MALAYSIAN JOURNAL OF
CYBERSECURITY AND APPLICATIONS

CYBERSECURITY GOVERNANCE GAPS IN EXTENDED REALITY (XR) SYSTEMS: A COMPARATIVE RISK MANAGEMENT ANALYSIS

Muhammad Syafiq[1], Ahmad Firdaus[1,2,*], Nurshahira Mohd[3], Noormadinah Allias[4], Mohd Zamri Osman[5] and Mohd Faizal Ab Razak[1]

[1]Faculty of Computing,
Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), Pekan, Pahang, Malaysia

[2]Centre For Artificial Intelligence & Data Science (CAIDAS),
Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), Lebuh Persiaran Tun Khalil Yaakob,
Gambang, Kuantan, Pahang, Malaysia

[3]Cybersecurity Malaysia (CSM),
Menara Cyber Axis, Jalan Impact, Cyberjaya, Selangor, Malaysia

[4]Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA (UiTM), 40450, Shah Alam Selangor

[5]Faculty of Computing,
Universiti Teknologi Malaysia (UTM), Johor Bahru, Johor, Malaysia

Emails: pez25025@adab.umpsa.edu.my[1], firdausza@umpsa.edu.my[2], nurshahira.mohd@cybersecurity.my[3], noormadinah@uitm.edu.my[4], mohdzamri.osman@utm.my[5], faizalrazak@umpsa.edu.my[6]

ABSTRACT

Extended Reality (XR), which covers Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), is being widely adopted across many application areas such as healthcare, education, and enterprise systems because it enables highly immersive and engaging user experiences for users today environments. However, XR systems operate in sensor-filled environments that continuously process sensitive biometric, behavioural, and spatial data from users. This continuous data collection introduces new cybersecurity and privacy risks that go beyond those found in traditional IT infrastructures systems overall. This study examines the mismatch between emerging cybersecurity threats in Extended Reality (XR) systems and established governance frameworks such as ISO/IEC 27001 and the NIST Cybersecurity Framework. Using an analytical comparison analysis, the study evaluates framework coverage against an XR threat taxonomy derived from existing literature. The findings identify three major residual risk gaps involving biometric data protection, avatar identity integrity, and adaptive incident response. Based on these findings, a Mitigation Readiness Toolkit is proposed to support XR-specific risk governance. The study contributes toward developing future XR-oriented cybersecurity frameworks. To conclude, XR tailored risk assessment approach is urgently needed which may inform the design of more advanced threat modelling methodologies and Mitigation Readiness Toolkit for safe and trust-worthy adoption of all immersive technologies.

Keywords: Extended Reality (XR); Cybersecurity; Risk Management; ISO/IEC 2700; NIST CSF; Biometric Data; Privacy; Threat Modelling; Virtual Reality (VR); Augmented Reality (AR).

1. INTRODUCTION

The rapid growth and ubiquitous use of Extended Reality (XR), encompassing Virtual Reality (VR), Augmented Reality (AR) and Mixed Reality (MR), is already revolutionizing human-computer interaction by generating new virtual environments called Metaverse [1], [2], [3]. This technology is revolutionizing diverse sectors, such as immersive learning and medical training to remote work and e-commerce, offering unprecedented levels of immersion and utility [1], [3]. But this paradigm shift in technology is based on the constant acquisition and processing of large quantities of real-time highly sensitive information such as eye tracking, face expressions, spatial mapping and biometric factors [4]. Since XR relies heavily on sensed environments, a new and

complicated attack surface appeared with deep fears regarding user privacy, data fidelity and system security [1], [5].The integration of physical world with digital world in hyper spatiotemporal, introduce massive privacy leaks and new type of security threats that come from the subatomic level technologies or introduces fundamentally new cyber-physical interaction environments [2].

The focus of this study is to address the gaps between the emerging threat paradigm for XR technologies and conventional cybersecurity risk management principles. Unlike legacy Information Technology (IT) systems, XR platforms capture sensitive information including but not limited to cognitive load and emotions which could potentially be employed for intrusive profiling purposes [6], [7]. Therefore, the traditional security mechanisms built for static data and perimeter-based defence have failed to protect against dynamic threats that are sensor rich such as the "Human Joystick Attack" or other threats that target BCI integration [5], [8]. Consequently, current security approaches must be re-evaluated. This is due to existing models simply fail to account for constant sensor data, the real-time blending of physical and virtual identities, and cyber-physical threats.

While existing literature has extensively documented XR threat taxonomies, including sensory manipulation and biometric exposure [2], [9], and has mapped these threats using structured frameworks such as STRIDE alongside the identification of requirements for AI-based defensive mechanisms [3], [10]. A significant research gap remains as current studies largely focus on identifying vulnerabilities rather than evaluating the governance structures meant to mitigate them. Specifically, traditional risk management standards such as ISO/IEC 27001 and NIST CSF rely on static, perimeter-based defences and credential-centric access controls. These legacy control objectives are fundamentally misaligned with the continuous, sensor-rich, and cyber-physical nature of XR [11], [12], [13]. Previous research has not explicitly mapped this discrepancy, leaving a critical void in understanding exactly how and where established frameworks fail when confronted with an XR-specific attack surface.

The significance of addressing this gap cannot be overstated as the inadequacy of existing frameworks results in unacceptable levels of residual risk which would compromise the safe and dependable uptake of XR technologies across critical sectors. The continuous and dynamic nature of sensor data streams combined with the possibility for cyber-induced physical harm, demands a security approach that is just as adaptable and multi-faceted as the technology itself [2], [14]. Hence, the goal of this study is to supplement an existing IT security policy with specific, evidence-based controls tailored to XR. This work is a step towards helping companies and regulator define the governance needed to responsibly deploy AR, VR and other immersive technologies.

Hence, this paper contributes to the literature by a thorough critical analysis of practical deficiencies in existing risk management approaches when confronted with XR threats. More precisely, the contributions are a comparative analysis of ISO/IEC 27001 and NIST CSF against a detailed XR threat taxonomy and the identification and formulation of three key residual risk gaps that persist under current practices which follow by the proposal of a Mitigation Readiness Toolkit, a set of evidence-based enhancements designed to address these gaps. The remainder of this paper is structured as follows: Section 2 details the research methodology. Section 3 presents the data analysis and results. Section 4 proposes the enhancements and Section 5 concludes the study with future work directions. The primary objective of this research consists of:

I. To evaluate how well existing cybersecurity frameworks such as ISO/IEC 27001 and NIST CSF address the unique risks posed by Extended Reality (XR) technologies.
II. To identify and analyse key gaps in current risk management approaches, particularly in handling real-time biometric data, avatar integrity, and dynamic incident response in XR environments.
III. To propose practical evidence-based improvements including a Mitigation Readiness Toolkit to strengthen XR-specific cybersecurity risk management.

2. METHODOLOGY

2.1 Manuscript Organization and Length

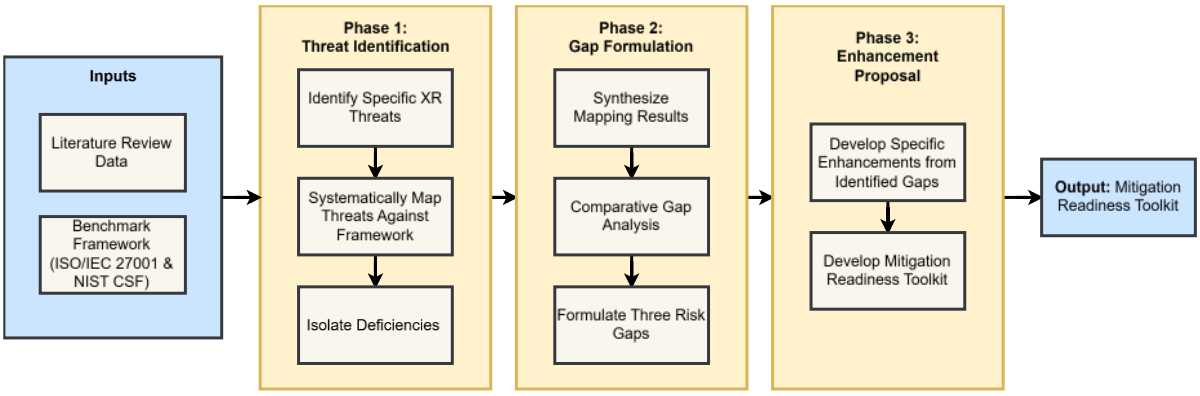

Fig. 1: The Methodological Framework Employed in This Study

Figure 1 illustrates the three-phase methodological framework adopted in this study. Phase 1 (Threat Identification) focuses on extracting specific XR threats from the literature and systematically mapping them against the benchmark frameworks. Phase 2 (Comparative Gap Analysis) synthesizes these mapping results to articulate three critical residual risk gaps. Finally, Phase 3 (Enhancement Proposal) translates these identified gaps into the Mitigation Readiness Toolkit. A comprehensive breakdown of the analytical procedures within each phase is detailed in Section 2.3.

A critical comparative approach serves as the foundation of the research methodology, employing an analytical comparison research design. The core purpose revolves around conducting a critical examination of the fitness-for-purpose of two cybersecurity risk management systems which are ISO/IEC 27001 standard and NIST Cybersecurity Framework (NIST CSF) in addressing the challenging and complex cybersecurity threats presented by Extended Reality (XR). This approach has been adopted because there are no dedicated XR security standards which have been adopted everywhere. In particular, the approach comprises a step-by-step comparison of security capabilities in such traditional frameworks with those present in a pre-identified taxonomy of threats and vulnerabilities specific to XR. The objective is to expose the limitations of traditional risk management and measurement in a sensor-rich and hyper-spatiotemporal realm such as XR. This process yields rigorous evidence base derived from systematic and comparative reflective analysis of established governance models, moving beyond anecdotal evidence.

2.2 Sections and Subsections

The study rests on two independent but complementary data sources as the basis for a strong and fact-based analysis:

1 Literature Review Data: To form the foundation of the XR threat taxonomy and gap analysis, a targeted literature search was conducted using three primary academic databases: Scopus, IEEE Xplore, and Web of Science. The search strategy employed the following query string to capture the intersection of immersive technology and security: ("XR" OR "extended reality" OR "virtual reality" OR "augmented reality" OR "mixed reality") AND ("cybersecurity" OR "information security" OR "data protection" OR "network security" OR "cyber defence"). To ensure the relevance and rigorous quality of the literature, inclusion criteria were strictly limited to peer-reviewed academic literature and authoritative technical standards papers published in the English language between the years 2020 and 2025. This targeted approach yielded a consisting of 26 academic and industry sources on XR cybersecurity, privacy, and the Metaverse [1]–[26]. The corpus provides the basis for identifying the current state of knowledge, constructing an XR threat taxonomy, and verifying that the research gaps identified in our Introduction indeed exist.

2 Primary Analysis Data: An illustrative, structured comparison of XR threats to existing frameworks comparisons with traditional frameworks and optional targeted improvement suggestions by the authors of

this paper. This initial study involved systematic decomposition of XR security concerns into a logically ordered list of threats and vulnerabilities.

The study employed three core conceptual toolkits the ISO/IEC 27001 and NIST CSF as Reference points, while a designated XR Threat Taxonomy for classifying threats such as Sensor Data Exploitation, Identity Spoofing was used as an analytical toolkit for performing an attack surface mapping analysis six identified Key Risk Areas for example Incident Response restrictions, weak data protection were used to measure both frameworks performance. The data used for the main analysis were obtained from a systematic review of XR system architectures, sensor data flows and reported security incidents to make sure that our findings are rooted in actual security concerns.

2.3 Analytical Procedure

The analytic process was comprised of three sequential phases to allow for a logical flow from identifying the threat through making evidence supported judgments:

Phase 1: Threat Identification and Comparative Mapping which this phase was organized in two initial analytical steps. Firstly, which are threat identification that extracted and classified the sensor-intensive threats of XR platforms by means the literature review and system analysis and secondly systematically mapping these identified XR threats towards ISO/IEC 27001 control objectives alongside NIST CSF. This was an important step to identify areas of overlap but, more importantly, establishing where the gaps are.

Phase 2: Gap Formulation and Thematic Analysis are the results of the comparative mapping that were integrated to articulate three main gaps causing unaddressed or residual risk in XR settings. The analysis subsequently used analytical comparison thematic approach to searching for key themes of inadequacy and residual risk within the six Key Risk Areas. This content analysis formed the basis of the Evaluation of Current Approaches and subsequent Evidence-Based Judgment.

Phase 3: Enhancement Proposal, which is the last phase, included the compilation of an evidence-based set of enhancements and a Mitigation Readiness Toolkit. This proposal focuses on attacking the 3 core gaps identified in Phase 2 and, ensuring that the proposed solutions are more closely associated to the research objectives while enabling the research to start building a stronger foundation for an XR-specific cybersecurity risk management model. This methodological framework and plan will enhance the strength and specificity of the findings.

3. DATA ANALYSIS AND RESULTS

The analysis conducted is an analytical comparison thematic cross-check of the XR threat landscape with control objectives defined by ISO/IEC 27001 and NIST CSF. The aim is to assess the analytical comparison gap how far traditional risk management methods are insufficient when applied on unprecedented sensor-rich Extended Reality. The results are outcomes of the systematic analysis, described in Section 4, where the study compared the mapping of the presented XR Threat Taxonomy with control objectives from both benchmark frameworks.

3.1 Comparative Analysis of Current Risk Management Approaches

Systematic comparison shows that although both ISO/IEC 27001 and NIST CSF are robust frameworks for a general information security landscape, they fall short when it comes to the XR dimension. For a systematic evaluation of the problem at hand, the comparative study applied the criteria of Severity, Likelihood, and Impact while examining the frameworks with reference to the XR threat taxonomy. Severity refers to the level of threat posed by the specific XR risk, including possible injuries and permanent biometric data leakage. Likelihood relates to the probability that the threat will succeed even when conventional perimeter defence and static control mechanisms are implemented. Impact refers to the overall effect of this framework weakness.

Six primary areas that the framework appears to fall short are highlighted by the analysis and listed in Table 1 below. The inadequacy is especially significant as it fails to address a multi-dimensional attack surface, as well as hyper-spatiotemporally of XR environments [2], [15]. More precisely in these traditional standards, the control also misses granularity to handle continuous, high-volume stream data and physical-virtual integration which are idiomatic of XR systems. Accordingly, reliance on these models leads to a significant remaining risk exposure that remains unaddressed.

Table 1: Identified Insufficiencies in Traditional Risk Management Frameworks vs. XR Threats

Insufficiency Area	Description of Gap in Traditional Frameworks	XR-Specific Threat Context
Incident Response Limitations	Emphasis is given to data breaches and system recovery, but only limited support is provided for the physical-virtual coordination.	Positional spoofing or malicious content injection can cause real-world physical harm or psychological distress [5], [8].
Access Control Gaps	Favours credentials-based IAM, few guidance about behavioural and biometric authentication.	XR relies on immutable biometric data for example eye-tracking for authentication, increasing risk of long-term identity compromise [4], [16], [17].
Data Protection Shortcomings	Concerns primarily data at rest and in transit which does not specify in detail sensor streams dynamic, continuous.	Real-time spatial mapping, voice samples, and eye-tracking logs require novel encryption and data minimization policies [1], [4].
Third-Party Vendor Dependencies	Promotes a general assessment of risk, does not adequately address the high degree of integration and complexity in some multi-tier XR value chain.	Vulnerabilities in hardware sensors, cloud APIs, or SDKs may propagate throughout the entire XR system [2], [18].
Threat Modelling Challenges	Existing models Traditional STRIDE model is not suitable for the multi-aspect attack surface.	XR spans physical space, cognitive perception, and real-time interaction, requiring new methodologies to map attack vectors [9], [15].
Regulatory and Ethical Complexities	Stresses compliance with existing laws which lacks guidance for emerging ethical issues for insurance mental privacy, manipulation.	The immersive nature of XR introduces novel ethical and regulatory challenges not covered by current data protection laws [1], [6], [7].

3.2 Results: Formulation of Evidence-Based Judgments on Residual Risk

The following three critical gaps were developed from the comprehensive review of the six insufficiencies outlined in Table 1. These gaps correlate to the most critical unaddressed risk areas in traditional frameworks when applied to XR environments and thereby undermine the fundamental security and privacy of users and the systems within XR. These three judgements then constitute the fundamental output of the comparative analysis, since they specify exactly where current risk management practice would fail.

3.2.1 Inadequate Protection of Real-Time Behavioural and Biometric Data

The first, and most critical, shortcoming concerns the safeguarding of the flowing real-time behavioural and biometric data obtained by XR devices [4]. In contrast with relatively static Personal Identifiable Information (PII), the data streams occurring in XR for examples eye tracking, gait analysis or physiological response are highly sensitive and continuous and often unalterable. The examination supports the notion that conventional data protection controls centred on encryption and access control of data-at-rest and in-flight are woefully insufficient for this dynamic type of data. The huge amount and high speed of this data stream, commonly known as “big data” in the context of XR, inundate traditional security boundaries [2]. Furthermore, the particular risk arises from re-identification and profiling [19]. Even if data are anonymized, the fine-grained information about behaviour can be used to infer cognitive states, health status and emotional responses which allow invasive surveillance or manipulation [6], [7]. This is further complicated by the fact that the data is often processed at the edge, introducing new vulnerabilities in the wireless communication networks that support XR [20]. As a result, there are no existing solutions to implement Sensor-Level Data Minimization, which necessitates a "privacy-by-design" strategy that is not specifically required by current regulation [1]. Additionally, dynamic encryption for safeguarding constant and high-volume sensor streams without intolerable latency overheads and irreversible Anonymization, as the high dimensionality of XR data makes traditional techniques ineffective [4], [10]. The

inability to protect data is a long-term issue since once biometric data is compromised, it cannot be reset and user identity would be permanently breached .

3.2.2 Insufficient Identity and Avatar Integrity Controls

The second important decision involves being open to potential threats on the user identity and integrity of avatars in immersive applications. XR platforms are based on avatars and behaviour biometrics for authentication, but traditional Identity & Access Management (IAM) controls are mainly credential based. Such imbalance leads to confident dislocation of trustworthy virtual interaction, with the growth of virtual economy and digital assets. In particular, the study shows that biometric eye-tracking authentication is less secure to identity faking and impersonation in XR [9], [21] . An adversary with such access to session or avatar can give false commands or discredit the user in a very realistic manner and in real time [5], [7]. The immersive nature of the Metaverse makes social engineering attacks particularly effective, as users are more susceptible to manipulation when their cognitive load is high [21].

The deficiency is in not accounting for continuous, passive authentication to mitigate session hijacking, and deepfake-style impersonation [3], [4]. Moreover, avatar trust checking becomes necessary to cryptographically verify mistreatment actions of an avatar and its look due to theft of virtual ownership. In this context, blockchain-based Decentralized Identifiers (DIDs) offer a robust contrast to traditional, centralized IAM [22]. By anchoring a user's digital identity to an immutable blockchain ledger, DIDs allow users to cryptographically prove ownership of their avatar without relying on a single central authority. This decentralized approach mitigates the risk of avatar spoofing and provides definitive proof in the investigation of virtual asset loss [17], [18] . In addition, non-repudiation for virtual transaction that is important to professional or financial XR applications holds true due to because actions taken by a compromised avatar may potentially disclaimed as made by its owner [2], [7]. Ultimately, these vulnerabilities highlight the profound shortcomings of traditional Identity and Access Management (IAM) in securing the fluid, multi-persona identities of the Metaverse, where a user's digital self is simultaneously a high-value asset and a significant vulnerability.

3.2.3 Limited Adaptive Incident Response Capabilities

The third evaluation illustrates the severe inadequacy of conventional ordinary Incident Response (IR) to respond to hybrid physical–virtual aspects that nonetheless underpin XR related incidents [5], [8]. Traditional IR approaches are centred on the digital cordon and system restore but, given hazards for immediate physical safety of the user in XR environments for example the “Human Joystick Attack” can cause a user to collide with physical objects [8]. What seems to be missing is an integrated approach for a coordinated response on both physical safety measures and digital security measures. The communication that would be established between IT security staff and physical security or emergency services is a process which has not been addressed by traditional IR plans and is becoming more prevalent in critical infrastructure situations where XR technologies are employed [23].

Further, this gap is reflected by absence of physical safety triggers in IR they are not capable to sensing and responding to physical safety indicators for example a rapid loss of balance which signals a cyber-induced physical threat [24], [25]. In addition, lack of sensor integrity verification in IRL critical for parsing through spoofed sensor data and legitimate sensor information [9] Also to be conceived the non-adaptive containment, where the compromised immersive environment is not autonomously and safely revert back to a benign state preventing imminent harm, causing an imperfect isolation [10], [17]. Therefore, these three judgments in aggregate indicate that applying the current risk management methodologies to XR places organisations at intolerable levels of residual risk unless some specific improvements are made. Ignoring these risks poses the risk of severe consequences, such as user injury or industrial systems crashing completely due to catastrophic failure to the XR environment.

4. PROPOSED TAXONOMY FOR XR THREAT MITIGATION

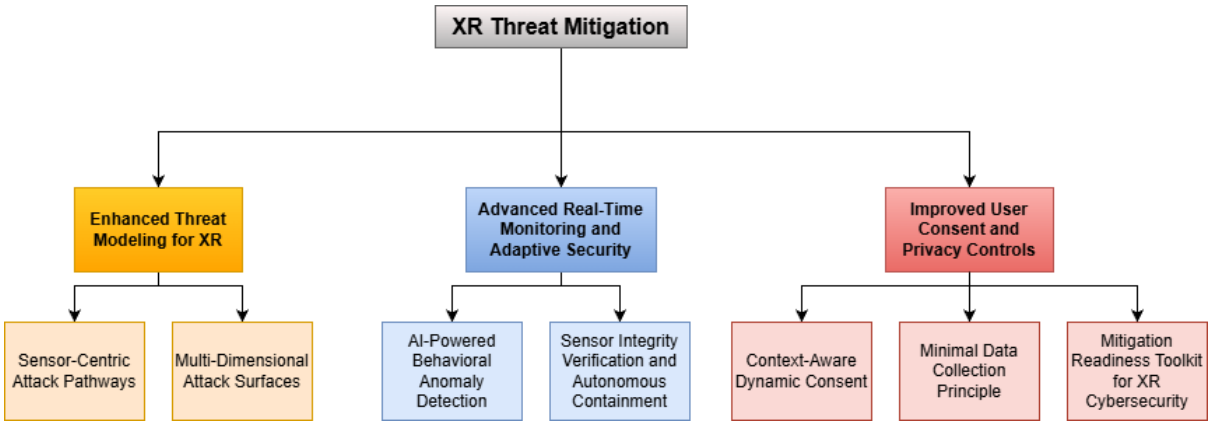

Fig. 2: Proposed Taxonomy for XR Threat Mitigation.

Figure 2 above describes the three main principles of the potential enhancing approach which are Enhanced Threat Modelling, Advanced Real-Time Monitoring, and Improved User Consent, along with specific sub-components. The systemic search for risk gaps in residual risk requires focused improvements to existing risk management. Unlike prior work that has been predominantly about logging XR threats, this section now describes series of practical steps together with the creation Mitigation Readiness Toolkit that aims to plug directly into the three key gaps identified in Section 3. These concepts stem from the idea that XR security needs to move beyond traditional defence of a perimeter, to a sensor-driven and adaptive multi-dimensional defence posture. It is important to note that this proposed taxonomy and the MRT function as a conceptual framework. Rather than being independently validated through empirical testing, these structures are synthesized directly from the vulnerabilities, threat models, and risk gaps identified in the curated literature reviews. They are designed to serve as a theoretical bridge between established cybersecurity principles and XR-specific threats.

Moreover, although most current security frameworks for XR only consider one aspect or a single technique, such as using certain cryptography techniques for securing blockchain assets [17] or deploying AI-based intrusion detection systems [3], the MRT model offers an integrated solution that is in harmony with risk governance principles. Contrary to the limited perspective of most technical frameworks, the MRT model is explicitly designed to fill in the blanks in the implementation of overall risk management frameworks such as ISO/IEC 27001.

4.1 Enhanced Threat and Modelling for XR

Traditional threat modelling approaches, such as Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege (STRIDE) are inherently limited in their software centred and boundary-oriented view, which is insufficient to address the dynamic, integrated nature of XR environments [9], [15]. Beyond technical vulnerabilities, the integration of Metaverse characteristics raises significant concerns regarding societal impact and human ethics [26]. Thus, a paradigm shift is necessary to take such a multi-dimensional sensor-centric perspective to adequately document the XR attack surface.

4.1.1 Sensor-Centric Attack Pathways

Threat modelling must start at the sensor level, for instance studying the data lifecycle from capture to processing, because this is, by far, the weakest link in extremely sensitive biometric information [4]. Three fundamental analytical elements are presented in this new approach. Input Integrity Analysis is indeed a preliminary step aimed at constantly enforcing the trustfulness of information coming from every sensor for instance camera, microphone eye-tracker before it will be exploited by the application layer. Second, there is a need to use Fusion Vulnerability Mapping to discover new attack vectors resulting from the fusion of data from different sources such as both coarse-grained positional information with biometric signatures leading to an aggregate determination of a user's identity and location simultaneously . Finally**,** Output Manipulation Analysis is essential for modelling attacks that manipulate the sensory output such as visual, auditory, haptic to cause physical or

psychological harm, such as the "Auditory Stimulus Attack" [5], [8]. This intense focus is a way of ensuring vulnerabilities are addressed all the way to the root cause rather than just at an application layer.

4.1.2 Multi-Dimensional Attack Surfaces

The modelling should take into consideration the three dimensions under which XR environment is defined, and that together define a unique attack surface. The first dimension is Physical Space, which has to do with threats that exploit the real world aspect as a context or sensory input alike mapping where a user lives or works, or even the infamous "Human Joystick Attack" [8].

The second type is Cognitive Perception Attacks, which the kind of threats from human psychological insecurities such as immersive phishing, trust manipulation and deep mental privacy attack [6], [7]. The third dimension of infrastructure risk is network and application layer risks, encompassing virtual asset and blockchain-related threats that should be considered in the total risk profile [3], [17]. By taking into account these three dimensions, the extended threat model provides a comprehensive image of risk which cannot be obtained by single models.

4.2 Advanced Real-Time Monitoring and Adaptive Security

The extensive, immersive and sensor-rich nature of XR environments requires a move from traditional periodic security assessment models to continuous collection of data, real-time analytics and response. XR systems are based on the continuous flow of spatial, biometric, and behavioural data where user context, condition environment, and interaction state change over time in a session.

Security threats such identity forgery, session stealing and malicious avatar behaviour, privacy revealing are all once-off or short-lived and may escape the static scheduling security checks. Ongoing identity assurance is not static. The key mission of continuous monitoring and analysis requires continually proving identity in operational terms, including behavioural re-authentication and contextual validation beyond the point of initial access control. In addition, the ability to make real-time analysis enables early anomaly detection, as well as adaptive incident response actions such as dynamic access revocation, session isolation, or sensor-level control. In the absence of these mechanisms, XR platforms are exposed to delayed detection and ineffective mitigation, which further increases the identified identity management and incident response gaps in this paper.

4.2.1 AI-Powered Behavioural Anomaly Detection

Security systems should be designed to integrate AI on learning user normal behaviour in the XR environment [3], [18]. Contrarily to traditional signature or network level analysis-based intrusion detection systems, such an approach allows for a continuous and passive verification of the user's identity based on distinguishing behavioural characteristics over time, such as movement speed and interaction rate. Any variation between this ground-truth value should trigger an alert automatically, and be a starting point for background passive identity verification and early detection of subtle presence of session hijacking or deepfake attack because of live impersonation [17]. This is to avoid long-term identity exposure due to non-revocable biometric information.

4.2.2 Sensor Integrity Verification and Autonomous Containment

To address the problem of adaptive incident response, two crucial elements should be designed. First, a security layer should be applied to Sensor Integrity Verification, which can constantly check whether the sensor data stream has not been tampered and are authenticated [9]. This encompasses protection mechanisms on the lower level hardware to avoid tampering of firmware and cryptographic verification of sensor inputs, therefore guaranteeing trustworthy data [2]. These are so-called prescribed immediate-response procedures capable of automatically and safely decontaminating an infected immersive environment and restoring it to a safe state. These strategies include using the “Safe Room” strategy, limiting sensory stimuli to reduce confusion [5], and taking the device off-line without disconnecting from the network to preserve the log file for analysis.

4.3 Improved User Consent and Privacy Controls

In order to minimise increasing risks that real time behavioural data is not adequately protected, the time for privacy controls to move from static forested legal agreements into dynamic foliage of such context-based systems would be now, as operating environment should evolve continuously even if it were conceived only at a singular point in time.

4.3.1 Context-Aware Dynamic Consent

Standard permission requests are often too broad to be effective in complex digital environments. Instead, users should interact with a system of context-aware dynamic consent. Users should receive permission requests which are as granular, timely and data specific [7], [19]. For example, a user should be prompted for their agreement to share eye tracking data only during the directly necessary time that the application is relying on it for a particular function, rather than via a general opt-in. This, in conjunction with Transparent Data Visualization where XR platforms support real-time and "transparent" visualization of collected data, along with by whom it is being collected and why provides the user greater influence and control [1]. By making this information easily visible in users, they can make informed decisions about their privacy. This shifts the user's consent from a kind of one-time decision to a process that is ongoing and active, so at any time in the future users regain control over their personal data.

4.3.2 Minimal Data Collection Principle

Privacy should not simply be a matter of personal choice for the user but built into technology by design. The principle of minimal data collection suggests that systems should be designed to gather only the absolute minimum amount of information required for a feature to work [4], [19]. In other words, software developers should be technically prohibited from getting at any extra data not needed to deliver the core experience. For instance, software developers should technically be deterred from getting at any extra data not necessary to delivery core experience. For example, if an app only needs to know where a user is pointing, it shouldn't have access to an entire video feed of the user's room.

This level of defence is optimal when it is right in the hardware. By controlling the flow of data with these hardware-level tools, the system whittles down what security experts refer to as the "attack surface," or ways hackers might steal sensitive information. This is of value too for businesses with a risk adverse profile, and culminate in the non-necessity to store large amounts of sensitive biometric data that they simply do not need. Through the creation of these borders by technical means, it will be possible to create an environment where convenience and privacy are no longer at odds with one another.

4.4 Mitigation Readiness Toolkit for XR Cybersecurity

These identified enhancements have culminated into the Mitigation Readiness Toolkit (MRT), which is a set of actionable security controls developed to specifically mitigate each of the three highest residual risks identified in Section 3. The toolkit offers a structured and evidence-based strategy for developing risk management practices tailored to the specific needs of XR environments. This toolkit aims to support enhanced mitigation of risks associated specifically with issues of identity protection, privacy preservation, and incident response.

Table 2: Mitigation Readiness Toolkit Components

Residual Risk Gap	Toolkit Component	Description
Inadequate Data Protection	**Sensor Data Protection and Lifecycle Security**	End-to-end encryption for dynamic sensor streams, hardware-enforced data minimization irreversible anonymization techniques [4], [19].
Insufficient Identity Integrity	**Continuous Identity and Avatar Integrity Verification**	AI-powered passive biometric authentication cryptographic signing of avatar actions session integrity monitoring [3], [16].
Limited Adaptive IR	**Real-Time Incident Response Capabilities**	Autonomous containment protocols sensor integrity validation tools integrated physical-virtual safety coordination [5], [8], [10].
Cross-Cutting	**Enhanced Threat Modelling**	Sensor-centric and multi-dimensional analysis to proactively identify novel attack vectors [9], [15].
Cross-Cutting	**Dynamic Consent Management**	Context-aware, granular consent mechanisms and transparent data visualization for users [1], [7].

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

This paper examines current cybersecurity risk management strategies such as ISO/IEC 27001 and NIST CSF, taking into account the extent to which they can be applied to XR cybersecurity framework. The findings indicate that the existing approaches, which are based on traditional perimeter protection models, are not sufficiently equipped to address with the constantly changing landscape of XR technologies. Furthermore, a key contribution of this study is the identification of three major residual risk areas. These involve inadequate protection of real-time behavioural and biometric data, limited safeguards for maintaining avatar integrity, and the absence of effective dynamic incident response mechanisms for cyber-physical attack scenarios. In addition, to address these limitations, this study introduces the Mitigation Readiness Toolkit (MRT) as a conceptual framework designed to bridge the identified gaps. By combining sensor-aware threat modelling, AI-driven anomaly detection, and consent-based security mechanisms, the toolkit offers a structured and adaptive approach for strengthening governance and enhancing security resilience in XR environments.

5.2 Limitation of the Study

This study has several limitations. First, the findings are based on analytical comparison supported by a systematic literature review and a structured examination of XR system architectures. Although rigorous and evidence-based, this work did not collect primary empirical data nor perform quantitative cost-benefit analysis of the controls being proposed. Additionally, the focus was specifically limited to gaps in ISO/IEC 27001 and NIST CSF and not a broad evaluation of all emerging XR specific security standards. These limitations illustrate the need for future research to address empirical validation and applied use of the toolkit components.

5.3 Future Work

To expand and develop the XR cybersecurity the study identified new research paths which are promising for the future, building up from our results. A primary goal is the experimental test of the theoretical model proposed by us in this work. In the future, there is a need for quantitatively validating Mitigation Readiness Toolkit using controls for examples AI-based behavioural anomaly detection implemented inside real-world XR test beds. This would enable the rigorous comparison of their effective reductions in residual risks. Meanwhile, there is an urgent need to implement a standardised XR threat modelling methodology. An architecture of this model would have to integrate the XR multimodal attack surface extending from cybersecurity domain across the physical and cognitive realm.

In addition, beyond initial technical validation there is a wider space for future work to explore issues relating economic, governance and organisation aspects of XR ecosystem. This can be achieved by combining a comparative framework analysis with more recent industry-suggested standards or guidelines for XR security, such as those published by the Metaverse Standards Forum. Furthermore, as the XR hardware platforms continue to advance towards increasingly complicated sensor-rich architectures, this study advocate future enhancement on mechanisms for reasoning and managing supply chain security risk and third-party vendor dependences with special emphasis in XR sensor-components, as well as firmware.

ACKNOWLEDGEMENT

This study was made possible thanks to a grant from Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) with grant number RDU240323.

REFERENCES

[1] P. Kourtesis, ‘A Comprehensive Review of Multimodal XR Applications, Risks, and Ethical Challenges in the Metaverse’, *Multimodal Technol. Interact.*, vol. 8, no. 11, 2024, doi: 10.3390/mti8110098.

[2] Y. Wang *et al.*, ‘A Survey on Metaverse: Fundamentals, Security, and Privacy’, *IEEE Commun. Surv. Tutor.*, vol. 25, no. 1, pp. 319–352, 2023, doi: 10.1109/COMST.2022.3202047.

[3] A. Awadallah *et al.*, ‘Artificial Intelligence-Based Cybersecurity for the Metaverse: Research Challenges and Opportunities’, *IEEE Commun. Surv. Tutor.*, vol. 27, no. 2, pp. 1008–1052, 2025, doi: 10.1109/COMST.2024.3442475.

[4] M. El-Hajj, 'Cybersecurity and Privacy Challenges in Extended Reality: Threats, Solutions, and Risk Mitigation Strategies', *Virtual Worlds*, vol. 4, no. 1, 2025, doi: 10.3390/virtualworlds4010001.
[5] K. Otsuka and A. Kanaoka, 'Auditory Stimulus Attack in XR: Stimulus Characteristics and Technical Background Considerations', in *Proc. - Int. Conf. Metaverse Comput., Netw. Appl., MetaCom*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 332–339. doi: 10.1109/MetaCom65502.2025.00059.
[6] T. Lahtinen, A. Costin, and G. Suarez-Tangil, 'Brain-Computer Interface Integration With Extended Reality (XR): Future, Privacy And Security Outlook', in *European Conf. Inf. Warfare Security, ECCWS*, Curran Associates Inc., 2024, pp. 265–271.
[7] G. Duggal, M. Garg, and A. Nigam, 'Dark Side of the Metaverse and User Protection', in *The Metaverse Dilemma: Challenges and Opportunities for Business and Society*, C. Krishnan, A. Behl, S. Dash, and P. D. Yadav, Eds, Emerald Publishing, 2024, pp. 269–284. doi: 10.1108/978-1-83797-524-220241016.
[8] P. Casey, I. Baggili, and A. Yarramreddy, 'Immersive Virtual Reality Attacks and the Human Joystick', *IEEE Trans. Dependable Secure Comput.*, vol. 18, no. 2, pp. 550–562, 2021, doi: 10.1109/TDSC.2019.2907942.
[9] E. M. Ghourab, M. Azab, D. Gracanin, O. Alhussein, M. Al-Qutayri, and S. Muhaidat, 'Extended Reality-Aware Wireless Communication Networks: A Systematic Literature Review', *IEEE Open J. Commun. Soc.*, vol. 6, pp. 7567–7588, 2025, doi: 10.1109/OJCOMS.2025.3599091.
[10] K. Lake *et al.*, 'Cybersecurity and Privacy Issues in Extended Reality Health Care Applications: Scoping Review', *JMIR XR Spat. Comput.*, vol. 1, 2024, doi: 10.2196/59409.
[11] V. Sobeslav and J. Horalek, 'Cybersecurity Baseline and Risk Mitigation for Open Data in IoT-Enabled Smart City Systems: A Case Study of the Hradec Kralove Region', *Sensors*, vol. 25, no. 16, p. 4966, Jan. 2025, doi: 10.3390/s25164966.
[12] C.-H. Min, D.-H. Kim, H. Yang, and J. Kwak, 'Towards Secure Legacy Manufacturing: A Policy-Driven Zero Trust Architecture Aligned with NIST CSF 2.0', *Electronics*, vol. 14, no. 20, p. 4109, Jan. 2025, doi: 10.3390/electronics14204109.
[13] O. Borchert, G. Howell, A. Kerman, S. Rose, and M. Souppaya, 'Implementing a zero trust architecture : high-level document', National Institute of Standards and Technology (U.S.), Gaithersburg, MD, NIST SP 1800-35, Jun. 2025. doi: 10.6028/NIST.SP.1800-35.
[14] S.-M. Park and Y.-G. Kim, 'A Metaverse: Taxonomy, Components, Applications, and Open Challenges', *IEEE Access*, vol. 10, pp. 4209–4251, 2022, doi: 10.1109/ACCESS.2021.3140175.
[15] X. Zhang, Y. Chen, L. Hu, and Y. Wang, 'The metaverse in education: Definition, framework, features, potential applications, challenges, and future research topics', *Front. Psychol.*, vol. 13, 2022, doi: 10.3389/fpsyg.2022.1016300.
[16] R. Acheampong, T. C. Balan, D.-M. Popovici, and A. Rekeraho, 'Embracing XR System Without Compromising on Security and Privacy', in *Lect. Notes Comput. Sci.*, Springer Science and Business Media Deutschland GmbH, 2023, pp. 104–120. doi: 10.1007/978-3-031-43401-3_7.
[17] D. Malve, C. Kishor Kumar Reddy, H. Meenal, and V. Hagaldivte, 'Cybersecurity vulnerabilities and threat mitigation strategies in the Metaverse', in *Defending the Metaverse: Cybersecurity Strategies for the Next Generation Internet*, R. Sheth, M. Ouaissa, M. Ouaissa, E. M. Onyema, and C. Parekha, Eds, CRC Press, 2025, pp. 94–114. doi: 10.1201/9781003581659-6.
[18] S. Qamar, Z. Anwar, and M. Afzal, 'A systematic threat analysis and defense strategies for the metaverse and extended reality systems', *Comput. Secur.*, vol. 128, 2023, doi: 10.1016/j.cose.2023.103127.
[19] A. Ishtaiwi, A. Al-Qerem, A. Aldweesh, and M. Alkasassbeh, 'A Framework for Addressing Cybersecurity Risks in the Metaverse Safeguarding Against Generative AI Threats', in *Examining Cybersecurity Risks Produced by Generative AI*, A. Almomani and M. Alauthman, Eds, IGI Global, 2025, pp. 381–400. doi: 10.4018/979-8-3373-0832-6.ch016.
[20] D. Gračanin, J. Park, and M. Eltoweissy, 'XR-CEIL: Extended Reality for Cybersecurity Experiential and Immersive Learning', in *Commun. Comput. Info. Sci.*, Springer Science and Business Media Deutschland GmbH, 2022, pp. 487–492. doi: 10.1007/978-3-031-06394-7_61.
[21] R. C. Sharma and A. Zamfiroiu, 'Cybersecurity Threats and Vulnerabilities in the Metaverse', in *Int. Conf. Intell. Metaverse Technol. Appl., iMETA*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/iMETA59369.2023.10294950.
[22] C. Mazzocca, A. Acar, S. Uluagac, R. Montanari, P. Bellavista, and M. Conti, 'A Survey on Decentralized Identifiers and Verifiable Credentials', *IEEE Commun. Surv. Tutor.*, vol. 27, no. 6, pp. 3641–3671, Dec. 2025, doi: 10.1109/COMST.2025.3543197.
[23] E. R. Noble, T. F. Ask, and B. J. Knox, 'Tailored Extended Reality Environments for Education and Training in Cybersecurity: Engagement Beyond Awareness', in *IEEE Global Eng. Edu. Conf., EDUCON*, IEEE Computer Society, 2025. doi: 10.1109/EDUCON62633.2025.11016494.

[24] M. Al-Emran and M. Deveci, 'Unlocking the potential of cybersecurity behavior in the metaverse: Overview, opportunities, challenges, and future research agendas', *Technol. Soc.*, vol. 77, 2024, doi: 10.1016/j.techsoc.2024.102498.
[25] M. Al-Emran *et al.*, 'Evaluating the barriers affecting cybersecurity behavior in the Metaverse using PLS-SEM and fuzzy sets (fsQCA)', *Comput. Hum. Behav.*, vol. 159, 2024, doi: 10.1016/j.chb.2024.108315.
[26] Y. Huang, Y. J. Li, and Z. Cai, 'Security and Privacy in Metaverse: A Comprehensive Survey', *Big Data Min. Anal.*, vol. 6, no. 2, pp. 234–247, 2023, doi: 10.26599/BDMA.2022.9020047.