impact). The scenario is synthesized from benchmark evidence and domain-specific threat assumptions; it is illustrative of the architectural class of cross-domain attack patterns Mythos-class capability enables, not a documented incident and not a kill-chain representation of any observed operation. Defender implications are developed in §5.2 (Attack-Path Generation Loop), §6.4 (Validation phase), and §8.3 (Red-Team Requirements).

Legacy-System Pressure Loop

Legacy systems become high-risk under AI-accelerated scanning because their attack surface — classical-crypto exposure, inability to support PQC drop-in replacements, firmware that cannot be re-flashed without certification cycles — is precisely the space the Exploit-Discovery Loop excels at probing. As migration proceeds for modern assets, the residual legacy population becomes proportionally more exposed: the same quantity of offensive compute, now pointed at a shrinking target set, yields higher per-target discovery density. This inverts the common planning assumption that partial migration reduces aggregate risk and makes the sequencing decisions developed in Section 7 — which systems migrate first, which are retired, which are isolated — load-bearing rather than cosmetic.

5.3. Combined Dynamics

The three acceleration loops and three stress loops do not operate in isolation; they interact through shared substrate — the same codebases, protocols, and capability surfaces — and produce a race condition whose outcome depends on the relative adoption velocity of defender and adversary. Where defenders have adopted the acceleration loops, migration compresses toward the trajectory shown in Figure 4's compressed track. Where defenders have not, the acceleration loops remain theoretical while the stress loops operate on real targets today, producing the widening defender–adversary gap described in §3.3 and bounded in §7.3.

For software-analytical phases — discovery, protocol decomposition, exploit-chain generation, and code modification — the binding constraint on the race is therefore no longer cryptographic availability (ML-KEM, ML-DSA, and SLH-DSA are standardized [1–3] and implementations exist) but defender organizational capacity to act on the output of AI-accelerated analysis at the tempo adversaries can sustain once equivalent capability reaches open-weight parity [10,12]. For embedded, regulated, and externally governed domains — FIPS 140-3 module validation, Authority to Operate renewals, CNSA 2.0 audit cadences, embedded cryptographic hardware replacement, and vendor certification cycles — the binding constraint remains non-compressible external cadence that frontier models do not accelerate; these are developed in §7.3 as the non-compressible components of the compressed-track scenario envelope. The software-analytical constraint is a governance problem before it is an engineering problem: the analysis capability defenders need is the same capability that restricted-access controls [6,15] are designed to limit, and the window during which Project Glasswing-class access regimes can preserve defender advantage is bounded by capability diffusion. The governance requirements that follow from this observation are developed in Section 8.

6. Migration Lifecycle Model

We propose a lifecycle model aligned with federal guidance [4,5,17] and Mythos-class AI capabilities. The five phases — Discovery, Planning, Execution, Validation, and Assurance — are summarized visually in Figure 3, which shows how Mythos capability acts on each phase through a distinct mode: acceleration during Discovery, analysis during Planning, automation during Execution, red-team pressure during Validation, and continuous analysis during Assurance. The phase ordering in Figure 3 is a logical sequence, not a calendar serialization; phase concurrency under Mythos-class capability is developed in §7.2 and visualized as overlapping bars in Figure 4. Each phase is described below with its inputs, principal activities, outputs, Mythos-specific role, human-in-the-loop requirements, and gate criteria for advancing to the next phase.

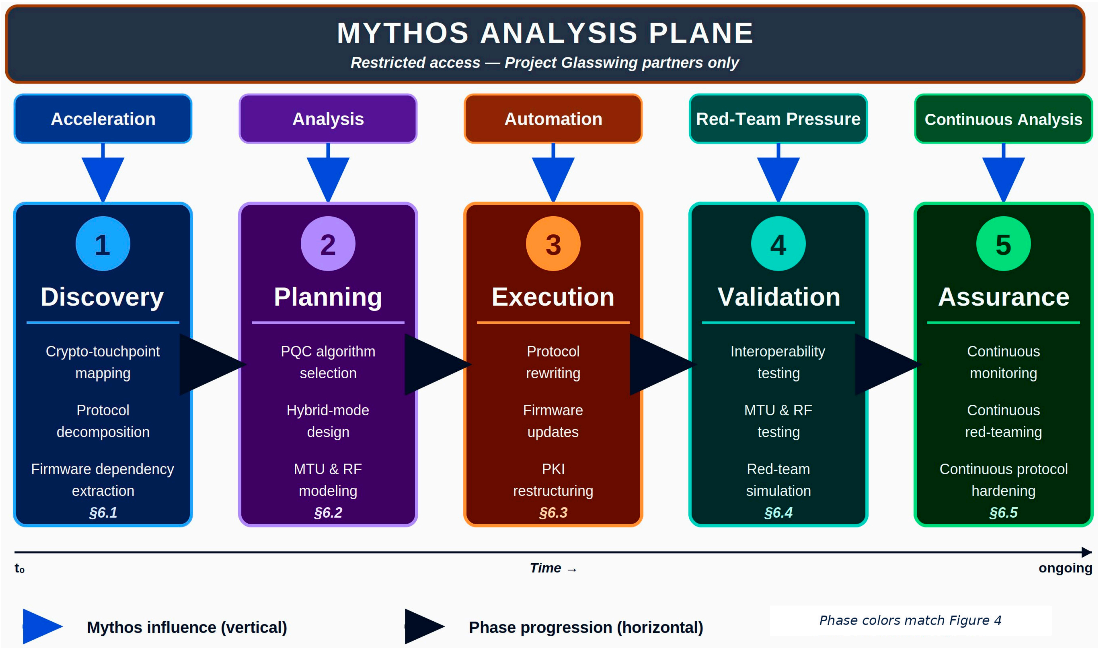


Figure 3. Mythos-class acceleration of the PQC migration lifecycle. The five phases each receive a distinct mode of Mythos capability — Acceleration, Analysis, Automation, Red-Team Pressure, Continuous Analysis — shown as top-tier badges. Each phase lists its principal activities and cross-references its §6 subsection; phase colors match Figure 4.

6.1. Phase 1: Pre-Migration Discovery

Traditional discovery requires manual inventory, protocol analysis, and firmware review. Mythos accelerates this phase by automating protocol decomposition, crypto-touchpoint mapping, and firmware dependency extraction.

Inputs to the Discovery phase are the organization's full software and firmware inventory, network architecture diagrams, PKI issuance data, and the cryptographic approval documentation held by the system authorization authority. Principal activities are crypto-touchpoint enumeration across all eight domains in Figure 1, protocol decomposition for every transport and application protocol in use, firmware dependency extraction across embedded platforms, and initial identification of downgrade surfaces. Outputs are a cryptographic bill of materials (CBOM) keyed to code location and deployment target, a touchpoint register ranked by exploitability and downstream fan-out, and a domain-by-domain exposure heat map that feeds Planning. Mythos-specific role: in a Glasswing-partner context, Mythos-class capability is projected to compress touchpoint discovery from a months-scale traditional task to a days-scale AI-augmented task and to produce ranked, weighted output rather than a flat inventory (§4.1; see §3.3 for the inferential basis). Human-in-the-loop requirements: CBOM attestation requires a cryptographic officer's sign-off because the model does not have regulatory responsibility for completeness. Gate criteria to advance: the CBOM must reach an auditable coverage threshold (typically 95% of in-scope systems), the touchpoint register must be stable across two successive Mythos-assisted passes, and the heat map must be reviewed and annotated by at least one domain owner per archetype.

6.2. Phase 2: Migration Planning

Planning requires selecting PQC algorithms per FIPS 203/204/205 [1–3], designing classical/PQC hybrid modes, updating PKI, and modeling MTU/RF impacts. Mythos can simulate handshake flows, message sizes, and failure modes.

Inputs to the Planning phase are the Discovery outputs plus the applicable federal guidance [1–5,16,17] and the bearer characteristics of each affected transport path. Principal activities are parameter-set selection per touchpoint (matching ML-KEM, ML-DSA, or SLH-DSA to use case and security level), hybrid-mode design where backward compatibility must be retained, PKI restructuring to handle post-quantum certificate chains, MTU and radio-frequency tolerance modeling against the sizing numbers in §4.2, and tiered prioritization that places crown-jewel assets and downgrade-vulnerable systems first. Outputs are a per-domain migration plan, an agility-target architecture specifying how the next cryptographic transition will be absorbed with lower cost, and a test-plan skeleton that feeds Validation. Mythos-specific role: Mythos-class handshake simulation short-circuits the traditional build-test-measure loop because a trained model can enumerate the realistic failure envelope for each candidate configuration before any code is written. Human-in-the-loop requirements: parameter-set selection is a risk-acceptance decision that a cryptographic authority must own; Mythos output is decision-support, not decision-authority. Gate criteria to advance: every Discovery-identified touchpoint must have an assigned parameter set, every bearer constraint must have a documented handshake-size budget, and the agility-target architecture must be endorsed by systems-engineering leadership.

6.3. Phase 3: Migration Execution

Execution includes updating libraries, rewriting protocols, updating firmware, and re-signing bootloaders. Mythos can assist with code transformation, protocol redesign, and dependency resolution, subject to governance constraints discussed in Section 8.

Inputs to the Execution phase are the Planning outputs, the code base(s) targeted for change, and the firmware images, bootloader trust anchors, and certificate-issuance tooling that must be modified. Principal activities are library substitution in cloud and enterprise code, protocol-stack rewriting at TLS/IKEv2/QUIC touchpoints, firmware updates, and re-signing of bootloaders across embedded and OT targets, PKI restructuring including intermediate-certificate reissuance, and cross-domain-gateway rule-set updates to tolerate the larger PQC message envelopes discussed in §4.4. Outputs are PQC-enabled code and firmware artifacts, an updated certificate hierarchy, and cross-domain-gateway rule sets that have been staged but not yet promoted to production. Mythos-specific role: Mythos-class code transformation is most valuable where the change is mechanical (library substitution, handshake-record-length adjustments, buffer resizing) and least valuable where the change requires judgment about backward compatibility, partner ecosystem readiness, or risk acceptance. Human-in-the-loop requirements: all code changes produced with Mythos assistance must carry human authorship attribution and pass the organization's existing secure-code-review process; no AI-generated code should enter a production crypto path without human review by an engineer qualified in the affected subsystem. Gate criteria to advance: build reproducibility across a clean environment, successful unit and integration testing against a reference PQC vector set, and a successful staged deployment to a non-production mirror of each target domain. Execution is where simultaneity pressure (§5.1, §7.1) bites hardest; Figure 4 shows the execution band as the dominant overlapping interval in the Mythos-compressed trajectory.

6.4. Phase 4: Validation and Testing

Validation requires interoperability testing, MTU/RF testing, and security testing. Mythos can generate test cases, attack paths, and failure scenarios; within Project Glasswing, partners have already used the model in this mode against critical open-source codebases [9,15].

Inputs to the Validation phase are the Execution-stage artifacts, plus the documented threat model and the partner-ecosystem interoperability requirements. Principal activities are interoperability testing against every partner endpoint the organization exchanges protocol traffic with, MTU and RF airtime testing against each bearer class identified in Planning, security testing, including red-team simulation of the cross-domain attack path illustrated in Figure 2, and regression testing of all downgrade-protection measures. Outputs are an interoperability matrix, a bearer-specific tolerance report, a red-team report covering at minimum the Figure-2 attack-pattern stages relevant to the organization's deployment, and a go/no-go recommendation per domain. Mythos-specific role: Mythos-class test-case generation is the primary defensive use of the model. The Frontier Red Team capability [9] that makes adversarial use dangerous is the same capability that makes defensive validation feasible at the scale PQC requires. Human-in-the-loop requirements: red-team findings require triage by qualified security engineers; the model produces candidate attack paths, not adjudicated risk. Gate criteria to advance: no open high-severity findings from the red-team pass, interoperability matrix above an agreed threshold (typically 99% for enterprise, higher for safety-critical OT), and bearer tolerance reports within the handshake-size budgets set in Planning.

6.5. Phase 5: Post-Migration Assurance

Assurance requires continuous monitoring, continuous red-teaming, and continuous protocol hardening. Mythos-class models can act as a continuous red-team engine and as a protocol-analysis engine, if access controls, audit logging, and use restrictions are in place.

Inputs to the Assurance phase are the production PQC deployment, the ongoing threat-intelligence feed, and the cryptographic-agility architecture delivered by Planning. Principal activities are continuous monitoring of protocol health and cryptographic-compliance metrics, continuous red-teaming that re-exercises the Figure-2 attack-pattern archetype as new capabilities become publicly available [10], continuous protocol hardening as weaknesses are discovered, and maintenance of the CBOM so that future primitives can be dropped in rather than bolted on. Outputs are operational: compliance reports, incident-response records, and a rolling backlog of hardening tasks. Mythos-specific role: Assurance is the phase where the defender-acceleration case for continued Mythos access is strongest, because the set of possible protocol configurations and adversary techniques grows faster than human review capacity. Human-in-the-loop requirements: every mitigation that reaches production passes through normal change management; Mythos output is a candidate set, never a direct production change. Gate criteria to sustain tracked metrics for mean-time-to-detect and mean-time-to-mitigate, a documented red-team re-run cadence aligned with public capability diffusion [10], and an agility-refresh cadence that ensures the next cryptographic transition is lower-cost than this one.

Figure 4 compares the phase concurrency of a traditional sequential migration against the Mythos-compressed parallel trajectory this lifecycle is designed to support. In the traditional model, each phase waits for the previous phase to finish, producing a five-to-ten-year cumulative duration [18,19], consistent with NIST's assessment that past cryptographic migrations have taken over a decade and that the PQC transition will likely take at least that long [23,24]. Under Mythos-class capability, the phases overlap: Discovery compresses and continues into Planning, Planning and Execution run in parallel across different domains, Validation begins on early-completing domains while Execution continues elsewhere, and Assurance starts as soon as the first production cut-over completes. Figure 4's adversary-capability window marks when open-weight models are projected to achieve parity with Mythos [10] and is the binding constraint on how long the compressed trajectory can be stretched.

7. Cost and Timeline Model

The cost and timeline model developed here translates the system dynamics of §5 into operational projections. Three cost drivers are identified (§7.1), a compressed-track migration timeline is proposed and bounded (§7.2), the methodology and limitations of the projection are

doi:10.20944/preprints202604.1744.v1

disclosed (§7.3), and an updated cost model appropriate to Mythos-era conditions is derived (§7.4). Figure 4 anchors the timeline argument with side-by-side traditional and compressed trajectories, and the adversary-capability window — set by the months-scale open-weight convergence horizon documented by CETaS [10] — defines the binding constraint on how long the compressed trajectory can be stretched.

7.1. Cost Drivers

Three cost drivers dominate PQC migration under Mythos-class adversarial pressure, and each multiplies the traditional cryptographic-transition cost signature in qualitatively distinct ways. The drivers operate on different lifecycle phases, concentrate in different domain archetypes, and require different executive responses. Their combined effect shapes the elevated-density profile visible in Figure 4's compressed-track summary panel and is the mechanism behind the updated cost model developed in §7.4.

Embedded-System Gravity

Embedded, firmware, and silicon-level cryptography impose cost burdens that scale non-linearly with migration scope. Firmware images typically require re-signing with updated root certificates, bootloaders must validate against new trust anchors, and hardware-backed keystores embedded in secure enclaves and tactical radios cannot be updated without vendor engagement (§4.3). Certification timelines compound the engineering cost: FIPS 140-3 module validation, NSA CNSA 2.0 audit cycles [4], and spectrum-allocation interoperability certification for tactical RF waveforms measure in years rather than months, and any firmware change that alters on-air signaling may require re-entry into the certification process. Hardware dependencies also gate cost directly — HSMs, trusted platform modules, and embedded secure elements with PQC-incompatible silicon require physical replacement on procurement and end-of-life cycles that are externally imposed and non-compressible (§7.3).

Simultaneity Pressure

Mythos-class capabilities compress adversarial timelines (§5.2), forcing defender migrations to run in parallel across domains that would traditionally have been sequenced. The cost consequence is a shift from baseline resource density to elevated peak density: staffing must scale to handle concurrent workstreams, validation infrastructure must operate continuously rather than at phase gates, and executive coordination overhead grows super-linearly with the number of simultaneous cutovers. The traditional sequential model — complete Discovery, then Planning, then Execution — amortizes specialist staff across phases; the compressed model requires discovery, planning, and execution teams operating concurrently on different domains, each at full allocation. This is the mechanism behind Figure 4's elevated-cost-density framing and the primary reason the compressed-track projection in §7.2 is not cheaper in aggregate even though it is shorter in calendar time.

Cross-Domain Complexity

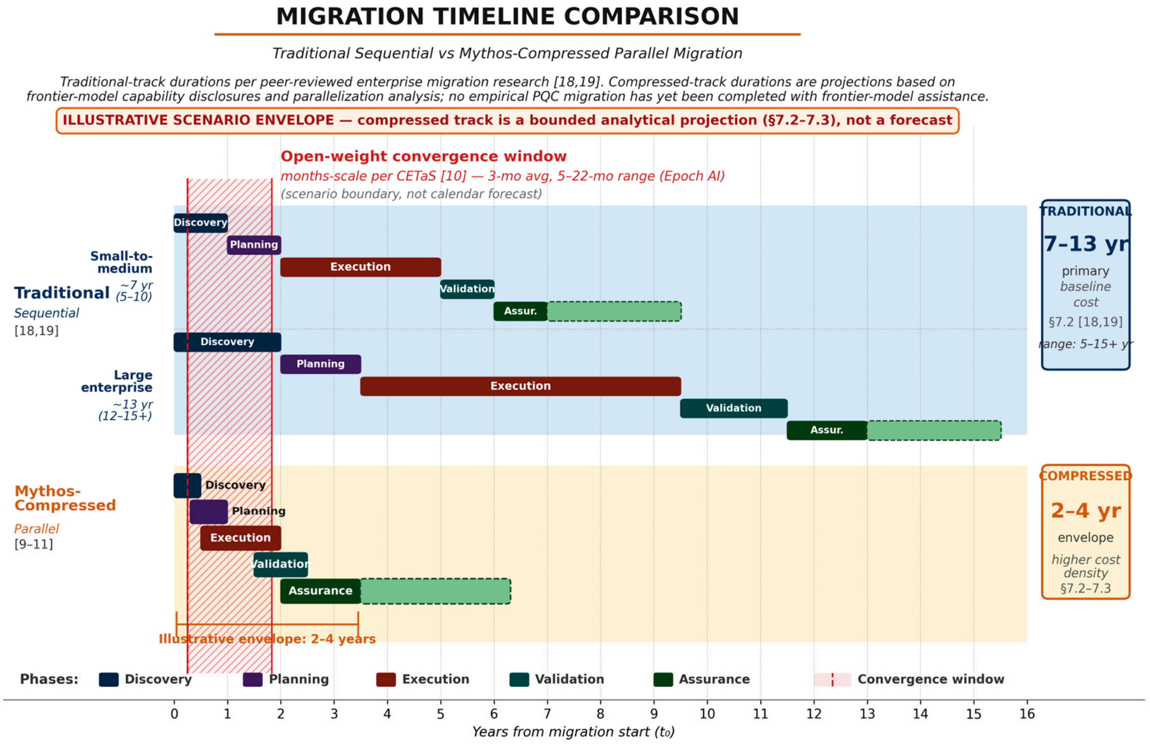
Cross-domain gateways and PKI identity systems incur high costs, as each change in a PQC primitive must be validated across all relevant protocols. Guard rule-set updates are byte-level work written against exact wire formats (§4.4), and the expanded message envelopes produced by ML-KEM and ML-DSA break parse boundaries that were stable under classical cryptography. PKI restructuring multiplies this: every issuing authority, intermediate, and end-entity certificate is a migration object, and hybrid PQC deployments require parallel trust chains until classical roots can be retired. The validation surface grows combinatorially with the product of primitive count, hybrid-mode count, and certificate-chain depth, which is why cross-domain complexity is the dominant line item in the §7.4 updated cost model rather than a secondary concern.

doi:10.20944/preprints202604.1744.v1

7.2. Timeline Compression

Traditional enterprise PQC migration estimates range from 5 to 10 years [18,19], within the decade-or-more timeframe characterized in NIST IR 8547 (Initial Public Draft) and the NCCoE Migration to PQC project [23,24]. Under Mythos-accelerated adversarial pressure [9–11], a 2–4-year effective window for migrating the most-exposed high-value systems is more consistent with the observed rate of offensive capability diffusion. This window is an analytic projection derived from the method described in §1.1: the lower bound anchors on the Frontier Red Team-documented capability already demonstrated [9]; the interior is constrained by the empirical migration cadence reported in [18,19]; and the upper bound reflects the latest plausible defender-only advantage horizon given the months-scale open-weight convergence framing documented by CETaS [10], which cites Epoch AI estimates of a three-month average capability lag between proprietary and open-weight frontier models (rising to five-to-twenty-two months in certain benchmarks). Because the specific cyber-offensive convergence time is not separately estimated in that analysis, the 4-year upper edge should be read as generous relative to CETaS's central estimate, not as a tight forecast. The transition from documented model capability [9–11] to the specific 2–4-year window is itself an analytical inference, informed by the compression-mechanism analysis in §7.3 and the traditional-baseline cadence in [18,19]; the window's specificity is a modeling choice about scenario bounds, not a number derivable from the cited capability evidence alone. The two trajectories are shown side-by-side in Figure 4, with the adversary-capability window — bounded above by this convergence horizon — providing the binding constraint on how long the compressed trajectory can be stretched. The projection does not imply that full migration can be completed in two to four years; rather, it implies that tiered prioritization — crown-jewel assets, downgrade-exposed bearers, and cross-domain gateways first — is now mandatory, and that the traditional sequential model is not feasible within the available window.

The traditional-track range in [18,19], consistent with the decade-scale cryptographic-migration timeframe characterized by NIST and the NCCoE Migration to PQC project [23,24], differentiates by organization size: small organizations typically require five to ten years for full PQC migration; large enterprises, with their higher touchpoint density, deeper PKI hierarchies, and greater cross-domain integration surface, typically require twelve to fifteen or more years. The compressed-track scenario envelope, therefore, locates different organization classes at different points within the 2–4-year range. A small federal civilian agency migrating a primarily cloud and enterprise portfolio — with a limited number of externally facing TLS terminators, a single PKI hierarchy, and minimal embedded or tactical RF exposure — can plausibly locate itself near the lower edge of the envelope, approximately two years, for its highest-exposure systems. A defense organization spanning tactical RF waveform stacks, satellite and space segments, multiple cross-domain gateways, and embedded cryptographic modules subject to independent certification cycles (§4.3, §7.1) will locate itself near the upper edge of the envelope, approximately four years, and that placement is achievable only with the governance restructuring developed in §8. The two-to-four-year window, therefore, is a scenario envelope that bounds where the highest-exposure subset of systems must land under Mythos-era adversarial pressure; it is not the calendar for completing migration in full, and it is neither a floor below which migration cannot be accelerated nor a ceiling above which migration cannot extend in practice — it is the range within which organizations across the size spectrum must locate their highest-exposure migration if they are to complete it before the adversary-capability window crosses (Figure 4).



Compressed phases overlap in parallel execution; vertical offset within the track shows simultaneity, not sequence.

Figure 4. PQC migration timeline and phase-concurrency comparison: traditional sequential migration against Mythos-compressed parallel migration. The traditional track (blue) shows two representative cases from the empirical range in [18,19] — a small-to-medium organization (~7 years, range 5–10) and a large enterprise (~13 years, range 12–15+). The compressed track (amber) shows the same five phases overlapping within the 2–4-year illustrative envelope; compressed-track durations are projections derived from [9–11] with methodology developed in §7.3. The hatched red column marks the open-weight convergence window per CETaS [10]; its position, breadth, and epistemic status are discussed in §7.2. Summary panels on the right show the corresponding duration ranges (7–13 years traditional, 2–4 years compressed envelope) and cross-reference the cost-model sections.

7.3. Methodology and Limitations of the Compressed-Track Projection

The traditional-track timeline in Figure 4 (5–10 years for small organizations; 12–15+ years for large) is empirically grounded in peer-reviewed enterprise migration research [18,19] and is consistent with the broader multi-year migration timeframe characterized by NIST IR 8547 (Initial Public Draft) and the NCCoE Migration to PQC project [23,24]. The compressed-track timeline (approximately three years, range 2–4) is a projection and warrants explicit methodological disclosure.

Mechanism. The compressed track does not assume autonomous frontier-model execution. It assumes AI-augmented teams operating under re-architected program governance, with compression deriving phase-by-phase from mechanisms of differing epistemic status. The five phases of the lifecycle in Section 6 compress unevenly: the strongest empirical case is in Discovery, the weakest empirical and strongest theoretical case is in Execution parallelism, and Assurance is not compressed in either track.

Discovery (strongest empirical case). Traditional cryptographic discovery — manual code review, protocol analysis, binary inspection, and vendor questionnaires — typically consumes approximately six to eighteen months at a large-enterprise scale, a practitioner-observed range synthesized in the enterprise-migration timeline analysis in [19] and used here as the traditional-baseline anchor against which AI-augmented compression is estimated; this range is an analytical inference from practitioner-synthesis work rather than an independent industry measurement or controlled benchmark, and readers should treat it accordingly. Frontier models have demonstrated the capability to read codebases at scale, identify cryptographic primitives, parse binaries, and produce Cryptographic Bills of Materials (CBOMs) programmatically. The Frontier Red Team benchmark [9] documents the underlying capability: Mythos Preview developed 181 working JavaScript shell exploits on Mozilla Firefox 147 (with register control in 29 additional trials) compared to two for Opus 4.6, and achieved tier-5 full control-flow hijack on ten separate fully patched targets across the OSS-Fuzz corpus where Sonnet 4.6 and Opus 4.6 each achieved only a single tier-3 crash and zero tier-5 hijacks. Both results require the same large-scale codebase reading and cryptographic logic identification that CBOM generation demands. This is the most empirically defensible of the five mechanisms: code-analysis throughput is directly measurable in AI-augmented pilots.

Planning and dependency mapping (moderate empirical case). Planning requires mapping which systems depend on which cryptographic primitives, identifying breakage modes, and sequencing migration safely — graph-reasoning work over complex dependency structures. Frontier models are demonstrably effective at cross-file reasoning and dependency analysis in general coding benchmarks, but PQC-specific planning workflows have not been directly measured. The capability transfer is plausible by analogy to neighboring software-engineering benchmarks, but empirically unverified for cryptographic-agility planning at enterprise scale.

Execution parallelism (weakest empirical case, strongest theoretical case). In the traditional sequential model, execution is bound by how many subsystems specialist teams can modify concurrently. A frontier model can, in principle, draft migration pull-requests for many subsystems simultaneously; the model is not the bottleneck on code production. The compression here is primarily theoretical: real serialization constraints — testing infrastructure capacity, staging environment availability, change-management approvals, certificate-authority coordination, and vendor release schedules — remain. The model speeds code generation, not these institutional processes.

Validation (moderate empirical case). Validation involves running tests, analyzing failures, and identifying regression paths. Frontier models demonstrably accelerate test generation and failure analysis for general-purpose software. Cryptographic validation, however, imposes domain-specific requirements — side-channel analysis, timing analysis, FIPS 140-3 validation — that are not general coding tasks. The compression here applies to the general-test-generation component but not to the cryptographic-specialist components, which remain human-gated and externally cadenced.

Assurance (not compressed). Assurance is a continuous post-migration activity in both trajectories. The compressed track reaches Assurance sooner but does not compress Assurance itself; Figure 4 shows Assurance extending beyond the primary-work window in both the traditional and compressed trajectories.

Non-compressible components. Three classes of critical-path activity are explicitly not compressed, and the projection does not assume their compression. Institutional change-management cycles — enterprise change advisory boards, production-change approval windows, coordinated rollout governance — operate on human-organizational clocks; frontier models do not accelerate. Regulatory and compliance timelines — FIPS 140-3 module validation, Authority to Operate renewals, CNSA 2.0 audit cadences, and the annual inventory and funding-assessment cycles established by OMB M-23-02 and the Quantum Computing Cybersecurity Preparedness Act [25] are externally imposed and time-bound. Hardware replacement schedules for cryptographic functionality embedded in silicon (HSMs, tactical-RF cryptographic modules, embedded IoT authentication) are bound by procurement and end-of-life cycles. The three-year point-estimate within the 2–4-year envelope reflects the duration achievable when AI-augmented software workstreams execute in parallel within institutional timelines that are themselves running concurrently, not sequentially; organizations facing more demanding certification, cross-domain integration, or hardware-replacement constraints locate themselves toward the four-year edge of the envelope, and organizations with narrower scope locate themselves toward the two-year edge.

Limitations. Four limitations bound the projection's epistemic status. (i) No empirical PQC migration has yet been completed with frontier-model assistance at enterprise scale; the compressed track is unobserved, and validation requires longitudinal case studies of AI-augmented migration programs that publish before/after throughput data. (ii) Capability transfer from benchmarked software-engineering tasks to PQC-specific sub-tasks (protocol downgrade analysis, cryptographic agility instrumentation, hybrid-deployment validation) is assumed by analogy to neighboring workflows; the transfer is plausible but unverified. (iii) The projection requires organizational readiness that most enterprises do not currently possess: inserting frontier-model assistance into an unchanged sequential-waterfall migration plan produces a fraction of the projected compression, and organizations that treat AI assistance as an add-on rather than a governance restructuring will likely realize timelines closer to the traditional track's lower bound. (iv) The same capability that enables the defender-side compression empirically compresses the offense side today, as documented in Section 5; if defenders do not adopt AI-augmented migration programs, Mythos-class models will compress only the offense side of the timing equation, widening rather than narrowing the defender–adversary gap.

Falsifiability. The projection makes falsifiable predictions suitable for empirical investigation: (a) discovery-phase duration in enterprises using AI-augmented cryptographic inventory should be measurable as a fraction of the 6–18 month traditional baseline, with early signal available from vendor-reported ACDI pilots under CISA's automated-inventory strategy [26]; (b) pull-request generation throughput for cryptographic-library replacement should be directly benchmarkable across AI-augmented and traditional engineering teams working comparable codebases; (c) rework rate (planning-to-execution defect leakage) should decline under AI-augmented programs if the simulation mechanism holds; (d) institutional-cycle duration (change-advisory, FIPS validation, ATO) should remain approximately constant across both tracks, and any observed compression there would indicate the projection is underestimating total compression rather than overestimating it. Organizations undertaking AI-augmented PQC migrations are encouraged to publish migration telemetry to enable empirical calibration of these projections in subsequent work.

7.4. Updated Cost Model

Costs increase due to:

- Accelerated timelines and the resulting concurrency premium.
- Increased testing at both interoperability and adversarial levels.
- Increased red-team requirements, including AI-assisted continuous testing.
- Cryptographic-agility investments that reduce the cost of the next transition.

The cost signature of a Mythos-compressed migration is therefore qualitatively different from the traditional sequential model: lower total calendar duration, higher peak resource density, and a materially larger validation and red-team line item than classical cryptographic transitions have required. Figure 4's summary panels make the comparison concrete — approximately seven years of primary work for small-to-medium organizations and approximately thirteen years for large enterprises at baseline resource density for the traditional trajectory, against the 2–4-year envelope at markedly higher resource density for the compressed trajectory — and Figure 2 illustrates the adversary side of the same ledger: testing-campaign costs under USD 20,000 in Anthropic's publicly disclosed examples [9] support the analytical inference, on a practitioner-grade comparison against standard-industry penetration-testing rates, of a cost order-of-magnitude for AI-augmented vulnerability discovery and exploit generation that shifts the defender's rational spending point upward on testing, validation, and continuous red-teaming; this is an economic-signal estimate, not a controlled per-exploit benchmark, because traditional and AI-augmented workflows produce different exploit classes against different target sets. The economic argument for the compressed trajectory is not that it is cheaper in aggregate but that it is the only trajectory that completes before the adversary-capability window in Figure 4 is crossed. Organizations that treat PQC migration as a classical cryptographic upgrade project, rather than a systems-engineering transformation under AI-accelerated adversarial pressure, will systematically under-resource the validation and assurance phases where the compressed trajectory's cost is concentrated.

8. Governance and Risk

The binding constraint developed in §5.3 — for the software-analytical phases where AI-augmentation applies — is the defender's organizational capacity to act on AI-accelerated analysis output at the tempo adversaries can sustain. Governance determines whether defenders can reach that capacity in time within the domains where it binds, while the non-compressible external cadence in embedded, regulated, and certification-bound domains continues to govern those parts of the migration regardless of frontier-model capability. Three governance concerns follow, each developed in the subsections below: access controls that determine who may use frontier-model capability, evaluation requirements that determine what must be verified before fielding, and red-team requirements that determine how capability is continuously stress-tested. Each concern inherits the core tension that restricting defender access in the name of safety also forecloses defensive use of the same capability [6,10,15].

8.1. Frontier-Model Access Controls

Anthropic's decision to restrict Mythos Preview to Project Glasswing partners [7,15] and to publish a system card for a non-generally available model [6] establishes a new operational pattern for frontier-model governance. The pattern requires four mutually reinforcing elements: controlled environments with attestable access controls, per-query logging with scope tagging adequate for post-facto review, model-use restrictions enforceable through both technical sandboxing and contractual terms, and coordinated disclosure windows that give defender organizations time to patch before findings propagate. The disclosure regime described in [9] uses SHA-3 hash commitments to preserve findings under coordinated disclosure without exposing operational detail, and the pattern is replicable by federal operators and other frontier labs. CETaS [10] and the World Economic Forum [12] both note that this regime has a limited horizon: as open-weight models converge on similar capability, access controls designed around commercial frontier labs lose their protective effect against adversaries who operate outside the commercial regime. Access-control governance is therefore a bridging mechanism rather than a durable solution, and the remaining time it provides must be spent building the evaluation and red-team capacity developed below.

8.2. Evaluation Requirements

Evaluations for frontier-model deployment in PQC migration contexts must cover four axes, each addressing a different failure mode. PQC-specific test suites are the first requirement and must cover ML-KEM, ML-DSA, SLH-DSA [1–3], and classical/PQC hybrid-deployment modes; evaluations that validate only individual primitives miss the hybrid-mode failure modes that dominate real deployment. Protocol-analysis benchmarks are the second requirement and must span the protocol families where PQC primitives land in practice: TLS, IPsec, SSH, DNSSEC, and the tactical RF waveform profiles described in §4.3. Exploit-generation controls are the third requirement; scaffold-level restrictions on autonomous execution, sandboxed runtimes, and per-target rate limits are necessary to prevent evaluation infrastructure from being repurposed as an offensive capability generator, a concern amplified by the benchmark evidence in [9]. Alignment evaluations consistent with Anthropic's Responsible Scaling Policy v3 [8] are the fourth requirement and must be extended beyond general-purpose alignment cases to cover cryptographic-tampering scenarios in which a model is asked to weaken rather than strengthen crypto during an apparent migration task, since the same capability that performs Validation Loop simulations can, under adversarial prompting, produce downgrade vectors against its own outputs.

8.3. Red-Team Requirements

Red-team requirements under Mythos-class conditions must shift from point-in-time engagements to continuous testing against an evolving PQC implementation surface. The compressed-track trajectory in Figure 4 is iterative: migrations proceed domain by domain, hybrid deployments evolve as classical roots are retired, and certificate chains shorten over time. A red team that tests the deployment once at the end of Execution cannot surface the transitional failure modes that exist only during the handoff between hybrid phases. Red-teams must also model attack paths that span the full set of reachable domains rather than the one under immediate migration; Figure 2 illustrates the cross-domain attack-pattern structure that single-domain red teams cannot reproduce, and §5.2 develops the Attack-Path Generation Loop as the adversary-side dynamic the red-team function is designed to counter. PQC-specific adversarial testing completes the requirement set: downgrade attacks against hybrid deployments, signature-substitution probes, and key-exchange downgrade sequences that exploit the transitional window in which both classical and PQC primitives are in production simultaneously. Together, these three requirements — continuity, cross-domain scope, and PQC-specific adversarial testing — define the red-team posture that completes the governance regime introduced in §8.1.

9. Conclusion

Mythos-class frontier models fundamentally alter PQC migration. They accelerate defensive migration through automated analysis while simultaneously accelerating adversarial exploitation. PQC migration becomes a race condition between system redesign and AI-accelerated attack development. This paper has provided a systems-engineering architecture, a lifecycle model, and a cost model to guide PQC migration under Mythos-class AI conditions. The central operational implication is domain-conditional: for the software-analytical phases of PQC migration — cryptographic discovery, protocol decomposition, exploit-chain reasoning, and code modification — the binding constraint has shifted from cryptographic availability toward the capacity of defender organizations to act on the output of AI-accelerated analysis at the tempo that adversaries, once they obtain equivalent capability through open-weight successors or leakage, will be able to sustain. For embedded, regulated, and externally governed domains — FIPS 140-3 module validation, Authority to Operate renewals, CNSA 2.0 audit cadences, embedded cryptographic hardware replacement, and vendor certification cycles — the binding constraint remains non-compressible external cadence, which frontier-model capability does not accelerate and which the compressed-track scenario envelope in §7.2 does not assume to compress.

Three findings anchor this conclusion. First, the Mythos benchmark evidence [9] — ten tier-5 control-flow hijacks on fully patched open-source targets, 181 working exploits on Mozilla Firefox 147 against two for Opus 4.6, and testing-campaign costs under USD 20,000 in Anthropic's publicly disclosed scaffolds and campaigns [9] — supports the analytical inference, developed in §5.2 and §7.4 on a practitioner-grade comparison against standard-industry penetration-testing and bug-bounty economics, that the cost-per-exploit signature of the adversary side of the equation has shifted by approximately one order of magnitude; a controlled benchmark against traditional programs is not attempted because the workflows produce different exploit classes against different target sets. Second, the compressed-track projection (2–4 years for highest-exposure systems) against the traditional 5–10 years for small organizations and 12–15+ years for large enterprises [18,19] — within the decade-scale migration timeframe characterized by NIST and the NCCoE [23,24] — is a scenario envelope derived from explicit bounding conditions, not a forecast, and is bounded explicitly in §7.3. Third, the governance regime developed in §8 is a bridging mechanism rather than a durable solution; its horizon is defined by open-weight capability diffusion rather than by policy choice.

Four limitations bound these findings and are developed in full in §7.3. No empirical PQC migration has yet been completed with frontier-model assistance at enterprise scale; capability transfer from benchmarked software-engineering tasks to PQC-specific workflows is plausible by analogy but unverified; realizing the compressed trajectory requires organizational restructuring that most enterprises do not currently possess; and if defenders do not adopt AI-augmented migration, Mythos-class models will compress only the offense side of the timing equation, widening rather than narrowing the defender–adversary gap. Future empirical work should prioritize longitudinal case studies of AI-augmented migration programs publishing before and after throughput data against the three compression mechanisms identified in §7.3 and should benchmark PQC-specific sub-task performance on the protocol families (TLS, IPsec, SSH, DNSSEC, tactical RF) named in §4.2.

The operational implication for federal program managers, enterprise CISOs, and critical-infrastructure operators is that PQC migration can no longer be planned as a sequential cryptographic upgrade running alongside business as usual. It must be planned as a tiered, parallelized, AI-augmented systems-engineering transformation operating under compressed calendar time, with explicit governance restructuring to absorb frontier-model capability into the program rather than layering it on top. Organizations that wait for Mythos-class capability to become generally available will find themselves operating against adversaries who have already obtained equivalent capability through open-weight successors, at which point the race condition modeled here will have resolved against them.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new datasets were generated; the analysis relies on publicly available documents.

Conflicts of Interest: The author declares no conflict of interest.

References

1. National Institute of Standards and Technology. FIPS 203: Module-Lattice-Based Key-Encapsulation Mechanism Standard (ML-KEM); NIST: Gaithersburg, MD, USA, 2024.
2. National Institute of Standards and Technology. FIPS 204: Module-Lattice-Based Digital Signature Standard (ML-DSA); NIST: Gaithersburg, MD, USA, 2024.
3. National Institute of Standards and Technology. FIPS 205: Stateless Hash-Based Digital Signature Standard (SLH-DSA); NIST: Gaithersburg, MD, USA, 2024.
4. National Security Agency. Announcing the Commercial National Security Algorithm Suite 2.0 (CNSA 2.0); NSA Cybersecurity Advisory; NSA: Fort Meade, MD, USA, 2022.
5. Office of Management and Budget. OMB M-23-02: Migrating to Post-Quantum Cryptography; Executive Office of the President: Washington, DC, USA, November 2022.
6. Anthropic. System Card: Claude Mythos Preview. 7 April 2026. Available online: https://www.anthropic.com/claude-mythos-preview-system-card (accessed 22 April 2026).
7. TechCrunch. Anthropic debuts preview of powerful new AI model Mythos in new cybersecurity initiative. 7 April 2026. Available online: https://techcrunch.com/2026/04/07/anthropic-mythos-ai-model-preview-security/ (accessed 22 April 2026).
8. Anthropic. Alignment Risk Update: Claude Mythos Preview. 7 April 2026. Available online: https://www.anthropic.com/claude-mythos-preview-risk-report (accessed 22 April 2026).
9. Carlini, N.; Cheng, N.; Lucas, K.; Moore, M.; Nasr, M.; Prabhushankar, V.; Xiao, W.; et al. (Anthropic Frontier Red Team). Assessing Claude Mythos Preview's Cybersecurity Capabilities. 7 April 2026. Available online: https://red.anthropic.com/2026/mythos-preview (accessed 22 April 2026).
10. Centre for Emerging Technology and Security (CETaS), Alan Turing Institute. Claude Mythos: What Does Anthropic's New Model Mean for the Future of Cybersecurity? April 2026. Available online: https://cetas.turing.ac.uk/publications/claude-mythos-future-cybersecurity (accessed 22 April 2026).
11. AI Security Institute (AISI). Our Evaluation of Claude Mythos Preview's Cyber Capabilities; UK Department for Science, Innovation and Technology: London, UK, 13 April 2026. Available online: https://www.aisi.gov.uk/blog/our-evaluation-of-claude-mythos-previews-cyber-capabilities (accessed 22 April 2026).

12. World Economic Forum. Anthropic's Mythos moment: how frontier AI is redefining cybersecurity. April 2026. Available online: https://www.weforum.org/stories/2026/04/anthropic-mythos-ai-cybersecurity/ (accessed 22 April 2026).
13. Fortune. Anthropic says testing Mythos, powerful new AI model, after accidental data leak reveals its existence. 26 March 2026. Available online: https://fortune.com/2026/03/26/anthropic-says-testing-mythos-powerful-new-ai-model-after-data-leak-reveals-its-existence-step-change-in-capabilities/ (accessed 22 April 2026).
14. Newton, C. Why Anthropic's new model has cybersecurity experts rattled. Platformer, April 2026. Available online: https://www.platformer.news/anthropic-mythos-cybersecurity-risk-experts/ (accessed 22 April 2026).
15. Anthropic. Project Glasswing. 7 April 2026. Available online: https://www.anthropic.com/project/glasswing (accessed 22 April 2026).
16. National Institute of Standards and Technology. NIST SP 800-208: Recommendation for Stateful Hash-Based Signature Schemes; NIST: Gaithersburg, MD, USA, 2020.
17. Cybersecurity and Infrastructure Security Agency. Zero Trust Maturity Model, Version 2.0; CISA: Arlington, VA, USA, April 2023.
18. Campbell, R. Synchronizing Concurrent Security Modernization Programs: Zero Trust, Post-Quantum Cryptography, and AI Assurance. Systems 2026, 14, 233. https://doi.org/10.3390/systems14030233
19. Campbell, R. Enterprise Migration to Post-Quantum Cryptography: Timeline Analysis and Strategic Frameworks. Computers 2026, 15, 9. https://doi.org/10.3390/computers15010009
20. Glazunov, S.; Brand, M.; Project Zero; DeepMind. From Naptime to Big Sleep: Using Large Language Models to Catch Vulnerabilities in Real-World Code. Google Project Zero, 1 November 2024. Available online: https://googleprojectzero.blogspot.com/2024/10/from-naptime-to-big-sleep.html (accessed 22 April 2026).
21. Bhatt, M.; Chennabasappa, S.; Nikolaidis, C.; Wan, S.; Evtimov, I.; Gabi, D.; Song, D.; Ahmad, F.; Aschermann, C.; Fontana, L.; et al. Purple Llama CyberSecEval: A Secure Coding Benchmark for Language Models. arXiv 2023, arXiv:2312.04724. https://doi.org/10.48550/arXiv.2312.04724
22. Defense Advanced Research Projects Agency (DARPA). AI Cyber Challenge (AIxCC) Final Competition Results; DARPA: Arlington, VA, USA, 8 August 2025. Available online: https://www.darpa.mil/news/2025/aixcc-results (accessed 22 April 2026).
23. Moody, D.; Perlner, R.; Regenscheid, A.; Robinson, A.; Cooper, D. Transition to Post-Quantum Cryptography Standards; NIST Internal Report (IR) 8547 (Initial Public Draft); National Institute of Standards and Technology: Gaithersburg, MD, USA, November 2024. https://doi.org/10.6028/NIST.IR.8547.ipd
24. National Cybersecurity Center of Excellence (NCCoE). Migration to Post-Quantum Cryptography Project; NIST: Gaithersburg, MD, USA, 2022–2026. Available online: https://www.nccoe.nist.gov/applied-cryptography/migration-to-pqc (accessed 22 April 2026).
25. U.S. Congress. Quantum Computing Cybersecurity Preparedness Act; Public Law 117-260; U.S. Government Publishing Office: Washington, DC, USA, 21 December 2022. Available online: https://www.congress.gov/117/plaws/publ260/PLAW-117publ260.pdf (accessed 22 April 2026).
26. Cybersecurity and Infrastructure Security Agency. Strategy for Migrating to Automated Post-Quantum Cryptography Discovery and Inventory Tools; CISA: Arlington, VA, USA, August 2024. Available online: https://www.cisa.gov/resources-tools/resources/strategy-migrating-automated-post-quantum-cryptography-discovery-and-inventory-tools (accessed 22 April 2026).