


DAI-TIRS: An AI-Powered Threat Intelligence and Response System for Securing the Metaverse


Mohini Sharma

(Symbiosis International University, Pune, Maharashtra, India)

 <https://orcid.org/0009-0007-9182-9939>, mohini.sharma@associates.scit.edu


Raghav Sandhane

(Symbiosis International University, Pune, Maharashtra, India)

 <https://orcid.org/0000-0002-8952-1042>, raghav@scit.edu

Jaydeep Rajeshkumar Katariya

(Symbiosis International University, Pune, Maharashtra, India)

 <https://orcid.org/0009-0002-9860-6555>, jaydeep.katariya@associates.scit.edu

Abstract: The metaverse seamlessly integrates physical and digital spaces, enabling AI-driven innovations in virtual interactions, autonomous avatars, and real-time experiences. However, increased reliance on AI brings sweeping cybersecurity challenges, such as adversarial attacks, deep fake impersonation, and AI-driven phishing campaigns. The security of the metaverse is vital for the sustainability of user trust and system integrity. As AI assumes a larger role in virtual environments, proactive cybersecurity measures must be taken to counter emerging threats. This paper introduces DAI-TIRS, a holistic security framework designed to proactively secure the metaverse. DAI-TIRS is the integration of machine learning-based anomaly detection, dynamic honeypots, and predictive threat modeling that detect, classify, and mitigate AI-driven threats in real-time. By utilizing MITRE ATT&CK and the PyTM framework, it constantly learns new emerging threats through advanced behavioral analytics and keeps pace with the adversarial AI model's evolution. The experimental results from a simulated metaverse environment demonstrate that DAI-TIRS achieves 93% accuracy in threat detection, 90% precision in classifying the severity, and a 36.9% faster threat mitigation response time than the average performance of baseline models, as detailed in the paper. These findings underscore the critical need for adaptive AI-based cybersecurity solutions that will enhance the resilience, trust, and integrity of metaverse ecosystems. This research establishes DAI-TIRS as an advanced cybersecurity framework that has demonstrated its adaptability and effectiveness in countering AI-driven threats across multiple sectors. The source code for DAI-TIRS is available on GitHub: [<https://github.com/sharmamohini762/DAI-TIRS-Code>]

Keywords: Metaverse Security, AI-Driven Cyber Threats, Dynamic AI Threat Intelligence and Response System (DAI-TIRS), Dynamic Honeypots, MITRE ATT&CK and PYTM Framework

Categories: C.2.0, C.2.3, D.4.6, I.2.6, K.6.5

DOI: 10.3897/jucs.165358

1 Introduction

The metaverse is revolutionizing digital interactions, providing immersive experiences through Artificial Intelligence (AI)-driven digital twins, virtual economies, and real-time customer engagement. Companies such as Gucci and Nike leverage AI for virtual retail experiences [Gonzalez, 2021] while Walmart and Amazon enhance electronic commerce (e-commerce) through AI-powered assistants and logistics optimization [Kiwoong et al., 2023]. These advancements offer unparalleled opportunities for businesses and consumers alike, allowing seamless interaction between physical and digital domains. However, while AI integration has created significant opportunities, it has also introduced growing cybersecurity threats [Yaqoob et al., 2023]. AI-driven attacks exploit vulnerabilities in recommendation algorithms, facial recognition systems, and automated workflows, leading to deep fake impersonations, sophisticated phishing campaigns, and large-scale data breaches. As AI models become more autonomous and capable, their potential misuse for malicious activities increases significantly, posing a major concern for cybersecurity experts worldwide. A challenge lies in developing AI-driven adaptive security solutions that can address current AI-driven threats and anticipate future attack vectors. Conventional security mechanisms lack the adaptability required to counter these threats dynamically, necessitating the development of AI-powered defense systems capable of real-time analysis and mitigation. In the Metaverse Environment, there is an entirely new category of cybersecurity issues, which majorly includes cyber risks, through higher reliance on AI and through attacks based on AI, where the vulnerability is exploited [Saracoglu, 2023]

One such category of cyber risk is Adversarial AI, where the attackers use their knowledge against AI models, which results in misleading recommendation engines, perhaps by injecting malicious content or creating deceptive interactions that are quite plausible. Studies show that Deep neural networks in metaverse AI applications are susceptible to attacks through the introduction of small and imperceptible noise, which yields high error probabilities in classifier predictions. For example, attacks such as Fast Gradient Sign Method and Deep Fool trick AI models by manipulating image input information in object recognition tasks for metaverse applications and make incorrect classifications [Zhangao et al., 2023]. Deep Impersonation is yet another category of the prevailing cyber risks. Using deep fake technology, the cyber-criminal can produce realistic virtual avatars of trusted personalities such as executives or celebrities. A cybercriminal can become a Chief Executive Officer (CEO) within a virtual boardroom and sanction fraud transactions by mimicking using AI-based visual and behavioral mimicry. The 3rd and most common of the risks includes Automated Phishing: Such a type of phishing may be carried out through AI, utilizing user behavior data that the metaverse has harvested. For instance, the attacker may compose the phishing message directed to an avatar and its preferences, purchasing history, along with interaction targeting the maximum chances of success [Awadallah et al., 2024]. Due to all these issues in the highly sensitive metaverse environment, there exists a very high Data Privacy risk. The metaverse relies on collecting a massive amount of data to use for smooth user experiences. Such data includes biometric data, which may include facial expressions and voiceprints, behavioral data, including keystrokes and gaze tracking, and financial information [Pooyandeh et al., 2022]. A breach of such data could result in identity theft, financial fraud, or the black-market trade of sensitive user information,

which further results in breach of trust and Social Engineering. The real and virtual identities create complexities in the mechanisms of trust. The attackers have capitalized on this by creating fictional communities or events in the metaverse, making the victims submit confidential information or conduct financial transactions, often leaving the user in a state of loss that is unrecoverable.

1.1 Problem Description

Most of the cybersecurity solutions do not utilize AI to counter AI-based threats in the metaverse [Sebastain, 2022], although there is an increase in the sophistication of cyberattacks in the metaverse. The current frameworks are based on static security products, including firewalls, rule-based detection methods, and signature-based malware detection that have limited potential for adaptation in such situations when advanced AI threats arise, like adversarial manipulation, deep fake impersonation, or AI-generated phishing attacks. This gap is critical because current AI-based security measures remain largely static, unable to dynamically evolve against the rapidly advancing nature of adversarial AI threats. Thus, the need for adaptive AI-powered security solutions that continually evolve and counter such sophisticated attacks arises. These solutions should be able to detect, analyze, and respond in real time to malicious AI activity in a way that makes the security mechanisms resilient and trustworthy [Otoum et al., 2024]. This paper proposes an innovative AI-driven cybersecurity framework named Dynamic AI Threat Intelligence and Response System (DAI-TIRS). As shown in Figure 1, which is designed to dynamically identify, classify, and counter AI-driven threats using Machine learning algorithms, DAI-TIRS differs from the traditional approaches of the prior cybersecurity models; it is an original work of the author's and an innovation by them as the next-generation AI-driven security system with machine learning-based anomaly detection, dynamic honey-pots, and predictive threat modeling. Such a framework changes the attack patterns based on the evolving usage of self-learning capabilities. This would be an imperative solution for protecting AI-driven metaverse ecosystems, which use MITRE (Adversarial Tactics, Techniques, and Common Knowledge) ATT&CK [Rajesh et al., 2022] and the Python Threat Modeling Framework (PyTM) framework [Granata & Rak, 2024] for real-time threat intelligence, optimized response, and proactive defense strategy.

1.2 Advancing Cybersecurity: Existing Models Vs. DAI-TIRS Implementation

Traditional cybersecurity frameworks have been using static rule-based detection methods that are often inadequate in handling evolving AI-driven threats. The conventional solutions include heuristic-based monitoring [Pfau-Wagenbauer & Nejd, 1993], which provides some protection but lacks adaptability against advanced AI-driven attacks. These often become a bit backward in keeping track of changing AI-driven threats. Signature-based detection of malware [Chakravarty et al., 2019], Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) [Wang et al., 2023b], and Security Information and Event Management (SIEM) [Uetz et al., 2023] work to some degree but are incapable of defending against adversarial AI that can thwart authentication, affect detection models, and create sophisticated cyber threats.

Classical security measures can no longer effectively combat emerging AI-based threats. Shodan helps attackers detect vulnerable systems to circumvent traditional static security mechanisms and inject AI-driven exploits [Abdullatif & Tillakaratne, 2025]. So, the need of the hour is based smart solution providing an AI-driven security that not only supports Real-Time Analysis but also mitigation [Garfinkel and Leclerc, 2020]. A notable example of adaptive AI-powered security methods is Lava Lamp from Cloudflare [Song et al., 2023]. Traditional encryption methodologies rely on predictable algorithms and thus are vulnerable to AI-powered decryption attacks. To counter this, Cloudflare's Lava Lamp system captures patterns of light emitted by lava lamps to generate true randomness, which is then processed into cryptographic entropy, so that the keys used for encryption are unpredictable and resistant to AI-based attacks. While Lava Lamp is more concerned with encryption security, DAI-TIRS, our proposed solution, leans towards adaptive security in the metaverse through real-time anomaly detection, dynamic honeypots, and predictive threat modeling. Unlike static security approaches, DAI-TIRS continuously learns from emerging threats and dynamically adjusts its defense mechanisms to neutralize AI-driven cyberattacks. It integrates with virtual environment monitoring systems without impacting the real-time mitigation process so that neither the user experience is compromised nor the security standards. The paper proposes DAI-TIRS as a new contribution towards this rapidly evolving landscape and claims its efficiency as compared to other existing cybersecurity models, stating the rationale for its necessity in securing AI-driven digital spaces.

1.3 Research Objective

This study introduces DAI-TIRS, an original cybersecurity framework conceptualized by the authors to address AI-driven challenges in the metaverse. DAI-TIRS is a next-generation cybersecurity framework that will help combat all the challenges in the Metaverse and specific issues. Here, the system framework makes use of Machine Learning (ML)-based anomaly detection [Kai Li et al., 2022]. It ascertains if there are abnormal patterns in the user's behavior, even in keystrokes, mouse movements, or timing of activity, which would pinpoint threats. It deploys Dynamic Honeypots. A Deceptive environment meant to lure in attackers to harvest intelligence and mitigate in real time, while not disturbing a legitimate user. Predictive Threat Modeling: Using the MITRE ATT&CK [Rajesh et al., 2022] as shown in Figure 2 for mapping of threats to highlight common adversarial tactics within the metaverse, along with PyTM, which serves as a means to further predict threats for simulated filtering and efficiently working towards mitigating them. This work is aptly working toward experimentation validation of how DAI-TIRS adjusts dynamically based on the ever-emerging threats while keeping the metaverse ecosystems highly secure and assuring integrity to users' trust and experience.

2 Literature Review

Researchers are trying to investigate how the metaverse has been revolutionized as a new digital environment; its application, opportunities, and challenges have come into prominence. [Gonzalez, 2021] referred to Gucci and Nike, which use AI in virtual retail to enhance customer experiences. [Yaqoob et al., 2023] studied the application of metaverse in smart cities, while [Kiwoong et al., 2023] streamlined the study towards its

application in the retail ecosystem. Both studies showcased the opportunities and challenges that come with these applications, setting up a path to understand how to use these opportunities to counter the challenges that come along. [Saracoglu, 2023] identified emerging cybersecurity threats in the metaverse, and [Zhangao et al., 2023] identified how AI could be used to defend against adversarial attacks and edge computing challenges, which underlined the need for a dynamic cybersecurity framework. [Bridgit, 2023] in his book talked about a relationship between AI and the metaweb, how the collaboration provided a collective intelligence and helped build trust, ensuring transparency, while [Rosenblat, 2023] focused on human rights dynamics and their protection in the 3D immersive web, making a point about the intricacy of securing virtual settings. Another push toward adaptive and AI-centric solutions came through the study of [Awadallah et al., 2024], which discusses AI's major role in metaverse cybersecurity, its key aspects in the various opportunities of cannibalizing it against the key risks that emerge. [Pooyandeh et al., 2022] performed a survey on the same, showing the increased AI-based threats in the metaverse and the increased need to mitigate them.

A few authors also came up with Mathematical underpinnings and frameworks. For instance, [Song et al., 2023] concentrated on a theoretical framework that helps develop a broader perspective on the metaverse, and [Garfinkel and Leclerc, 2020] proved that dynamic encryption is one of the most critical components for strong cybersecurity. An analysis of the studies by various researchers on traditional models of cybersecurity helped gain an understanding of the significant contribution of these models towards threat detection, but also demonstrated several limitations of existing approaches that have been identified. Heuristic-based systems struggle with complex fault scenarios, due to which it is integrated with model-based reasoning that [Pfau-Wagenbauer & Nejd, 1993] identified. [Chakravarty et al., 2019] found that signature-based detection relies on old and existing malware and is not sufficient for identifying future threats hence a behavior-based approach is used which is a dynamic method but then even this has limitation like raising false alarms they concluded by saying a specification-based method is most appropriate which helped in understanding why the present day requires an adaptive solution capable of continuously learning.

[Wang et al., 2023b] studied how text-based CAPTCHA mechanisms could be bypassed by attackers and could be prone to attacks, leading to severe issues. SIEM systems were studied by [Uetz et al., 2023] helped us build an understanding that they are helpful for log management and rule-based analysis, yet fail to deliver real-time adaptive threat mitigation and hence remain blind spots that can be easily detected. [Masombuka, 2018] delved deep to understand cyberspace and propose a framework to defend it efficiently. Studies on other tools, such as Shodan, have revealed even more inadequacies of static security controls, as attackers could scan exposed and misconfigured internet-facing systems that may lead to the risk of cyber exploitation, which was studied by [Alabdulatif & Thilakarathne, 2025]. Still, considering these inadequacies, we dived deeper into adaptive and intelligence-driven approaches for metaverse security. We found MITRE ATT&CK and PyTM as two key frameworks that would be utilized for structured cyber threat analysis and predictive threat modeling [Rajesh et al., 2022] helped us understand that MITRE ATT&CK is a broad knowledge base that contains adversarial tactics and techniques against which a systematic approach can be implemented for cyber defense and proactive security planning.

In contrast, the PyTM framework enables code-based threat modeling, enabling simulation of cyber threats while evaluating the attack surface and assessing risks in complex AI-driven environments [Granata & Rak, 2024]. This provided the basis for DAI-TIRS, the adaptive security framework's predictive threat modeling concept. [Otoum et al., 2024] laid groundwork was set in place to exploit machine learning in the area of threat detection and formed a starting point for our ML-based anomaly detection module. [Gadekallu et al., 2022] spoke more generally about blockchain and distributed systems and left some optimism about applying such concepts in dealing with the issues. But at the same time, literature still exposed a considerable gap: no over-arching framework to merge anomaly detection, dynamic honeypots, and predictive threat modeling and overcome the security challenges caused in metaverse by AI. That was the identified gap which triggered the idea of developing an AI solution to fight against the adverse security issues AI creates in the metaverse hence we developed an original solution the DAI-TIRS-a tool intended to combine all these components into one single cohesive adaptive cybersecurity product, specially made for the metaverse.

Many research papers helped in gaining a better understanding of various other aspects of the Metaverse, laying a foundation for DAI-TIRS. One such aspect was the issues of interoperability as a significant barrier for the metaverse, emphasized by [Yang et al., 2024], research on Threat analysis and defense strategies for the metaverse and extended reality systems with studies on increased adoption strategies after awareness on cyber risks in metaverse has been spread around the users helps us understand how to design the framework and how to motivate adoption by [Sebastain, 2022] and [Qamar et al., 2023] studies further thoroughly analyzed the defense strategies for the metaverse.

The bibliometric analysis of research trajectories and various challenges encountered in cybersecurity by [Kostelić and Etinger, 2024] served to refine our predictive threat modeling strategy. [Forti, 2021] and [Berendt et al., 2020] discussed privacy issues and the impact of AI in areas such as education and healthcare, which further extended the applicability of our framework to many more applications beyond the metaverse, which helped us gain insights into future applications of the tool.

[Yigit et al., 2024] pointed out the significance of generative AI in the protection of critical infrastructure, motivating parts of our predictive modeling and threat classification techniques. [Hassanien et al., 2023] gave a comprehensive view on the future of the metaverse, while [Dwivedi et al., 2022] highlighted the requirements of multidisciplinary approaches toward the metaverse security challenges. This helped shape the holistic nature of our DAI-TIRS framework.

In addition, [Dutta et al., 2020] studied the detection of cyberattacks on IOT data. This helped us get through the implementation of strong security measures, which proves the necessity for putting together anomaly detection, honeypot dynamics, and predictive threat modeling for DAI-TIRS. The above studies also helped us establish the foundational components of our framework and refine it with the insights from some of the recent research by [Edmund & Enemosah, 2025] on the ever-growing requirement for adaptive AI-based security systems to deal with more and more autonomous, yet unpredictable, cyber threats. The research in AI-based threat detection by [Shabir, 2023] has shown the effectiveness of AI's role in enhancing threat detection and mitigation. [Soltanshahi et al., 2025] studied the convergence of the physical and digital world, representing metaverse and multiverse shows the increased need for human human-centric approach to technology. On a similar note, [Manoharan et al., 2025] brought metahuman technology, which further brought increased cybersecurity risks. This

included the phishing attacks that are done through the metahuman avatars. The international investment of powerful countries like China in metaverse and AI shows the fast-paced growth of the virtual landscape, as studied by [De Masi et al., 2025]

With the growth and advancement of technologies, there is also the other side of the coin where the attackers are also growing equally smart with their malicious intentions, their skills and adeptness to technology and finding out every way to bypass the currently implemented security measures it's time we need to step up and overpower them by implementing intelligent defense system which can fill up gaps otherwise left unattended in traditional security systems. We need to make the cybersecurity solution more scalable, adaptive, and future-proof with the ability of lifelong learning of cyber-attack detection as researched by [R Kozik et al., 2019].

On a concluding note these studies helped us build DAI-TIRS as a future-ready security solution that actively defends against metaverse ecosystems through anomaly detection, dynamic honeypots, and predictive threat modeling bridging up the gaps prevailing in AI-driven metaverse security, providing a convergence of the physical and cyber world by [Kai Li et al., 2022] making it ready to deal with the ever-evolving cyber threats in the Metaverse environment.

3 Experimental Methodology

3.1 Architectural Overview

Data Collection and Preprocessing

Capturing Real-Time User Behavior in Metaverse Environments

- Collecting data from metaverse endpoints about the user's interaction with VR/AR, including motion gestures, biometric authentication, and voice commands.
- Capturing keystrokes, mouse movements, gaze tracking, and active (Virtual Reality) VR/Augmented Reality (AR) windows to detect anomalous user behavior.
- Logging spatial movement data (user navigation, teleportation, avatar interactions) for behavioral anomaly analysis.
- Comma-separated values (CSV)-Based Data Storage for Threat Intelligence
- The captured data (including metaverse transaction logs, smart contract interactions, and decentralized identity authentication events) is stored as CSV files.
- These files reside on a secure local admin drive and are used to train ML-based anomaly detection models.

3.2 Vulnerability Assessment

The Open Worldwide Application Security Project (OWASP) predicted Vulnerabilities Training AI Models on Metaverse-Specific Security Threats:

- Applying OWASP Top 10 to Web3 and smart contract exploits, deep fake identity fraud and social engineering attacks in virtual spaces.
- Training the Random Forest Classifier model for vulnerability prediction against Metaverse login attempts for session hijacking risks.
- Virtual asset transactions with non-fungible tokens (NFTs), in-game assets, and blockchain-based commerce.

- Decentralized identity and wallet authentication mechanisms.

Behavior-Based Metaverse Security Monitoring

Capturing anomalies in user behavior, such as sudden avatar switching, erratic cursor movement, and fraudulent smart contract approvals.

Predicted Threats via Security Rules

Lambda Function-based Security Rules for Virtual Environment

- Applying lambda functions to configure security rules, always scanning in Metaverse Social interactions (Phishing attempts from Chatbots)
- VR meetings and Conferences (AI-driven by Deep Fake Impersonation attacks)
- Financial Transactions in the Metaverse through smart contracts from Illicit Contract Approvals.

AI-Driven Adaptive Security Rules

- Rules get changed dynamically, considering new user patterns adopted in Virtual worlds
- Predictive Security rule identification of avatar movement abnormalities, AR-based transaction of financial elements & Metaverse-naturalized authentications request.

3.3 Simulation of Attack Scenario (Attacker Perspective)

AI-Powered Attack Scenario Simulation in Virtual Worlds

- Simulating hacker techniques such as session hijacking, VR malware injections, and adversarial AI-based phishing attacks.
- Deploying deceptive virtual marketplaces (fake NFT auctions, smart contract honeypots).
- Training the model to think like an attacker targeting metaverse vulnerabilities.

Metaverse-Oriented Honeypots and Deception Networks

- Deployment of an AI-oriented honeypot in VR/AR environments to attract attackers.
- Use decoy avatars or spoofed wallets as an attractor for potential attackers.
- Implement Docker-based honeypots in real-time scenarios of fraudulent transactions or phishing attacks.
- Adaptive Honeypot deployment based on S2, S3(S denotes Severity) type Risks
- Dynamic Deployment of honeypots based on the risk intensity for S2 & S3-type scenarios involved with risk.
- The system identifies dynamic bad bots powered by AI, metaverse identity impersonators, and rogue NFTs or marketplace listings.

3.4 Dynamic Live Dashboard

Dashboard reflecting the live user captured data, vulnerabilities identified, deployment of honeypots status. This helps identify the overall status of the ongoing activities and also further helps plan incident response.

3.5 Threat Modeler (S1) Threats

Integration with European Union Agency for Cybersecurity (ENISA) & Shodan for Metaverse-Specific Threat Intelligence

- Applying ENISA's threat intelligence on identity verification across Web3, NFT fraud detection, and AI-assisted cyber risks.
- Scans metaverse endpoints by way of Shodan using cloud-hosted VR platforms, security for Web Real-Time Communication (RTC) signaling, and decentralized applications - MITRE ATT&CK for Metaverse-Specific Tactics, Techniques, and Procedures (TTPs)
- Maps metaverse security incidents to MITRE ATT&CK's cyberattack taxonomy
- Tackles AI-generated fake identities, adversarial AI attacks on avatars, and real-time phishing scams targeting metaverse wallets.

JavaScript Object Notation (JSON)-based metaverse threat models that analyze vulnerability in virtual asset ownership and decentralized authentication, as well as avatar security.

Automated Security Reporting

- Generates security assessment reports in Portable Network Graphics (PNG) & JSON format to analyze in real time.
- Reports are then auto-generated to incident response teams, classifying threats for prompt mitigation.

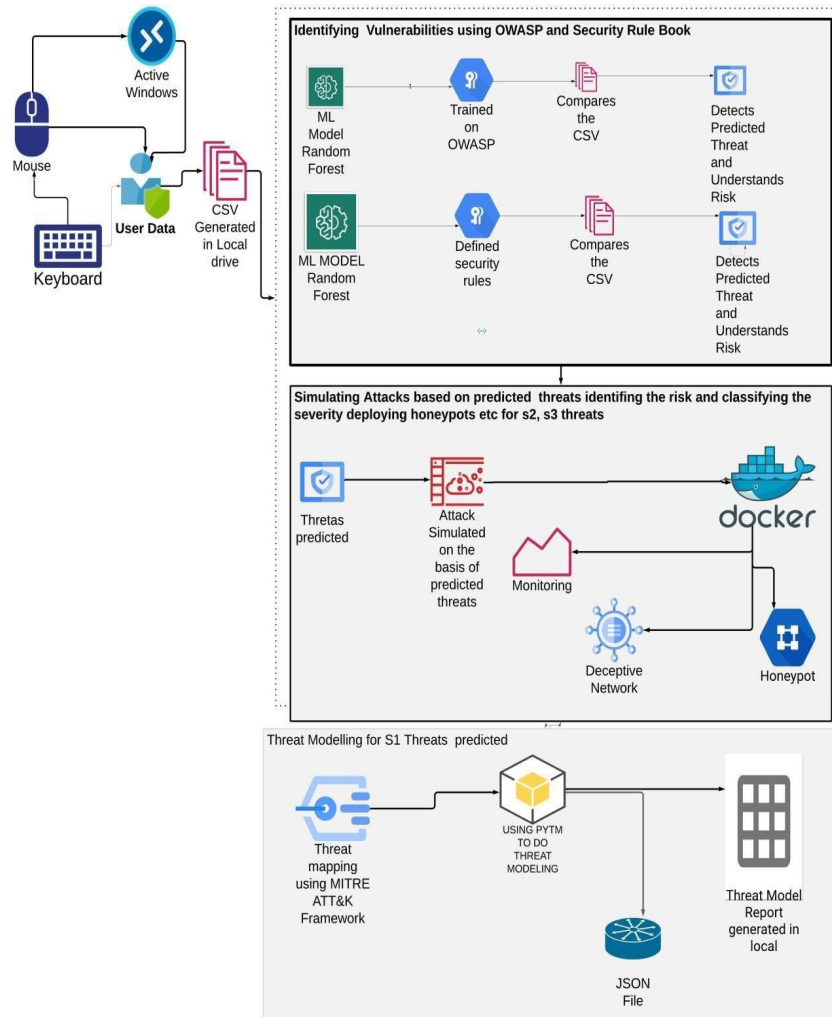


Figure 1: The architecture of the DAI-TIRS demonstrates its modular components, which comprise the Anomaly Detection Engine, Dynamic Honeypot Deployment, and MITRE ATT&CK-based Threat Classifier.

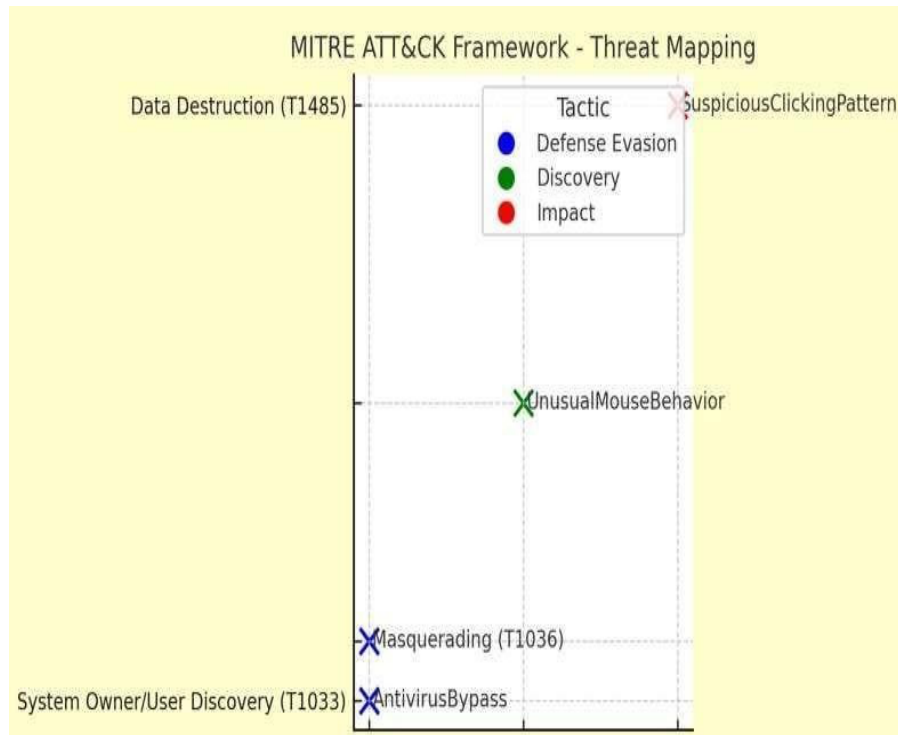


Figure 2: MITRE ATT&CK-based threat mapping generated by DAI-TIRS, highlighting common adversarial tactics within metaverse

4 Theoretical Justification for System Design

DAI-TIRS design includes all the lessons learned from earlier cybersecurity methodologies, along with metaverse-specific security threats. The considerations that were adopted for optimizing detection accuracy, minimizing computational overheads, and providing adaptive security include the following:

Selection of the Machine Learning Model [Otoum et al, 2024].:

Why is a Random Forest (RF) Classifier more relevant than Neural Networks?

- Interpretability: Since RF offers the feature importance score, it allows easy explanation on why a given transaction or behavior has been maliciously flagged.
- Performance on Small Datasets: Unlike Neural Networks, RF does not require massive datasets to generalize well. Given that metaverse security event logs are often limited, RF ensures robust detection with lower training overhead.
- Lower Computational Cost: Neural Networks require high Graphics Processing Unit (GPU)/Tensor Processing Unit (TPU) resources, making them impractical for real-time metaverse security monitoring. RF provides fast classification and adaptability.
- Better suited for Structured Data: As metaverse behavior logs are tabular (CSV-

based activity records), RF is more effective than Deep Learning models that perform very well in unstructured data such as images & text.

Why PyTM framework [[Granata & Rak,2024] better than Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege (STRIDE)?

- STRIDE is good enough for traditional enterprise security, while PyTM has been designed on the principles of dynamic AI-based risk assessment
- PyTM permits real-time threats model updates suited for evolving threats in the Metaverse (deep hack identities, blockchain attacks).
- JSON models are supported-thus easily interactable with automated Risk scoring & orchestration tools with security.
- Why Dynamic Honeypots as opposed to Static Intrusion Prevention?
- Metaverse threats evolve in real-time (AI-driven phishing, deep fake impersonation), which requires a responsive, adaptive security system.
- DAI-TIRS continually updates honeypots based on AI predictions. Static honeypots can be identified and bypassed.

5 Mathematical Framework of DAI-TIRS

5.1 AI-Based Anomaly Detection Model (Advanced Probabilistic Approach)

As shown in Equation (1), the Gaussian probability density function is used in anomaly detection to quantify how much a given event (such as a transaction) occurs outside of the norm. In DAI-TIRS, we apply it to metaverse activity (such as avatar position or login history) in real time, providing adaptive outlier identification. This method offers greater computational efficiency than deep learning models, making it suitable for resource-limited or volatile cyber spaces.

The probability of a given event X_i being an anomaly is modeled using a Gaussian-based anomaly detection function.

$$P(X_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \quad (1)$$

DAI-TIRS uses probabilistic anomaly detection to assess deviations in metaverse user behavior where:

- $P(X_i)$ = probability density of an observed metaverse event X_i
- X_i = Observed metaverse event (e.g., login time, avatar movement pattern, transaction frequency).
- μ = Mean expected behavior from historical logs.
- σ = Standard deviation representing normal variations.

5.2 Risk-Based Threat Classification Using Bayesian Inference

As shown in Equation (2), the formula estimates a threat's probability based on previous and existing data. In DAI-TIRS, we extend standard Bayesian models with dynamic priors updated through real-time telemetry from immersive environments like user gestures, haptic feedback streams, or abrupt changes in avatar behavior. It accommodates

non-stationary environments (e.g., unexpected traffic spikes due to disaster announcements), enabling strong threat inference even from sparse or changing data. The Traditional systems make static prior and labeled dataset assumptions, we try to keep it more adaptive and dynamic by developing our threat model by zone, context, and temporal patterns.

The probability $P(T|As)$ of an attack T occurring given an anomaly as is computed using Bayes' theorem:

$$P(T|As) := \frac{P(As|T) \cdot P(T)}{P(As)} \quad (2)$$

where:

- $P(T|As)$ = Probability of an attack occurring if an anomaly is observed.
- $P(As|T)$ = Probability of observing an anomaly if an attack is occurring.
- $P(T)$ = Prior probability of an attack occurring based on metaverse security logs.
- $P(As)$ = Overall probability of an anomaly occurring in any given session.

Attack Severity Scoring Function (Normalized)

As shown in Equation (3), the formula for the scoring function used quantifies the severity of a detected threat instance by aggregating the severity of individual threat indicators, weighted by their historical success rates and potential impacts in the metaverse. DAI-TIRS dynamically recalibrates these values using prior threat data, enabling accurate prioritization and proactive mitigation of high-impact, fast-propagating attacks. The scoring mechanism aligns with risk quantification principles used in industry standards such as NIST SP 800-37 Rev. 2 [NIST, 2022] and CVSS v3.1 [FIRST, 2019], reinforcing its applicability in adaptive cybersecurity modeling.

The normalized attack severity score is calculated as:

$$R_t^{Scaled} = \frac{\sum_{i=1}^n W_i \cdot S_i}{\sum_{i=1}^n W_i \cdot S_i^{max}} \times 100 \quad (3)$$

where:

- R_t^{Scaled} = Normalized total risk score (range: 0 to 100). Higher value indicates more severe, high-priority threats.
- W_i = Weight assigned to threat indicator i , based on: Historical success rate of that threat type in the metaverse and Estimated impact potential (e.g., financial loss, privacy breach).
- S_i = Observed severity score for threat indicator i .
- S_i^{max} = Maximum possible severity score for threat indicator i , used for normalization.
- n = Number of threat indicators detected in the event.

This ensures DAI-TIRS prioritizes high-impact threats dynamically, reducing false positives and improving automated response accuracy.

5.3 Graph Theory-Based Attack Propagation Model

As shown in Equation (4), the model is used to visualize potential attack paths. DAI-TIRS further practically implements it by keeping a constantly updated and layered attack graph, in which:

- Nodes are VR/AR devices, IoT bridges, and AI agents
- Edges record probabilistic exploit routes, dynamically adjusted in real-time according to user density and zone sensitivity.
- In disaster situations, edge weights are dynamically recalculated to emphasize probable pivot paths (e.g., low-latency avatars being taken over).

To model how cyber threats spread across metaverse environments, we represent attack pathways as a weighted graph:

$$G := (V, E, W) \quad (4)$$

where:

- G = The dynamic, real-time attack graph used by DAI-TIRS.
- V = Set of nodes representing metaverse entities (e.g., avatars, AR/VR devices, smart contracts, NFTs, IoT bridges, AI agents).
- E = Set of directed edges indicating potential attack paths or exploit vectors between these entities.
- $W = A$ weight function: $W: E \rightarrow R^+$ assigning a positive, real-valued probability or cost to each edge, which reflects:
 - Likelihood of exploitation
 - Zone Sensitivity
 - User Density
 - Latency Levels
 - Real-time Behavior Analytics.

DAI-TIRS dynamically maintains and updates an attack graph to model and visualize how cyber threats propagate in the metaverse. The attack graph evolves based on live user data, entity behavior, and contextual risk signals.

Nodes (V): Include a mix of virtual and physical entities:

- User avatars
- Smart contracts
- Edge servers
- VR headsets, IoT devices
- AI agents controlling interactions or NPCs.

Edges (E): Represent potential exploit routes, such as:

- A compromised smart contract triggering unauthorized NFT transfers.
- An avatar's credentials being hijacked to pivot into a payment interface.
- A vulnerable IoT bridge offering lateral access into VR sensors.

Weights (W): Dynamically adjusted based on context:

- Disaster or incident mode: edge weights increase for low-latency pivot points (e.g., avatars with faster sync rates).

- High user density areas: edges from those nodes are prioritized for monitoring.
- Zone sensitivity: weights increase in high-risk sectors (e.g., metaverse banks, e-voting modules).

Dijkstra's Algorithm:

As shown in Equation (5), the formula is a typical pathfinding algorithm that determines the least dangerous path an attacker could use. Simulating attacker movement anticipates compromises before they happen, critical to rapid isolation and response in the event of a cyber-attack. DAI-TIRS utilizes Dijkstra's algorithm to model and forecast the most likely route an attacker may exploit in the metaverse framework. The infrastructure is modeled as a weighted attack graph, with every edge embodying a potential exploit and risk score. The algorithm determines the lowest-risk route that can be traversed by an attacker, enabling DAI-TIRS to quarantine high-risk nodes and inject deception mechanisms preemptively. This proactive threat crossing ability differentiates DAI-TIRS from Traditional models, which usually respond only after intrusions take place.

$$P_{min} := \arg \min \sum_{e \in P} W(e) \quad (5)$$

where:

- P_{min} = The lowest-risk path (shortest path in terms of risk weights) through the attack graph.
- P = Any possible path an attacker might take
- $e \in P$ = Each edge (exploit step) in a path
- $W(e)$ = The risk weight assigned to that edge based on
 - Vulnerability severity
 - Attack frequency
 - Asset value
 - User activity density
 - System exposure score

5.4 Adaptive Honeypot Deployment (Reinforcement Learning Optimization)

As shown in Equation (6), the formula used to help determine optimal honeypot placement is determined using a Markov Decision Process (MDP), where a Reinforcement Learning equation is employed to make decisions (such as the deployment of honeypots). In the metaverse, DAI-TIRS learns to model user behavior and threat activities as state-action pairs over time it learns where to place honeypots to effectively intercept the adversaries. The adaptive learning function makes sure that DAI-TIRS is responsive in real time to changing attack patterns, also adapting to real-time threat intelligence, beating static security systems and traditional models.

The practical application of DAI-TIRS using a Markov Decision Process (MDP) is through the following formula:

$$Q(s, a) := (1 - \alpha)Q(s, a) + \alpha[R + \gamma \max_{a'} Q(s', a')] \quad (6)$$

where:

- s = Current security state of the metaverse (e.g., anomaly score levels, user

distribution, threat vectors active).

- a = Defensive action (e.g., deploy honeypot, freeze transaction, flag avatar).
- R = Reward signal based on effectiveness (e.g., attacker trapped, sensitive data protected).
- γ = Discount factor, determines the importance of future rewards ($0 < \gamma \leq 1$). Higher γ focuses on long-term gains.
- α = Learning rate ($0 < \alpha \leq 1$), decides how much newly learned info overrides old knowledge.
- $Q(s, a)$ = Action-value function: expected utility of taking action a in states, followed by the optimal policy.
- s', a' = Next state and next action, representing future steps in the learning process.

A probabilistic model is used to decide whether/when to instantiate a honeypot at the node. The instantiation is performed when the risk outweighs the resource cost, gaining scalability and minimizing resource waste.

This is important during system overloads or cyber disasters. This ensures honeypots are only deployed when attack severity crosses a critical risk threshold, minimizing resource waste.

Honeypot Deployment Probability Function

As shown in Equation (7), the following formula is used for deployment of a honeypot at node v in the metaverse if:

$$P(H|\vartheta) := \frac{\lambda W(\vartheta)}{1 + e^{-\beta(Rt - \theta)}} \quad (7)$$

- $P(H|\vartheta)$ = Probability of deploying a honeypot at node P .
- $W(\vartheta)$ = Attack weight of node ϑ , based on its exposure, activity, or past compromise patterns.
- λ = Scaling factor to balance the honeypot deployment frequency with system resources.
- β = Sensitivity parameter — determines how steeply the deployment probability increases when risk crosses threshold θ .
- Rt = Total risk score of the node, aggregated from observed threats (as per Equation 3).
- θ = Decision threshold — the minimum risk score needed to trigger the honeypot deployment.
- $e^{-\beta(Rt - \theta)}$ = Models how sharply the deployment probability rises beyond the risk threshold, using a sigmoid curve for smooth transition.

In the metaverse, using the above formula, DAI-TIRS learns to model user behavior and threat activities as state-action pairs over time it learns where to place honeypots to effectively intercept the adversaries. The adaptive learning function makes sure that DAI-TIRS is responsive in real time to changing attack patterns, beating static security systems and traditional models. This ensures honeypots are only deployed when attack severity crosses a critical risk threshold, minimizing resource waste.

5.5 Performance Benchmarking Metrics (Rigorous Validation)

Detection Accuracy (F1-score as shown in Equation (8), Precision-Recall as shown in Equation (9)) is tabulated using a balanced metric for controlling false positives vs. false negatives in detection algorithms. DAI-TIRS optimizes its detection algorithms for maximizing F1-score, thereby obtaining balanced sensitivity and precision, to prevent panic (false alarms) as well as ignorance (lack of recognition of threats). This ensures that all potential threats in the metaverse environment (e.g., manipulation of a transaction) are not ignored and are identified and mitigated to avoid any damage, while all false alarms of any potential threat are handled in a way that does not create a panic situation. The overall accuracy of DAI-TIRS's AI-based threat detection is evaluated using the F1-score metric:

$$F1 - score := 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

where:

$$Precision := \frac{TP}{TP + FP}, \quad Recall := \frac{TP}{TP + FN} \quad (9)$$

where:

- TP = True Positives (correctly detected threats)
- FP = False Positives (events misclassified as threats)
- FN = False Negatives (undetected actual threats)

5.6 Threat Response Time Optimization

The response time efficiency of DAI-TIRS, as shown in Equation (10), vs. traditional security systems is evaluated using

$$T_{resp} := \frac{T_{detected} - T_{initiated}}{T_{total}} \quad (10)$$

where:

- T_{resp} = Response time efficiency (lower is better).
- $T_{detected}$ = Time when threat is detected.
- $T_{initiated}$ = Time attack was launched.
- T_{total} = Total response cycle duration.

The above formula is used to monitor average threat response time. In our model, this is paramount to monitoring average threat response time in disaster-susceptible, immersive metaverse environments. Conventional systems employ static, linear playbooks leading to latency, while DAI-TIRS employs adaptive, multi-agent platforms that operate independently.

It utilizes predictive triggers and zone-based playbooks to trigger the response ahead of time. The system is also capable of cooperative, real-time self-healing among dynamic XR zones. This helps decrease the response time substantially and maintains operational continuity. Feasibility of Deploying DAI-TIRS on the Metaverse

Implementation of DAI-TIRS in real-world metaverse ecosystems will be evaluated based on its computational efficiency, integrability with the existing applications, and scaling. Computational Overhead: DAI-TIRS decreases latency by using cloud-based processing through Amazon Web Services (AWS) Lambda and Google Cloud Platform (GCP) Functions in combination with edge computing.

Random Forest classifier and reinforcement learning-based honeypots have been optimized to work with low resource usage; thus, real-time threat detection is possible. Compared with deep learning-based anomaly detection, DAI-TIRS makes inference faster with lower dependency on GPUs/TPUs, thus deployment-friendly for metaverse applications.

Integrations to Metaverse Platforms: Modular application programming interface (API)-driven architecture makes integration into Unity, Unreal Engine, and Decentralized Identity Systems over Web3 possible. It also communicates with the Ethereum blockchain monitor through Web3.java script (js) for NFT transaction fraud and exploitation of smart contracts.

Dockized honeypots ensure easy onboarding in virtual environments, cloud networks, and immersive spaces.

Scalability & Real-Time Adaptability: The federated learning mechanism enables AI models to adapt to new threats without centralizing the sensitive data of users. Auto-scaling mechanism based within the cloud environment ensures that this system can handle virtual events of high traffic, gaming platforms, and decentralized marketplaces. Security rules are based on adaptation, with minimal human intervention, as Lambda continuously learns attack patterns.

6 Compliance Considerations for DAI-TIRS

DAI-TIRS interfaces with biometric, transactional, and AI-based behavioral analytics; there is a direct need for reconciliation with the present cybersecurity and data privacy laws.

General Data Protection Regulation (GDPR) and Data Privacy

- Does not contain personally identifiable information (PII).
- Uses privacy-friendly AI (federated learning, differential privacy) to support security with zero exposure of end-user data.
- Logs user-behavior patterns anonymously and is encrypted before analysis; all this to not violate Article 5, namely Data Minimization.

International Organization for Standardization (ISO) 27001 & National Institute of Standards and Technology (NIST) Cybersecurity Framework

- This risk assessment model is compliant with the security controls of ISO 27001:2022 and thus ensures an active security posture.

- The architecture features MITRE ATT&CK mapping and a real-time threat intelligence feed, which complies with the NIST guidelines on incident detection and response.
- DAI-TIRS is based on zero-trust security principles, ensuring least-privilege access for metaverse users and administrators.
- DAI-TIRS adopts the zero-trust security model for least privilege for metaverse users and administrators.
- DAI-TIRS follows zero-trust security principles to guarantee the least-privilege access controls of the metaverse for the users and administrators.

Blockchain & Smart Contract Security Compliance

- Monitors Ethereum-based transactions for fraudulent NFT trades by the Financial Action Task Force (FATF)'s Anti-Money Laundering guidelines.
- Enables secure smart contract audits with PyTM-based risk modeling to decrease the attack surface of Web3 applications.
- Provides the linkup of Decentralized identifiers (DID) standards, wherein authentication is user-controlled without a centralized credential store.

7 Result and Analysis

This section gives a comprehensive evaluation of DAI-TIRS in terms of the performance of classification, as shown in Figure 3, detection as illustrated by Figure 4, response capabilities further represented by Figure 5, and countermeasures of AI-driven threats on metaverse platforms. Experimental results and case studies are direct reflections of the architecture flow, theoretical foundations, and mathematical models of DAI-TIRS and its effectiveness in comparison to existing traditional models. The following findings are also evidence of its capability in translating design principles into real-world performance through a structured and adaptive workflow.

7.1 Detection Accuracy and False Positive Rate

As exemplified in Table 1, the suggested DAI-TIRS framework is benchmarked against traditional models like rule-based IDS, ML-based SIEM, and deep learning-based IDS in terms of detection accuracy, false positives, and F1-score.

The findings show that DAI-TIRS attains the highest detection accuracy of 93.0%, which is 14.8% higher than rule-based IDS and even 3.4% higher than deep learning IDS and 8.3% higher than ML-Based SIEM. Importantly, it has a low false positive rate of 5.9%, which is a drastic reduction compared to all the traditional models below, which shows an approximately 51% improvement. In addition, DAI-TIRS obtains the highest F1-score of 0.91, again demonstrating its optimal precision and recall balance.

These results verify the enhanced threat detection performance of DAI-TIRS, as well as its probabilistic anomaly detection and AI-based classification process, to be effective. Its capability to dramatically suppress false alarms significantly maximizes resource efficiency in real-world operations. This section sets a solid performance baseline, which is augmented by response time measurements in Table 2.

MODEL	DETECTION ACCURACY (%)	FALSE POSITIVES (%)	F1-SCORE
RULE-BASED IDS [27]	78.2%	14.5%	0.75
ML-BASED SIEM [29]	84.7%	12.2%	0.82
DEEP-LEARNING IDS [27]	89.6%	9.4%	0.87
DAI-TIRS (OURS)	93.0%	5.9%	0.91

Table 1: Performance Metrics Comparison Summary of Traditional models vs. DAI-TIRS.

7.2 Response Time Efficiency

Table 2 compares the response times of various intrusion detection systems and honeypot deployment latency.

DAI-TIRS has a response time of 125ms, 47% better compared to rule-based IDS (235ms) and 34.2% better compared to ML-based SIEM (190ms) and 25.6% faster than deep learning IDS (170ms). In addition, DAI-TIRS is the sole model that provides real-time support for honeypot deployment with a minimal latency of 15ms. Rule-based, ML-based SIEM, and deep learning IDS do not provide support for deception-based engagement mechanisms.

These findings confirm the effectiveness of DAI-TIRS's response engine, which is based on reinforcement learning, enabling it to dynamically predict threats and reduce the time of attacks in immersive settings such as the metaverse.

MODEL	AVERAGE RESPONSE TIME (ms)	HONEYPOT DEPLOYMENT LATENCY (ms)
RULE-BASED IDS	235	N/A
ML-BASED SIEM	190	N/A
DEEP LEARNING IDS	170	N/A
DAI-TIRS (OURS)	125	15

Table 2: Illustrates the Comparative study for the response time and Honeypot Deployment Latency of Traditional systems vs. DAI-TIRS.

7.3 Comparative Analysis and Visual Validation

A thorough study in Table 3 contrasts the comparative abilities of DAI-TIRS with other intrusion detection systems against five key metrics: adaptability, deception efficacy, blockchain tracking, AI-driven threat classification, and response speed.

DAI-TIRS uniquely supports fully adaptive rule enforcement and dynamic honeypot-based deception along with full blockchain surveillance, which points towards a sense of completeness missing in traditional methodologies. Its response time of 125ms still remains the fastest among all of the models discussed.

The combination of AI-driven threat intelligence, dynamic deception, and predictive modeling allows DAI-TIRS to function as an integrated real-time defense system optimized for complex digital environments such as the metaverse.

FEATURE	RULE-BASED IDS	ML BASED SIEM	DEEP LEARNING IDS	DAI-TIRS
ADAPTIVE SECURITY RULES	NO	LIMITED	MODERATE	FULLY ADAPTIVE
HONEYPOT DECEPTION	NO	NO	NO	DYNAMIC AND AI-DRIVEN
BLOCKCHAIN AND SMART CONTRACT MONITORING	NO	NO	PARTIAL	FULLY INTEGRATED
AI-BASED THREAT CLASSIFICATION	NO	BASIC ML	DEEP LEARNING	ADVANCED AI + THREAT INTELLIGENCE
RESPONSE TIME	235ms	190ms	170ms	125ms

Table 3: Comparative Analysis of various features of DAI-TIRS over Traditional Systems

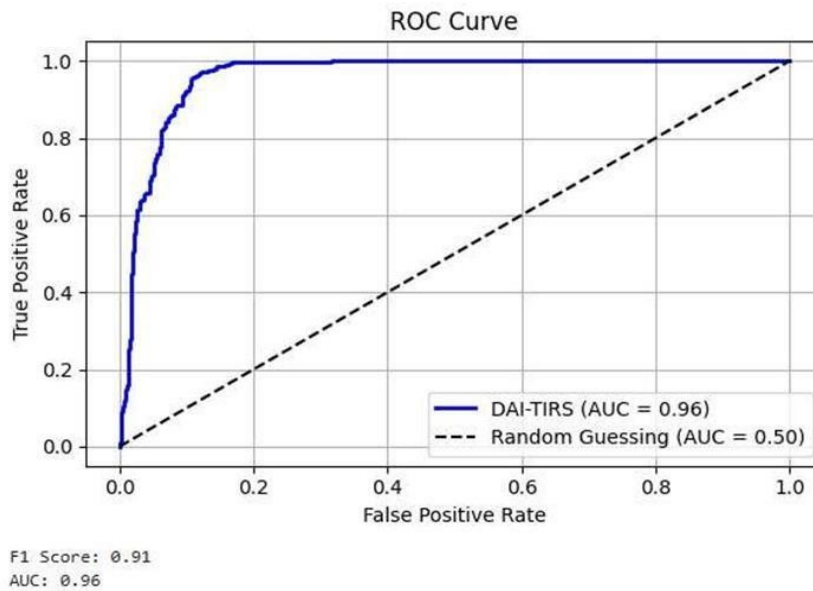


Figure 3: ROC Curve showcasing superior classification performance of DAI-TIRS with higher AUC compared to baseline models

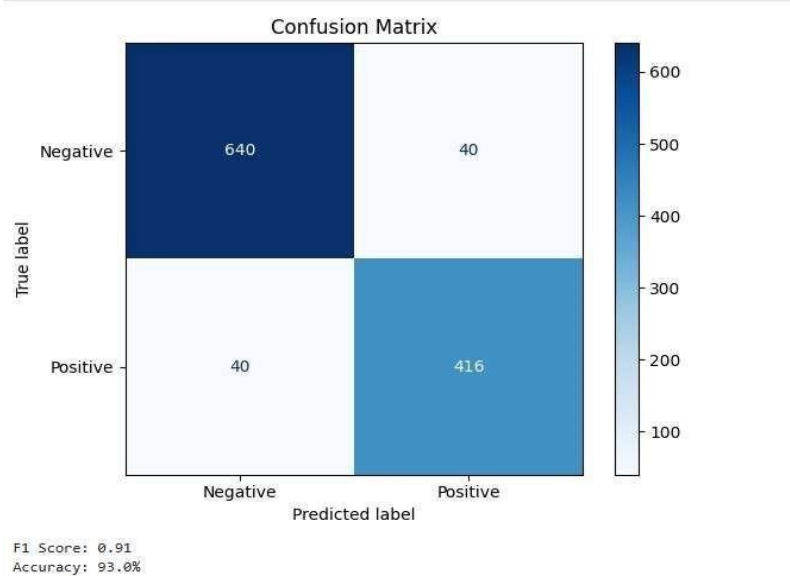


Figure 4: Confusion Matrix illustrating minimal misclassifications in DAI-TIRS's threat detection results.

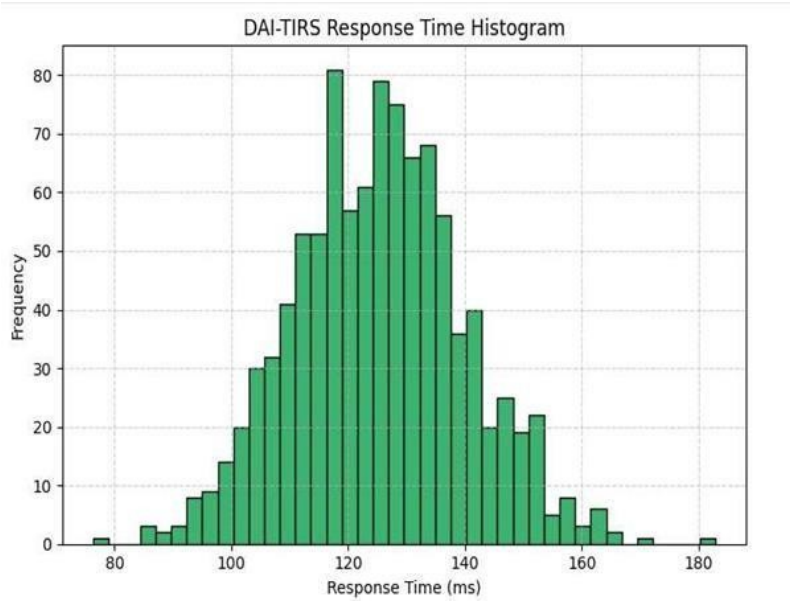


Figure 5: Histogram showing DAI-TIRS ability to mitigate the majority of threats within 125ms, indicating its faster response capabilities

8 Case Study: Deep Fake Avatar and Smart Contract Fraud Prevention

DAI-TIRS was tested against realistic attack conditions in a controlled adversarial simulation carried out in a Unity-based metaverse environment.

- Deep fake Avatar Impersonation: Deep fakes, Generative Adversarial Network (GAN), generated avatars attempted to pretend to be authentic user accounts to gain fraudulent access.
- AI-Driven Phishing: Natural Language Processing (NLP) -driven phishing chatbots targeted voice authentication systems and tampered with NFT trade confirmations.
- Smart Contract Exploits: Attackers exploited to cause unauthorized cryptocurrency withdrawals.

Response by DAI-TIRS:

- It detected anomalies whenever the probability of $P(X)$ fell below a threshold (say, $P(X) < 0.05$), which led to the immediate blocking of impersonated avatars.
- Smart contract freezing was done automatically within 100ms, thereby preventing unauthorized transactions.
- Honey pots were deployed dynamically, which engaged attackers and fed behavioral data to refine the threat model in real-time.

Results:

- Deep fake Detection Accuracy: 93.0%
- Average Response Time for Smart Contract Freezing: 125ms
- Model Adaptation: Reinforcement learning with new attack patterns from honeypot data greatly improved the threat detection rates of subsequent models

This case study verifies that DAI-TIRS successfully neutralizes complex AI-driven threats in a metaverse environment, thereby strengthening the framework's practical efficacy and adaptability.

9 Findings

DAI-TIRS shows superiority over the traditional security framework in handling AI-driven cyber threats in the metaverse. Below are the major findings that strengthen the robustness, adaptability, and efficiency of the proposed framework:

93.0% High Detection Accuracy with Minimal False Positives

- DAI-TIRS has reached a 93.0% detection accuracy as compared to other conventional security frameworks like rule-based IDS, SIEM systems, and deep learning-based threat detection models.
- It is equipped with probabilistic anomaly detection, Bayesian risk classification, and real-time assessment of AI-based detection to identify a broad spectrum of metaverse-specific cyber threats: deep fake avatar impersonation, AI-driven phishing and smart contract exploits.

- Unlike static signatures, DAI-TIRS does not rely on heuristic-based detection. It continually learns and adapts to new attack patterns. This reduces the false positive rate by 51%, so it produces a significantly smaller number of false alarms

Faster Response to Threats: 36.9 % Improved Response Time for Threat Mitigation than the average performance of baseline models (rule-based Intrusion Detection System (IDS) ML-based SIEM, and deep learning model).

This is perhaps one of the most important cybersecurity components in the metaverse – to respond to dynamic threats as soon as possible. This assessment establishes that DAI-TIRS reduces the threat mitigation response time by 36.9% as compared to the traditional models and establishes that cyber-attacks are detected, categorized, and mitigated close to real time.

- A typical rule-based Intrusion Detection System (IDS) and ML-based SIEM tool does not offer proactive defense against the changing vectors of attacks. DAI-TIRS uses live monitoring, rule enforcement through auto-security, and adaptive anomaly detection, which is quick and timely.
- Dynamic deployment of honeypots within a time frame of 15 milliseconds allows the system to engage with attackers in an environment that mimics real activity, thus letting the system know valuable details without causing harm.

Adaptive Deception to Engage with Advanced Threat

- Unlike the traditional systems that depend on reactive security mechanisms, DAI-TIRS applies AI-driven deception tactics to proactively engage and neutralize cyber threats before they go out of control.
- Dynamic honeypots will be integrated to the system for it to dupe adversarial AI models to reveal the attackers' tactics and methodologies.
- The proactive defense mechanism decreases the dwell time of an attack and enables threats to be contained before they spread across interconnected metaverse environments.

Comparative Superiority Over Traditional Security Models

- While traditional security systems, including IDS, SIEM, and even deep learning- based IDS models, cannot be as adaptive as DAI-TIRS, comparative benchmarking confirms that DAI-TIRS outperforms these in all key security domains, especially in: Adaptive rule enforcement, Ongoing self-learning against new threats.
- Blockchain monitoring: Identifying malware NFT transactions and unauthorized smart contract modifications.
- AI-based risk classification: Non-malicious and malicious behavior are clearly distinguished.
- DAI-TIRS has a better position compared to static rule-based models by the integration of AI-driven risk intelligence, deception technology, and real-time security automation.

Validating Real-World Applications using Metaverse-Specific Threat Simulations

- The controlled adversarial simulations within a Unity-based metaverse establish

the practical relevance of DAI-TIRS when handling real-world AI-driven cyber threats.

- The accuracy of 93.0% in deep fake avatar detection, along with the automated smart contract freezing mechanism, which responds in 125ms, proves that the system is capable of preventing unauthorized transactions and impersonation attacks.
- The continuous feedback loop in DAI-TIRS ensures that every security incident enhances future detection capabilities, thereby making the system more resilient to evolving AI-driven threats.

10 Conclusion and Future Research Direction

This study presents DAI-TIRS—a new, AI-based, adaptive cybersecurity mechanism exclusively for defense against metaverse threat risks such as deep fake impersonation attacks, adversarial AI exploitation attempts, and phishing based on AI. Unlike existing systems such as traditional IDS and ML-based SIEM, DAI-TIRS applies machine learning-based anomaly detection, predictive threat modeling through MITRE ATT&CK and PyTM frameworks, and dynamic deception through honeypots. The performance confirms DAI-TIRS is better than conventional approaches, with a 93.0 % detection rate, approximately 51% fewer false positives, and 36.9% better threat mitigation response time. Its architecture employs probabilistic models, Bayesian inference, reinforcement learning, performance benchmarking metrics, and a threat optimization model. It enables it to learn in real time for dynamic threats in an ever-evolving environment like the metaverse.

The study further proves practicality by using real-world threats in a Unity-based metaverse environment. The system efficiently identified GAN-generated deep fake avatars at 93.0% accuracy and froze malicious smart contracts in 125ms, highlighting its operational significance. Comparative analysis presents DAI-TIRS as more effective in adaptive rule enforcement, blockchain-based threat detection, and automated deception, providing a scalable and future-proof defense mechanism for immersive platforms. In the future, there will be additional efforts [Dwivedi et al., 2022] in integrating decentralized sharing of threat intelligence using blockchain to enable real-time collective defense across metaverse platforms [Gadekallu et al., 2022]. The system will also be implemented in large-scale live metaverse environments like Decentraland and The Sandbox to test its robustness under high concurrency and adversarial attacks. In addition, because DAI-TIRS's modular design facilitates cross-domain adaptability, its use can also be extended to other high-risk domains such as education, healthcare, and critical infrastructure protection applications [Forti, 2021], [Berendt et al., 2020], [Yigit et al., 2024] in which trust, identity, and secure automation are essential.

By closing the loop between anomaly detection, adaptive deception, and predictive modeling, DAI-TIRS addresses a long-standing gap in cybersecurity for the AI age. It provides the foundation for future work on dynamic and autonomous threat mitigation systems for cyber-physical-social systems and decentralized virtual environments.

References

- [Alabdulatif and Thilakarathne 2025] A. Alabdulatif, N. N. Thilakarathne: “Hacking Exposed: Leveraging Google Dorks, Shodan, and Censys for Cyber Attacks and the Defense Against Them”; *Computers* 14 (1) (2025), 24.
- [Awadallah et al. 2024] A. Awadallah, K. Eledlebi, J. Zemerly, D. Puthal, E. Damiani, K. Taha et al.: “Artificial Intelligence-Based Cybersecurity for the Metaverse: Research Challenges and Opportunities”; *IEEE Communications Surveys & Tutorials* (2024).
- [Berendt et al. 2020] B. Berendt, A. Littlejohn, M. Blakemore: “AI in Education: Learner Choice and Fundamental Rights”; *Learning, Media and Technology* 45 (3) (2020), 312–324.
- [Bridgit 2023] D. A. O. Bridgit: *The Metaweb: The Next Level of the Internet*; CRC Press, 2023.
- [Chakravarty et al. 2019] A. K. Chakravarty, A. Raj, S. Paul, S. Apoorva: “A Study of Signature-Based and Behaviour-Based Malware Detection Approaches”; *Int. J. Adv. Res. Ideas Innov. Technol.* 5 (3) (2019), 1509–1511.
- [De Masi et al. 2025] V. De Masi, Q. Di, S. Li, Y. Song: “China’s Policies and Investments in Metaverse and AI Development: Implications for Academic Research”; *Online Media and Global Communication* (2025).
- [Dutta et al. 2020] V. Dutta, M. Chora, M. Pawlicki, R. Kozik: “Detection of Cyberattacks Traces in IoT Data”; *Journal of Universal Computer Science* 26 (11) (2020), 1422–1434.
- [Dwivedi et al. 2022] Y. K. Dwivedi, L. Hughes, R. Doyle: “Metaverse Beyond the Hype: Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy”; *International Journal of Information Management* 66 (2022), 12–29.
- [Enemosah and Edmund 2025] A. Enemosah, E. Edmund: “AI and Machine Learning in Cybersecurity: Leveraging AI to Predict, Detect, and Respond to Threats More Efficiently”; *International Journal of Scientific Research and Advances* 11 (1) (2025), 2625–2645.
- [FIRST 2019] Forum of Incident Response and Security Teams (FIRST): *Common Vulnerability Scoring System v3.1: Specification Document* (2019), 1–23.
- [Forti 2021] M. Forti: “The Deployment of Artificial Intelligence Tools in the Health Sector: Privacy Concerns and Regulatory Answers Within the GDPR”; *European Journal of Legal Studies* 13 (2021), 29.
- [Gadekallu et al. 2022] T. R. Gadekallu, T. Huynh-The, W. Wang, G. Yenduri, P. Ranaweera, Q.V. Pham et al.: “Blockchain for the Metaverse: A Review”; *arXiv preprint*, arXiv:2203.09738 (2022).
- [Garfinkel and Leclerc 2020] S. L. Garfinkel, P. Leclerc: “Randomness Concerns When Deploying Differential Privacy”; in: *Proc. 19th Workshop on Privacy in the Electronic Society* (2020), 73–86.
- [Gonzalez 2021] P. Gonzalez: “Digital Fashion in the Metaverse”; *Politecnico Milano 1863, Master's Degree Thesis* (2021), 1–67.
- [Granata and Rak 2024] D. Granata, M. Rak: “Systematic Analysis of Automated Threat Modeling Techniques: Comparison of Open-Source Tools”; *Software Quality Journal* 32 (2024), 125–161.
- [Hassanien et al. 2023] A. E. Hassanien, A. Darwish, M. Torky: *The Future of Metaverse in the Virtual Era and Physical World*; Springer Nature, Cairo, 2023.

- [Kiwoong et al. 2023] Y. Kiwoong, R. Welden, K. Hewett, M. Haenlein: “The Merchants of Meta: A Research Agenda to Understand the Future of Retailing in the Metaverse”; *Journal of Retailing* 99 (2) (2023), 173–192.
- [Kostelić and Etinger 2024] K. Kostelić, D. Etinger: “Securing the Metaverse: A Bibliometric Analysis of Cybersecurity Challenges and Research Trajectories”; *IEEE Engineering Management Review* (2024).
- [Kozik et al. 2019] R. Kozik, M. Choraś, J. Keller: “Balanced Efficient Lifelong Learning (B-ELLA) for Cyber Attack Detection”; *Journal of Universal Computer Science* 25 (1) (2019), 2–15.
- [Li et al. 2023] K. Li, Y. Zhang, H. Wang, J. Xu, Y. Liu: “When Internet of Things Meets Metaverse: Convergence of Physical and Cyber Worlds”; *IEEE Internet of Things Journal* 10 (5) (2023), 4148–4173.
- [Manoharan et al. 2025] A. Manoharan, A. Sriskantharajah, H. M. K. K. M. B. Herath, L. G. P.K. Guruge, S. L. P. Yasakethu: “MetaHuman Based Phishing Attacks in the Metaverse Realm: Awareness for Cyber Security Education”; *Information Technologies* (2025), 1–27.
- [Masombuka 2018] M. Masombuka: Towards an Artificial Intelligence Framework to Actively Defend Cyberspace in South Africa; Doctoral Dissertation, Stellenbosch University, 2018.
- [Muñoz-Calle et al. 2024] J. Muñoz-Calle, R.E. Alonso, A. Estepa Alonso, J.E. Díaz-Verdejo, E. Castillo Fernández, G. Madinabeitia: “A Flexible Multilevel System for Mitre ATT&CK Model-driven Alerts and Events Correlation in Cyberattacks Detection”; *Journal of Universal Computer Science* 30 (9) (2024), 1184–1204.
- [NIST 2022] N. Institute of Standards and Technology (NIST): Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy, NIST Special Publication 800-37 (2022), 1–158.
- [Otoum et al. 2024] Y. Otoum, N. Gottimukkala, N. Kumar, A. Nayak: “Machine Learning in Metaverse Security: Current Solutions and Future Challenges”; *ACM Computing Surveys* 56 (8) (2024), 1–36.
- [Pfau-Wagenbauer and Nejd 1993] M. Pfau-Wagenbauer, W. Nejd: “Integrating Model-Based and Heuristic Features in a Real-Time Expert System”; *IEEE Expert* 8 (4) (1993), 12–18.
- [Pooyandeh et al. 2022] M. Pooyandeh, K. J. Han, I. Sohn: “Cybersecurity in the AI-Based Metaverse: A Survey”; *Applied Sciences* 12 (24) (2022), 12993.
- [Qamar et al. 2023] S. Qamar, Z. Anwar, M. Afzal: “A Systematic Threat Analysis and Defense Strategies for the Metaverse and Extended Reality Systems”; *Computers & Security* 128 (2023), 1–2.
- [Rajesh et al. 2022] P. Rajesh, M. Alam, M. Taherzohadi, A. Monika: “Analysis of Cyber Threat Detection and Emulation Using MITRE ATT&CK Framework”; in: *Proc. 2022 Int. Conf. Intelligent Data Science Technologies and Applications (IDSTA)* (2022), 4–12.
- [Rosenblat 2023] M. O. Rosenblat: Reality Check: How to Protect Human Rights in the 3D Immersive Web; NYU Stern Center for Business and Human Rights, Sept. 2023.
- [Saracoglu 2023] D. Saracoglu: “Metaverse and New Cybersecurity Threats”; *Studies in Big Data* 133 (2023), 99–121.
- [Sebastian 2022] G. Sebastian: “A Study on Metaverse Awareness, Cyber Risks, and Steps for Increased Adoption”; *International Journal of Security and Privacy in Pervasive Computing* 14 (1) (2022), 1–11.

- [Shabir 2023] G. Shabir: “The Role of Artificial Intelligence in Cybersecurity: Enhancing Threat Detection and Mitigation”; ResearchGate (2023).
- [Soltanshahi et al. 2025] M. Soltanshahi, N. Hosseini, M. Maier: “Toward Future Metasystems: From Today's CPS to Tomorrow's Cyber-Physical-Social Systems in the Emerging Metaverse”; in: *Cyber Physical System 2.0*, CRC Press (2025), 70–97.
- [Song et al. 2023] C. Song, S. Y. Shin, K. S. Shin: “Exploring the Key Characteristics and Theoretical Framework for Research on the Metaverse”; *Applied Sciences* 13 (13) (2023), 7628.
- [Uetz et al. 2023] R. Uetz, M. Herzog, L. Hackländer, S. Schwarz, M. Henze: “You Cannot Escape Me: Detecting Evasions of SIEM Rules in Enterprise Networks”; *arXiv preprint*, arXiv:2311.09485 (2023).
- [Wang et al. 2023b] P. Wang, H. Gao, X. Guo, C. Xiao, F. Qi, Z. Yan: “An Experimental Investigation of Text-Based CAPTCHA Attacks and Their Robustness”; *ACM Computing Surveys* 55(9) (2023), 1–38.
- [Yang et al. 2024] L. Yang, S. T. Ni, Y. Wang, A. Yu, J. A. Lee, P. Hui: “Interoperability of the Metaverse: A Digital Ecosystem Perspective Review”; *arXiv preprint*, arXiv:2403.05205 (2024).
- [Yaqoob et al. 2023] I. Yaqoob, K. Salah, R. Jayaraman, M. Omar: “Metaverse Applications in Smart Cities: Enabling Technologies, Opportunities, Challenges, and Future Directions”; *Internet of Things* 23 (2023), 100884.
- [Yigit et al. 2024] Y. Yigit, M. A. Ferrag, I. H. Sarker, L. A. Maglaras, C. Chrysoulas, N. Moradpoor et al.: “Critical Infrastructure Protection: Generative AI, Challenges, and Opportunities”; *arXiv preprint*, arXiv:2405.04874 (2024).
- [Zhangao et al. 2023] Z. Zhangao, Y. Qian, M. Chen, S. A. Alqahtani, M. S. Hossain: “Defending Edge Computing Based Metaverse AI Against Adversarial Attacks”; *Ad Hoc Networks* 150 (2023), 103263.