virtual worlds

MDPI

Article

Securing Virtual Reality: Threat Models, Vulnerabilities, and Defense Strategies

Andrija Bernik [1,*], Igor Tomicic [2] and Petra Grd [2]

[1] Department of Multimedia, University North, 42000 Varazdin, Croatia
[2] Faculty of Organization and Informatics, University of Zagreb, 42000 Varazdin, Croatia; igor.tomicic@foi.unizg.hr (I.T.); petra.grd@foi.unizg.hr (P.G.)
* Correspondence: abernik@unin.hr

Abstract

As virtual reality technologies evolve toward widespread adoption in education, industry, and social communication, their increasing complexity exposes new and often overlooked security challenges. Immersive environments collect continuous multimodal data, including motion tracking, gaze, voice, and biometric indicators that extend far beyond traditional computing attack surfaces. This paper synthesizes recent research (2023–2025) on cybersecurity, privacy, and behavioral safety in virtual reality (VR) systems, identifies the main vulnerabilities, and proposes a unified defense architecture: the three-layer VR Security Framework (TVR-Sec). Through comparative review and conceptual integration of 31 peer-reviewed studies, three interdependent protection domains emerged: (1) System Integrity, securing hardware, firmware, and network communications against spoofing and malware; (2) User Privacy, ensuring the ethical management of biometric and behavioral data through federated learning and consent-based control; and (3) Socio-Behavioral Safety, addressing harassment, manipulation, and psychological exploitation in shared virtual spaces. The framework situates VR security as a multidimensional adaptive process that combines technical hardening with human-centered defense and ethical design. By aligning cyber–human protections through an AI-driven monitoring and policy engine, TVR-Sec advances a holistic paradigm for securing future immersive ecosystems.

Keywords: virtual reality; cybersecurity; threat modeling; privacy; biometric data; human–computer interaction; behavioral safety

Check for updates

Academic Editor: Rosario Schiano Lo Moriello

Received: 18 December 2025
Revised: 25 February 2026
Accepted: 2 March 2026
Published: 10 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Virtual Reality (VR) has rapidly progressed from experimental visualization technology to a mature platform that supports education, industrial design, healthcare, and social collaboration [1,2]. Modern head-mounted displays, high-fidelity sensors, and haptic interfaces now deliver immersive experiences that blur the boundary between the digital and the physical. However, as VR systems become increasingly networked and data-rich, they also inherit and amplify the vulnerabilities of conventional computing environments [3]. The same attributes that make VR powerful (embodiment, interactivity, and sensory realism) introduce unique attack surfaces that traditional cybersecurity frameworks were never designed to address [4,5].

Recent research has exposed weaknesses ranging from biometric data leakage and motion-tracking inference attacks to manipulation of the man-in-the-room and virtual harassment [6–9]. Experimental analysis of immersive platforms such as Oculus, Vive,

and OpenXR has revealed susceptibility to spoofed sensor streams, malicious plug-ins, and unauthorized environment capture [10]. These risks are compounded by the proliferation of cross-reality ecosystems, where VR interacts with AI-driven avatars, Internet-of-Things (IoT) devices, and cloud data services [11]. Yet, the current literature on VR security remains fragmented. Technical studies typically focus on encryption and network integrity [6,11,12]; human-factor research emphasizes privacy and psychological safety [13,14]; and ethical analyses highlight data ownership and manipulation concerns [15]. Weerasinghe et al. [16] explicitly integrates user-safety tools and system-level considerations in social VR. Consequently, stakeholders—from developers to educators-lack a coherent model that bridges the gap between hardware hardening and soft-skills safety.

This paper addresses that gap by synthesizing 31 recent peer-reviewed, open-access studies published between 2023 and 2025. Through cross-disciplinary thematic analysis, recurring vulnerabilities are identified and organized into three interdependent security domains: System Integrity, User Privacy, and Socio-Behavioral Safety. Building on these insights, the Tri-Layer VR Security Framework (TVR-Sec) is introduced, a holistic socio-technical model that integrates hardware and firmware protection, privacy-preserving data governance, and real-time behavior-aware safety mechanisms.

The contribution of this work is trifold: (i) Mapping the contemporary VR threat landscape and revealing how technical, informational, and behavioral vulnerabilities intersect; (ii) Synthesizing fragmented research perspectives into a coherent structure that reflects the socio-technical nature of immersive systems; and (iii) Proposing TVR-Sec, an adaptive, AI-assisted framework that unifies integrity, privacy, and behavioral safety controls into a single defense architecture suitable for diverse VR applications.

By situating VR security as a continuous negotiation between technological reliability and human well-being, this paper lays the groundwork for future empirical validation and standardization efforts. The goal is not only to secure immersive environments technically but also to enhance trust, autonomy, and safety for all participants.

The remainder of this paper is organized as follows. Section 2 describes the methodology of the literature synthesis, including data sources, selection criteria, and qualitative coding procedures. Section 3 shows the results of the comparative review and identifies cross-domain threat patterns in contemporary VR systems. Section 4 introduces the Tri-Layer VR Security Framework (TVR-Sec), detailing its conceptual rationale, architectural components, and theoretical implications. Section 5 discusses practical challenges, ethical considerations, alignment with existing standards, regulations, and future research directions. Finally, Section 6 concludes the paper.

2. Methodology of Synthesis

2.1. Purpose and Scope

This study adopts a conceptual synthesis approach, combining systematic review techniques with interpretive analysis to map current trends in VR security and privacy research. The goal was not to quantify findings statistically but to identify recurring vulnerabilities, conceptual gaps, and defense strategies across diverse disciplines: computer science, cybersecurity, psychology, and human-computer interaction. This study is intentionally designed as a conceptual–integrative synthesis, rather than an exhaustive systematic review. The objective is to identify recurring cross-domain patterns that emerge at the intersection of immersive system integrity, privacy risk, and socio-behavioral harm—domains that are typically analyzed in isolation. Accordingly, the corpus was restricted to highly relevant, peer-reviewed, open-access studies that explicitly examine security threats or defenses in immersive XR contexts. While only the most analytically central studies are cited explicitly in the main text, the qualitative synthesis and bibliometric validation were informed by a

broader screened corpus of 32 studies. The remaining studies, which reinforce but do not independently alter the conceptual conclusions, are documented in Appendix B Table A2 for transparency and reproducibility.

Given the rapid evolution of immersive technologies, the synthesis focused on studies published between January 2023 and November 2025, a period marked by the emergence of AI-enhanced XR ecosystems and growing attention to ethical design [6,7,10]. Although the primary search window focuses on 2023–2025 to capture recent advances in AI-enabled immersive systems, the synthesis is conceptually informed by earlier foundational research demonstrating that movement, gaze, and avatar behavior function as biometric identifiers, and that immersive social environments produce distinct psychological risks. These foundational insights underpin the threat categories analyzed here, while the restricted timeframe ensures relevance to contemporary XR architectures and governance challenges. The analysis sought to bridge two often-disconnected research communities: those developing technical countermeasures and those investigating social and psychological safety in virtual environments.

2.2. Literature Sources and Search Strategy

To conduct the literature review, the databases Web of Science (WoS), Scopus, IEEE, and Science Direct were used. Boolean queries were constructed along three thematic axes: (a) system security, (b) data privacy, and (c) user safety.

Representative search strings included:

- "virtual reality" AND ("cybersecurity" OR "threat model" OR "attack surface");
- "XR" AND ("privacy" OR "biometric data" OR "identity leakage");
- "social virtual worlds" AND ("harassment" OR "manipulation" OR "behavioral safety");
- "metaverse" AND ("trust framework" OR "defense strategy" OR "risk assessment").

Filters applied: 2023–2025; English language, Open access, Research articles, reviews, or framework papers. After deduplication, 93 records remained. Screening titles, abstracts, and methodology sections for relevance yielded 31 eligible studies (Figure 1).

Searches were conducted primarily within titles, abstracts, and author-provided keywords across Web of Science, Scopus, IEEE Xplore, and MDPI databases to ensure thematic precision. Full-text searches were applied selectively when database functionality permitted. After de-duplication, records underwent title/abstract screening to remove clearly irrelevant items (e.g., VR content without a security/privacy/safety focus). The remaining articles were retrieved in full text and screened against the inclusion criteria. When multiple versions existed (preprint + journal), the peer-reviewed version was retained; if only a preprint existed, but it introduced a novel XR threat/defense and met inclusion criteria, it was retained. Disagreements were resolved by re-reading the full text and documenting the exclusion reason (e.g., non-XR setting, non-security focus, inaccessible full text). The final query execution date was recorded, and duplicate records were removed prior to screening. Screening proceeded in two stages: (i) title and abstract screening for relevance, followed by (ii) full-text screening against inclusion criteria. Preprints were included only when they introduced novel XR security mechanisms and demonstrated significant scholarly uptake.

Final query string: TITLE-ABS-KEY(("virtual reality" OR VR OR "extended reality" OR XR OR metaverse OR "social VR") AND (security OR cybersecurity OR privacy OR threat* OR vulnerab* OR attack* OR defense* OR "sensor spoofing" OR "biometric inference" OR harassment OR moderation)) AND (PUBYEAR > 2022).

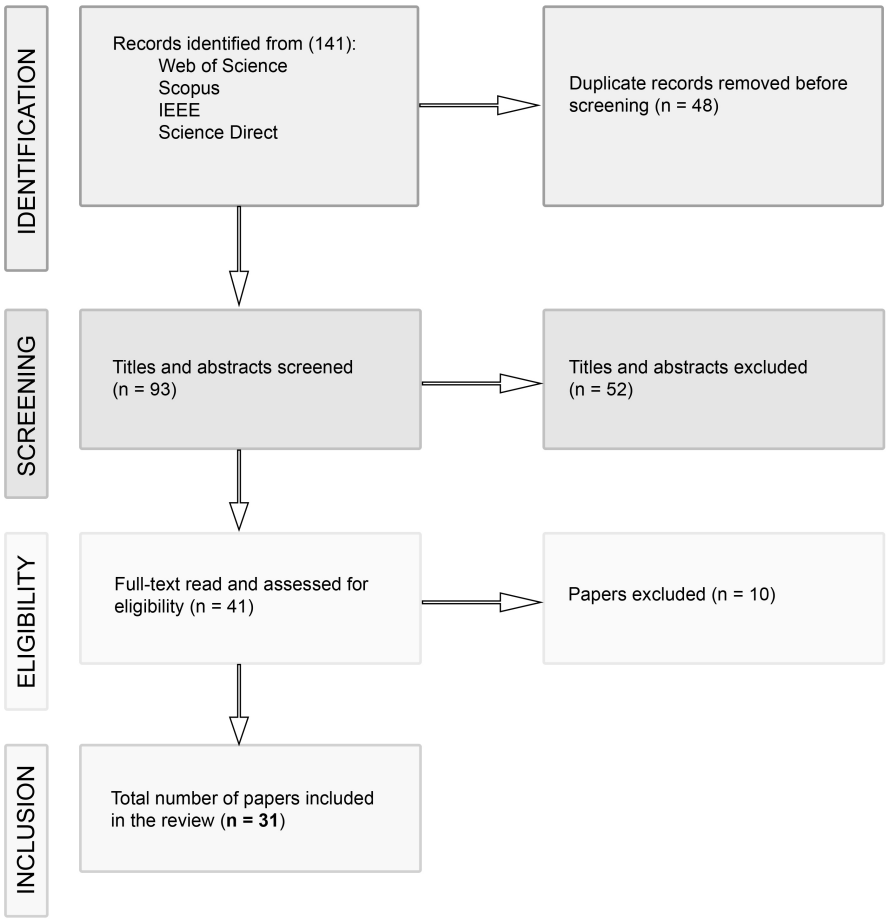


Figure 1. Literature review diagram.

2.3. Inclusion and Exclusion Criteria

The selected works shared three defining characteristics: (1) they explicitly examined security, privacy, or trust within immersive VR/AR/XR contexts; (2) they offered empirical evidence, technical modeling, or conceptual frameworks rather than anecdotal commentary; and (3) they were openly accessible to ensure verifiability and alignment with open-science principles.

Each paper was reviewed in detail and coded along six dimensions: system architecture, attack type, security goal, defense mechanism, evaluation method, and outcome. This qualitative coding process, performed in NVivo 2024, enabled thematic comparison and the identification of conceptual convergence. Iterative axial coding exposed three dominant clusters recurring across technical and human-factor literature:

- System Integrity-safeguarding hardware, firmware, and network components from malicious tampering [6,10,13];
- User Privacy-protecting biometric, locational, and behavioral data from unauthorized collection or profiling [7,8,16];
- Socio-Behavioral Safety-mitigating psychological manipulation, harassment, and social engineering within shared virtual spaces [9,14].

These categories mirror the structure of the classical cybersecurity triad (Confidentiality–Integrity–Availability) but extend it to include human experience and ethical dimensions inherent to immersive systems.

Inclusion criteria:Studies were included if they: (i) addressed VR/AR/MR/XR systems or metaverse/social VR settings; (ii) analyzed security, privacy, trust, or safety threats

(technical or socio-behavioral) OR proposed a defense/mitigation; (iii) provided empirical evidence (measurements, user studies), formal models (threat taxonomies), or peer-reviewed synthesis; (iv) were available as full text (open access preferred) and written in English; and (v) fell within 2023–2025, with foundational pre-window exceptions included only when directly grounding a threat class used in the framework. Figure 2 visualizes the citation-to-framework mapping used during selective coding, showing how each included study (1–31); Table A1) contributes evidence to the System Integrity, User Privacy, or Socio-Behavioral Safety layer (provided below).

Exclusion criteria: Studies were excluded if they (i) addressed only non-immersive 2D systems, (ii) focused purely on VR application content without security/privacy/safety implications, (iii) were editorials without technical or analytical contribution, or (iv) lacked retrievable full text.

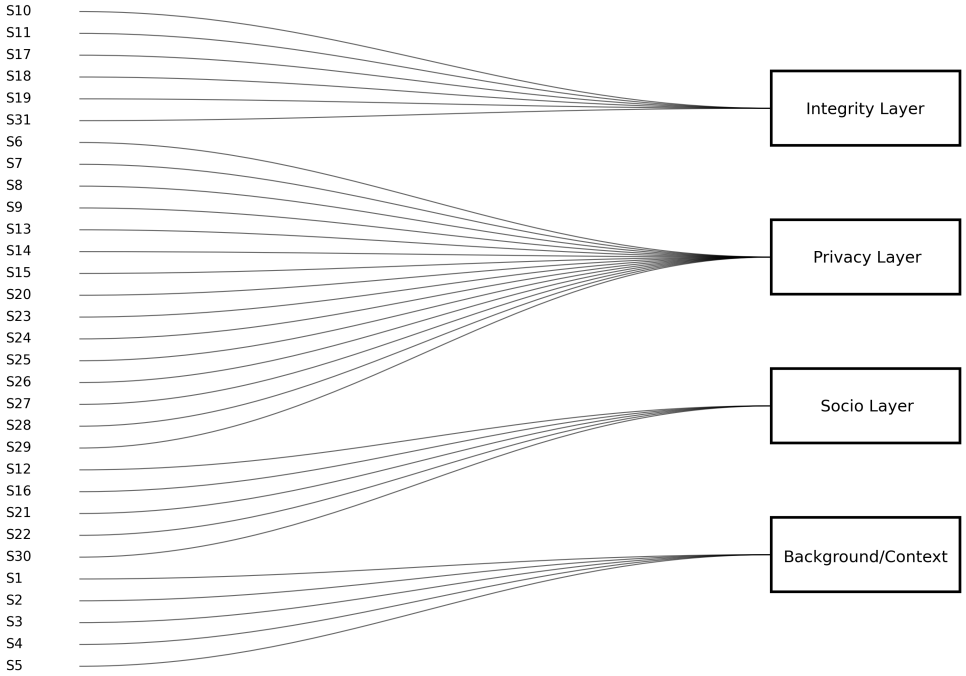


Figure 2. Citation-to-framework mapping for the reviewed corpus (1–31).

2.4. Bibliometric Snapshot

To complement qualitative synthesis, a bibliometric snapshot was performed using Bibliometrix (R) to analyze keyword co-occurrence and thematic clustering across the 31-study corpus. After importing metadata (BibTeX/CSV), we calculated (i) keyword frequency, (ii) keyword co-occurrence adjacency matrices, and (iii) clustering via Louvain community detection (default Bibliometrix network clustering) to identify thematic groups. Thematic evolution was inspected using the year as a temporal attribute to confirm the persistence of clusters over time. Metadata fields (title, keywords, publication year) were extracted and analyzed to identify dominant research themes. Geographic provenance was operationalized as the corresponding author's affiliation country; if a corresponding author was not indicated, the first author's affiliation country was used. Country values were obtained from the article metadata available in the publisher record (or bibliographic export when available) and normalized to a single-country label per study for aggregation. The resulting geographic distribution of the included corpus (1–31) is visualized in Figure 3.

For sources lacking standardized export (e.g., arXiv), metadata were manually appended to the same schema (title, year = arXiv, keywords assigned from abstract terms) to ensure consistent processing. Results indicate three stable clusters corresponding to (i) de-

vice and sensor integrity, (ii) biometric privacy and data inference, and (iii) socio-behavioral risk and moderation in immersive environments. The co-occurrence network consistently yielded three dense keyword communities across multiple clustering runs: (1) integrity-focused terms (sensor, firmware, spoofing, authentication, network), (2) privacy-focused terms (biometrics, gaze, motion data, inference, identifiability, differential privacy, federated learning), and (3) socio-behavioral safety terms (harassment, moderation, trust, embodiment, governance). Stability was assessed by re-running clustering after removing the 5 most frequent generic terms (e.g., "VR", "metaverse") and observing that the community structure remained unchanged. These bibliometric results align with the three-layer structure we derived qualitatively. It refers to the network modularity and cluster separability observed in keyword co-occurrence graphs (high within-cluster density and limited between-cluster ties) and the persistence of these clusters across sensitivity checks. This provides quantitative corroboration that the tri-layer structure is not arbitrary but reflects dominant research themes. Figure 4 visualizes these stable clusters across publication years, with each study positioned by its dominant TVR-Sec layer.

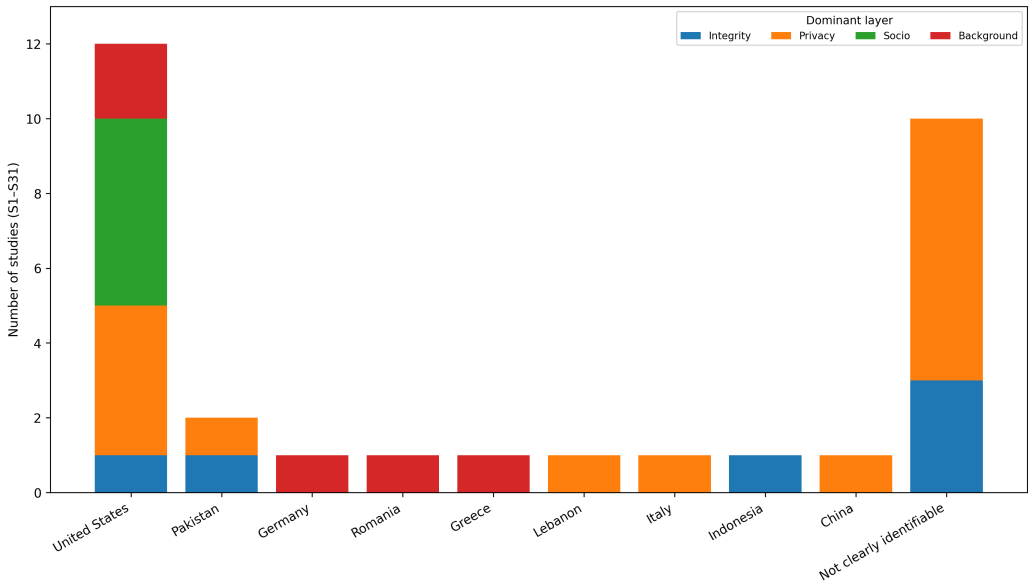


Figure 3. Geographic distribution of included studies (1–31).

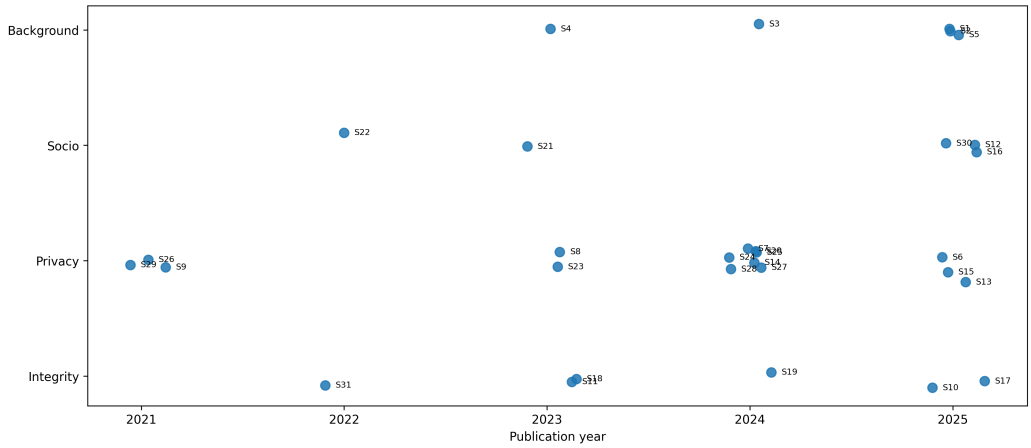


Figure 4. Cluster visualization of included studies.

3. Results of the Comparative Review

3.1. Overview of the Security Landscape

Across the thirty-one studies reviewed, researchers agree that immersive systems introduce a unique, multi-layered threat surface. The convergence of hardware sensors, continuous user tracking, and networked collaboration produces vulnerabilities that traditional IT security models only partially address. Approximately two-thirds of the analyzed papers presented experimental or simulation data, while the remainder offered conceptual frameworks or systematic reviews. Together, they portray an ecosystem where technical integrity, personal privacy, and psychological safety are interdependent dimensions of protection.

Most empirical investigations focus on two technological layers—device integrity and data protection. El-Hajj [6] characterized VR headsets as "perceptual endpoints," vulnerable to sensor spoofing, firmware tampering, and unencrypted wireless links. Recent analyses identified weaknesses in unsecured data channels and inadequate authentication in peer-to-peer VR sessions [11]. Confirmed that insecure communication protocols could permit packet interception during peer-to-peer sessions, while [10] demonstrated that manipulated calibration signals could disorient users physically. Complementary research on data leakage, for example, Nair et al. [8] and Giaretta [7], showed that motion traces and gaze trajectories can reveal user identity or emotional state with up to 95% accuracy, even when datasets are anonymized.

Beyond infrastructure, a growing body of behavioral research highlights social and psychological vulnerabilities inherent in immersive environments. Recent studies documented psychological harm from immersive harassment, including avatar intrusion and unwanted proximity in shared VR spaces [12]. Clark et al. [14] and Rahman [15] warned that the same sensory realism that enhances empathy can also enable emotional manipulation or persuasion. These works collectively emphasize that technical fortification alone cannot ensure user safety; psychological integrity must also be protected. Figure 5 presents a Sankey-style thematic evolution view showing how the included studies accumulate over time and converge into the dominant TVR-Sec layers identified through selective coding.

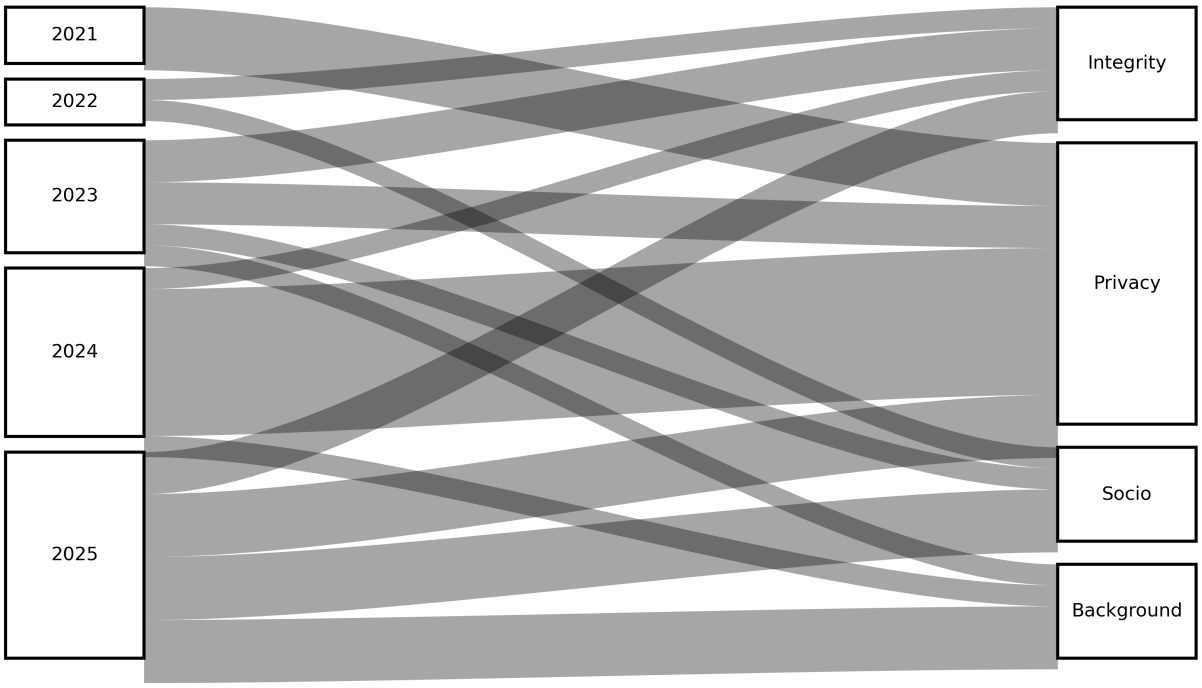


Figure 5. Thematic evolution of the evidence base (year to TVR-Sec layer).

3.2. Patterns and Cross-Domain Insights

When viewed together, the studies reveal several converging insights:

- Threat Interdependence. A breach at one level often cascades across others. Compromised sensors (system integrity) can leak behavioral data (privacy) that later enables targeted manipulation (behavioral safety).
- Data Centralization Risks. Many commercial VR systems rely on cloud processing for tracking and analytics, creating single points of failure. Works by Weerasinghe et al. [16] and Nair et al. [8] propose federated or edge-based learning as partial solutions.
- Underdeveloped Human-Centered Defenses. Only a minority of studies integrate psychological-safety mechanisms, such as spatial-boundary enforcement or real-time moderation, into the same architectural layer as encryption and authentication [9,14].
- Lack of Standardized Evaluation Metrics. Security performance, latency impact, and user comfort are measured inconsistently, complicating cross-study comparisons [7,10].

This shows that existing VR security research is fragmented but complementary: engineers focus on technical reliability, while social science researchers emphasize experience and ethics. Synthesizing these perspectives reveals that a truly secure virtual environment must treat hardware resilience, data governance, and social trust as parts of the same continuum.

3.3. The Fragmentation Problem

Synthesizing these perspectives reveals a critical fragmentation: technical studies rarely account for psychological harm, while behavioral studies often ignore hardware root causes. This disconnection highlights the necessity for a model that treats hardware resilience, data governance, and social trust as a continuum rather than isolated silos. A unified framework must therefore bridge the gap between hardware reliability and human experience.

These domains are conceptually intertwined rather than hierarchical. Technical compromise undermines privacy; privacy loss fuels social exploitation; and social manipulation, in turn, erodes user trust, making security compliance less effective. Recognizing this interdependence lays the groundwork for the Tri-Layer VR Security Framework (TVR-Sec) presented in the next section—a model that unifies these elements into a single, adaptive defense architecture.

In immersive systems, user-generated telemetry, pose, gaze, interaction timing, and spatial behavior are treated as ground-truth input for analytics and adaptive control. Socio-behavioral manipulation, such as harassment or coercive interaction, can distort this telemetry by inducing stress, avoidance, or erratic behavior, thereby corrupting downstream datasets. Consequently, behavioral harm does not merely affect user experience but can propagate into system-level decision processes, undermining data integrity and adaptive reliability.

4. The Tri-Layer VR Security Framework (TVR-Sec)

4.1. Conceptual Rationale

The synthesis of vulnerabilities across 31 reviewed studies demonstrates that immersive environments require multidimensional defense architectures. Conventional cybersecurity models, rooted in confidentiality, integrity, and availability, are insufficient for VR, where threats span hardware reliability, data privacy, and psychosocial well-being [6,7,9]. The Tri-Layer VR Security Framework (TVR-Sec) extends traditional paradigms by intro-

ducing three mutually reinforcing domains: 1. System Integrity Layer; 2. User Privacy Layer; 3. Socio-Behavioral Safety Layer. Together, they form an adaptive security ecosystem capable of defending both the technical substrate and the human participant in virtual spaces. Figure 6 illustrates the conceptual architecture, showing the interactions among the system integrity, user privacy, and socio-behavioral safety layers through a unified AI-driven monitoring and policy engine with continuous feedback across device, data, and behavioral streams.

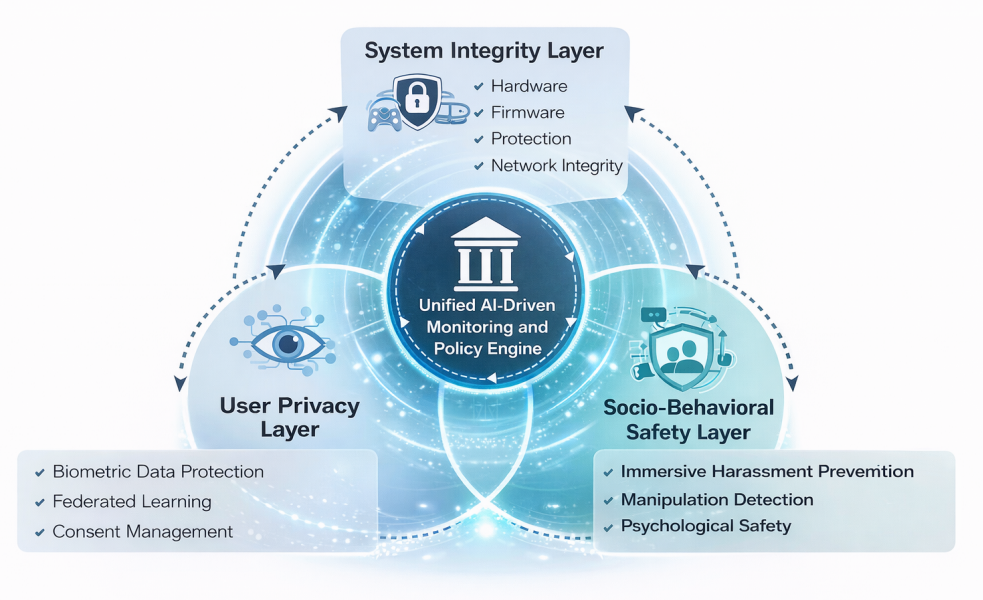


Figure 6. Conceptual architecture of the Tri-Layer VR Security Framework (TVR-Sec).

4.2. Structural Overview and Interdependence of Layers

At its core, TVR-Sec functions as a closed-loop defense system. Sensor data, user actions, and network activity flow into a central inference engine that correlates anomalies between layers. If irregular latency or motion inconsistency is detected (integrity alert), the system checks whether correlated anomalies appear in data access patterns (privacy alert) or in user behavior (safety alert). This feedback structure prevents isolated vulnerabilities from cascading into systemic breaches.

Cross-Layer Signaling: Crucially, these layers must communicate via low-latency signals. For example, the System Integrity layer must immediately flag a 'sensor drift' event to the Socio-Behavioral layer. If the system detects a hardware anomaly that could cause physical disorientation, the AI moderator should preemptively pause complex social interactions. This prevents technical faults from being exploited as behavioral vulnerabilities (e.g., users being unable to react to harassment due to tracking lag).

Table 1 shows the connection between layers, primary protected assets, threat classes, and defense mechanisms:

- System Integrity protects the technological backbone. It preserves the fidelity of perceived reality by validating firmware integrity, encrypting sensor communication, and monitoring for hardware spoofing. A compromised headset or controller can distort not only visuals but trust itself.
- User Privacy defends the informational dimension. VR devices continuously collect gaze, gesture, and voice data, forms of "digital physiology." The framework promotes privacy-by-design approaches such as edge analytics and federated learning, ensuring that sensitive data are processed locally rather than centralized in the cloud [8,15,16].

- Socio-Behavioral Safety secures the experiential layer. Social VR spaces can reproduce real-world power dynamics and misconduct. By incorporating AI-assisted moderation, avatar authentication, and personal-space enforcement, this layer translates ethical principles into real-time behavioral protection [12,14,16].

Table 1. Layers, assets, threats, and defense mechanisms.

Layer	Primary Assets Protected	Threat Classes	Defense Mechanisms	Representative Evidence
System Integrity	Hardware, firmware, network stack	Sensor spoofing, malware injection, network intrusion	Secure boot, firmware attestation, end-to-end encryption, anomaly detection	[6,10,12]
User Privacy	Identity, biometric and behavioral data	Profiling, data leakage, inference attacks	Differential privacy, on-device analytics, federated learning, consent management	[7,8,16]
Socio-Behavioral Safety	Psychological state, social interaction	Harassment, manipulation, misinformation	AI-assisted moderation, avatar authentication, ethical design policies	[9,14,15]

These layers operate in synchrony. Technical failures can escalate into emotional distress; breaches of privacy can erode social safety; and psychological manipulation can compromise data integrity. Hence, the TVR-Sec framework insists on a holistic view: a secure virtual environment is one where users feel safe, respected, and in control.

4.3. Implementation and Theoretical Implications

Although conceptual, TVR-Sec is technologically feasible by leveraging existing XR middleware. Engines such as Unity or Unreal handle rendering and interaction, while the framework would integrate platform-level security services (e.g., secure enclaves, TPM- or TrustZone-based integrity APIs) alongside OpenXR extensions for device abstraction. Its architecture can be realized through modular components that combine encryption protocols [12], federated learning modules for privacy-preserving personalization [16], and real-time behavioral analytics for safety monitoring [9]. Interoperability is supported through existing XR and telemetry standards, for example, OpenXR for device interfaces, glTF/USD for scene representation, and emerging ISO/IEC and IEEE frameworks for cross-platform data exchange, which provide the necessary scaffolding for integrating heterogeneous hardware and applications.

Beyond its technical utility, TVR-Sec contributes theoretically by reframing VR security as a **s**ocio-technical system rather than a purely computational one. It positions security as a dynamic process of negotiation between machine and human agency—between algorithmic control and ethical responsibility. In doing so, it bridges the gap between cybersecurity engineering and digital well-being research, proposing a new paradigm where safeguarding information and safeguarding experience are two sides of the same principle: trust. Future validation of TVR-Sec should operationalize each layer into measurable variables: integrity (latency, firmware validation success rate), privacy (data minimization ratio), and safety (incident reduction metrics). Establishing these indicators would enable empirical testing and benchmarking against conventional XR security models.

Publicly observable VR platforms already deploy isolated security mechanisms corresponding to individual TVR-Sec layers, yet rarely integrate them holistically. For example, enterprise XR deployments increasingly rely on secure boot, firmware attestation, and runtime verification to protect device integrity, while social VR platforms implement personal-space boundaries, user reporting, and AI moderation to address behavioral harm. However, privacy-preserving governance of biometric and interaction data, such as gaze, motion, and avatar behavior, remains inconsistently implemented across platforms. Taken together, these examples show why an integrated approach is needed.

5. Discussion and Future Work

Implementing TVR-Sec introduces significant architectural challenges, primarily regarding latency and user agency. Socio-behavioral safety requires real-time analysis of voice and gesture, which competes with rendering tasks for bandwidth. Furthermore, aggressive System Integrity checks may break immersion, the quality VR seeks to sustain. Future implementations must balance the 'Safety vs. Immersion' trade-off, potentially by offloading behavioral analysis to edge processors. When sensors capture movement, gaze, or voice, they do not only transmit data-they translate human presence into machine-readable form. Consequently, any weakness in this chain threatens both system reliability and user dignity.

The proposed Tri-Layer VR Security Framework (TVR-Sec) addresses this interdependence by merging System Integrity, User Privacy, and Socio-Behavioral Safety into a unified structure. In this model, protecting the system also means protecting the person: hardware and network resilience sustain trust in perception, privacy safeguards ensure ethical control of biometric data, and behavioral protections preserve the sense of safety that allows users to explore immersive spaces without fear of manipulation or harm.

This holistic orientation shifts cybersecurity from a reactive task to a living process of adaptive negotiation. Security becomes dynamic, continuously monitoring risks and rebalancing protection across layers in real time. Similar to the concept of "autonomous computing," the framework encourages context-aware defense, enabling systems to recognize irregular patterns and automatically adjust encryption strength or moderation thresholds [11]. Such adaptability ensures that safety evolves alongside experience rather than obstructing it.

Equally important is the call for interoperability and transparency. Current VR platforms often operate within closed ecosystems, limiting data sharing about vulnerabilities and best practices. Open standards such as xAPI and ISO/IEC 27032 [17] could form a common language for security events, enabling collective defense across devices and vendors. Transparency at the user level is equally critical—people must know what information their systems collect, why it is needed, and how they can control or revoke access [16]. Additionally, compliance with emerging governance standards such as AI risk management and the EU General Data Protection Regulation (GDPR) is essential for ensuring lawful and ethical deployment of adaptive VR defenses. Integrating these norms into the TVR-Sec policy layer would align the framework with real-world legislative contexts.

The ethical dimension of immersive protection is inseparable from its technical one. Biometric and behavioral data hold predictive power: they can infer emotion, stress, or intent [7,8]. This sensitivity demands ethical oversight, particularly as AI-driven personalization blurs the line between assistance and surveillance. Socio-behavioral safety measures, such as AI moderation or avatar authentication, must therefore be guided by fairness and respect for user autonomy [12,14,16]. At a deeper level, as Rahman [15] observes, virtual experiences challenge our moral understanding of what constitutes harm or

responsibility. If the mind perceives the virtual as real, then ethical design must protect the virtual self as carefully as the physical one.

Future research should pursue empirical validation of the TVR-Sec framework in real-world contexts, assessing how adaptive protection influences user trust, latency, and engagement. Longitudinal studies could evaluate how continuous monitoring affects comfort and agency over time. Technical advances in edge processing and federated learning [16] may further support privacy-preserving analytics without compromising immersion. Ultimately, securing virtual reality requires not just advanced encryption or regulation, but a redefinition of safety as the equilibrium between technological integrity and human well-being.

6. Conclusions

This study synthesizes contemporary research on virtual reality security to propose the Tri-Layer VR Security Framework (TVR-Sec)—a model that unites the technological, informational, and psychological dimensions of protection into a single adaptive ecosystem. By integrating System Integrity, User Privacy, and Socio-Behavioral Safety, the framework responds to the evolving threat landscape where hardware vulnerabilities, data exposure, and behavioral manipulation intersect. The conclusions presented are grounded in recurring patterns identified across the reviewed corpus, in which system integrity failures frequently coincided with privacy leakage, and both amplified socio-behavioral risk. These interdependencies were consistently observed across application domains, supporting the need for an integrated, multi-layer security model rather than isolated defensive measures.

Findings across thirty-one open-access studies confirm that immersive systems face unique risks: sensor spoofing and firmware attacks compromise realism [6,10]; biometric and motion-tracking data enable identity inference and profiling [7,8]; and social VR platforms expose users to harassment or emotional coercion [9,14]. Addressing these threats requires coordination across technical and ethical boundaries. TVR-Sec provides coordination through an AI-driven policy engine that interprets real-time telemetry while respecting user autonomy.

Beyond a defense model, TVR-Sec represents a paradigm shift in how security is conceptualized. It extends the traditional cybersecurity triad—confidentiality, integrity, and availability—by embedding empathy, transparency, and psychological safety as core design principles. It invites collaboration among engineers, ethicists, and policymakers to create immersive systems that are not only resilient but humane. Future implementations should focus on interoperable standards, user-driven consent interfaces, and continuous validation in diverse application domains.

In the emerging metaverse, the measure of progress will not be immersion alone, but trust-the confidence that virtual worlds are safe spaces for exploration, creation, and learning. The TVR-Sec framework offers a foundation for that trust, guiding the next generation of immersive technologies toward security that is adaptive, ethical, and profoundly human. Moving forward, future iterations of the TVR-Sec framework should be implemented as reference architectures within open-source XR engines (e.g., Unity Security SDKs). Pilot evaluations in education, healthcare, and social collaboration could empirically verify its adaptability and ethical robustness. Such deployments would bridge the current conceptual–practical gap and facilitate standardization efforts in immersive cybersecurity.

Author Contributions: Conceptualization, A.B.; methodology, A.B., I.T. and P.G.; software, I.T. and P.G.; validation, A.B.; formal analysis, A.B., I.T. and P.G.; investigation, P.G.; resources, A.B.; writing—original draft preparation, A.B.; writing—review and editing, A.B., I.T. and P.G.; visualization, A.B. and P.G.; supervision, A.B.; project administration, P.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the 435 article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A. Literature Corpus

To ensure the synthesis captures layer-specific evidence beyond the paper's core citation set (ID1–ID16), we added a targeted corpus of XR/metaverse security and social-VR safety studies (ID17–ID32) spanning surveys and systematic reviews, empirical attack papers (sensor/eye-tracking/keystroke inference), and HCI studies on harassment and moderation. This triangulation provides direct empirical grounding for each TVR-Sec layer and reduces reliance on purely conceptual claims by anchoring the framework to validated threat mechanisms and evaluated defense/mitigation approaches across device, data, and behavioral streams.

Table A1. Reviewed literature corpus used for qualitative synthesis and bibliometric analysis.

ID	Reference	Core Focus	Research Method	Mapped TVR-Sec Layer(s)	DOI/Identifier
[1]	Surridge et al. (2025) VR to improve gait/balance in Parkinson's: scoping review (Virtual Worlds 4(2):13)	VR rehab efficacy for PD gait/balance	Scoping review	R, V	10.3390/virtualworlds4020013
[2]	Vola et al. (2025) 360° video VR and undergrad performance/engagement (Virtual Worlds 4(2):14)	360° video impact on engagement/performance	Empirical study (education intervention)	R, V	10.3390/virtualworlds4020014
[3]	Conrad et al. (2024) Learning effectiveness of immersive VR: systematic review (Computers & Education: X Reality 4:100053)	Learning outcomes + engagement in IVR	Systematic review	R, V	10.1016/j.cexr.2024.100053
[4]	Gligorea et al. (2023) Adaptive learning using AI in e-learning: literature review (Education Sciences 13:1216)	AI-driven adaptive learning (non-XR core)	Literature review	T	10.3390/educsci13121216
[5]	Lampropoulos (2025) AI + AR/VR in education: trends & perspectives (MTI 9(2):11)	Convergence of AI with XR in education	Review	T, V, R	10.3390/mti9020011
[6]	El-Hajj (2025) Cybersecurity & privacy challenges in XR (Virtual Worlds 4:1)	XR threat landscape + mitigations	Survey/review	T, V, Sec	10.3390/virtualworlds4010001
[7]	Giaretta (2025) Security & privacy in VR: literature survey (Virtual Reality 29:10)	VR security/privacy overview	Literature survey	T, V, Sec	10.1007/s10055-024-00860-4
[8]	Nair et al. (2023) Identify 50,000+ VR users from head/hand motion (USENIX Security)	Motion biometrics re-identification	Conference paper (ML + dataset analysis)	R, T, Sec	arXiv:2302.08927

Table A1. *Cont.*

ID	Reference	Core Focus	Research Method	Mapped TVR-Sec Layer(s)	DOI/Identifier
[9]	Zhang et al. (2021) ReAvatar de-anonymization via avatar motion (ACM CCS 2021 Poster)	De-anonymization attack via correlated motion	Conference poster/paper	V, R, Sec	10.1145/3460120.3485345
[10]	Liu et al. (2025) False Reality: sensor-induced Human–VR vulnerability (arXiv)	Sensor attack surface/induced vulnerabilities	Preprint	T, R, Sec	arXiv:2508.08043
[11]	Qamar et al. (2023) Threat analysis & defenses for Metaverse/XR (Computers & Security 128:103127)	Systematic threats + defenses	Threat analysis/survey	T, V, Sec	10.1016/j.cose.2023.103127
[12]	Trapp et al. (2025) Understanding harassment in social VR (IEEE VRW)	Harassment mechanisms + study design	Workshop paper (study design)	V, R, Sec	10.1109/VRW66409.2025.00043
[13]	Rehman et al. (2025) Avatar privacy: biometric + inferred data (Frontiers in VR)	Avatar data categories + privacy risks	Conceptual/review	V, R, Sec	10.3389/frvir.2025.01432
[14]	Clark et al. (2024) Behavioral identity detection in VR (Frontiers in VR)	Behavioral re-identification threats	Empirical/analysis (likely)	R, V, Sec	10.3389/frvir.2024.00987
[15]	Rahman (2025) Federated Learning survey (privacy-preserving) (arXiv)	Privacy-preserving collaborative ML	Survey	T, Sec	arXiv:2504.17703
[16]	Weerasinghe et al. (2025) Safety tools in social VR (IEEE TVCG)	Adoption/effectiveness of safety tools	Empirical user study	V, R, Sec	10.1109/TVCG.2025.3456107
[18]	Rahartomo et al. (2025) Metaverse security & privacy systematic review (Computers & Security 140:104602)	SLR on security/privacy research	Systematic review	T, V, Sec	10.1016/j.cose.2025.104602
[19]	Ren et al. (2024) DPGazeSynth: differentially private synthetic gaze (Information Sciences 649:120720)	Privacy-preserving gaze analytics	Method + evaluation	R, T, Sec	10.1016/j.ins.2024.120720
[20]	Ni et al. (2024) Sensor security in VR (ACM MobiSys)	Sensor exploits + defenses	Conference paper/survey+systems	T, R, Sec	10.1145/3643832.3661389
[21]	Schulenberg et al. (2023) Women's safety & harassment in social VR (PACM HCI)	Harassment experiences/risks for women	Qualitative (interviews)	V, R, Sec	10.1145/3610027
[22]	Freeman et al. (2022) Harassment & safety in social VR (PACM HCI)	User experiences + mitigation strategies	Qualitative/mixed (platform study)	V, R, Sec	10.1145/3512932
[23]	Wang et al. (2025) ChatGPT-integrated feedback aids in VR learning (EIT)	LLM feedback improving SRL/HOTS in VR	Empirical study	T, V, R	10.1007/s10639-025-13557-x
[24]	Crogman et al. (2025) VR/AR/MR in experiential learning (Education Sciences 15)	Experiential learning transformation	Conceptual/review	V, R	10.3390/educsci15030303
[25]	Muzata et al. (2024) Engineering education via VR: SLR (Virtual Worlds 3:480–505)	VR impact in engineering education	Systematic literature review	V, R	10.3390/virtualworlds3040026

Table A1. *Cont.*

ID	Reference	Core Focus	Research Method	Mapped TVR-Sec Layer(s)	DOI/Identifier
[26]	Hur (2025) Preservice teachers' intentions to adopt VR (Virtual Worlds 4)	Adoption factors in teacher education	Mixed-methods	R, V	10.3390/virtualworlds4020012
[27]	Lopes et al. (2025) Gamified VR exposure for arachnophobia (Virtual Worlds 4)	VR therapy feasibility	Pilot study	R, V	10.3390/virtualworlds4020015
[28]	Craig et al. (2025) VR cognitive–motor assessment in elite athletes (Virtual Worlds 4)	Validity/reliability of VR testing	Validation study	R, V, T	10.3390/virtualworlds4040046
[29]	Gentile et al. (2025) VR in speech therapy students' training (Virtual Worlds 4)	VR-based training in speech therapy edu	Scoping review	R, V	10.3390/virtualworlds4030037
[30]	Fernandez et al. (2025) Virtual acoustic environment rehearsal (Virtual Worlds 4)	VR acoustics rehearsal/performance	Experimental/applied study	V, R	10.3390/virtualworlds4030035
[31]	Tracy et al. (2025) Planogen: procedural generation framework for VR research envs (Virtual Worlds 4)	Research platform tooling	Systems/framework	T, V	10.3390/virtualworlds4030033
[32]	Bertolasi et al. (2025) HoloLens 2 hand tracking evaluation (Virtual Worlds 4)	Tracking accuracy + kinematic features	Evaluation/benchmark	T, R	10.3390/virtualworlds4030031
[33]	Gehrke et al. (2025) Neuroadaptive Haptics: RL from ratings + neural signals (arXiv)	Adaptive XR via neuro/explicit feedback	Method + experiments	T, R, V	arXiv:2504.15984

TVR-Sec layer key (used in the table), T = Technical/data/sensors/AI/ML pipelines/system design, V = Virtual layer/environment/avatars/social VR interactions/content, R = Real layer/user factors (learning, health, behavior, physiology)/real-world outcomes, Sec = Security–privacy–safety governance/threats/mitigations/controls.

Appendix B. Qualitative Coding

Table A2 illustrates representative examples of open, axial, and selective coding applied to three studies, demonstrating how threat and defense concepts were systematically mapped to the System Integrity, User Privacy, and Socio-Behavioral Safety layers of the TVR-Sec framework.

Table A2. Example of qualitative coding for TVR-Sec tri-layer structure.

Integrity	Privacy	Socio-Behavioral Safety
Paper: Ni (MobiSys 2024) [20]	Paper: Ren et al. (DPGazeSynth, Info Sciences 2024) [19]	Paper: Schulenberg et al. (CSCW 2023) [21]
Open coding (extract concepts): "motion sensor exploitation" "optical sensor leakage" "signal injection" "sensor spoofing" "mitigation strategies: validation/filtering/attestation"	Open coding: "eye-tracking data" "gaze trajectory" "differential privacy" "synthetic data generation" "utility-preserving privacy"	Open coding: "embodied harassment" "unwanted proximity" "gendered abuse patterns" "user mitigation strategies" "platform governance needs"
Axial coding Group open codes under "Device/Sensor Attack Surface" Link "signal injection" → "tracking corruption" Link "mitigation strategies" → "integrity assurance mechanisms"	Axial coding Group "gaze trajectory" + "identity inference risk" → "Biometric privacy leakage" Group "DP synthesis" + "utility preservation" → "Privacy defense strategy"	Axial coding Group harassment forms → "Immersive misconduct vectors" Group mitigation strategies → "Safety tool effectiveness" Link governance needs → "Policy-level intervention"

Table A2. *Cont.*

Integrity	Privacy	Socio-Behavioral Safety
Selective coding Dominant theme = System Integrity Layer Secondary mapping = Privacy (if optical/eye tracking leakage is central)	Selective coding Dominant theme = User Privacy Layer	Selective coding Dominant theme = Socio-Behavioral Safety Layer

References

1. Surridge, R.; Stilp, C.; Johnson, C.; Brumitt, J. The Use of Virtual Reality to Improve Gait and Balance in Patients with Parkinson's Disease: A Scoping Review. *Virtual Worlds* **2025**, *4*, 13. [CrossRef]
2. Vola, E.; Stoltz, R.; Schumpert, C.A. Impacts of Virtual Reality Experiences: Enhanced Undergraduate Student Performance and Engagement with use of 360-degree video. *Virtual Worlds* **2025**, *4*, 14. [CrossRef]
3. Conrad, M.; Kablitz, D.; Schumann, S. Learning effectiveness of immersive virtual reality in education and training: A systematic review of findings. *Comput. Educ. X Real.* **2024**, *4*, 100053. [CrossRef]
4. Gligorea, I.; Cioca, M.; Oancea, R.; Gorski, A.T.; Gorski, H.; Tudorache, P. Adaptive learning using artificial intelligence in e-learning: A literature review. *Educ. Sci.* **2023**, *13*, 1216. [CrossRef]
5. Lampropoulos, G. Combining artificial intelligence with augmented reality and virtual reality in education: Current trends and future perspectives. *Multimodal Technol. Interact.* **2025**, *9*, 11. [CrossRef]
6. El-Hajj, M. Cybersecurity and Privacy Challenges in Extended Reality: Threats, Solutions, and Risk Mitigation Strategies. *Virtual Worlds* **2025**, *4*, 1. [CrossRef]
7. Giaretta, A. Security and Privacy in Virtual Reality: A Literature Survey. *Virtual Real.* **2025**, *29*, 10. [CrossRef]
8. Nair, V.; Miller, E.; Davis, D. Unique Identification of 50,000+ Virtual Reality Users from Head & Hand Motion Data. In Proceedings of the USENIX Security Symposium, Anaheim, CA, USA, 9–11 August 2023.
9. Zhang, Y.; Meng, Y.; Wang, H. ReAvatar: Virtual Reality De-anonymization Attack Through Correlated Avatar Motion. In Proceedings of the ACM Conference on Computer and Communications Security (CCS), Virtual, 15–19 November 2021.
10. Liu, Y.; Zhou, K.; Park, S. False Reality: Uncovering Sensor-Induced Human–VR Interaction Vulnerability. *arXiv* **2025**, arXiv:2508.08043. [CrossRef]
11. Qamar, S.; Anwar, Z.; Afzal, M. A Systematic Threat Analysis and Defense Strategies for the Metaverse and Extended Reality Systems. *Comput. Secur.* **2023**, *128*, 103127. [CrossRef]
12. Trapp, H.; Chen, K.S.; Martin, L.J. Towards Understanding Harassment in Social Virtual Reality. In Proceedings of the IEEE VR Workshops, Saint Malo, France, 8–12 March 2025.
13. Rehman, S.; Patel, L.; Kumar, G. Rethinking Privacy for Avatars: Biometric and Inferred Data in the Metaverse. *Front. Virtual Real.* **2025**, *6*, 1520655. [CrossRef]
14. Clark, J.L.; Hollander, P.; Tanaka, K. Privacy Threats of Behavioral Identity Detection in VR. *Front. Virtual Real.* **2024**, *5*, 1197547. [CrossRef]
15. Rahman, R. Federated Learning: A Survey on Privacy-Preserving Collaborative Intelligence. *arXiv* **2025**, arXiv:2504.17703. [CrossRef]
16. Weerasinghe, A.; Johnson, J.D.; Fox, R.M. Beyond Mute and Block: Adoption and Effectiveness of Safety Tools in Social VR. *IEEE Trans. Vis. Comput. Graph.* **2025**, *31*, 3275–3284. [CrossRef] [PubMed]
17. *ISO/IEC 27032:2023*; Cybersecurity—Guidelines for Internet Security; International Organization for Standardization (ISO); International Electrotechnical Commission (IEC). ISO: Geneva, Switzerland, 2023.
18. Rahartomo, F.; Setiawan, I.; Nugraha, A. Metaverse Security and Privacy Research: A Systematic Review. *Comput. Secur.* **2025**, *140*, 104602. [CrossRef]
19. Ren, J.; Liu, X.; Wang, Y.; Zhang, Q. DPGazeSynth: Differentially Private Synthetic Gaze Data for Privacy-Preserving Virtual Reality Analytics. *Inf. Sci.* **2024**, *649*, 120720. [CrossRef]
20. Ni, Y.; Zhao, R.; Chen, Y.; Kotz, D. Sensor Security in Virtual Reality: Threats, Exploits, and Defenses. In Proceedings of the ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), Tokyo, Japan, 3–7 June 2024. [CrossRef]
21. Schulenberg, K.; McVeigh-Schultz, J.; Freeman, G.; DiSalvo, B. Women's Safety and Harassment in Social Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* **2023**, *7*, 1–25. [CrossRef]
22. Freeman, G.; Maloney, D.; McVeigh-Schultz, J. Harassment and Safety in Social Virtual Reality: User Experiences and Mitigation Strategies. *Proc. ACM Hum.-Comput. Interact.* **2022**, *6*, 1–25. [CrossRef]
23. Wang, W.S.; Lin, C.J.; Lee, H.Y.; Huang, Y.M.; Wu, T.T. Enhancing Self-Regulated Learning and Higher-Order Thinking Skills in Virtual Reality: The Impact of ChatGPT-Integrated Feedback Aids. *Educ. Inf. Technol.* **2025**, *30*, 19419–19445. [CrossRef]

24. Crogman, H.T.; Cano, V.D.; Pacheco, E.; Sonawane, R.B.; Boroon, R. Virtual Reality, Augmented Reality, and Mixed Reality in Experiential Learning: Transforming Educational Paradigms. *Educ. Sci.* **2025**, *15*, 303. [CrossRef]
25. Muzata, A.R.; Singh, G.; Stepanov, M.S.; Musonda, I. Immersive Learning: A Systematic Literature Review on Transforming Engineering Education Through Virtual Reality. *Virtual Worlds* **2024**, *3*, 480–505. [CrossRef]
26. Hur, J.W. Investigating Factors Influencing Preservice Teachers' Intentions to Adopt Virtual Reality: A Mixed-Methods Study. *Virtual Worlds* **2025**, *4*, 12. [CrossRef]
27. Lopes, I.; Simões de Almeida, R.; Veloso Gomes, P.; Marques, A. A Gamified Virtual Reality Exposure Therapy for Individuals with Arachnophobia: A Pilot Study. *Virtual Worlds* **2025**, *4*, 15. [CrossRef]
28. Craig, C.; Noble, E.; Parra, M.A.; Grealy, M.A. Advancing Cognitive–Motor Assessment: Reliability and Validity of Virtual Reality-Based Testing in Elite Athletes. *Virtual Worlds* **2025**, *4*, 46. [CrossRef]
29. Gentile, F.; Wanke, M.; Mueller, W.; Hochuli, E. Virtual Reality in Speech Therapy Students' Training: A Scoping Review. *Virtual Worlds* **2025**, *4*, 37. [CrossRef]
30. Fernandez, C.; Lawless, M.S.; Poirier-Quinot, D.; Katz, B.F.G. Virtual Acoustic Environment Rehearsal and Performance in an Unknown Venue. *Virtual Worlds* **2025**, *4*, 35. [CrossRef]
31. Tracy, K.; Kouzelis, L.R.; Dari, R.; Spantidi, O. Planogen: A Procedural Generation Framework for Dynamic VR Research Environments. *Virtual Worlds* **2025**, *4*, 33. [CrossRef]
32. Bertolasi, J.; Garcia-Hernandez, N.V.; Memeo, M.; Guarischi, M.; Gori, M. Evaluation of HoloLens 2 for Hand Tracking and Kinematic Features Assessment. *Virtual Worlds* **2025**, *4*, 31. [CrossRef]
33. Gehrke, L.; Koselevs, A.; Klug, M.; Gramann, K. Neuroadaptive haptics: A proof-of-concept comparing reinforcement learning from explicit ratings and neural signals for adaptive XR systems. *Front. Virtual Real.* **2025**, *6*, 1616442. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.