Scalable Secure Platform for XR

Rebecca Acheampong
Dept. of Automatic and Information Tech.
Transilvania University of Brasov
Brasov, Romania
rebecca.acheampong@unitbv.ro

Bogdan Valentin Floricescu
Dept. of Electronics and Computers
Transilvania University of Brasov
Brasov, Romania
bogdan.floricescu@student.unitbv.ro

Ionut Alexandru Oprea
Dept. of Electronics and Computers
Transilvania University of Brasov
Brasov, Romania
ionut-alexandru.oprea@student.unitbv.ro

Alexandre Rekeraho
Dept. of Electronics and Computers
Transilvania University of Brasov
Brasov, Romania
alexandre.rekeraho@unitbv.ro

Vladut Gabriel Anghel
Dept. of Electronics and Computers
Transilvania University of Brasov
Brasov, Romania
vladut.anghel@student.unitbv.ro

Gabriel Danciu
Dept. of Electronics and Computers
Transilvania University of Brasov
Brasov, Romania
gabriel.danciu@unitbv.ro

Ioana Corina Bogdan
Dept. of Electronics and Computers
Transilvania University of Brasov
Brasov, Romania
corina.bogdan@unitbv.ro

George Stefan Ionescu
Dept. of Electronics and Computers
Transilvania University of Brasov
Brasov, Romania
stefan.ionescu@student.unitbv.ro

Abstract*—Today, interest in real-time immersive interaction in training and remote collaboration applications requires scalable and secure platforms. In this paper, a distributed microservices-based architecture is presented to support high-performance and low-latency interactions over 5G networks. The system integrates essential key technologies such as Nakama for secure authentication, MageAI pipelines for efficient data processing, and WebSocket-based communication for real-time interactivity. The platform is designed to support high availability and scalability, based on containerized deployments, load balancing, and fault-tolerant mechanisms. In addition, security aspects are integrated at multiple levels, ensuring protected user authentication, ensuring a balance between usability and security, enabling real-time user engagements while maintaining a robust security posture. Furthermore, by embedding DevSecOps principles into the platform's development lifecycle through Kubescape, we have established a security-first approach—ensuring that security is proactively integrated rather than reactively mitigated. While existing literature primarily focuses on security methods such as encryption, our approach not only prioritizes security monitoring and proactive risk assessment, but also introduces adaptive learning capabilities through MageAI, fostering a user-friendly and dynamically secured environment that evolves alongside emerging threats. Although the system is designed and structured to support immersive applications such as VR training environments, its flexible design enables adaptation to various interactive real-time scenarios. This work contributes to a modular, secure, and high-performance infrastructure, establishing a strong foundation for seamless integration into next-generation immersive applications. Future work will focus on integrating the Virtual Reality Training application into the platform architecture to validate the capabilities of the system.

***Index Terms*—Virtual Reality, Augmented Reality, XR, Scalability, Security.**

I. Introduction

Immersive technologies seamlessly merge the real and virtual worlds, creating remarkable, realistic and interactive experiences that are revolutionizing education, healthcare, industry, and defense [1] with a commitment to the progress of society. As a fundamental component of digital transformation, these technologies enable businesses, to improve user engagement by increasing operational efficiency and safety [2], driving innovation in their products and services. Essential to these technologies' success is the integration of solutions capable of transmitting and processing data in real time, ensuring interactions such as rendering virtual reality (VR) environments, displaying live augmented reality (AR) annotations, or enabling synchronous multiplayer experiences. However, obtaining this level of interactivity at scale requires overcoming considerable technical challenges, including synchronization, security, and scalability. The available platforms are still struggling to balance these mandatory requirements, leading to compromised user experiences and multiple unexpected security vulnerabilities [3].

High latency can interrupt the flow of user experiences, leading to detrimental effects. For example, lag in headsets can cause motion sickness, reducing the effectiveness of training simulations [4], or even creating vulnerabilities related to security, leading attackers to intercept or exploit delayed data transmission. Thus, maintaining low latency minimizes the delay between data transmission and its processing or display. Maintaining low latency is a critical step in ensuring that all data are transmitted, processed, and displayed with minimal delay, preserving the integrity and quality of immersive interactions. To achieve this, the core infrastructure must utilize

effective communication protocols, integrate strong security measures that do not compromise safety, and have network architectures capable of meeting rigorous latency requirements. These solutions ensure safety and real-time interactivity, which are essential for delivering immersive experiences at scale.

As presented, industry trends highlight the growing adoption of immersive technologies in diverse fields. According to Gartner's 2024 research, about 70% of the top ten Fortune 500 firms are expected to integrate augmented connected worker strategies by 2027. Furthermore, the report predicts that by 2028 45% of organizations with more than 500 employees will use AI-powered avatars to enhance human capacity [5]. These anticipations highlight the increasing reliance on immersive applications, leading to a critical demand for high-speed, high-bandwidth networks capable of supporting real-time interactions and expansion.

As the number of users and interactions grows, the supporting infrastructure must also be designed to scale dynamically. Scalability is a basis for ensuring that immersive platforms can handle increasing workloads without compromising performance [6]. Horizontal scaling, which involves distributing workloads across multiple servers or regions, and vertical scaling, which focuses on enhancing the capacity of individual servers, are both essential strategies. By leveraging these approaches, the infrastructure can dynamically allocate resources to meet demand, ensuring consistent performance even during peak usage. In other words, this scalability is crucial for delivering seamless, real-time experiences to a growing global user base, making it a foundational requirement for next-generation immersive platforms.

In response to these challenges and opportunities, our study proposes a scalable and secure microservices-based system architecture to meet the growing demand for low-latency real-time immersive interactions. The platform leverages modern and cutting-edge technologies such as 5G networks, containerization, data pipelines, and WebSocket-based real-time communication. The platform integrates security components to ensure security is maintained throughout the development lifecycle of real-time immersive applications. Given this, the platform ensures high availability, robust security, and efficient data processing. This makes it ideally suited for next-generation applications in VR, AR, and real-time collaboration, addressing the needs of businesses and users alike in an increasingly immersive digital landscape.

The study is organized as follows: Section 2 presents research and relevant studies on immersive platform architecture. In Section 3, the architecture of the proposed system is presented. Section 4 discusses the security integration into the platform architecture, and Section 5 concludes the study.

II. Extended Reality Related Studies

Several studies investigated the design of real-time interaction platforms, concentrating on aspects related to scalability, communication efficiency, and security. This section reviews relevant works and essential details about their strengths and limitations.

Some existing immersive application back-end architectures typically rely on monolithic or client-server models, which can lead to scalability bottlenecks and increased latency when handling large-scale interactions [7]. The increasing user demand for real-time interactivity reveals the limitations of monolithic architectures, decreasing their effectiveness for highly dynamic and resource-intensive applications [8]. To overcome these issues, research has increasingly focused on cloud-based architectures that offer high computing power and scalable storage. Through these kinds of architectures, XR platforms can improve Quality of Service (QoS) and optimize computational resource allocation for immersive environments [9]. However, the remote location of cloud servers can introduce high latencies, reducing the effectiveness of real-time rendering and interactions. Although cloud solutions efficiently manage large-scale data processing and storage, they do not always meet the stringent Quality of Experience (QoE) demands of immersive applications.

An alternative to these issues involves combining cloud and edge computing to enhance low-latency rendering and real-time data processing [10]. This hybrid model leverages edge servers closer to end-users to reduce latency. However, prior studies have not integrated dynamic load balancing and backup recovery systems, which are critical for ensuring system availability and reliability under varying workloads. Immersive interactions require the smooth transmission of human behaviors, rendered virtual scenes, and interactive data between various components and end users. However, the diverse nature and high-bandwidth demands of the immersive ecosystem can create significant communication and computation overheads. These overheads introduce challenges in meeting the low latency and high bandwidth requirements of interactive platforms [11]. To address these issues, researchers have proposed 6G network technology for faster data processing and ultra-low latency in real-time interactive environments [11]. While future 6G may reach well above a frequency of 100 GHz, compared to 5G's range of 24 GHz to 100 GHz, it requires advanced hardware, complex infrastructure, and costly deployment [12]. The practical availability of 6G technology remains limited, making 5G the more viable option for current real-time immersive platforms.

The security factors encountered in virtual reality and immersive back-ends are often treated as secondary concerns, leading to vulnerabilities in data transmission, authentication, and access control mechanisms. With the increasing integration and interest of immersive applications in the enterprise and consumer domains, protecting sensitive user data is becoming a critical priority [13]. Data encryption reflects a major concern to prevent unauthorized access and interception of real-time interactions [14]. Authentication and access control mechanisms are required to securely verify a user and to restrict unauthorized access to sensitive data. Although some existing platforms integrate basic security measures, most of them do not include essential aspects related to end-to-end encryption, secure authentication protocols, immersive infrastructure monitoring, or resilient access control mechanisms.

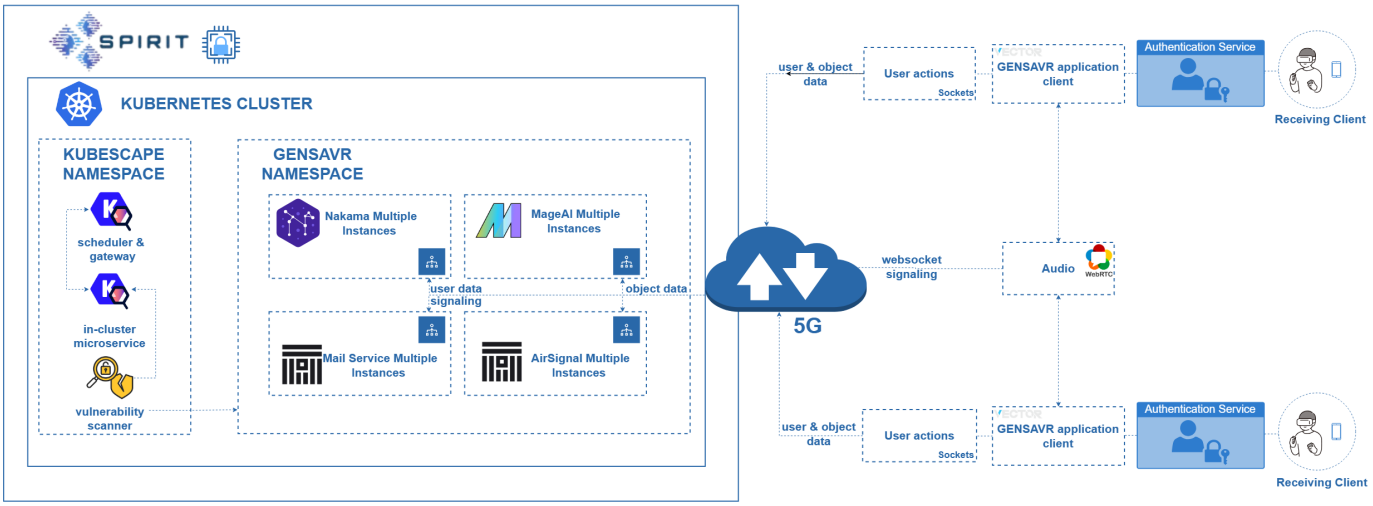

Fig. 1. Proposed Architecture for scalable and secure immersive platform.

However, knowing that immersive applications continue to expand rapidly in enterprises and training sectors, future research must prioritize the integration of security-enhanced architectures that maintain real-time performance without introducing additional latency.

III. XR System Architecture

This section describes the XR platform structure taking into consideration aspects related to general architecture, components of the platform, MageAI integration used for data real-time adaptive processing, and the simultaneous communications protocol that handles the data exchange for a smooth experience without latencies.

A. Architecture Overview

The proposed platform adopts a microservices-based architecture to ensure scalability, security, and modularity in real-time immersive applications. Unlike monolithic architectures, where all functionalities are tightly coupled [15], microservices break the system into independent services, each responsible for a specific function [16]. This modular approach allows dynamic scaling, independent updates, and improved fault isolation, making it ideal for real-time applications requiring high availability. Each microservice in the platform is designed to operate independently while interacting through well-defined APIs and message queues, ensuring seamless communication. The system is composed of the following core services:

- Authentication and User Management (Nakama Service).
- Data Processing Pipeline (MageAI Service).
- Real-Time Communication (AirSignal WebSocket Service).
- Email and Notification Service (MailService).
- Load Balancer and API Gateway.

A proposed architecture for the scalable and secure immersive platform is given in Fig. 1.

Explanations on the platform architecture and details are given in the next part of this work (Components of the Platform).

B. Components of the Platform

1) Containerization and Orchestration for Dynamic User Load Management: The platform leverages containerization with Docker and orchestration with Kubernetes (K8s) to ensure efficient deployment, scalability, and fault tolerance. These technologies provide a modular, resilient, and high-performance infrastructure capable of dynamically adjusting resources to handle fluctuating user loads in real-time immersive environments.

a) Containerization with Docker: Docker is used to encapsulate application components into lightweight self-contained containers that package dependencies, configurations, and runtime environments. This ensures that services run consistently across development, testing, and production environments. The key benefits of using Docker for the foundational components of the platform are portability and consistency, resource efficiency utilization, security isolation, and rapid deployment [17]. Each core microservice, such as Nakama, MageAI, and the rest, is containerized using Docker, allowing for multi-container applications with defined networks, scalability, and isolated instances for load balancing.

b) Orchestration with Kubernetes (K8s): While Docker provides containerization, Kubernetes (K8s) orchestrates the deployment and scaling of these containers across a cluster of servers, ensuring high availability and efficient resource allocation [18]. Our platform leverages Kubernetes for auto-scaling, load balancing, updates, rolling updates, and rollbacks.

2) Nakama: Shared experiences among users remain the primary objective in real-time immersive environments. Nakama, well known for its low latency performance and high throughput capabilities allowing for hundreds of thousands of concurrent players in real-time multiplayer games [19] becomes a choice. Given Nakama's comprehensive backend

solution for real-time multiplayer game development and the available tools for real-time networking, matchmaking, leaderboards, data storage, and more, it provides quick data synchronization for immersive environments requiring persistent backend features [14]. The platform employs Nakama for a real-time backend solution for handling multi-user interactions, matchmaking, persistent user data, and managing user account security.

C. MageAI integration for adaptive real-time data processing

To enhance the real-time immersive interaction, the platform integrates MageAI, data pipeline framework that processes, transforms, and optimizes data in real time. MageAI is an open-source tool designed to simplify the building, deployment, and management of data pipelines for machine learning and analytics. Also, MageAI's integration allows adaptive data processing, ensuring that immersive applications will dynamically respond to user interactions under environmental conditions.

The flow in Fig. 2 uses Mage AI to adapt learning experiences in a VR environment based on user interactions. The process starts with data collection, where user input is gathered and processed through pipeline selection. The chosen pipeline undergoes transformation and optimization to refine the learning content. The optimized data is then rendered and integrated into the environment, ensuring a responsive and personalized experience.

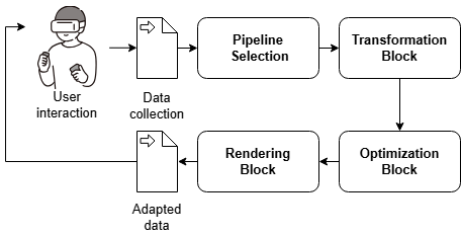

Fig. 2. Adaptive Learning Flow in VR Using MageAI.

D. Real-time communication protocol

To enable smooth real-time interactions within the platform, multiple communication protocols are integrated, including WebSockets and WebRTC-based peer-to-peer communication. These protocols work together to ensure low latency, high reliability, and secure data transmission across immersive applications such as VR training, industrial simulations, and remote collaboration [9].

IV. Security Integration

This section discusses the integration of security mechanisms into the platform's architecture. The integration of the security solutions strengthens the architecture's framework, ensuring a modular, scalable, and resilient security model. These enhancements significantly improve the platform's security posture, making it more resistant to threats while maintaining optimal performance and usability. The implemented solutions focus on three key areas:

- Multi-Layered Authentication System for secure access control to verify user identities and prevent unauthorized usage.
- Real-time surveillance, logging mechanisms, and standardized compliance measures that work together to detect anomalies and strengthen system resilience.

A. Authentication with Nakama

Built on Nakama's robust backend framework, the authentication mechanism employs a multi-layered approach, integrating traditional login methods, email/password authentication, and device-based authentication to ensure session continuity across devices. Also, the implementation of advanced token management is included, which utilizes short-lived tokens and automatic refresh mechanisms for enhanced security. The authentication system consists of six key scenes, each one designed to facilitate different aspects of user authentication and session management. These scenes and their workflow are illustrated in Fig. 3.

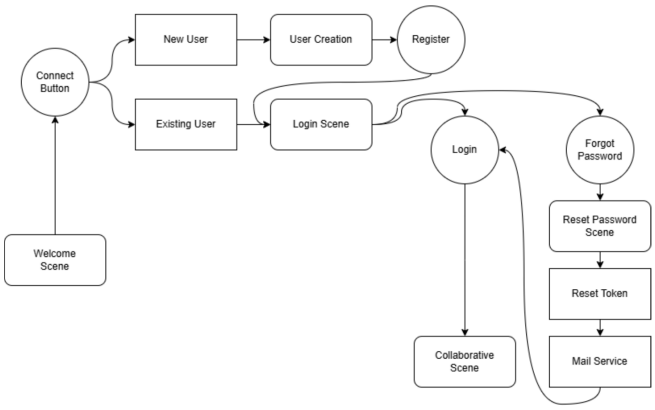

Fig. 3. Overview of Authentication System Operation and Design.

B. Authentication and Session Management

1) Email Authentication: The main authentication mechanisms implemented to authenticate users are email authentication and device-based authentication. The traditional email and password authentication ensures a secure and user-friendly login experience. Given the importance of protecting user credentials, Nakama's built-in security mechanisms were leveraged to hash and salt passwords before storing them, preventing plaintext storage, and also to store user profiles in the backend, allowing centralized user account management. For a user to create an account, Fig. 4 illustrates the process. An existing user clicking on the Existing User button is directed to the login scene. While the new user clicking on the New User button is directed to the user registration scene. Fig. 5 demonstrates Nakama's backend user management of users who have registered.

2) Session Management: During this work, we implemented multiple authentication mechanisms, each performing a particular role in authentication and session man-

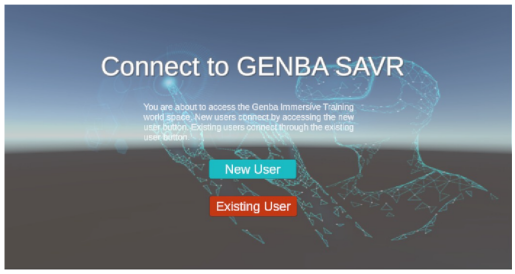

Fig. 4. Scene Connecting Users to their Accounts.

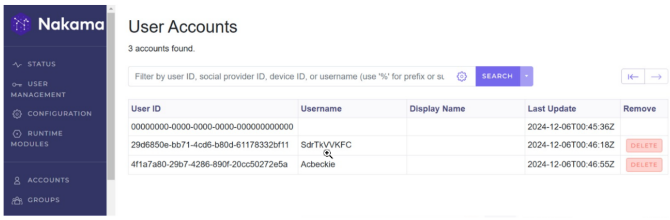

Fig. 5. Nakama's Dashboard Monitoring of Users Accounts.

agement. Since authentication needs to persist across multiple scenes, a Singleton pattern is used to manage the Nakama session. The singleton pattern is implemented with a combination of device-based authentication using `client.AuthenticateDeviceAsync(deviceId)`, a short-lived token, and automatic session refresh to ensure that users maintain an active session without frequent re-authentication. Also, the system continuously monitors the session state. Fig. 6 demonstrates the singleton session that is maintained throughout the lifecycle of immersive interaction.

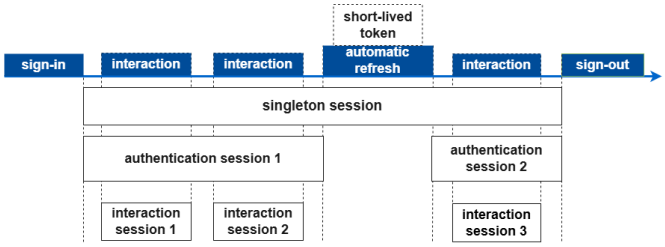

Fig. 6. Singleton Session.

C. Infrastructure Monitoring and Management

Kubescape is an open-source Kubernetes security framework designed to enhance security visibility and compliance across containerized applications. The integration of Kubescape into the platform was vital to safeguard the infrastructure. It provides actionable insights into vulnerabilities, misconfigurations, and also ensures compliance with security standards. Kubescape's ability to proactively identify and mitigate potential security threats allows the platform to maintain a strong security posture throughout its Kubernetes-based deployment. Actually, it acts as a first line of defense, continuously scanning and hardening the infrastructure against potential attacks. Moreover, Kubescape uses a vulnerability scanner to scan the Kubernetes cluster for gaps. In Fig. 7 is shown the Kubescape workflow.

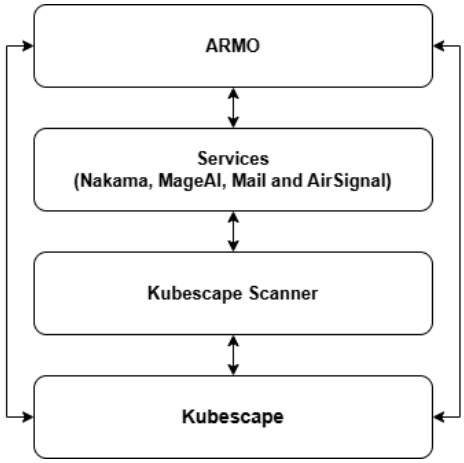

Fig. 7. Kubescape Monitoring Workflow.

To complete Kubescape, ARMO provides an interactive security dashboard that offers real-time insights into security vulnerabilities within the Kubernetes cluster. By visualizing detected risks and misconfigurations, the ARMO dashboard enhances security monitoring and incident response, allowing administrators to take proactive measures efficiently.

1) Kubescape Security Testing: Leveraging on the Kubescape scanner and the ARMO platform, vulnerability tests were performed to expose vulnerable access to the systems. The purpose of this assessment was to discover and identify vulnerabilities in infrastructure and to introduce DevSecOps into immersive platform operations and development, embedding security as a fundamental component of the workflow. During the tests, the results show that, while there are a few critical and high-risk concerns, the majority of issues are medium-risk, which still require significant attention for maintaining system security and integrity; this is illustrated in Fig. 8. In terms of risk categories, the largest number of issues is related to Workloads (17 issues), reflecting potential vulnerabilities in application deployments and resource management. Both the attack path and RBAC configuration contribute to six risks each, primarily emphasizing threats associated with unauthorized access and privilege escalation.

These categories collectively emphasize the need for robust workload management, access controls, secure network configurations, and safeguarding sensitive data to minimize potential attack surfaces. External-facing workloads often contain critical security flaws that can be exploited by attackers to gain unauthorized access, while weak access controls and misconfigurations frequently lead to privilege escalation and data exposure. Unencrypted storage and improper handling of sensitive credentials remain widespread issues, increasing the likelihood of data breaches.

These threats highlight the need for robust security measures, including strict access controls and secure workload

Fig. 8. Armo Dashboard Displying the Security Scan Result of the infrastructure.

configurations. In addition, enforcing the principles of least privilege and implementing strong authentication mechanisms are essential to reducing the attack surfaces.

Together, Kubescape and ARMO ensure that the immersive platform operates in a secure and compliant Kubernetes environment, mitigating potential security risks at every stage of the application lifecycle.

V. Conclusions

By adopting a containerized architecture, our scalable immersive platform achieves greater flexibility, simplified management, and optimized resource utilization, making it well-suited for real-time immersive applications. This infrastructure integrates modern technologies, including containerization, orchestration, MageAI pipelines, high-speed networking, and scalable backend systems, to support seamless and responsive experiences. However, such an infrastructure is inherently prone to security vulnerabilities, making security a critical aspect that cannot be overlooked. To address these challenges, we implemented a comprehensive, multi-layered security approach, incorporating strong authentication mechanisms, real-time backend security, infrastructure monitoring, and proactive vulnerability management. Our session management framework ensures a balance between usability and security, enabling real-time user engagements while maintaining a robust security posture. Furthermore, by embedding DevSecOps principles into the platform's development lifecycle through Kubescape, we have established a security-first approach, ensuring that security is proactively integrated rather than reactively mitigated. While existing literature primarily focuses on security methods such as encryption, it often lacks emphasis on architecture monitoring within immersive platform development lifecycles. Our approach not only prioritizes security monitoring and proactive risk assessment but also introduces adaptive learning capabilities through MageAI, fostering a user-friendly and dynamically secured environment that evolves alongside emerging threats.

Acknowledgment

This work was funded by the EU Horizon Europe SPIRIT project ("Scalable Platform for Innovations on Real-time Immersive Telepresence") under the grant number 101070672.

References

[1] N. Chhabhaiya et al., Virtual and Augmented Reality Applications: A Broader Perspective Review, Journal of Data Science and Intelligent Systems, 2024, Vol. 00(00) 1–12, DOI:10.47852/bonviewJDSIS32021064.

[2] G. Lampropoulos et al., Examining the Role of Augmented Reality and Virtual Reality in Safety Training. Electronics 2024, 13, 3952, 2024, https://doi.org/10.3390/electronics13193952.

[3] A. Alismail et al., A Systematic Literature Review on Cybersecurity Threats of Virtual Reality (VR) and Augmented Reality (AR), In: Jacob, I.J., Kolandapalayam Shanmugam, S., Izonin, I. (eds), Data Intelligence and Cognitive Informatics, Algorithms for Intelligent Systems, Springer, 2023.

[4] E. Chang et al., Virtual Reality Sickness: A Review of Causes and Measurements, International Journal of Human–Computer Interaction, 36(17), 1658–1682, 2020, https://doi.org/10.1080/10447318.2020.1778351.

[5] "Hype Cycle for the Future of Work, 2024." Accessed: Feb. 03, 2025. [Online]. Available: https://www.gartner.com/en/documents/5598759.

[6] J. Liu, Parallel and Distributed Immersive Real-Time Simulation of Large-Scale Networks, in Parallel and Distributed Computing, A. Ros, Ed., InTech, 2010, doi: 10.5772/9453.

[7] J. Chen, F. Qian, and B. Li, An Interactive and Immersive Remote Education Platform based on Commodity Devices, in IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 1–2, 2021, doi: 10.1109/INFOCOMWKSHPS51825.2021.9484492.

[8] Microservices vs. monolithic architecture — Atlassian. Accessed: Feb. 04, 2025. [Online]. Available: https://www.atlassian.com/microservices/microservices-architecture/microservices-vs-monolith.

[9] L. Bao, S. V.-T. Tran, T. L. Nguyen, H. C. Pham, D. Lee, and C. Park, Cross-platform virtual reality for real-time construction safety training using immersive web and industry foundation classes, Autom. Constr., vol. 143, p. 104565, Nov. 2022, doi: 10.1016/j.autcon.2022.104565.

[10] L. Ismail and R. Buyya, Metaverse: A Vision, Architectural Elements, and Future Directions for Scalable and Realtime Virtual Worlds, Aug. 24, 2023, arXiv: arXiv:2308.10559. doi: 10.48550/arXiv.2308.10559.

[11] S. Doroudian, Collaboration in Immersive Environments: Challenges and Solutions, Jan. 16, 2025, arXiv: arXiv:2311.00689. doi: 10.48550/arXiv.2311.00689.

[12] 5G vs. 6G: The main advantages and drawbacks. Accessed: Feb. 13, 2025. [Online]. Available: https://www.lifecycle-software.com/resources/5g-vs-6g-main-advantages-drawbacks.

[13] S. Doroudian, Collaboration in Immersive Environments: Challenges and Solutions, Jan. 16, 2025, arXiv: arXiv:2311.00689. doi: 10.48550/arXiv.2311.00689.

[14] R. Kumar Yekollu et al., Securing the Virtual Realm: Strategies for Cybersecurity in Augmented Reality (AR) and Virtual Reality (VR) Applications, in 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Oct. 2024, pp. 520–526. doi: 10.1109/I-SMAC61858.2024.10714591.

[15] L. De Lauretis, From Monolithic Architecture to Microservices Architecture, in 2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), Oct. 2019, pp. 93–96. doi: 10.1109/ISSREW.2019.00050.

[16] S. Li et al., Understanding and addressing quality attributes of microservices architecture: A Systematic literature review, Inf. Softw. Technol., vol. 131, p. 106449, Mar. 2021, doi: 10.1016/j.infsof.2020.106449.

[17] M., Sobieraj and D., Kotyński, Docker Performance Evaluation across Operating Systems. Appl. Sci. 2024, 14, 6672. https://doi.org/10.3390/app14156672.

[18] V. R. S. B., Soma, Container Orchestration with Kubernetes, Journal of Artificial Intelligence Machine Learning and Data Science, 2024.

[19] Benchmarks - Heroic Labs Documentation. Accessed: Feb. 21, 2025. [Online]. Available: https://heroiclabs.com/docs/nakama/getting-started/benchmarks/.

[20] K. Lakkaraju, J. Whetzel, and N. Fabian, Experiences with SIGNAL: Opportunities and challenges with distributed wargaming, SAND2021-9920R, 1821850, 699880, Aug. 2021. doi: 10.2172/1821850.