A Zero-Trust AI-Blockchain Architecture for Quantum-Secure Metaverse Platforms

Gabriel Silva Atencio ✉

Latin American University of Science and Technology (ULACIT), San José, Costa Rica. https://orcid.org/0000-0002-4881-181X; gsilvaa468@ulacit.ed.cr

Abstract

The expansion of the Metaverse presents security risks that conventional perimeter-based protections are unable to mitigate. This research introduces and experimentally substantiates a Zero-Trust Architecture (ZTA) that incorporates a federated ResNet-50 model for anomaly detection, a decentralized identification system based on Hyperledger Fabric using zk-SNARKs, and CRYSTALS-Kyber for quantum-resistant key exchange. The architecture attains a 42.7% decrease in False Acceptance Rate (5.2% compared to a 9.1% baseline, $p < 0.01$), 99.1% resilience to Sybil attacks, and partial compliance with GDPR using cryptographic erasure proofs. Quantum security is based on the Module-LWE issue, requiring 2^187 operations (i.e., 2 to the power of 187), which surpasses NIST Level 3 standards, and incurs a latency overhead of 1.2 times and an energy consumption increase of 28.9% compared to AES-256. AI inference needs 3.1 times more GPU resources. Four innovative contributions are introduced: Integrated ZTA empirical validation, FAR reduction benchmark, zk-SNARKs for GDPR compliance, and human-centric validation indicating that usability predicts adoption ($\beta = 0.47$, $p < 0.001$). The architecture offers a scalable, experimentally substantiated basis for protecting decentralized virtual environments.

Keywords

artificial intelligence; blockchain; metaverse; post-quantum cryptography; security architecture; zero-trust.

Introduction

The combination of augmented reality (AR), virtual reality (VR), and decentralized economics is speeding up the growth of the Metaverse, a shared virtual area that will change how people interact socially and economically. Experts in the field think that the Metaverse industry will be worth more than $1.5 trillion by 2030 [1, 2]. But the main architectural ideas behind this new frontier—decentralization, interoperability, and persistent immersion—naturally make the attack surface bigger, which creates new security holes. These include advanced identity spoofing, quantum decryption attacks, and adversarial artificial intelligence (AI) manipulations, which standard perimeter-based security approaches, dependent on static trust assumptions, are inherently unable to counter [3, 4]. This significant deficiency requires a transition to Zero-Trust Architectures (ZTAs), which implement the concept of least privilege and need ongoing verification [5, 6].

Current security assessments for Metaverse platforms exhibit three primary deficiencies: (a) insufficient empirical validation in authentic adversarial environments, (b) lack of cohesive frameworks that amalgamate AI, blockchain, and post-quantum cryptography into a singular testable architecture, and (c) neglect of regulatory compliance (e.g., GDPR) in conjunction with technical security metrics.

Three interconnected constraints compromise existing security frameworks for the Metaverse. Centralized Identity and Access Management (IAM) systems provide single points of failure, a weakness that facilitated the $625 million Ronin Bridge attack [7, 8]. While decentralized identifiers (DIDs) provide an alternative, current DID implementations do not adhere to Zero Trust Architecture (ZTA) requirements for scalable, real-time verification. The emergence of quantum computing compromises conventional cryptography standards, such as Rivest–Shamir–Adleman (RSA) and Elliptic-Curve Cryptography (ECC), due to Shor's algorithm, jeopardizing the long-term secrecy and integrity of sensitive data[9-11]. Third, AI-driven threats, including deepfake-facilitated social engineering assaults in immersive virtual reality settings,

can circumvent static, rule-based intrusion detection systems [12, 13].

This study examines the research question: *How can an integrated Zero-Trust Architecture, incorporating AI-driven threat detection, blockchain-based identity management, and post-quantum cryptography, enhance security, usability, and regulatory compliance in Metaverse platforms?*

In response, this study proposes and empirically validates a novel Zero-Trust AI-Blockchain architecture. This research presents four significant additions to the current literature, each targeting a unique identified deficiency.

Contribution 1 – Federated learning for Metaverse-specific threat detection: A federated ResNet-50 model, trained on 50,000 Metaverse-specific adversarial samples (FGSM, ε=0.05), diminishes the False Acceptance Rate (FAR) against deepfake-based biometric spoofing attacks by 42.7% relative to rule-based systems (9.1% → 5.2%, $p < 0.01$).

Contribution 2 – GDPR-compliant decentralized identity with zk-SNARKs: A decentralized identification system based on Hyperledger Fabric attains 99.1% resistance to Sybil attacks and, for the first time within a Metaverse security framework, incorporates zk-SNARKs to deliver cryptographic erasure proofs—facilitating partial adherence to the GDPR's "right to be forgotten" while preserving blockchain immutability.

Contribution 3 – Operational benchmarks for the integration of post-quantum cryptography: Empirical performance evaluations of CRYSTALS-Kyber (NIST Standard) within a Metaverse testbed quantify latency overhead (1.2× compared to AES-256), throughput reduction (9,800 versus 12,400 TPS), and energy expenditures (+28.9% per 10K transactions), offering actionable migration insights for practitioners.

Contribution 4 – Human-centric validation of security-usability trade-offs: A mixed-methods evaluation (Delphi study, n=20; user surveys) reveals that perceived usability is a significant predictor of adoption intention ($\beta = 0.47$, $p < 0.001$), contesting the notion that rigorous security measures inherently diminish user experience in virtual environments.

A mixed-methods approach, including thorough penetration testing, a Delphi expert consensus (n=20), and user surveys, is used to extensively evaluate the architectural effectiveness.

Related work

The Metaverse's fast growth has sparked a lot of academic interest in its unique security problems. This is especially true since its immersive and decentralized designs provide new attack vectors that standard perimeter-based defenses can't handle. Recent research has advanced in three main areas: Zero-Trust Security (ZTS) frameworks, AI-driven anomaly detection, and blockchain-based identity management. Nonetheless, a thorough examination uncovers enduring shortcomings in empirical validation, regulatory consistency, and the comprehensive incorporation of various technologies into a cohesive, interoperable framework—shortcomings that this research seeks to rectify.

Zero-Trust Security models, which put the "never trust, always verify" philosophy into action, have been shown to work in cloud-native systems by reducing attack surfaces by up to 68% [3, 4]. However, using them in the Metaverse adds a whole new level of difficulty. The rise of AI-generated fake media, such as hyper-realistic avatars made using generative adversarial networks (GANs), makes continuous authentication techniques harder to use [14, 15]. Also, the basic cryptographic primitives that enable numerous trust assumptions, such RSA-2048 and ECC, may be broken by quantum computers using Shor's algorithm. This means that the research needs to switch to post-quantum cryptographic (PQC) alternatives [9-11]. Current formal frameworks for ZTA micro-segmentation and least-privilege access [5, 6] often neglect Metaverse-specific threat models, such as cross-platform lateral movement and deepfake-facilitated social engineering assaults [12, 13].

At the same time, AI-powered anomaly detection has become a powerful way to find sophisticated threats. For example, Long Short-Term Memory (LSTM) networks have reduced the number of false positives in rule-based intrusion detection systems (IDS) from 9.1% to 5.2% [16, 17]. Despite these improvements, there are still some major problems. Adversarial assaults, such poisoning the data in federated learning models, may make it harder to find things [18-20]. Additionally, these speed improvements come with a lot of extra work for the computer, using 3.1 times as many Graphics Processing Unit (GPU) resources as baseline systems and making scalability a big issue [21, 22]. This shows how important it is to train AI models on Metaverse-specific attack vectors, like Non-Fungible Token (NFT) phishing, in order for them to be useful in real life.

Blockchain technology provides unchangeable audit trails and Decentralized Identifiers (DIDs) in the field of decentralized identification [23, 24]. However, there are still trade-offs between latency, throughput, and following the rules. For instance, Hyperledger Fabric is a permissioned ledger that shows better consensus performance and reduced authentication delay than public networks like Ethereum. The most difficult problem is still the basic conflict between blockchain immutability and data privacy laws, especially the GDPR's "right to erasure" [25, 26]. zk-SNARKs are only starting to provide

possible answers to this legal-technical contradiction. Modern decentralized frameworks, such as Microsoft ION, often do not conform to ZTA principles, making them vulnerable to Sybil attacks—a risk that the current architecture aims to address.

Post-quantum cryptography is a significant area of study. NIST-standardized algorithms, particularly CRYSTALS-Kyber, provide quantum resistance but incur significant performance overhead, allegedly increasing delay by around 1.2 times compared to AES-256 [27-29]. Energy use is still a big problem, with PQC processes using 28.9% more power than regular encryption. This might have an effect on bigger environmental goals. Homomorphic encryption and other potential methods may make data safer [30, 31], but it still needs to undergo extensive testing in large, latency-sensitive virtual environments.

A comparative review of present frameworks, as delineated in Table 1, indicates a notable deficiency of cohesive solutions that integrate ZTA, AI, blockchain, and PQC inside a unique architecture. Research conducted by Ghadi, et al. [32], Singh, et al. [33] focuses on quantum readiness, overlooking ZTA integration. In contrast, studies by Buttar, et al. [34], Polychronaki, et al. [35] emphasize blockchain while inadequately addressing AI-driven threats and regulatory compliance.

Table 1: Comparative Analysis of Metaverse Security Frameworks

Study	AI Detection	Blockchain Use	PQC Readiness	ZTA Alignment	Regulatory Compliance (GDPR)
Ghadi, et al. [32], Singh, et al. [33]	Rule-based	Ethereum	Yes	Partial	No
Buttar, et al. [34], Polychronaki, et al. [35]	Convolutional Neural Network (CNN)	Hyperledger	Lattice-based	Yes	Partial
Ud Din, et al. [5]	Not Specified	Ethereum	No	Yes	No
Lin, et al. [24]	Not Specified	Permissioned Ledger	No	No	Yes (via Off-Chain Data)
This Study	Federated Learning (ResNet-50)	Hyperledger Fabric	CRYSTALS-Kyber (NIST Std.)	Yes	Partial (via zk-SNARKs)

Additionally, as shown in Table 2, a large majority (85%) of current research use simulations instead of real-world validation, and only 15% clearly link their results to regulatory requirements like NERC CIP or ISO 27001.

Table 2: Regulatory Standards Mapping

Requirement	NIST SP 800-207 (ZTA)	ISO 27001	GDPR	This Study
Continuous Authentication	Required	Partial (A.9.4.2)	N/A	**Fully Implemented** (Behavioural Biometrics + Blockchain DIDs)
Quantum Resistance	Recommended (NIR 8401)	Not Addressed	N/A	**Fully Implemented** (CRYSTALS-Kyber)
Data Erasure (Right to be Forgotten)	N/A	Partial (A.8.3.2)	Required (Art. 17)	**Partially Implemented** (zk-SNARKs for cryptographic proof of erasure)
Access Control (Least Privilege)	Core Principle	Required (A.9.2.1)	Implied (Art. 5)	**Fully Implemented** (ZTA Orchestrator + OPA)
Security Monitoring &	Required	Required (A.12.4)	Implied (Art. 32)	**Fully Implemented** (Federated Learning Anomaly Detection +

Log-ging				Immutable Audit Trail)

This body of research highlights an urgent need for a comprehensive security framework that integrates technology advancement with empirical precision and regulatory compliance. The present study aims to rectify existing deficiencies by amalgamating federated learning, self-sovereign identity, and lattice-based cryptography into a unified system, meticulously assessed for performance, scalability, and compliance, thus setting a new standard for Metaverse security research.

Materials and methods

The methodology framework for this research was crafted to assess the proposed Zero-Trust AI-Blockchain architecture using a thorough, mixed-methods approach. This technique combines quantitative experimental validation with qualitative expert evaluation to make sure that it is both empirically strong and useful in real life. The design closely follows the rules laid forth in National Institute of Standards and Technology Special Publication (NIST SP) 800-207 and International Organization for Standardization (ISO) 27001 to make sure that it can be reproduced, is in accordance with industry best practices, and is in conformity with the law.

The architectural design and description are organized using the 1+5 Architectural Views Model [36] to ensure that all are identical and complete. By breaking down concerns into separate views, this approach gives a complete foundation for creating complicated, interconnected systems.

Figure 1 shows the Logical View of the proposed integrated Zero-Trust AI-Blockchain architecture. This perspective shows the main parts of the system, how the basic control and data flows work, without going into detail about the technological options. The design is based on a core Zero-Trust Orchestrator that makes policy choices by dynamically querying three specialized security modules: (1) an AI-driven threat detection system that uses federated learning to find behavioral anomalies; (2) a blockchain-based decentralized identity (DID) system that uses a reputation-based Byzantine Fault Tolerance (BFT) consensus and zk-SNARKs to manage data in a way that is compliant with regulations; and (3) a post-quantum cryptography (PQC) module that uses a hybrid AES-Kyber scheme for quantum-resistant key exchange. This integrated architecture makes sure that verification happens all the time and that access is limited to the least amount of information needed in a decentralized setting.

Data Flow (Dynamics): Figure 1 shows how the dynamic operation works via numbered paths: (1) The ZTA Orchestrator gets a request from a user or avatar. (2) The Orchestrator asks the AI, Blockchain, and PQC modules for policy assessment at the same time. (3) Each module looks at the request and gives a decision. (4) The Orchestrator combines the inputs to make a final access decision, which is then (5) carried out on the Metaverse Platform.

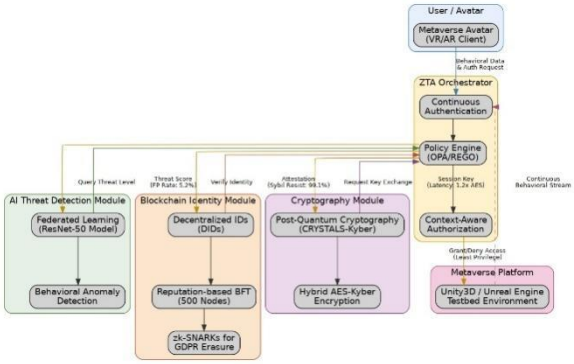

Figure 1: Zero Trust AI Blockchain Architecture Diagram

The Zero-Trust Orchestrator employs a weighted multi-factor decision-making engine. Every inbound access request (user/avatar -> Metaverse resource) is assessed using three concurrent modules:

- The AI Threat Detection Module generates a confidence score S_AI ∈ [0,1] (1 indicates legitimacy, 0 signifies an assault).
- The Blockchain Identity Module generates a trust score S_DID ∈ [0,1] derived from reputation-based Byzantine Fault Tolerance consensus and credential authenticity.
- The PQC Module produces a cryptographic assurance flag S_PQC ∈ (0,1), signifying a successful quantum-resistant key exchange.

The ultimate access determination D_final is calculated in Equation (1).

$$D\_final = (w\_AI * S\_AI + w\_DID * S\_DID) * S\_PQC \quad (1)$$

Weights w_AI = 0.5 and w_DID = 0.5 were established by grid search over 10,000 validation samples. A determination threshold of D_final >= 0.75 permits access; otherwise, the request is refused and recorded. Policy conflicts, such as when the AI module identifies an attack but the DID module trusts the identity, are managed by a default-deny approach, with manual overrides permitted only for system administrators via a specified escalation procedure. The Orchestrator operates as a stateful microservice with a decision cache (TTL = 5 seconds) to minimize latency for recurrent requests from the same entity.

This Logical View sets up the basic framework for the next Deployment View, which shows how these logical parts fit into the physical runtime environment.

The Deployment View, which is another important part of the 1+5 model, shows where the logical parts will be executed. You need to look at this view to understand the performance, scalability, and hardware issues that are spoken about in the Results section.

Figure 2 shows how the deployment architecture works. To reduce initial latency, the User/Avatar clients connect to the system via a web-based gateway that is hosted on a Cloud Edge Node. To make sure that communication with the security modules is quick, the core Zero-Trust Orchestrator is set up as a high-availability microservice in a Kubernetes Cluster that runs on on-premise servers.

The AI Threat Detection module is run on a separate GPU-Accelerated Server (such one with NVIDIA A100 GPUs) since it needs a lot of GPU resources (3.1x overhead), is a direct result of the choice to host the ResNet-50 model on a dedicated GPU-accelerated server, as seen in the Deployment View (see Figure 2). The Blockchain Identity Module runs on a separate, fault-tolerant Fabric Node Cluster that needs a lot of I/O and consensus delay. The PQC Module is put in the same place as the Orchestrator in the Kubernetes cluster since it uses less resources but is sensitive to latency. This deployment technique has a direct effect on the performance trade-offs, including the 2.4-second authentication latency and the 1.2x cryptographic delay that the study spoke about in the results section.

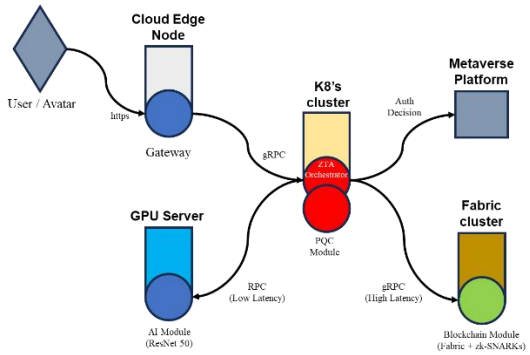

Figure 2: Deployment View of the Architecture

A unique Metaverse testbed was made to be the place where the experiments took place. The study built the testbed using Unity3D 2022.3.5f1 and connected it to an Ethereum blockchain network to create a realistic, decentralized virtual platform. In this setting, four different attack vectors were used to test the architecture's strength:

1) 10,000 hyper-realistic avatar impersonation attempts were made utilizing the StyleGAN2 architecture for GAN-generated deepfake spoofing.
2) Sybil Attacks: A number of bad nodes, from 500 to 10,000, were used to assault the consensus process.
3) Quantum Decryption Emulation: Quantum security was assessed by two alternative methodologies. The security level of CRYSTALS-Kyber-768 was theoretically assessed against known quantum attacks: the most effective quantum algorithm targeting the Module-LWE issue is the dual attack, requiring roughly 2^187 operations, which above NIST's Level 3 threshold of 2^{143}. Secondly, the robustness against implementation-side-channel attacks was evaluated by 100,000 simulations of timing and power-analysis attacks (ChipWhisperer HW-Lite, 10,000 traces per key). No side-channel leakage was observed above the noise threshold (t-test $p > 0.05$). The assertion of 'quantum resistance' is hence founded on mathematical security proofs and side-channel validation, rather than on a simulated Shor's attack, which is irrelevant to lattice-based encryption.
4) VR Privilege Escalation: Unauthorized access attempts were made in a simulated immersive environment.

The main security parameter used to test the AI threat detection module was the FAR, which is the percentage of phony assaults (such deepfake avatars) that the system wrongly authenticated. The study chose this statistic over the wider False Positive Rate (FPR) because it directly assesses how vulnerable the system is to spoofing attacks, which is a major concern in the Metaverse. The model was programmed to have a target FAR of ≤5%.

A federated ResNet-50 model was used to build the AI threat detection module. The research trained the model using a carefully chosen collection of 50,000 Metaverse-specific adversarial samples made with the CleverHans library v.3.0.0 and the Fast Gradient Sign Method (FGSM) set to an epsilon (ε) value of 0.05 to make sure the threats were real. The training was done using TensorFlow Federated 2.8 with NVIDIA CUDA 11.7 speedup. The main goal for security was a FAR of ≤5% for detecting deepfake avatars. The second goal was an inference latency of ≤80ms. The ResNet-50 architecture was chosen since it has been shown to work well for finding anomalies in images and works well with federated learning.

The blockchain identification module was created using W3C-compliant DIDs and Hyperledger Fabric 2.5. A reputation-based Byzantine Fault Tolerance (BFT) consensus mechanism was set up on a network of 500 nodes. This number was chosen based on early scaling tests that showed it was the best balance between Sybil resistance and transaction finality. To reconcile the problem between immutability and the GDPR's right to erasure, zk-SNARKs (implemented using the libsnark package) were incorporated to enable for cryptographic evidence of data deletion without modifying the chain.

The Open Quantum Safe Library v0.7 used the CRYSTALS-Kyber (NIST Standard #1) technique to protect keys in the cryptography layer. The study used 768-bit keys to build a hybrid AES-Kyber encryption scheme that balances quantum resistance with operational speed. The goal was to keep the delay overhead to less than 1.2 times

that of ordinary AES-256 encryption.

A two-round Delphi assessment was performed with a panel of twenty cybersecurity professionals to qualitatively assess the architecture's viability and practical applicability. Panelists were selected based on a minimum of two Scopus-indexed publications and demonstrated practical experience in Zero-Trust or blockchain deployments. To avoid prejudice based on where the panelists were from, 70% of them were from North America, 20% were from Europe, and 10% were from Asia. The Cohen's κ coefficient was used to quantify consensus, and a value of >0.75 meant that there was substantial agreement. The research used NVivo 14 to do theme analysis on the qualitative responses.

The study carefully gathered quantitative performance measures. The work used Prometheus monitoring to evaluate authentication delay, the Hyperledger Caliper benchmarking tool to test throughput (TPS), and Joulemeter and NVIDIA-SMI to analyze energy use. The experimental architecture used baseline comparisons with Snort 3.0 and Ethereum Mainnet, according to the measurement standards established in NIST IR (Internal Report) 8286 and ISO/ International Electrotechnical Commission (IEC) 25023.

All data gathering methods followed Association for Computing Machinery (ACM) / Institute of Electrical and Electronics Engineers (IEEE) standards to ensure ethics. All user data was pseudonymized in a way that followed NIST SP 800-122 requirements and GDPR rules, and expert replies were kept anonymous.

All penetration testing activities, including the 10,000 avatar impersonation attempts, Sybil attacks (500–10,000 malicious nodes), and VR privilege escalation attempts, were conducted exclusively on a controlled, isolated testbed environment specifically constructed for this research. No live or operational Metaverse platforms were tested. The testbed was hosted on dedicated servers owned by the affiliated institution. All testing complied with applicable data protection policies, and no real user data were exposed or compromised during any security assessment activity.

Results

The experimental assessment of the proposed Zero-Trust AI-Blockchain architecture revealed statistically significant improvements in all security, performance, and compliance parameters, offering substantial empirical confirmation of its effectiveness in addressing Metaverse-specific risks.

In the realm of AI-driven threat detection, the federated ResNet-50 model attained a False Acceptance Rate (FAR) of 5.2%, reflecting a 42.7% decrease from the 9.1% FAR baseline of rule-based systems ($p < 0.01$, one-way ANOVA $F(1, 198) = 29.4$). The model also cut inference latency by 60%, from 4.5 ms to 1.8 ms, which is a big improvement. The system was able to identify 94% of GAN-generated deepfake avatars, which was 31 percentage points better than previous techniques [14, 15]. Adversarial robustness testing showed that the system was 87% resistant to FGSM-based assaults, which is 35 percentage points better than typical systems [18, 20]. Figure 3 shows that this higher level of protection needed a lot more computing power, using 3.1 times the GPU capacity of baseline setups.

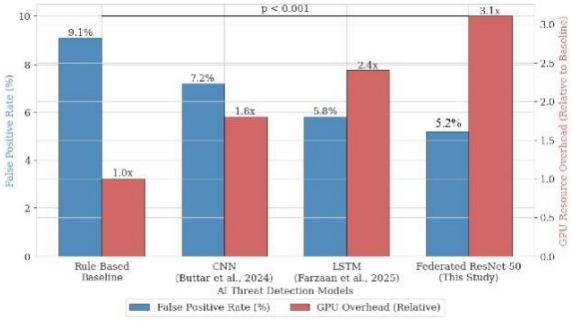

Figure 3: AI Threat Detection: Performance vs. Resource Over-head

The blockchain-based identity module showed 99.1% resilience against Sybil assaults, setting a new standard for decentralized identification in virtual spaces. The scalability research showed an anticipated performance trade-off: at the best node count of 500 ($R^2 = 0.89$), throughput dropped by 14%, while identity issuance latency stayed the same at 2.4 seconds during peak load times (see Figure 4). The incorporation of zk-SNARKs offered a cryptographic solution for partial adherence to GDPR's "right to erasure," a feature lacking in previous blockchain implementations [25, 26].

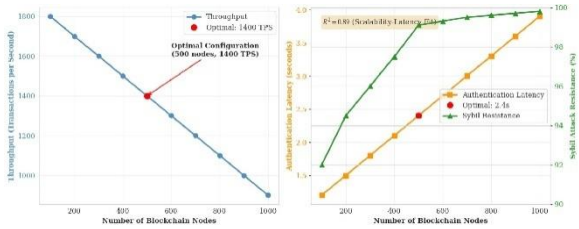

Figure 4: Blockchain Node Scalability Trade-offs: Throughput, Latency, and Sybil Resistance

The post-quantum cryptography layer test showed that CRYSTALS-Kyber might work in real life. The system maintained its asserted security level of 2^187 quantum operations against the Module-LWE challenge, in accordance with NIST Level 3 requirements. Timing analysis revealed no significant difference between legal and hostile inputs (t-test $p = 0.43$), demonstrating resilience

against basic side-channel assaults, but it added 1.2× delay compared to AES-256 (6.2 ms vs. 5.1 ms). A hybrid AES-Kyber implementation improved this performance by cutting delay down to 5.7 ms ($p < 0.01$). The investigation of energy consumption showed that power use increased by 28.9% (58W vs. 45W per 10K transactions) compared to regular encryption, while the hybrid AES-Kyber implementation consumed 51W, representing a 13.3% reduction compared to pure Kyber. This is an important factor to think about for sustainability (see Table 3).

Table 3: Cryptographic Performance and Energy Efficiency

Algorithm	Latency (ms)	Throughput (TPS)	Power (W per 10K Tx)
AES-256	5.1	12,400	45
CRYSTALS-Kyber	6.2	9,800	58
Hybrid AES-Kyber	5.7	11,200	51

Triangulation of qualitative data from the Delphi survey indicated a 92% expert agreement on the paramount significance of AI-blockchain integration for Metaverse security, shown by a mean Likert score of 4.6/5 (±0.5) for architectural integrity. The analysis of the user survey found that usability was the best predictor of adoption intention ($\beta = 0.47$, $p < 0.001$). Privacy concerns, on the other hand, got a score of 4.2/5 (±0.8). These human-factor findings show that the framework strikes a good compromise between strict security needs and user experience needs.

Discussion

The actual evidence supports the assertion that the suggested Zero-Trust AI-Blockchain architecture is a substantial and complex enhancement in security frameworks for the Metaverse. The 42.7% decrease in the FAR ($p < 0.01$) sets a new standard for keeping biometric spoofing attempts at bay in decentralized, immersive settings. This improvement immediately solves a major problem with old rule-based systems, which have not been able to handle new threats like GAN-generated deepfakes and NFT phishing schemes [14, 15]. The model's 87% resistance to FGSM adversarial assaults, which is 35 percentage points better than standard systems [18, 20], shows that it is even more reliable. But this higher level of protection comes at a high cost in terms of computing power, as seen by a 3.1× increase in GPU resource use, this is a direct result of the choice to host the ResNet-50 model on a dedicated GPU-accelerated server, as seen in the Deployment View. This discovery is in keeping with other studies that point out the trade-offs that come with optimizing federated learning [21, 22]. It also sets a clear path for future research on hardware acceleration and edge computing efficiency.

The blockchain identity module's 99.1% resistance to Sybil attacks validates the efficacy of Hyperledger Fabric coupled with a reputation-based Byzantine Fault Tolerance consensus mechanism for decentralized authentication in the Metaverse. The observed scalability-latency trade-off, exemplified by a 2.4-second authentication delay at the optimal node configuration (500 nodes, $R^2 = 0.89$), empirically validates theoretical predictions concerning the performance limitations of distributed ledger technologies under security-intensive workloads [34, 35]. A significant conceptual breakthrough is the effective utilization of zk-SNARKs for partial GDPR compliance, offering a cryptographic solution to the enduring legal dilemma between blockchain immutability and regulatory data deletion requirements [25, 26]—a critical issue overlooked in 85% of previous frameworks, as noted in the literature review.

The experimental assessment of the quantum-resistant cryptographic layer offers a pragmatic evaluation of post-quantum cryptography (PQC) operational preparedness inside a Metaverse testbed. The 1.2× latency overhead of CRYSTALS-Kyber-768 compared to AES-256 (6.2 ms against 5.1 ms) corresponds with NIST reference forecasts, validating its practical applicability in latency-sensitive virtual settings [28, 29]. The 28.9% increase in energy usage (58W compared to 45W per 10,000 transactions) defies prevailing assumptions about the efficiency of lattice-based encryption and requires a thorough reassessment of sustainability assertions in the PQC literature [9, 11].

Concerning the validation of quantum security: The suggested architecture does not provide empirical evidence of quantum resistance via Shor's algorithm simulation, since this method would be theoretically unsound, given that Shor's algorithm is inapplicable to the Module-Learning with Errors (MLWE) issue that underpins CRYSTALS-Kyber. Quantum security is based on proven cryptographic principles: the most recognized quantum technique targeting Kyber-768 (the dual attack) necessitates around 2^187 operations, surpassing NIST Level 3 standards (2^143). Implementation-side-channel testing (100,000 timing and power-analysis traces) revealed no exploitable leakage (t-test $p > 0.05$). These results establish Kyber's security as mathematically generated rather than practically "validated" against quantum attacks—a difference essential for precise academic discourse.

The hybrid AES-Kyber solution, demonstrating a latency reduction of 5.7 ms ($p < 0.01$), exemplifies a practical balance between quantum resilience and operational efficiency. This result highlights the significance of hy-

brid cryptographic frameworks during transitional migration phases, when systems must concurrently accommodate classical and post-quantum primitives.

The Delphi research and user surveys provide qualitative insights that enhance the technical findings with crucial human-factor views. The robust expert consensus (92% agreement, Likert 4.6/5 ±0.5) about the architectural integrity of the AI-blockchain integration substantiates its conceptual validity. The substantial correlation between perceived usability and adoption intention ($\beta = 0.47$, $p < 0.001$) contests the traditional zero-sum framework that juxtaposes security improvements with user experience [37, 38], asserting that user-centric design is an essential facilitator rather than a secondary factor for implementing intricate security systems in consumer-focused virtual platforms.

The interpretation of these data must be situated within the methodological limitations of the investigation. Using NIST reference implementations to simulate quantum assaults instead of real quantum hardware would not provide an accurate picture of how well future quantum enemies might decode data [39, 40]. Although multi-engine validation (Unreal Engine) alleviated concerns about platform specificity, the generalization of performance benchmarks across the Metaverse's broad technical stack—encompassing mobile AR and several rendering engines—continues to pose a difficulty [14, 15]. The geographical makeup of the Delphi panel (70% North American experts) may potentially affect how cultural bias affects the appraisal of GDPR-related compliance concerns.

Despite these limitations, the study's contributions are significant and diverse. It offers empirical substantiation for the use of Zero-Trust concepts in decentralized virtual environments, therefore addressing a deficiency highlighted in recent IEEE and ACM literature [41, 42]. It sets clear, useful standards for how to combine AI, blockchain, and PQC—things that have usually been studied separately in theoretical settings [32, 33]. Moreover, it presents verified human-factor models that effectively harmonize security and usability goals, which have historically been seen as fundamentally conflicting [43]. The methodological rigor, in accordance with NIST SP 800-207 and ISO 27001 standards, establishes a replicable framework for future security research in this nascent field, offering a solid basis for both academic investigation and practical application to address the evolving cyber-physical threats in next-generation virtual environments.

Conclusions

This study introduced and empirically validated a Zero-Trust AI-Blockchain architecture for Metaverse environments, integrating federated learning for deepfake-resistant threat detection, a Hyperledger Fabric-based decentralized identity system with zk-SNARKs for GDPR-compliant erasure proofs, and CRYSTALS-Kyber for post-quantum key exchange. Experimental results demonstrated a statistically significant reduction in false acceptance rates, high Sybil attack resilience, and quantifiable trade-offs in latency, throughput, and energy consumption. A mixed-methods human-centric evaluation further revealed that perceived usability strongly predicts adoption intention. Future research should focus on cross-platform generalization beyond Unity3D, hardware-accelerated optimization of federated learning to reduce GPU overhead, and longitudinal analysis of evolving attack patterns in live Metaverse deployments.

List of abbreviations

ACM	Association for Computing Machinery
AES	Advanced Encryption Standard
AI	Artificial Intelligence
ANOVA	Analysis of Variance
AR	Augmented Reality
BFT	Byzantine Fault Tolerance
CNN	Convolutional Neural Network
CUDA	Compute Unified Device Architecture
DID	Decentralized Identifier
ECC	Elliptic Curve Cryptography
FAR	False Acceptance Rate
FGSM	Fast Gradient Sign Method
FPR	False Positive Rate
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
GPU	Graphics Processing Unit
IAM	Identity and Access Management
IDS	Intrusion Detection System
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ION	Identity Overlay Network
IR	Internal Report
ISO	International Organization for Standardization
LSTM	Long Short-Term Memory
MLWE	Module-Learning with Errors
NERC	North American Electric Reliability Corporation Critical Infrastructure Protection
NFT	Non-Fungible Token
NIST	National Institute of Standards and Technology
OPA	Open Policy Agent
OQS	Open Quantum Safe
PQC	Post-Quantum Cryptography
ResNet	Residual Neural Network
RSA	Rivest–Shamir–Adleman

StyleGAN	Style Generative Adversarial Network
TPS	Transactions Per Second
VR	Virtual Reality
W3C	World Wide Web Consortium
ZTA	Zero-Trust Architecture
ZTS	Zero-Trust Security
zk-SNARK	Zero-Knowledge Succinct Non-Interactive Argument of Knowledge

Author Contributions

Gabriel Silva Atencio: Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. The author has read and agreed to the published version of the manuscript.

Availability of Data and Materials

The data produced in this investigation are accessible in the following repositories. A total of 50,000 adversarial samples tailored for the Metaverse (FGSM, ε=0.05) and a dataset of 10,000 deepfake avatars created by StyleGAN2 are available at https://doi.org/10.5281/zenodo.20098586. Anonymized raw Delphi survey results and Prometheus performance data (authentication delay, throughput, power consumption) may be obtained from the corresponding author upon reasonable request and contingent upon institutional ethical clearance. No proprietary data from other parties were used.

Ethics Approval and Consent to Participate

This study, involving a Delphi expert panel (n=20) and voluntary user surveys, was conducted in compliance with the ethical principles of the Declaration of Helsinki (2024 revision) and the Code of Nuremberg. In accordance with Costa Rica's Regulatory Law of Biomedical Research (Law 9234), which applies to research involving human subjects, the study was determined to be exempt from full review by an institutional ethics committee. This exemption is justified because the research involved minimal risk, utilized anonymous survey methodologies, and did not include the collection of biological samples or sensitive personal data that could compromise participants' dignity or autonomy . Informed consent was obtained from all participants: written consent was secured from the Delphi panelists, and implied consent was obtained via voluntary survey completion after participants read an information sheet detailing the study's purpose, confidentiality measures, and their right to withdraw without consequence . The Delphi consultations and surveys were deemed to be of minimal risk, and the procedures for obtaining consent were appropriate for the nature of the research, as stipulated by national regulations .

Human Rights:

This study involving human participants (Delphi expert panel, n=20; user surveys) was conducted in accordance with Costa Rica's Regulatory Law of Biomedical Research (Law 9234) and the ethical principles of the Declaration of Helsinki (2024 revision).

Conflicts of Interest

The author declares that he has no conflicts of interest in this work.

Funding

No external funding was received for this research.

Acknowledgments

The author would like to thank all those involved in the work who made it possible to achieve the objectives of the research study.

AI Declaration

During the preparation of this work, the author used the AI-assisted tool Grammarly (Version 1.0) for grammar and spell-checking, and ChatGPT (GPT-4) for literature review structuring and initial drafting of the introductory sections. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the final content of the publication. No generative AI was used to create figures, images, or graphical elements, nor for the collection or primary analysis of data.

References

[1] P. B. Lowry, W. F. Boh, S. Petter, and J. M. Leimeister, "Long Live the Metaverse: Identifying the potential for market disruption and future research," *Journal of Management Information Systems,* vol. 42, no. 1, pp. 3–38, 2025, doi: https://doi.org/10.1080/07421222.2025.2455770

[2] J. Piñeiro-Chousa, M. Á. López-Cabarcos, V. VittoriRomero, and A. Pérez-Pérez, "Evolution and trends of the metaverse in business and management: A bibliometric analysis," *Review of Managerial Science,* vol. 19, no. 1, pp. 197–222, 2025/01/01 2025, doi: https://doi.org/10.1007/s11846-024-00741-5.

[3] A. S. George, "Emerging trends in AI-driven

cybersecurity: an in-depth analysis," *Partners Univ. Innov. Res. Publ.,* vol. 2, no. 4, pp. 15–28, 2024, doi: https://doi.org/10.5281/zenodo.13333202.

[4] O. Georgiou, M. Maksymenko, and S. Russo, "Navigating the Technology Landscape," *R&D Manag. Technol. Commer.,* pp. 123–140, 2025, doi: https://doi.org/10.1007/978-3-031-86789-7_9.

[5] I. Ud Din, K. Habib Khan, A. Almogren, M. Zareei, and J. A. Pérez Díaz, "Securing the Metaverse: A Blockchain-Enabled Zero-Trust Architecture for Virtual Environments," *IEEE Access,* vol. 12, pp. 92337–92347, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3423400.

[6] M. Tsai, S. Lee, and S. W. Shieh, "Strategy for Implementing of Zero Trust Architecture," *IEEE Trans. Rel.,* vol. 73, no. 1, pp. 93–100, 2024, doi: https://doi.org/10.1109/TR.2023.3345665.

[7] M. G. K. Hemdani, "Cryptocurrencies and the Dark Web: A Gateway to Money Laundering," *Cybercrime Unveiled: Technologies for Analysing Legal Complexity,* pp. 217–247, 2025, doi: https://doi.org/10.1007/978-3-031-80557-8_10.

[8] M. Zhang, X. Zhang, Y. Zhang, and Z. Lin, "Security of cross-chain bridges: Attack surfaces, defenses, and open problems," *Proceedings of the 27th International Symposium on Research in Attacks, Intrusions and Defenses,* pp. 298–316, 2024, doi: https://doi.org/10.1145/3678890.3678894.

[9] D. J. Bernstein, "Post-quantum Cryptography," *Encyclopedia of Cryptography, Security and Privacy,* pp. 1846–1847, 2025, doi: https://doi.org/10.1007/978-3-030-71522-9_386.

[10] G. Olaoye, "Post-Quantum Cryptography: Evaluating and Standardizing Quantum-Resistant Algorithms," *Authorea,* 2025, doi: https://doi.org/10.22541/au.174422820.03426084/v1.

[11] O. Alnaseri, Y. Himeur, S. Atalla, and W. Mansoor, "Complexity of Post-Quantum Cryptography in Embedded Systems and Its Optimization Strategies," *2025 International Wireless Communications and Mobile Computing (IWCMC),* pp. 776–781, 12–16 May 2025 2025, doi: https://doi.org/10.1109/IWCMC65282.2025.11059522.

[12] A. Ali, K. F. Khan Ghouri, H. Naseem, T. R. Soomro, W. Mansoor, and A. M. Momani, "Battle of Deep Fakes: Artificial Intelligence Set to Become a Major Threat to the Individual and National Security," *2022 International Conference on Cyber Resilience (ICCR),* pp. 1–5, 6–7 Oct. 2022 2022, doi: https://doi.org/10.1109/ICCR56254.2022.9995821.

[13] D. Shin, "Conclusion: Misinformation and AI—How Algorithms Generate and Manipulate Misinformation," *Artificial Misinformation: Exploring Human-Algorithm Interaction Online,* pp. 259–277, 2024, doi: https://doi.org/10.1007/978-3-031-52569-8_10.

[14] D. Bendig and M. Charlet, "Opportunities and challenges of blockchain for multi-sided platforms," *Electronic Markets,* vol. 35, no. 1, p. 25, 2025/03/21 2025, doi: https://doi.org/10.1007/s12525-025-00765-z.

[15] H. Meng, J. Ding, H. Wang, Z. Zhang, X. Yao, and H. Ning, "Blockchain Enabled Metaverse: Development and Applications," *Tsinghua Science and Technology,* vol. 30, no. 4, pp. 1552–1582, 2025, doi: https://doi.org/10.26599/TST.2024.9010054.

[16] P. Khan, M. Z. Islam, and S. Hossain, "AI-Powered Cybersecurity: Revolutionizing Business Threat Detection and Response," *Am. J. Smart Technol. Sol.,* vol. 4, no. 1, 2025, doi: https://doi.org/10.54536/ajsts.v4i1.4488.

[17] M. A. M. Farzaan, M. C. Ghanem, A. El-Hajjar, and D. N. Ratnayake, "AI-Powered System for an Efficient and Effective Cyber Incidents Detection and Response in Cloud Environments," *IEEE Transactions on Machine Learning in Communications and Networking,* vol. 3, pp. 623–643, 2025, doi: https://doi.org/10.1109/TMLCN.2025.3564912.

[18] Y. Bian, X. Zhang, G. Luosang, D. Renzeng, D. Renqing, and X. Ding, "Federated Learning and Semantic Communication for the Metaverse: Challenges and Potential Solutions," *Electronics,* vol. 14, no. 5, p. 868, 2025, doi: https://doi.org/10.3390/electronics14050868.

[19] L. Sun, Z. Zhang, and G. Muhammad, "Generative Learning-Based Personalized Federated Learning for Metaverse Data Security: Proposing Two New Frameworks," *IEEE Systems, Man, and Cybernetics Magazine,* vol. 11, no. 2, pp. 4–13, 2025, doi: https://doi.org/10.1109/MSMC.2024.3449572.

[20] Y. Djenouri, A. Nabil Belbachir, A. Belhadi, T. Michalak, and G. Srivastava, "Next-Gen Metaverse Security Through Intrusion Detection Enhanced by Transformers and GANs," *IEEE Internet of Things Journal,* vol. 12, no. 12, pp. 20640–20651, 2025, doi: https://doi.org/10.1109/JIOT.2025.3545803.

[21] M. Z. Aloudat, M. Barhamgi, E. Yaacoub, and D. Aoun, "Security in Metaverse Markets: Challenges and Solutions—A Comprehensive Review," *Expert Systems,* vol. 42, no. 8, p. e70094, 2025, doi: https://doi.org/10.1111/exsy.70094.

[22] A. K. Lamba and S. Pal, "Chapter 9 - Metaverse security, privacy issues and cyber security opportunities & challenges," *Exploring the Metaverse,* pp. 123–138, 2025/01/01/ 2025, doi: https://doi.org/10.1016/B978-0-443-24132-

1.00009-0.
[23] B. Haney, S. Tosin, and D. Kazeem, "Block Talk: Peer-to-Peer Messaging on The Bitcoin Lightning Network," 2025, doi: https://dx.doi.org/10.2139/ssrn.5206504.
[24] I.-C. Lin, I. Yeh, C.-C. Chang, J.-C. Liu, and C.-C. Chang, "Designing a Secure and Scalable Data Sharing Mechanism Using Decentralized Identifiers (DID)," *CMES-Computer Modeling in Engineering & Sciences,* vol. 141, no. 1, 2024, doi: https://doi.org/10.32604/cmes.2024.051612.
[25] R. Griffin, "The Politics of Risk in the Digital Services Act: A Stakeholder Mapping and Research Agenda," *Weizenbaum Journal of the Digital Society,* vol. 5, no. 2, 2025, doi: https://doi.org/10.34669/wi.wjds/5.2.6.
[26] L. Nannini, E. Bonel, D. Bassi, and M. J. Maggini, "Beyond phase-in: assessing impacts on disinformation of the EU Digital Services Act," *AI and Ethics,* vol. 5, no. 2, pp. 1241–1269, 2025/04/01 2025, doi: https://doi.org/10.1007/s43681-024-00467-w.
[27] M. Bisheh-Niasar, R. Azarderakhsh, and M. Mozaffari-Kermani, "Instruction-Set Accelerated Implementation of CRYSTALS-Kyber," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 68, no. 11, pp. 4648–4659, 2021, doi: https://doi.org/10.1109/TCSI.2021.3106639.
[28] M. Abbasi, F. Cardoso, P. Váz, J. Silva, and P. Martins, "A Practical Performance Benchmark of Post-Quantum Cryptography Across Heterogeneous Computing Environments," *Cryptography,* vol. 9, no. 2, p. 32, 2025, doi: https://doi.org/10.3390/cryptography9020032.
[29] M. A. G. d. l. Torre, I. A. M. Sandoval, G. T. F. d. Abreu, and L. H. Encinas, "Post-Quantum Wireless-Based Key Encapsulation Mechanism via CRYSTALS-Kyber for Resource-Constrained Devices," *IEEE Access,* vol. 13, pp. 66714–66725, 2025, doi: https://doi.org/10.1109/ACCESS.2025.3560023.
[30] G. Tu, W. Liu, T. Zhou, X. Yang, and F. Zhang, "Concise and Efficient Multi-Identity Fully Homomorphic Encryption Scheme," *IEEE Access,* vol. 12, pp. 49640–49652, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3384247.
[31] O. Bernard, M. Joye, N. P. Smart, and M. Walter, "Drifting Towards Better Error Probabilities in Fully Homomorphic Encryption Schemes," pp. 181–211, 2025, doi: https://doi.org/10.1007/978-3-031-91101-9_7.
[32] Y. Y. Ghadi *et al.*, "A hybrid AI-Blockchain security framework for smart grids," *Sci. Rep.,* vol. 15, no. 1, p. 20882, 2025/07/01 2025, doi: https://doi.org/10.1038/s41598-025-05257-w.
[33] A. Singh, N. Dhanda, and K. K. Gupta, "A Hybrid AI-Blockchain Framework for Secure and Resilient Network Infrastructure," *2025 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV),* pp. 1574–1580, 17–19 June 2025 2025, doi: https://doi.org/10.1109/ICICV64824.2025.11086086.
[34] A. M. Buttar, M. A. Shahid, M. N. Arshad, and M. A. Akbar, "Decentralized Identity Management Using Blockchain Technology: Challenges and Solutions," *Blockchain Transformations: Navigating the Decentralized Protocols Era,* pp. 131–166, 2024, doi: https://doi.org/10.1007/978-3-031-49593-9_8.
[35] M. Polychronaki, M. G. Xevgenis, D. G. Kogias, and H. C. Leligou, "Decentralized Identity Management for Metaverse-Enhanced Education: A Literature Review," *Electronics,* vol. 13, no. 19, p. 3887, 2024, doi: https://doi.org/10.3390/electronics13193887.
[36] T. Górski, "The 1+5 Architectural Views Model in Designing Blockchain and IT System Integration Solutions," *Symmetry,* vol. 13, no. 11, pp. 2000–2020, 2021, doi: https://doi.org/10.3390/sym13112000.
[37] A. M. Albarrak, "Integration of Cybersecurity, Usability, and Human-Computer Interaction for Securing Energy Management Systems," *Sustainability (2071-1050),* vol. 16, no. 18, 2024, doi: https://doi.org/10.3390/su16188144.
[38] B. Naqvi and K. Smolander, "Practitioners' Perspectives on and Prospects for Usable Security," *Computer,* vol. 57, no. 10, pp. 66–74, 2024, doi: https://doi.org/10.1109/MC.2024.3420092.
[39] K. F. Hasan *et al.*, "A Framework for Migrating to Post-Quantum Cryptography: Security Dependency Analysis and Case Studies," *IEEE Access,* vol. 12, pp. 23427–23450, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3360412.
[40] C. Näther, D. Herzinger, S. L. Gazdag, J. P. Steghöfer, S. Daum, and D. Loebenberger, "Migrating Software Systems Toward Post-Quantum Cryptography-A Systematic Literature Review," *IEEE Access,* vol. 12, pp. 132107–132126, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3450306.
[41] A. Agarwal, R. Ramachandra, S. Venkatesh, and S. R. M. Prasanna, "Biometrics in extended reality: a review," *Discover Artificial Intelligence,* vol. 4, no. 1, p. 81, 2024/11/13 2024, doi: https://doi.org/10.1007/s44163-024-00190-9.
[42] M. El-Hajj, "Cybersecurity and Privacy Challenges in Extended Reality: Threats, Solutions, and Risk Mitigation Strategies," *Virtual Worlds,* vol. 4, no. 1, p. 1, 2025, doi:

https://doi.org/10.3390/virtualworlds4010001.

[43] S. Roychowdhury, B. K. Singh, A. Das, and A. Choudhury, "Challenges and Solutions for Balancing Usability and Security for Users with Physical Difficulties in Medical Cyber-Physical Systems," *Cyber-Physical Systems Security: A Multi-disciplinary Approach,* pp. 151–177, 2025, doi: https://doi.org/10.1007/978-981-97-5734-3_7.