

Artificial Intelligence-Based Cybersecurity for the Metaverse: Research Challenges and Opportunities

Thange Sakshi¹, Prof. V. N. Dhakane², Prof. P. P. Randive³, Dr. H. B. Jadhav⁴

Student, Computer Engineering¹,
Guide, Computer Engineering²
Assistant Professor, Computer Engineering³,
HOD, Computer Engineering⁴
Adsul Technical Campus, Ahilyanagar, India

Abstract: *The metaverse connects immersive interfaces, virtual environments, artificial intelligence, extended reality, cloud and edge computing, digital twins, communication networks, and blockchain-based services. These technologies can support education, entertainment, commerce, training, and digital ownership, but they also create a broad, evolving cybersecurity surface. Identity theft, privacy loss, abnormal network activity, deepfake-based impersonation, and digital-asset fraud can occur across interconnected layers. This paper presents an independently written, student-oriented study of AI-assisted cybersecurity for the metaverse. It focuses on intelligent authentication, AI-based intrusion detection, privacy-aware learning, and protection of blockchain and NFT ecosystems. The paper describes a layered framework in which relevant security data are collected, processed with suitable AI methods, converted into risk decisions, and followed by proportionate responses. The study also discusses data quality, latency, interoperability, privacy, model attacks, and the need for human oversight. The central view is that AI should complement established cybersecurity controls rather than replace them.*

Keywords: Metaverse, Artificial Intelligence, Cybersecurity, Machine Learning, Continuous Authentication, Intrusion Detection, Blockchain, NFT Security, Privacy, Federated Learning.

I. INTRODUCTION

The metaverse is an interconnected digital environment, not a single application. It can combine avatars, virtual worlds, immersive devices, sensors, digital twins, cloud services, edge computing, AI, and blockchain. Because these components exchange information continuously, a security weakness in one component can influence another. Research on metaverse security therefore considers identity, privacy, infrastructure, access control, and digital assets as connected concerns [1], [2].

AI is useful in this setting because it can learn patterns from large amounts of security information. Instead of depending only on fixed rules, an AI-supported system can examine changes in user behaviour, network events, or transactions and identify activity that deserves further investigation. However, AI introduces its own risks, including poor-quality training data, false alarms, privacy concerns, adversarial manipulation, and model poisoning. The practical goal is therefore to combine AI with conventional controls, privacy protection, measurable evaluation, and human supervision.

This paper is newly composed for a student research context using the uploaded survey as the technical basis. The wording, explanations, organisation, and examples are independently written; standard terms such as AI, ML, IDS, blockchain, NFT, VR, AR, and XR remain because they are established technical terminology.



II. MAJOR CYBERSECURITY THREATS

The metaverse does not create every cyberattack from scratch. Instead, it combines familiar attacks with new risks created by immersion, continuous interaction, digital identities and decentralised assets. The source groups important concerns around data, identity, privacy, digital wellbeing, regulation, blockchain and NFTs.

A. Identity and impersonation: Users often enter virtual spaces through avatars. If credentials, biometric templates or behavioural information are stolen, an attacker may imitate the user. AI-generated deepfakes can make an artificial avatar appear convincing, increasing the difficulty of identity verification.

B. Data and privacy exposure: XR devices and connected sensors can collect facial features, movement patterns, eye activity and other behavioural information. Such information can be more sensitive than ordinary account data because it can reveal personal characteristics and habits. AI systems also require data for training, creating an additional privacy-management problem.

C. Network and availability attacks: The metaverse depends on continuous communication between devices, edge nodes, cloud systems and virtual worlds. DoS/DDoS attacks, malware, man-in-the-middle attacks and advanced persistent threats can interrupt services or manipulate information. The dynamic environment makes fixed rules less effective against unfamiliar behaviour.

D. Digital-twin integrity attacks: A digital twin depends on data coming from physical sensors. If that input is altered, the virtual representation may no longer match reality. The source discusses data tampering, man-in-the-middle and overlay attacks as examples of threats to digital integrity.

E. NFT and blockchain fraud: Virtual economies may use cryptocurrencies and NFTs to represent ownership. Threats include phishing, fraudulent marketplace activity, wallet compromise, smart-contract weaknesses and NFT plagiarism. AI can be used to identify suspicious transactions and verify digital content.

III. LITERATURE SURVEY

1. Pooyandeh, M., Han, K. J., & Sohn, I. (2022) – “Cybersecurity in the AI-Based Metaverse: A Survey” This study provides a comprehensive overview of cybersecurity issues in the AI-based Metaverse. It focuses on user identification, biometric authentication, DDoS attacks, privacy, and data security. The study highlights the potential of artificial intelligence and machine-learning techniques for detecting and preventing cyber threats in the Metaverse.

2. Wang, Y., et al. (2023) – “A Survey on Metaverse: Fundamentals, Security, and Privacy”

This survey examines the architecture, security, and privacy challenges of the Metaverse. Major issues such as privacy leakage, security breaches, scalability, authentication, and interoperability are discussed. The authors emphasise the need for advanced security mechanisms to build a secure and trustworthy Metaverse.

3. Li et al. (2024) – “Integrated Cybersecurity for Metaverse Systems Operating with Artificial Intelligence, Blockchains, and Cloud Computing”

This research investigates the integration of AI, blockchain, and cloud computing for Metaverse cybersecurity. AI can support threat detection, while blockchain can improve data integrity and security. However, the study identifies challenges related to cloud security, complex infrastructure, and regulatory requirements.

4. Awadallah, A., Eledlebi, K., Zemerly, M. J., et al. (2025) – “Artificial Intelligence-Based Cybersecurity for the Metaverse: Research Challenges and Opportunities”

This study is directly related to the proposed topic. It discusses AI-based approaches for user authentication, intrusion detection systems, blockchain security, and NFT security in the Metaverse. The authors also identify challenges related to AI algorithms, datasets, performance evaluation, and future research opportunities.

5. “Mapping the Literature on Factors Influencing Cybersecurity in the Metaverse” (2026)

This bibliometric study analyses existing research on Metaverse cybersecurity and identifies major research areas such as machine learning, IoT, secure data processing, access control, blockchain, authentication, AI, AR, and quantum technologies. The study suggests that more research is required on human factors, governance, privacy, and emerging security technologies.



Research Gap

The reviewed studies show that AI has significant potential for intrusion detection, authentication, anomaly detection, privacy protection, and digital-asset security in the Metaverse. However, challenges remain in real-time threat detection, privacy-preserving AI, adversarial attacks, scalability, interoperability, and governance. These gaps provide opportunities for developing more robust AI-based cybersecurity solutions for the Metaverse.

IV. PROPOSED AI-ASSISTED SECURITY FRAMEWORK

- The proposed framework contains four cooperating layers. The first is identity protection, where credentials can be supplemented with carefully selected behavioural or biometric evidence. The aim is not to collect every possible signal, but to use the minimum information required for a security decision.
- The second layer analyses network and user behaviour. Relevant events are converted into features and processed using classification or anomaly-detection models. The output is treated as a risk signal rather than an automatic declaration of malicious intent.
- The third layer focuses on digital assets. Transaction history, marketplace behaviour, NFT metadata, wallet events, and smart-contract information can be examined for unusual patterns. High-risk activity can be challenged with additional verification or sent for human review.
- The fourth layer manages privacy and response. It limits collection and retention of sensitive information and can use privacy-preserving approaches such as federated learning when appropriate. High-impact automated actions should remain subject to human oversight.

V. SYSTEM WORKFLOW

A simplified workflow is: Data Collection → Preprocessing → Feature Extraction → AI Analysis → Risk Assessment → Security Response → Continuous Monitoring.

- During collection, only information relevant to the security objective should be obtained.
- Preprocessing handles noise and incomplete records.
- Feature extraction converts raw events into useful representations. AI models then classify known patterns or detect unusual activity.
- The risk engine combines model output with security rules and context. Responses may include additional verification, transaction review, device isolation, alert generation, or continued access when risk is low. Continuous monitoring checks whether the response solved the issue without unnecessary interruption.



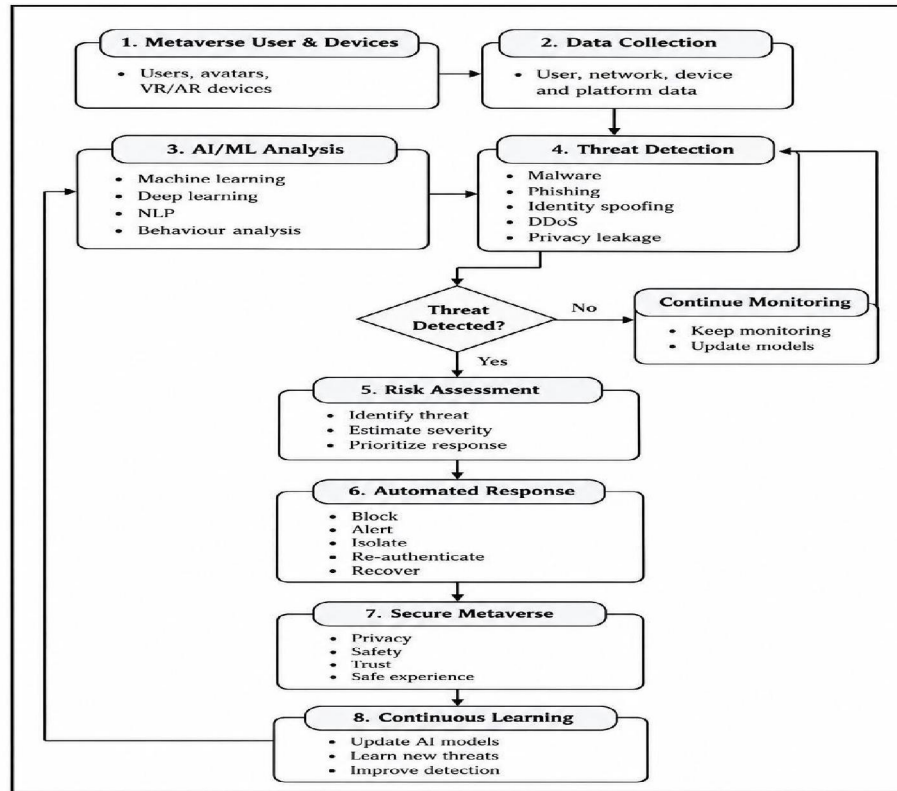


Fig. 1: AI-Based Cybersecurity System Workflow for the Metaverse

VI. CHALLENGES AND LIMITATIONS

- AI models need representative datasets, but metaverse platforms differ in hardware, users, interaction styles, and policies. A model that performs well in one environment may not transfer directly to another. Latency is also important because XR applications depend on fast responses; large AI models can increase computation and energy use.
- Privacy is another limitation. More data may improve a model while simultaneously increasing privacy exposure. Federated learning can reduce raw-data sharing, but it does not automatically remove privacy or security threats. AI models can also be attacked through manipulated inputs or poisoned training data.

Finally, interoperability remains difficult because different virtual worlds may use different identity systems, data formats, and security policies.

VII. FUTURE RESEARCH OPPORTUNITIES

- Future research can develop lightweight AI models for XR and edge devices and evaluate them using accuracy, latency, memory, energy, and privacy measures.
- Privacy-preserving continuous authentication is another promising direction, especially when multimodal signals can be processed locally.
- Researchers also need metaverse-specific cybersecurity datasets containing realistic authentication, network, behavioural, and digital-asset events while protecting participants' privacy. Standard benchmarks would make



comparisons more meaningful. Explainable AI is important because administrators should understand why a system raised a high-risk alert.

- Further work can also study fairness, robustness, adversarial resistance, and human-in-the-loop decision-making.
- A broader research direction is an integrated risk model that connects identity security, intrusion detection, and digital-asset protection instead of treating them as independent modules.

VIII. AI –BASED INTRUSION DETECTION

- A metaverse intrusion detection system may need to examine network traffic together with device events, application logs, account activity, and virtual-world changes.
- A hybrid design can combine fixed rules for known high-confidence events with machine learning for changing patterns.
- Useful evaluation measures include precision, recall, false-positive rate, false-negative rate, detection latency, and AUC where appropriate. Authentication tasks may additionally use false acceptance and false rejection measures.
- Concept drift is another concern because normal behaviour can change when users move between platforms, devices, or activities. Adaptive learning and periodic validation can help maintain model reliability.

IX. EXPECTED BENEFITS

- An AI-assisted layered design can potentially reduce dependence on one-time authentication, identify unusual activity, prioritise alerts, and provide additional evidence for digital-asset security. It can also help security teams process large event streams.
- These benefits depend on training data, model reliability, thresholds, privacy controls, and integration with conventional cybersecurity. The intended principle is balance: security should reduce meaningful risks without becoming excessively intrusive, computationally expensive, or difficult to use.

X. CONCLUSION

The metaverse combines immersive devices, digital identities, networks, AI services, cloud and edge infrastructure, and blockchain assets. This creates a broad attack surface in which identity theft, privacy loss, network attacks, impersonation, and digital-asset fraud can occur across connected layers. AI can assist by learning patterns, detecting anomalies, supporting continuous authentication, and analysing suspicious transactions.

AI is not a complete cybersecurity solution. Data quality, privacy, computational requirements, interoperability, false alarms, model attacks, and human oversight remain important. A practical approach should therefore combine AI with established security controls and privacy-preserving mechanisms. Lightweight models, continuous authentication, secure federated learning, stronger NFT protection, common benchmarks, and explainable security decisions provide useful directions for future research.

REFERENCES

- [1] A. Awadallah et al., “Artificial Intelligence-Based Cybersecurity for the Metaverse: Research Challenges and Opportunities,” *IEEE Communications Surveys & Tutorials*, vol. 27, no. 2, 2025, pp. 1008–1045, doi: 10.1109/COMST.2024.3442475.
- [2] “SoK: On the Security of Non-Fungible Tokens,” *Blockchain: Research and Applications*, vol. 6, no. 2, 2025, 100268, doi: 10.1016/j.bcr.2024.100268.
- [3] “Machine Learning and Deep Learning Architectures for Intrusion Detection System (IDS): A Survey,” *Proc. ICTEST*, 2024, doi: 10.1109/ICTEST60614.2024.10576052.



- [4] "Using Motion Forecasting for Behavior-Based Virtual Reality (VR) Authentication," Proc. IEEE AIxVR, 2024, doi: 10.1109/AIxVR59861.2024.00012.
- [5] P. Mell and D. Yaga, "Non-Fungible Token Security," NIST Interagency/Internal Report 8472, National Institute of Standards and Technology, 2024, doi: 10.6028/NIST.IR.8472.
- [6] Y. Wang et al., "A Survey on Metaverse: Fundamentals, Security, and Privacy," IEEE Communications Surveys & Tutorials, vol. 25, no. 1, 2023, pp. 319–352, doi: 10.1109/COMST.2022.3202047.
- [7] Y. Liu et al., "A Survey on Security and Privacy Threats to Federated Learning," Proc. NaNA, 2021, doi: 10.1109/NaNA53684.2021.00062.
- [8] S. Manimaran, V. N. Sastry, and N. P. Gopalan, "VRCAuth: Continuous Authentication of Users in Virtual Reality Environment Using Head-Movement," Proc. ICCES, 2020, pp. 518–523, doi: 10.1109/ICCES48766.2020.9137914.
- [9] B. McMahan et al., "Communication-Efficient Learning of Deep Networks from Decentralised Data," Proc. AISTATS, 2017, pp. 1273–1282.

