Zero-Trust AI: A Quantum-Resilient Framework for the Enterprise

Welcome to AI Unraveled, your daily briefing on the real-world business impact of AI.

Introduction: The Converging Storm of AI and Quantum Threats

We stand at a pivotal moment in technological history. Artificial Intelligence is no longer a futuristic concept but a core enterprise asset, driving innovation in fields from drug discovery and financial modeling to logistics and aerospace.[1] This new paradigm, however, introduces a new, highly sophisticated attack surface, fundamentally different from traditional software vulnerabilities.[4] Simultaneously, the dawn of quantum computing promises to shatter the very foundations of our current digital security, rendering the cryptographic algorithms that protect our global economy obsolete.[5]

This report dissects a dual-pronged, existential threat to enterprise AI. The first prong consists of intrinsic AI vulnerabilities—attacks that exploit the unique logic and lifecycle of machine learning models themselves, turning their greatest strengths into critical weaknesses.[9] The second is the quantum decryption threat: the impending obsolescence of classical cryptography that protects AI models, their invaluable training data, and the communication channels they depend upon.[5]

Legacy, perimeter-based security is fundamentally obsolete in this new era.[14] The "trust but verify" model, where entities inside the network are implicitly trusted, is a catastrophic liability when faced with AI's distributed nature and the looming power of quantum decryption. To secure the future of AI, enterprises must architect a new security paradigm from the ground up: a **Quantum-Resilient Zero-Trust AI** framework. This framework is not a single product but a strategic approach built upon three foundational pillars: **Zero-Trust MLOps**, **Verifiable Provenance**, and **Crypto-Agility & Quantum Resistance**.

Part 1: The New Attack Surface: AI in a Post-Quantum World

This section establishes the complete threat landscape, moving from current, tangible risks to

the more complex, future-facing dangers posed by the convergence of AI and quantum computing.

Section 1.1: The Intrinsic Vulnerabilities of Enterprise AI

Before an organization can protect its AI assets, it must first understand how they can be compromised. Unlike traditional software vulnerabilities such as buffer overflows or SQL injections, AI models possess a unique set of weaknesses intrinsically tied to their statistical nature and their profound reliance on data.[4] These are not mere bugs in code; they are exploitable features of the machine learning process itself.

Data Poisoning

Data poisoning is an insidious attack that targets the very foundation of an AI model: its training data.

- **Mechanism:** In a data poisoning attack, an adversary intentionally injects malicious, manipulated, or mislabeled data into a model's training dataset.[9] This corruption can occur through various vectors, including insider attacks by malicious employees with legitimate data access, supply chain compromises where third-party data sources are tainted, or unauthorized access to data pipelines via phishing or lateral movement within a network.[17] The attack occurs during the model's training phase, shaping how it learns from the very beginning.[18]
- **Impact:** The consequences of a poisoned model can range from subtle to catastrophic. A successful attack can introduce systemic biases, degrade the model's overall accuracy, or create dangerous real-world misclassifications, such as causing an autonomous vehicle's model to mistake a stop sign for a yield sign.[19] More alarmingly, data poisoning can be used to implant hidden backdoors—vulnerabilities that remain dormant during testing and only activate when presented with a specific, attacker-defined trigger post-deployment.[18] The infamous case of Microsoft's Tay chatbot, which was trained on public Twitter data and quickly began generating offensive content after being targeted by trolls, serves as a stark real-world example of a model being poisoned by its data inputs.[16]

Model Inversion and Inference Attacks

These attacks exploit a trained model to leak sensitive information about the private data it was trained on, representing a severe breach of data confidentiality.

- **Mechanism:** Adversaries with query access to a model can use its outputs to reverse-engineer and reconstruct sensitive information from the original training data.[9]

This category includes several distinct attack types. **Membership inference** aims to determine if a specific individual's data was part of the training set.[10] **Attribute inference** goes further, attempting to deduce sensitive attributes (e.g., medical conditions, financial status) about the data subjects.[24] These attacks are particularly effective against "overfitting" models, which have essentially memorized portions of their training data rather than learning generalized patterns.[23]

- **Impact:** A successful model inversion attack can lead to catastrophic privacy violations, exposing highly sensitive personal information such as medical records, financial transactions, or even facial images.[23] This not only causes immense harm to individuals but also exposes the organization to severe regulatory penalties under frameworks like GDPR and other data protection laws.[23] Furthermore, these attacks can be used to expose corporate trade secrets or copyrighted material that was inadvertently included in the training data.[23]

Evasion Attacks (Adversarial Examples)

Evasion attacks are designed to fool a fully trained and deployed model at the moment of inference, causing it to make a specific, incorrect prediction.

- **Mechanism:** An attacker makes subtle, often human-imperceptible modifications to input data to cause the model to produce a wildly incorrect output.[9] These crafted inputs are known as "adversarial examples." The perturbations can be as minor as changing a few pixels in an image or adding invisible noise to an audio file.[10] These attacks can be **targeted**, designed to force a specific incorrect classification (e.g., misidentifying a specific person), or **non-targeted**, aiming to cause any incorrect output to degrade the model's reliability.[10]
- **Impact:** In high-stakes, safety-critical systems, the consequences of evasion attacks can be dire. Researchers have famously demonstrated the ability to cause an AI system in a self-driving car to misclassify a stop sign as a speed limit sign by applying a few strategically placed stickers.[9] Similar attacks could compromise facial recognition systems, medical diagnostic tools, or fraud detection models, leading to physical harm, financial loss, and a complete erosion of trust in AI-driven systems.[10]

Model Theft and IP Leakage

This attack vector targets the AI model itself as a valuable piece of intellectual property.

- **Mechanism:** Also known as model extraction, this attack involves an adversary repeatedly querying a model's publicly accessible API to gather enough input-output pairs to train a functionally equivalent replica, or "clone," of the model.[9] The attacker does not need access to the model's source code, architecture, or training data; they

steal its learned functionality through black-box interaction.[9]

- **Impact:** Model theft represents a direct and significant loss of competitive advantage and valuable intellectual property.[9] A successfully stolen model can be reverse-engineered, analyzed for weaknesses, or incorporated into a competitor's product offering, undermining the massive investment required to develop and train enterprise-grade AI.[4]

LLM-Specific Threats

The rise of Large Language Models (LLMs) has introduced new, conversation-based attack vectors.

- **Prompt Injection:** Malicious instructions are cleverly embedded within user prompts to bypass the LLM's safety controls and guardrails.[4] A successful prompt injection can trick the model into revealing sensitive data, generating malicious content, or executing unauthorized actions on behalf of the user.[9]
- **Hallucination Abuse:** Threat actors can exploit an LLM's tendency to "hallucinate" or generate plausible but false information. By registering domains, creating fake scholarly articles, or seeding the web with content that aligns with potential hallucinations, attackers can legitimize and amplify misinformation, poisoning the information ecosystem that both humans and future AIs rely on.[4]

The following table provides a consolidated overview of these AI-specific attack vectors and their direct impact on the enterprise.

Attack Vector	Mechanism	Target Stage (MLOps)	Enterprise Impact
Data Poisoning	Injecting malicious or mislabeled data into the training set.[9]	Data Collection & Training	Compromised decision-making, systemic bias, hidden backdoors, reputational damage, regulatory fines.[4]
Model Inversion & Inference	Reverse-engineering model outputs to reconstruct sensitive training data.[9]	Inference	Catastrophic data breaches (PII, PHI), exposure of trade secrets, severe regulatory violations (GDPR), loss of customer trust.[23]
Evasion Attacks	Making subtle changes to input data to cause misclassification at	Inference	Physical safety hazards (autonomous systems), financial fraud, failed

	inference time.[10]		security authentication, compromised medical diagnoses.[9]
Model Theft / IP Leakage	Repeatedly querying a model's API to create a functional replica.[9]	Deployment & Inference	Direct loss of valuable intellectual property, erosion of competitive advantage, enabling competitor analysis of model weaknesses.[4]
Prompt Injection	Embedding malicious instructions in user prompts to bypass safety controls.[9]	Inference	Unauthorized data exfiltration, execution of malicious code, generation of harmful content, bypassing of safety protocols.[4]

Section 1.2: The Quantum Decryption Threat: Shattering Classical Defenses

The AI-specific vulnerabilities detailed above are often mitigated by a foundational layer of classical security controls, which themselves depend on the presumed strength of modern encryption. The advent of quantum computing poses a direct and existential threat to this foundation, threatening to neutralize the cryptographic protections that safeguard AI models, data, and communications.[7]

The Arrival of "Q-Day"

The day a quantum computer can break today's standard public-key encryption is often referred to as "Q-Day." This event is predicated on the practical implementation of a specific quantum algorithm.

- **Shor's Algorithm:** In 1994, mathematician Peter Shor developed a quantum algorithm that, when executed on a sufficiently powerful quantum computer, can solve the two core mathematical problems that underpin virtually all modern public-key cryptography: integer factorization and the computation of discrete logarithms.[5] This means that widely used asymmetric algorithms—including RSA, Elliptic Curve Cryptography (ECC), Diffie-Hellman (DH), ECDH, and ECDSA—will be rendered insecure.[5] A calculation that would take the world's most powerful classical supercomputers billions of years to complete could be solved by a cryptographically relevant quantum computer (CRQC) in

a matter of hours or days.[6]

- **Timeline:** While a CRQC capable of running Shor's algorithm at scale does not exist today, the consensus among experts is that such a machine is no longer a distant theoretical possibility. Projections place its arrival within the next decade, with many estimates converging around the 2030-2035 timeframe.[5] However, the threat is not defined by the date a CRQC is switched on; the danger is already present.

"Harvest Now, Decrypt Later" (HNDL): The Immediate Danger

The most critical and immediate quantum threat facing enterprises today is a strategy known as "Harvest Now, Decrypt Later" (HNDL).

- **Mechanism:** This long-term attack strategy involves adversaries exfiltrating and storing large volumes of encrypted data *today*.[13] These threat actors, often nation-states with significant resources, are not attempting to decrypt the data now. Instead, they are patiently stockpiling it, betting on the future arrival of a CRQC to unlock this harvested trove of information.[37]
- **Targeted AI Assets:** For enterprises investing heavily in AI, the prime targets for HNDL attacks are high-value, long-lifecycle data assets whose confidentiality must be maintained for years or even decades. This includes:
    - **Proprietary Training Data:** The curated datasets that represent an organization's core data advantage. This could be sensitive customer personally identifiable information (PII), financial records, patient health information (PHI), or geological survey data.[32]
    - **Trained Model Weights and Parameters:** The final, trained model is the distilled intellectual property of the organization. Exposing these parameters would allow competitors to perfectly replicate a model that cost millions to develop.
    - **Corporate Trade Secrets:** Sensitive strategic documents, M&A plans, or proprietary research and development data that might be fed into an internal LLM for analysis are high-value targets.[32]
    - **Secure Communications:** Encrypted communications channels (using protocols like HTTPS and VPNs) that carry discussions about AI strategy, development progress, and vulnerability management are also targets for harvesting.[7]

The HNDL threat model forces a fundamental paradigm shift in how organizations must approach data security. It introduces the concept that all encrypted data now has a potential "expiration date." The security of data can no longer be assessed solely by the strength of its current protection; it must be evaluated based on its required confidentiality lifetime versus the projected arrival of a CRQC. For an AI model trained on sensitive medical data that must remain confidential for 20 years to comply with regulations, protection via classical public-key cryptography is already insufficient. The security protecting that data has a "best before" date that may have already passed, creating an undeniable and immediate urgency to adopt quantum-resistant cryptography for all long-term, high-value AI assets.

Section 1.3: The Quantum-AI Multiplier Effect

The relationship between artificial intelligence and quantum computing is not one-sided. While quantum computing threatens the cryptographic foundation of AI security, the two technologies can also be combined to create more powerful and sophisticated attacks against AI models themselves.[38] This convergence creates a multiplier effect, amplifying existing risks and introducing entirely new classes of threats.

Quantum Machine Learning (QML) as an Offensive Tool

The same properties that make quantum computing a powerful tool for scientific discovery can be co-opted by adversaries to enhance their attack capabilities.

- **Accelerated Malicious AI:** Quantum computing's inherent ability to process vast, complex datasets and solve difficult optimization problems can be used to accelerate the training of malicious AI models.[1] An attacker with access to quantum resources could potentially develop more effective malware, discover novel software vulnerabilities faster than defenders can patch them, or generate ultra-realistic deepfakes for advanced social engineering and disinformation campaigns with unprecedented speed and scale.[40]

Quantum Adversarial Machine Learning (QAML): A New Class of Threat

Quantum Adversarial Machine Learning (QAML) is a nascent but critical field of research that studies the vulnerabilities of quantum machine learning systems. Crucially, it also explores how quantum algorithms could be leveraged to conduct more effective attacks against classical machine learning models.[46]

- **Potential Mechanisms:** While much of this research is still in the theoretical and early experimental stages, it suggests that quantum algorithms may be able to explore the high-dimensional, complex parameter spaces of neural networks more efficiently than classical methods.[53] This could enable the generation of adversarial examples that are more subtle, more potent, or require far fewer queries to the target model, making them harder to detect and defend against. Early proof-of-concept studies have already demonstrated that quantum classifiers are vulnerable to adversarial examples, mirroring the same fundamental weaknesses found in their classical counterparts.[47]
- **Future Threat:** This line of research implies a significant evolution in the threat landscape. Future AI defenses cannot be limited to robust cryptography alone; they must also anticipate that the very nature of adversarial attacks will become more sophisticated, potentially becoming quantum-enhanced.

This convergence of AI and quantum computing creates a complex, self-reinforcing feedback loop. Attackers may one day use quantum computing to enhance AI-driven attacks (QAML), while defenders will look to leverage quantum-enhanced AI for more powerful threat detection and response.[40] This dynamic elevates the cybersecurity arms race from the familiar domain of classical software to a new, quantum-physical level. Long-term security strategies must therefore account for an adversary who can not only decrypt classically protected data but also probe and manipulate model behavior with a level of sophistication previously unimaginable. This necessitates building defenses that are not only cryptographically agile but also algorithmically agile, capable of adapting to new and unforeseen attack vectors born from this powerful technological convergence.

Part 2: The Three Pillars of Quantum-Resilient Zero-Trust AI

In the face of the complex and converging threats outlined in Part 1, a new security paradigm is required. This section presents an actionable framework for defense, directly addressing the identified threats through a holistic strategy built on three essential and interdependent pillars.

Pillar 1: Zero-Trust MLOps — Securing the AI Lifecycle

The traditional "castle-and-moat" security model, which trusts entities once they are inside the network perimeter, is fundamentally broken and dangerously inadequate for modern IT environments.[14] A Zero-Trust Architecture (ZTA) provides the necessary replacement, operating on the core principle of "never trust, always verify".[14] It assumes no implicit trust and continuously validates every user, device, and application at every access request. Applying this rigorous philosophy to the entire Artificial Intelligence lifecycle—a practice known as MLOps (Machine Learning Operations)—is the first and most critical line of defense against both intrinsic AI vulnerabilities and external threats.[64] This pillar focuses on securing the *process* and *environment* where AI is developed, trained, deployed, and maintained.

Applying Core Zero-Trust Tenets to MLOps

- **Continuous Verification ("Never Trust, Always Verify"):** In a Zero-Trust MLOps pipeline, every entity—whether a data scientist, an automated CI/CD service, a data pipeline, or an inference API—must be strictly and continuously authenticated and authorized for every single action it attempts to perform. Implicit trust based on network location is eliminated.

    - **Implementation:** This is achieved through the enforcement of strong, risk-based multi-factor authentication (MFA) for all human users and the use of workload identity federation for non-human services.[4] For example, a CI/CD runner authenticates using a short-lived token to access a specific resource rather than using a static, long-lived secret key that could be compromised.[69]
- **Least-Privilege Access:** This principle mandates that every entity is granted the absolute minimum level of permissions required to perform its specific, authorized task—and nothing more.
    - **Implementation:** Access controls must be granular and context-aware. A data labeling service should only have read-access to the raw, unprocessed data, not the trained model or the production environment. An inference server should have execute-only access to the model artifact, preventing it from modifying the model's weights.[17] This is enforced through a combination of Role-Based Access Control (RBAC), which defines permissions based on job function, and Attribute-Based Access Control (ABAC), which allows for dynamic, context-sensitive policies (e.g., granting access based on time of day, device health, and geographic location).[67]
- **Micro-segmentation:** This involves logically dividing the MLOps environment into small, isolated security zones to limit the "blast radius" of a potential breach.
    - **Implementation:** The data ingestion pipeline, the experimental sandbox for data scientists, the model training environment, the model registry, and the production deployment servers should all reside in separate, isolated network segments.[59] If a vulnerability is exploited in an open-source library used during data preprocessing, micro-segmentation contains the breach to that specific segment, preventing the attacker from moving laterally to compromise the central model registry or exfiltrate sensitive production data.[65]
- **Assume Breach:** This mindset shift requires architecting the entire system with the assumption that an attacker is already inside the network. Security focus moves from prevention alone to rapid detection and response.
    - **Implementation:** This necessitates comprehensive, real-time monitoring and logging of all activities across every segment of the MLOps pipeline.[4] Advanced, AI-powered security analytics tools are used to establish baselines of normal behavior and automatically detect anomalies. For instance, such a system could flag a data scientist who suddenly attempts to download an entire training dataset at an unusual hour, or an inference API that is being queried with a high frequency of near-identical inputs—a pattern indicative of a model inversion or theft attack.[59]

The implementation of a Zero-Trust framework is not merely a generic security best practice; it serves as a direct and powerful operational antidote to many of the specific AI vulnerabilities discussed previously. There is a clear causal relationship between the failures of traditional, perimeter-based security models and the success of these novel attacks. For instance, data poisoning attacks, which often rely on an attacker gaining access to modify training data, are directly mitigated by the principles of least-privilege access and

micro-segmentation, which prevent a compromised component or user from accessing and writing to the core training dataset repository.[17] Similarly, model theft via API querying is made significantly more difficult by continuous verification with strong identity controls and rate limiting, while the associated anomalous query patterns can be detected by a system built on the "assume breach" principle.[14] By hardening the entire lifecycle, Zero-Trust MLOps provides a targeted, highly effective defense against the unique ways in which AI systems are attacked.

Pillar 2: Verifiable Provenance — Building an Immutable Chain of Trust

While Zero-Trust MLOps secures the *environment* and controls *access*, a critical question remains: how can the organization trust the *artifacts* themselves? The threats of data poisoning and surreptitious model tampering highlight the need for an unbreakable, cryptographically verifiable audit trail for every asset that moves through the AI pipeline. This is the essence of verifiable provenance—creating an immutable chain of trust for data and models.

Blockchain-Enabled Audit Trails

Blockchain technology, with its inherent properties of decentralization, immutability, and transparency, offers a powerful tool for establishing verifiable provenance in MLOps.

- **Mechanism:** By leveraging a permissioned (private) blockchain, an organization can create a shared, tamper-evident ledger for its entire AI lifecycle.[72] Every significant event—such as the ingestion of a new dataset, the completion of a preprocessing step, the successful training of a model, the creation of a new model version, and its deployment to production—is recorded as a transaction on the chain, complete with a secure timestamp.[75]
- **Implementation:** Storing large datasets or model files directly on a blockchain is impractical. Instead, cryptographic hashes (unique digital fingerprints) of these assets are recorded on-chain.[76] This provides an immutable record that can be used for integrity verification. At any point, an auditor or an automated system can re-calculate the hash of a deployed model and compare it to the hash stored on the blockchain for that version. A mismatch proves that the asset has been tampered with.[72] Furthermore, smart contracts—self-executing code on the blockchain—can be used to automate compliance checks, such as verifying that a model was trained on an approved dataset before allowing it to be deployed.[72]

Secure Multi-Party Computation (SMPC) for Privacy-Preserving AI

In scenarios involving collaborative AI development or the use of sensitive data from multiple

sources, SMPC provides a cryptographic method for computation without centralized trust.

- **Mechanism:** SMPC is a subfield of cryptography that enables multiple parties to jointly compute a function over their combined private data without any of the parties having to reveal their individual inputs to one another.[78]
- **As a Zero-Trust Tool:** SMPC is the technological embodiment of the "never trust, always verify" principle applied at the data level. It allows for collaboration in zero-trust environments. For example, a consortium of hospitals could collaboratively train a powerful cancer detection model on their collective patient data. Using SMPC, the model could learn from all the data without any single hospital—or any central server—ever gaining access to the sensitive, unencrypted patient records from the other institutions. While historically computationally expensive, recent advancements are making SMPC increasingly practical for complex AI models like Transformers and LLMs, enabling secure and private inference and training.[79]

These advanced cryptographic technologies are the technological manifestation of Zero Trust for the assets themselves. While the principles of Zero Trust define *how* access should be controlled, blockchain and SMPC provide the cryptographic proof needed to *verify* the integrity and confidentiality of the data and models. The verification step in Zero Trust is thus extended from just the identity of the user or service to the very artifacts they are interacting with. Traditional audit logs can be altered or deleted by a privileged attacker, but a blockchain-based log is immutable, providing a single, non-repudiable source of truth for auditing and forensics.[72] Similarly, SMPC enables collaboration without requiring trust in the other parties, using the cryptographic protocol itself to enforce the rules of engagement. These technologies are not mere add-ons; they are the enforcement mechanisms that transform Zero Trust for AI from a policy-based ideal into a cryptographically guaranteed reality.

Pillar 3: Crypto-Agility & Quantum Resistance — Future-Proofing the Foundation

The first two pillars, while powerful, are built upon a foundation of cryptography used for authentication, digital signatures, and data encryption. If that cryptographic foundation can be shattered by a quantum computer, the entire security structure collapses. This third pillar ensures that the bedrock of the Zero-Trust AI framework is quantum-resilient by design and agile enough to adapt to future threats.

Mandating Post-Quantum Cryptography (PQC)

The transition to quantum-resistant cryptography is no longer a theoretical exercise; it is an active, ongoing global initiative.

- **The NIST Standards:** The U.S. National Institute of Standards and Technology (NIST)

has completed its multi-year process to solicit and standardize a new generation of public-key cryptographic algorithms. In 2024, it published the first final standards: **ML-KEM** (formerly CRYSTALS-Kyber) for Key Encapsulation Mechanisms (general encryption), and **ML-DSA** (formerly CRYSTALS-Dilithium) and **SLH-DSA** (formerly SPHINCS+) for digital signatures.[8] These algorithms are based on different families of mathematical problems, such as those found in lattice-based and hash-based cryptography, which are believed to be computationally difficult for both classical and quantum computers to solve.[3]

- **Implementation:** A quantum-resilient framework mandates that all cryptographic operations within the MLOps pipeline and the broader enterprise must migrate to these NIST-approved PQC standards. This includes securing data in transit with PQC-enabled TLS, encrypting data at rest, digitally signing all software artifacts (code, containers, models), and authenticating users and services.[82]

Crypto-Agility as a Core Principle

Given that the field of post-quantum cryptography is still relatively new, it is crucial to build systems that can adapt if new vulnerabilities are discovered.

- **Concept:** Crypto-agility is the technical and architectural capability of a system to switch out cryptographic algorithms, keys, and protocols quickly and efficiently without requiring a complete system redesign.[33] The PQC landscape will continue to evolve. A crypto-agile architecture ensures that if a weakness is discovered in ML-KEM in five years, the organization can seamlessly transition to a new standard with minimal operational disruption.
- **Roadmaps:** The urgency and feasibility of this transition are being demonstrated by major technology companies and government bodies. Microsoft, for example, has published a quantum-safe roadmap targeting full transition by 2033, with early adoption starting in 2029.[97] NIST is also providing detailed transition guidance to help organizations plan their migration.[100]

PQC-Adapted Hardware Security Modules (HSMs)

HSMs are the physical root of trust for cryptographic keys, and their adaptation to the post-quantum era is critical.

- **Role of HSMs:** HSMs are dedicated hardware devices that securely generate, manage, and store cryptographic keys, performing sensitive operations within a hardened, tamper-resistant boundary.[96]
- **PQC Challenges & Adaptations:** PQC algorithms often have significantly larger key and signature sizes compared to their classical counterparts. This places new demands on the limited storage and computational resources of HSMs. To address this, modern

PQC-ready HSMs are being designed with new strategies, such as storing a much smaller cryptographic "seed" and deriving the full private key on-demand within the secure boundary. This approach balances the need for security with performance and storage constraints.[102] Utilizing PQC-adapted HSMs is essential for protecting the master private keys used for signing models, issuing certificates, and securing the entire Zero-Trust framework against both current and future threats.[96]

Ultimately, the entire Zero-Trust model hinges on the ability to cryptographically verify identities, data integrity, and secure communications. The "verify" step is not a matter of policy or opinion; it is a concrete cryptographic operation, such as checking a digital signature or completing a TLS handshake. As established, the classical algorithms currently used for these operations are demonstrably vulnerable to a future quantum adversary. Therefore, building a security framework on classical cryptography is akin to building on a foundation that is known to be crumbling. One cannot truly "verify" with a broken tool. Post-Quantum Cryptography provides the resilient algorithms needed to ensure the long-term integrity of these verification processes. Integrating PQC into every authentication, signing, and encryption operation is not just an upgrade; it is an absolute prerequisite for a Zero-Trust model to have any lasting meaning or validity in the quantum era. It future-proofs the very act of verification itself.

Conclusion: Building the Quantum-Ready AI Enterprise

The technological landscape is being reshaped by two powerful and converging forces. Artificial Intelligence has become a primary driver of enterprise value, but it has also introduced a novel and complex attack surface. Simultaneously, the steady advance of quantum computing is placing the classical cryptographic systems that protect our digital world on an expiring timeline. The converged threat is clear: AI models are high-value targets with unique vulnerabilities, and the tools we use to protect them are becoming obsolete.

In this new reality, incremental security updates and perimeter-based defenses are no longer sufficient. The only viable path forward is a proactive, multi-layered strategy that rebuilds security from the ground up on a foundation of explicit trust. This report has outlined such a strategy, built upon three essential and mutually reinforcing pillars:

1. **Zero-Trust MLOps:** Securing the AI development and deployment lifecycle by enforcing continuous verification, least-privilege access, and micro-segmentation.
2. **Verifiable Provenance:** Establishing an immutable, cryptographically-guaranteed chain of trust for all AI assets—from data to models—using technologies like blockchain and Secure Multi-Party Computation.
3. **Crypto-Agility & Quantum Resistance:** Future-proofing the entire security foundation by migrating to NIST-standardized Post-Quantum Cryptography and building systems that can adapt to the threats of tomorrow.

Adopting this framework is not merely a technical update; it is a fundamental business strategy. It is about protecting the crown jewels of the 21st-century enterprise: its data and its intelligence. Organizations that begin the journey to build this resilience now will not only secure themselves against the clear and present danger of "Harvest Now, Decrypt Later" attacks and the sophisticated AI threats of the future, but will also establish a foundation of trust that enables safer, more ambitious, and more powerful AI innovation.[8]
The time to act is now. The threat is active, the standards are available, and industry leaders are already executing their transition roadmaps.[97] The journey to a quantum-resilient, Zero-Trust posture for AI begins today with a comprehensive inventory of all cryptographic assets and AI systems, and a strategic commitment from leadership to embed these three pillars into the very fabric of the organization's technology stack and security culture.

Works cited

1. Harnessing the complementary power of AI and Quantum Computing | Global Policy Watch, accessed on October 15, 2025, https://www.globalpolicywatch.com/2025/10/harnessing-the-complementary-power-of-ai-and-quantum-computing/
2. The Relationship Between AI and Quantum Computing | CSA - Cloud Security Alliance, accessed on October 15, 2025, https://cloudsecurityalliance.org/blog/2025/01/20/quantum-artificial-intelligence-exploring-the-relationship-between-ai-and-quantum-computing
3. Quantum Leap in Finance: Economic Advantages, Security, and Post-Quantum Readiness - arXiv, accessed on October 15, 2025, https://arxiv.org/html/2508.21548v1
4. AI Security: Using AI Tools to Protect Your AI Systems - Wiz, accessed on October 15, 2025, https://www.wiz.io/academy/ai-security
5. Quantum Computing and the Risk to Classical Cryptography, accessed on October 15, 2025, https://www.appviewx.com/blogs/quantum-computing-and-the-risk-to-classical-cryptography/
6. Shor's Algorithm - Classiq, accessed on October 15, 2025, https://www.classiq.io/insights/shors-algorithm
7. What Is Quantum Computing's Threat to Cybersecurity? - Palo Alto Networks, accessed on October 15, 2025, https://www.paloaltonetworks.com/cyberpedia/what-is-quantum-computings-threat-to-cybersecurity
8. Future-proofing cybersecurity: Understanding quantum-safe AI and how to create resilient defenses - ITPro, accessed on October 15, 2025, https://www.itpro.com/technology/future-proofing-cybersecurity-understanding-quantum-safe-ai-and-how-to-create-resilient-defences
9. Top 10 AI Security Risks (and How to Protect Your Systems ..., accessed on October 15, 2025, https://mindgard.ai/blog/top-ai-security-risks
10. What Are Adversarial AI Attacks on Machine Learning? - Palo Alto Networks,

accessed on October 15, 2025, https://www.paloaltonetworks.com/cyberpedia/what-are-adversarial-attacks-on-AI-Machine-Learning
11. Decrypting the Future: Quantum Computing and The Impact of Grover's and Shor's Algorithms on Classical Cryptography - EasyChair, accessed on October 15, 2025, https://easychair.org/publications/preprint/NJCx/open
12. Shor's Algorithm – Quantum Computing's Breakthrough in Factoring - SpinQ, accessed on October 15, 2025, https://www.spinquanta.com/news-detail/Shor-s-Algorithm-Quantum-Computing-s-Breakthrough-in-Factoring
13. Warning: Quantum computers to soon crack modern encryption - Information Age | ACS, accessed on October 15, 2025, https://ia.acs.org.au/article/2025/warning--quantum-computers-to-soon-crack-modern-encryption.html
14. What is Zero Trust? - Guide to Zero Trust Security - CrowdStrike, accessed on October 15, 2025, https://www.crowdstrike.com/en-us/cybersecurity-101/zero-trust-security/
15. 10 Zero Trust Solutions for 2025 - SentinelOne, accessed on October 15, 2025, https://www.sentinelone.com/cybersecurity-101/identity-security/zero-trust-solutions/
16. Adversarial Machine Learning - CLTC UC Berkeley Center for Long-Term Cybersecurity, accessed on October 15, 2025, https://cltc.berkeley.edu/aml/
17. What is AI data poisoning? - Cloudflare, accessed on October 15, 2025, https://www.cloudflare.com/learning/ai/data-poisoning/
18. What Is Data Poisoning? [Examples & Prevention] - Palo Alto Networks, accessed on October 15, 2025, https://www.paloaltonetworks.com/cyberpedia/what-is-data-poisoning
19. What Is Data Poisoning? | IBM, accessed on October 15, 2025, https://www.ibm.com/think/topics/data-poisoning
20. What Is Data Poisoning? - CrowdStrike, accessed on October 15, 2025, https://www.crowdstrike.com/en-us/cybersecurity-101/cyberattacks/data-poisoning/
21. What Is Adversarial Machine Learning? Types of Attacks & Defenses - DataCamp, accessed on October 15, 2025, https://www.datacamp.com/blog/adversarial-machine-learning
22. ML03:2023 Model Inversion Attack | OWASP Foundation, accessed on October 15, 2025, https://owasp.org/www-project-machine-learning-security-top-10/docs/ML03_2023-Model_Inversion_Attack
23. Model inversion and membership inference: Understanding new AI security risks and mitigating vulnerabilities - Hogan Lovells, accessed on October 15, 2025, https://www.hoganlovells.com/en/publications/model-inversion-and-membership-inference-understanding-new-ai-security-risks-and-mitigating-vulnerabilities
24. 7 Serious AI Security Risks and How to Mitigate Them | Wiz, accessed on October 15, 2025, https://www.wiz.io/academy/ai-security-risks

25. Model Inversion: The Essential Guide | Nightfall AI Security 101, accessed on October 15, 2025, https://www.nightfall.ai/ai-security-101/model-inversion
26. www.paloaltonetworks.com, accessed on October 15, 2025, https://www.paloaltonetworks.com/cyberpedia/what-are-adversarial-attacks-on-AI-Machine-Learning#:~:text=An%20adversarial%20AI%20attack%20is,changes%20to%20the%20input%20data.
27. Adversarial Attacks Explained (And How to Defend ML Models Against Them) - Medium, accessed on October 15, 2025, https://medium.com/sciforce/adversarial-attacks-explained-and-how-to-defend-ml-models-against-them-d76f7d013b18
28. Cybersecurity in the Quantum Era: Assessing the Impact of Quantum Computing on Infrastructure - arXiv, accessed on October 15, 2025, https://arxiv.org/html/2404.10659v1
29. From Quantum Hacks to AI Defenses – Expert Guide to Building ..., accessed on October 15, 2025, https://thehackernews.com/2025/09/from-quantum-hacks-to-ai-defenses.html
30. Cyber Insights 2025: Quantum and the Threat to Encryption - SecurityWeek, accessed on October 15, 2025, https://www.securityweek.com/cyber-insights-2025-quantum-and-the-threat-to-encryption/
31. How Quantum Computing Will Upend Cybersecurity | BCG - Boston Consulting Group, accessed on October 15, 2025, https://www.bcg.com/publications/2025/how-quantum-computing-will-upend-cybersecurity
32. Harvest Now, Decrypt Later: The Quantum Threat That Has Already Begun - CYPFER, accessed on October 15, 2025, https://cypfer.com/harvest-now-decrypt-later-the-quantum-threat-that-has-already-begun/
33. Quantum Computing Threat Forces Crypto Revolution in 2025 | eSecurity Planet, accessed on October 15, 2025, https://www.esecurityplanet.com/cybersecurity/quantum-computing-threat-forces-crypto-revolution-in-2025/
34. theNET | Future-proofing using post-quantum cryptography - Cloudflare, accessed on October 15, 2025, https://www.cloudflare.com/the-net/security-signals/post-quantum-era/
35. What Is Post-Quantum Cryptography? | NIST, accessed on October 15, 2025, https://www.nist.gov/cybersecurity/what-post-quantum-cryptography
36. Quantum is coming — and bringing new cybersecurity threats with it - KPMG International, accessed on October 15, 2025, https://kpmg.com/xx/en/our-insights/ai-and-technology/quantum-and-cybersecurity.html
37. [2503.15678] Cyber Threats in Financial Transactions -- Addressing the Dual Challenge of AI and Quantum Computing - arXiv, accessed on October 15, 2025, https://arxiv.org/abs/2503.15678
38. 'Harvest Now, Decrypt Later' Attacks in the Post-Quantum, AI Era ..., accessed on

October 15, 2025, https://www.eetimes.eu/harvest-now-decrypt-later-attacks-in-the-post-quantum-and-ai-era/
39. Understanding the 'Harvest Now, Decrypt Later' Threat and the Protective Shield of Microsharding - ShardSecure, accessed on October 15, 2025, https://shardsecure.com/blog/understanding-the-harvest-now-decrypt-later-threat-and-the-protective-shield-of-microsharding
40. Quantum computing's potential impact on AI and cybersecurity - Delinea, accessed on October 15, 2025, https://delinea.com/blog/quantum-computing-the-impact-on-ai-and-cybersecurity
41. The Growing Impact Of AI And Quantum On Cybersecurity - Cognitive World, accessed on October 15, 2025, https://cognitiveworld.com/articles/2025/08/10/the-growing-impact-of-ai-and-quantum-on-cybersecurity
42. Understanding the Impact of Quantum Computing and AI on Cybersecurity - Wisconsin Bankers Association, accessed on October 15, 2025, https://www.wisbank.com/understanding-the-impact-of-quantum-computing-and-ai-on-cybersecurity/
43. How can AI and quantum computing improve cybersecurity, and what are the challenges?, accessed on October 15, 2025, https://www.researchgate.net/post/How_can_AI_and_quantum_computing_improve_cybersecurity_and_what_are_the_challenges
44. What Is Quantum Computing? | IBM, accessed on October 15, 2025, https://www.ibm.com/think/topics/quantum-computing
45. The Emerging Potential for Quantum Computing in Irregular Warfare, accessed on October 15, 2025, https://irregularwarfarecenter.org/publications/insights/the-emerging-potential-for-quantum-computing-in-irregular-warfare/
46. A schematic illustration of quantum adversarial machine learning. (a ..., accessed on October 15, 2025, https://www.researchgate.net/figure/A-schematic-illustration-of-quantum-adversarial-machine-learning-a-A-quantum_fig1_343509626
47. Universal adversarial examples and perturbations for quantum classifiers - PMC, accessed on October 15, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC9796671/
48. Quantum adversarial machine learning | Phys. Rev. Research, accessed on October 15, 2025, https://link.aps.org/doi/10.1103/PhysRevResearch.2.033212
49. [2001.00030] Quantum Adversarial Machine Learning - arXiv, accessed on October 15, 2025, https://arxiv.org/abs/2001.00030
50. Quantum Adversarial Machine Learning: Status, Challenges and Perspectives, accessed on October 15, 2025, https://www.computer.org/csdl/proceedings-article/tps-isa/2020/854300a128/1qyxBYROyXK
51. Quantum Adversarial Machine Learning for Defence and Military Systems -

ADSTAR Summit, accessed on October 15, 2025, https://admin.adstarsummit.com.au/uploads/Usman_M_Quantum_Adversarial_Machine_Learning_for_Defence_and_Military_Systems_61a113746a.pdf
52. Quantum Machine Learning - arXiv, accessed on October 15, 2025, https://arxiv.org/html/2506.12292v1
53. Quantum Generative Adversarial Learning - DSpace@MIT, accessed on October 15, 2025, https://dspace.mit.edu/handle/1721.1/117204
54. Quantum machine learning - Wikipedia, accessed on October 15, 2025, https://en.wikipedia.org/wiki/Quantum_machine_learning
55. Quantum-Enhanced Machine Learning for Cybersecurity: Evaluating Malicious URL Detection - MDPI, accessed on October 15, 2025, https://www.mdpi.com/2079-9292/14/9/1827
56. Future of Quantum Computing - arXiv, accessed on October 15, 2025, https://arxiv.org/html/2506.19232v1
57. Accelerating the drive towards energy-efficient generative AI with quantum computing algorithms - arXiv, accessed on October 15, 2025, https://arxiv.org/html/2508.20720v1
58. Quantum algorithms: A survey of applications and end-to-end complexities - arXiv, accessed on October 15, 2025, https://arxiv.org/abs/2310.03011
59. Zero-Trust AI Security Framework Protecting Enterprise Data - Agentra, accessed on October 15, 2025, https://www.agentra.io/blog/ai-security/zero-trust-ai-security-framework/
60. What Is Zero Trust? | IBM, accessed on October 15, 2025, https://www.ibm.com/think/topics/zero-trust
61. What Are the Core Principles of Zero Trust Security? - Own Data, accessed on October 15, 2025, https://www.owndata.com/blog/what-are-the-core-principles-of-zero-trust-security
62. 7 Core Principles of Zero Trust Security | NordLayer Learn, accessed on October 15, 2025, https://nordlayer.com/learn/zero-trust/principles/
63. NIST Zero Trust: Principles, Components & How to Get Started - Tigera, accessed on October 15, 2025, https://www.tigera.io/learn/guides/zero-trust/nist-zero-trust/
64. Security for MLOps: How to Safeguard Data, Models, and Pipelines Against Modern AI Threats - DS Stream, accessed on October 15, 2025, https://www.dsstream.com/post/security-for-mlops-how-to-safeguard-data-models-and-pipelines-against-modern-ai-threats
65. Fortifying MLOps: A Guide to Protecting Pipelines, Models, and Data from AI-Specific Threats | by DS STREAM | Sep, 2025 | Medium, accessed on October 15, 2025, https://medium.com/@ds_stream/fortifying-mlops-a-guide-to-protecting-pipelines-models-and-data-from-ai-specific-threats-8dbc220038f2
66. (PDF) Securing AI-Driven MLOps Pipelines with Zero-Trust ..., accessed on October 15, 2025, https://www.researchgate.net/publication/388659706_Securing_AI-Driven_MLOp

s_Pipelines_with_Zero-Trust_Architectures
67. Zero Trust MLOps with OpenShift Platform Plus - Red Hat, accessed on October 15, 2025, https://www.redhat.com/en/blog/zero-trust-mlops-with-openshift-platform-plus
68. Understanding AI risks and how to secure using Zero Trust - LevelBlue, accessed on October 15, 2025, https://levelblue.com/blogs/security-essentials/understanding-ai-risks-and-how-to-secure-using-zero-trust
69. Implementing Zero Trust in CI/CD Pipelines: A Secure DevOps ..., accessed on October 15, 2025, https://medium.com/globant/implementing-zero-trust-in-ci-cd-pipelines-a-secure-devops-approach-37324d84870b
70. Best Practices for Implementing Zero Trust Architecture in DevOps Pipelines - Collabnix, accessed on October 15, 2025, https://collabnix.com/best-practices-for-implementing-zero-trust-architecture-in-devops-pipelines/
71. Stop Model Inversion and Inference Attacks Before They Start - Galileo AI, accessed on October 15, 2025, https://galileo.ai/blog/prevent-model-inversion-inference-attacks
72. Enhancing MLOps with Blockchain: Decentralized Security for AI Pipelines - International Journal of Artificial Intelligence, Data Science, and Machine Learning, accessed on October 15, 2025, https://ijaidsml.org/index.php/ijaidsml/article/download/168/148
73. Enhancing MLOps with Blockchain: Decentralized Security for AI Pipelines - ResearchGate, accessed on October 15, 2025, https://www.researchgate.net/publication/393031206_Enhancing_MLOps_with_Blockchain_Decentralized_Security_for_AI_Pipelines
74. Innovative Journal of Applied Science Auditable AI Pipelines: Logging and Verifiability in ML Workflows, accessed on October 15, 2025, https://ijas.meteorpub.com/1/article/download/133/59
75. Blockchain for Provenance and Traceability in 2025 - ScienceSoft, accessed on October 15, 2025, https://www.scnsoft.com/blockchain/traceability-provenance
76. (PDF) Blockchain-enabled Audit Trails for AI Models - ResearchGate, accessed on October 15, 2025, https://www.researchgate.net/publication/395415248_Blockchain-enabled_Audit_Trails_for_AI_Models
77. Data Provenance on the Blockchain: Establishing Trust and Traceability in a Digital World, accessed on October 15, 2025, https://tokenminds.co/blog/knowledge-base/data-provenance-on-the-blockchain
78. Secure Multi-Party Computation for Machine Learning: A Survey - ResearchGate, accessed on October 15, 2025, https://www.researchgate.net/publication/379843467_Secure_Multi-Party_Computation_for_Machine_Learning_A_Survey
79. Secure Multiparty Generative AI - arXiv, accessed on October 15, 2025,

https://arxiv.org/html/2409.19120v1
80. Daily Papers - Hugging Face, accessed on October 15, 2025, https://huggingface.co/papers?q=Secure%20Multi-Party%20Computation%20(MPC)
81. MPC-Minimized Secure LLM Inference - OpenReview, accessed on October 15, 2025, https://openreview.net/forum?id=beAIX6RjsW
82. Post-quantum cryptography (PQC) | Google Cloud, accessed on October 15, 2025, https://cloud.google.com/security/resources/post-quantum-cryptography
83. Post-quantum cryptography - Wikipedia, accessed on October 15, 2025, https://en.wikipedia.org/wiki/Post-quantum_cryptography
84. Post-Quantum Cryptography and Quantum-Safe Security: A Comprehensive Survey - arXiv, accessed on October 15, 2025, https://arxiv.org/html/2510.10436v1
85. Hybrid Horizons: Policy for Post-Quantum Security - arXiv, accessed on October 15, 2025, https://arxiv.org/html/2510.02317v1
86. A Survey of Post-Quantum Cryptography Support in Cryptographic Libraries - arXiv, accessed on October 15, 2025, https://arxiv.org/abs/2508.16078
87. [2507.21151] NIST Post-Quantum Cryptography Standard Algorithms Based on Quantum Random Number Generators - arXiv, accessed on October 15, 2025, https://arxiv.org/abs/2507.21151
88. NIST Post-Quantum Cryptography Standardization - Wikipedia, accessed on October 15, 2025, https://en.wikipedia.org/wiki/NIST_Post-Quantum_Cryptography_Standardization
89. NIST's first post-quantum standards - The Cloudflare Blog, accessed on October 15, 2025, https://blog.cloudflare.com/nists-first-post-quantum-standards/
90. IR 8547, Transition to Post-Quantum Cryptography Standards | CSRC, accessed on October 15, 2025, https://csrc.nist.gov/pubs/ir/8547/ipd
91. NIST Releases First 3 Finalized Post-Quantum Encryption Standards, accessed on October 15, 2025, https://www.nist.gov/news-events/news/2024/08/nist-releases-first-3-finalized-post-quantum-encryption-standards
92. NIST's PQC Standardization Explained - Full Conversation - YouTube, accessed on October 15, 2025, https://www.youtube.com/watch?v=E5GrgcyoA3o
93. Post-Quantum Cryptography Review in Future Cybersecurity Strengthening Efforts | Scientific Journal of Engineering Research - PT. Teknologi Futuristik Indonesia, accessed on October 15, 2025, https://journal.futuristech.co.id/index.php/sjer/article/view/35
94. Post-Quantum Cryptography (PQC) Meets Quantum AI (QAI), accessed on October 15, 2025, https://postquantum.com/post-quantum/pqc-quantum-ai-qai/
95. Zero Trust and PQC Build a Stronger Security Foundation - GDIT, accessed on October 15, 2025, https://www.gdit.com/perspectives/latest/zero-trust-and-pqc-build-a-stronger-security-foundation/
96. Post-Quantum Crypto Agility - Thales, accessed on October 15, 2025, https://cpl.thalesgroup.com/encryption/post-quantum-crypto-agility
97. Microsoft Sets Roadmap for Security in the Quantum Era - Voices For Innovation,

accessed on October 15, 2025, https://www.voicesforinnovation.org/executive_briefing/microsoft-sets-roadmap-for-security-in-the-quantum-era/
98. Post-quantum resilience: building secure foundations - Microsoft On ..., accessed on October 15, 2025, https://blogs.microsoft.com/on-the-issues/2025/08/20/post-quantum-resilience-building-secure-foundations/
99. Microsoft Maps Path to a Quantum-Safe Future, accessed on October 15, 2025, https://thequantuminsider.com/2025/08/20/microsoft-maps-path-to-a-quantum-safe-future/
100. NIST Cybersecurity Center Outlines Roadmap for Secure Migration - The Quantum Insider, accessed on October 15, 2025, https://thequantuminsider.com/2025/09/19/nist-cybersecurity-center-outlines-roadmap-for-secure-migration/
101. Post-Quantum Cryptography HSM - Securosys, accessed on October 15, 2025, https://www.securosys.com/en/hsm/post-quantum-cryptography
102. Adapting HSMs for Post-Quantum Cryptography - IETF, accessed on October 15, 2025, https://www.ietf.org/archive/id/draft-reddy-pquip-pqc-hsm-00.html