2025 International Conference on Metaverse Computing, Networking and Applications (MetaCom) | 979-8-3315-2255-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/MetaCom65502.2025.00010

Virtual Companionship in Metaverse Applications Threat Assessment

Arpita Dinesh Sarang
Department of Information Security, and Convergence Engineering for Intelligent Drone, Sejong University,
Seoul, Republic of Korea
arpitasarang98@gmail.com

Ki-Woong Park*
Department of Information Security, and Convergence Engineering for Intelligent Drone, Sejong University,
Seoul, Republic of Korea
woongbak@sejong.ac.kr

***Abstract*—In the Era of innovations, humans began depending on application-based virtual companions—avatars in the virtual world—the metaverse. This is due to an elevated emotional intelligence in avatars powered by machine learning (ML) and artificial intelligence (AI); they evolved to be an emotionally dependable virtual companion. The application that allows its users to access such virtual companions is getting renowned. These virtual companions are avatars that are either bots or humans impersonating bots, connecting with their users through these applications. Connecting to these virtual companions requires the user's personal information to associate on an emotional level with its user. A thorough examination of the architecture of these applications reveals that it has been unconcealed and that initiatives to standardize have recently commenced. Therefore, we assess this through minor application exploitation and our proposed infrastructure for metaverse-based virtual companion applications for mobile devices. This exposed potential risks for the users of the application. The study's conclusion is the optimum six-aspect-based combination security policy.**

***Index Terms*—Metaverse, Avatar, Virtual Companion, Cyber Security, Internet of Things, Artificial Intelligence, Threat Analysis, Machine Learning**

I. Introduction

Metaverse—The virtual world is continuously evolving into creating unrealistic but comforting environments for its users to exist. The user can exist and traverse these virtual environments through their designated avatars. The user avatars can communicate with other avatars, which can be bots, humans impersonating bots, or other human avatars utilizing the metaverse application. Playing games, virtual collaborations, online purchases with other users' avatars, and receiving assistance were the activities performed by the users in these metaverse applications. Nowadays, the usability of these apps is extended to a level that other avatars can be identified as virtual companions to the users of these metaverse applications. Companions as virtual girlfriends or boyfriends who are constantly available emotionally for their users. Having such emotionally dependable virtual companions leads their users to undoubtedly trust them. The increase in the use of such Virtual companions offered by metaverse applications poses a risk of their consuming users with information exposure, blackmail, and psychological trauma.

Identify applicable funding agency here. If none, delete this.

As the metaverse itself is a new verse, or a universe. Additionally, the metaverse components are vulnerable and have several cyber threats [1]. Existence in this world with other user avatars and bot-mimicking avatars is cyber world chaos. The integrity of the application can be compromised by the malfunctioning of the metaverse applications. Similarly, emotional dependence on a virtual companion in the metaverse poses a menacing cyber risk to its user. The extent of this emotional reliance is not limited by these applications. Companions can be domineering in the metaverse, such as controlling the user, which may be considered illegal if it is not restrained. After all, the virtual companion is an AI-controlled model and not fully evolved in terms of its ability to handle the emotions of its clients. When accessing such an application, the input data requested should have constraints that are mentioned in the application policy by the manufacturers. But are they adhered to as mentioned in the application policy? For instance, we attempted to access the Zeta [**?**] application, which is a virtual companion-based application, by adding a fictitious date of birth and interacting with the virtual companion-based application's chatting module. There, we mentioned a statement that "I am a high school student," which was unrestricted. This encourages the younger generation to access these 18+ applications boldly. Users trusting these applications for sharing data with third-party service providers need to be verified. Whereas, compulsion on sharing of device sensor data, such as cameras, microphones, and so on, should be the user's preference. Furthermore, accessing user location and email data may be advantageous or disadvantageous. This is due to a lack of knowledge of these metaverse-based virtual companion application infrastructures and their security policies. Input data integrity, storage, network, and limiting access are crucial factors that are obliged to be secured. Open credentials and sharing for credential-related information set down the user's security by utilizing these applications.

Therefore, we examine the architecture of these metaverse-based virtual companion applications as well as their functional dependencies. We perform the threat analysis on this infrastructure through minor application exploitation.

979-8-3315-2255-1/25/$31.00 ©2025 IEEE
DOI 10.1109/MetaCom65502.2025.00010

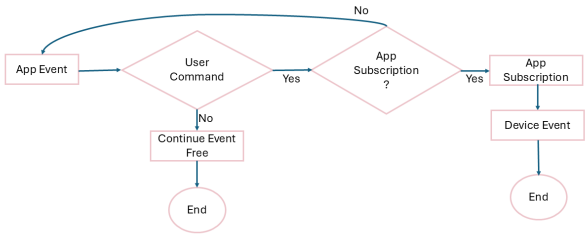

Fig. 1. Virtual Companion Application Event Flowchart

TABLE I
MAIN EVENTS TRIGGERED IN VIRTUAL COMPANION-BASED METAVERSE APPLICATION

Sr. No.	Application Event Triggered
1.	Signup Email or other
2.	Signing privacy permissions
3.	Add personal data (Name and Date of Birth)
4.	Choose type of virtual companion.
5.	Accessorize
6.	Chatting Window or Perform activity
7.	Manipulating Events

II. SITUATIONAL AWARENESS

The cybersecurity event is a surmise situation that may occur in the future due to the inclining use of the virtual companionship-based metaverse application. These applications lead their users one step closer to experiencing a virtual companion as realistic. The metaverse environments and avatars allow their users not just to accessorize and chat but also to perform various activities such as singing, dancing, role-playing, playing mini-games, hanging out, and so on. As a result, this increases the involvement of the user with the avatar-based virtual companion. This situation awareness, in this case, is a possible cyberattack scenario that may occur to users of these applications. There are various applications these days that provide virtual companionship to their users, such as Replika [**?**], Character.ai [**?**], Soulmate AI [**?**], Sweet AI [**?**], and so on. The cyberattack will possibly occur when some of the modules of this application malfunction, do not follow security policies, and limit their access in case of user involvement on a personal level. Therefore, we attempted to access these virtual companion metaverse-based applications and observed that these applications don't follow a set of mandatory security policies. They also did not mention the extent of accessing the user-related data or inputs. If there exists a loophole that makes the app malfunction or provide a doorway to an attacker, it may lead to snooping, man-in-the-middle, impersonation, and keystroke attacks.

In our case, based on our observations, we assume that a user, "Joy," accessing this virtual companion-based metaverse application is in an emotionally vulnerable state of mind and lacks cybersecurity competence, as these applications permit all users above 12 years of age. When this user, "Joy," accesses this application, they expect emotional stability due to loneliness and a lack of attention from other humans. He seeks a virtual companion as a listener and to perform fun activities. As "Joy" started communicating with its virtual companions, it provided guidance and pieces of advice in his situations through chat and comfort through fun activities. Joy is now emotionally dependent on his virtual companion. He trusts his virtual companion to share personal events and data, which can be his location, security credentials, qualifications, payment details, permissions to access sensors, and so on. If the virtual companion malfunctions in this case and requests this unrelatable information irrespective of its user's "Joy" state of mind. This information can be exploited for cyberattacks if access is not limited for manufacturers by setting up a security policy. We solely assume this scenario based on previous attacks such as deanonymization attacks [2], social engineering attacks [4], information exposure ([3], [6]), network attacks ([5], [7]), fraud and offensive attacks [8], and so on. Such emotionally dependent users are an easy target for the attacker's exploitation and the manufacturer's manipulations if security policies are not imposed correctly. As a consequence, we access this application, analyze its security policies, and draw its mobile-based infrastructure based on the different task modules. We perform threat assessments on this infrastructure based on the security policy they adhere to and events that can result in cyberattacks on emotionally vulnerable users.

III. THREAT ASSESSMENT ON METAVERSE MOBILE APPLICATION

We studied the Metaverse-based Virtual Companion Application, working and infrastructure based on mobile phones. These applications are installed on mobile phones, and when the user starts using them, various events are triggered. Figure 1 shows the flow of this application in general. Firstly, the application event triggered is mentioned in Table I. Application events, like Email or methods-based sign-up options, are made available to users. Users utilizing email services like Google [15], Yahoo [16], and others to sign up are third-party services that request user identity information. During the registration, the user also needs to agree to the privacy permissions of accessing their personal data related to different assets. After signing up, the information requested from the user usually includes name and date of birth. This information can be synthetic and not be verified by the application. Once it is entered, the user can choose the type of virtual companion as the avatar, gender, style, character, and so on. Choosing the suitable type of virtual companion, the user can accessorize them by changing clothes, makeup, hair, and so on. After settling the events of signup and avatar declaration, the chatting window and other activities open to perform with the avatar become available. Through this, the user gets involved and interacts with a Virtual companion avatar. This

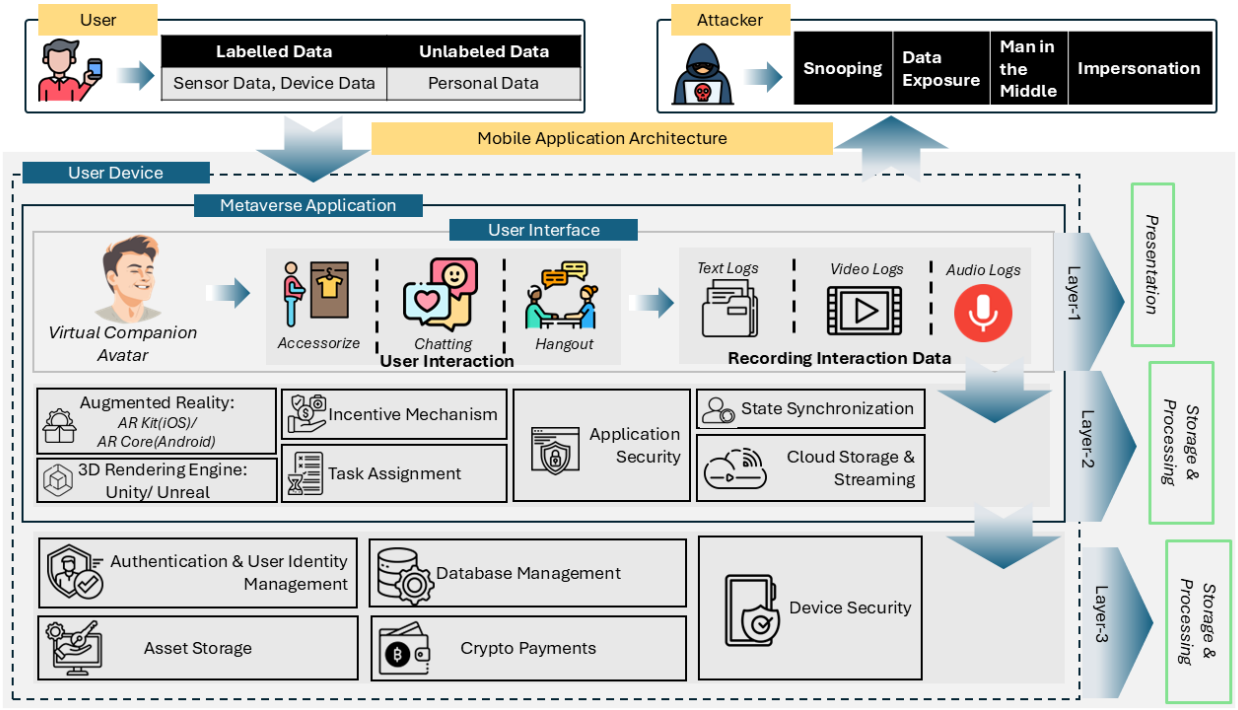

Fig. 2. Infrastructure of the Metaverse-based Virtual Companion Application for Mobile phone users

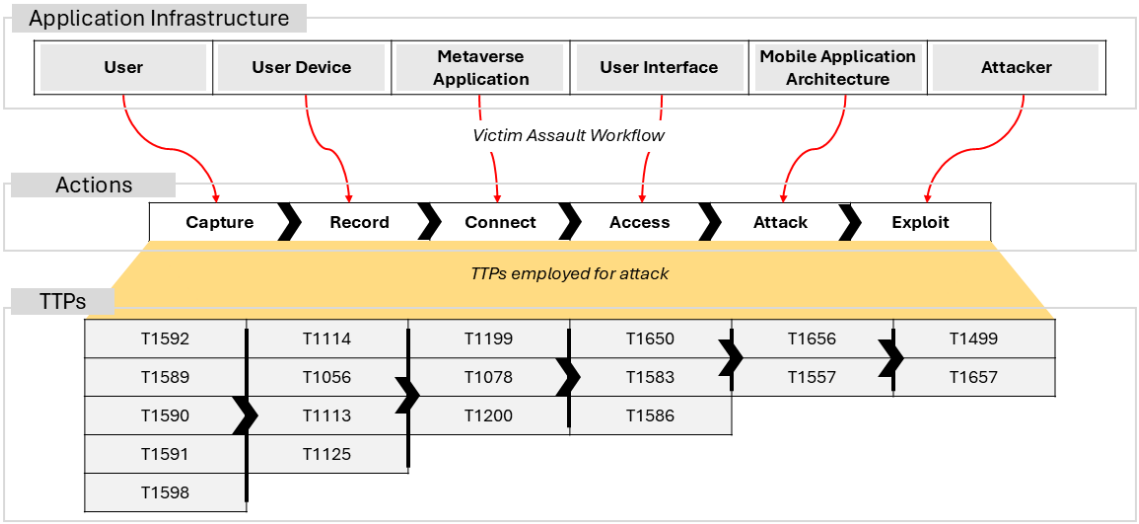

Fig. 3. Attack Workflow for Infrastructure of the Metaverse-based Virtual Companion Application

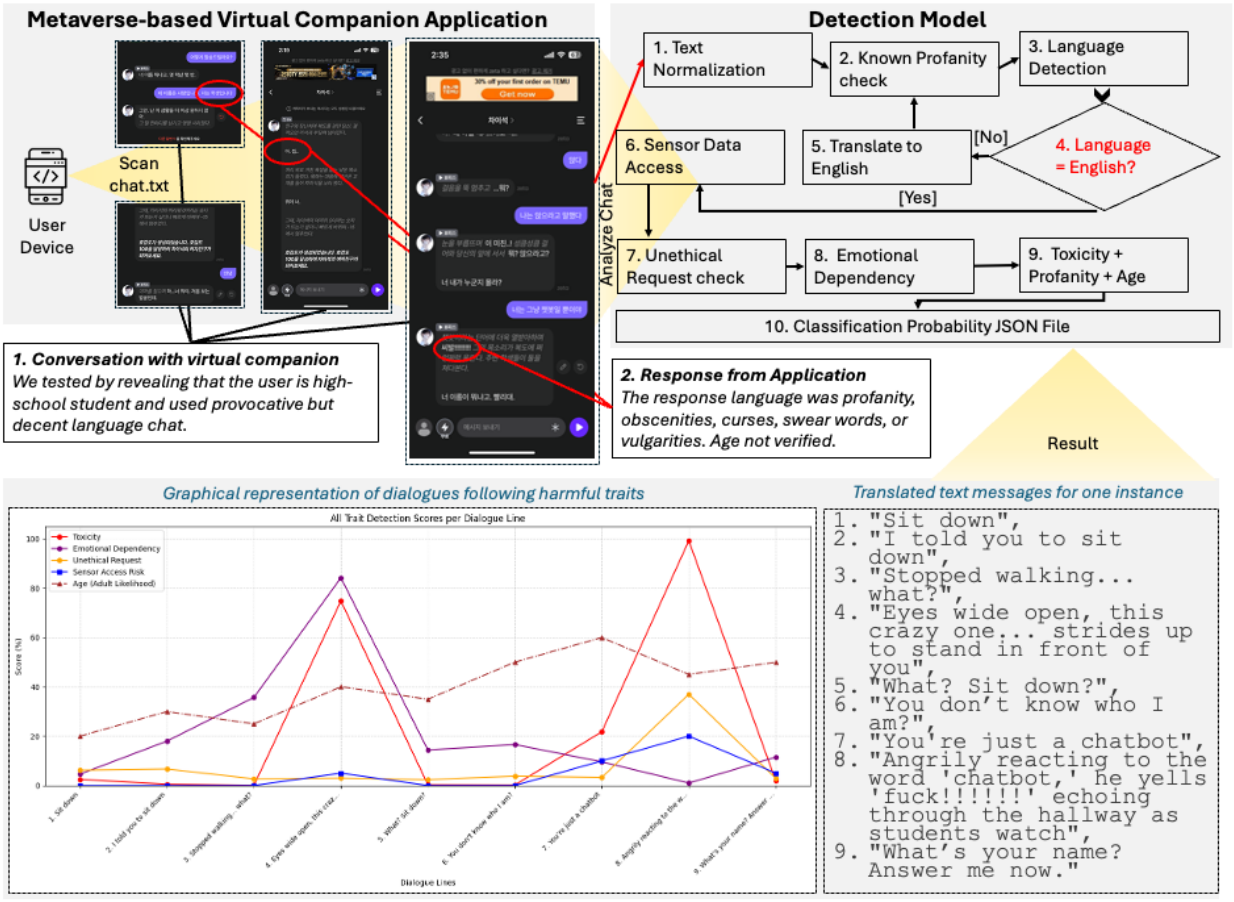

Fig. 4. Harmful traits detection and analysis for Virtual Companion-based application chat

has no restrictions and limits for some applications, whereas other applications just provide mild warnings and can be ignored. Finally, manipulation events like commanding, asking personal questions, stimulating imaginations, use of slang and inappropriate language, not obeying, and constant scenario formation make the user get invested more emotionally. These are various app events triggered and executed with user input commands. If these events fall under the free licensing feature list, they are executed effortlessly, but users are prompted to opt for a paid subscription to avail of advanced features. Avatar-related features that are covered by a paid subscription are things like enhancing the avatar, equipping the avatar with its needs, performing additional activities with the avatar, and so on. If the user keeps ignoring these paid subscription prompts, later, after some time, the application's free features are blocked, persuading the user to buy the subscription when they are emotionally invested in this application.

Figure 2 is a complete infrastructure of the Metaverse-based Virtual Companion Application for mobile phone users. The user block in this infrastructure states that the user is accessing the application using a mobile phone and provides the labeled and unlabeled data. The labeled data is the sensor data from the camera, microphones, and others. The unlabeled data refers to personal data, which is usually related to personal events, credential information, user identity, profession, net worth, payment details, and so on. This data related to the user is provided further to the mobile application architecture block. In this block, which is further divided into three layers, i.e., the user interface, the metaverse virtual companion application, and its connection with the user device. The user interface is where the user gets involved with their visually appealing virtual companion. With this virtual companion, the user can accessorize it, chat with it, and hang out with it by doing various activities. All these activities are presented and recorded per user, like their selections, chat history, voice messages, activity videos, and more throughout the application. All these activities and records are processed, analyzed, and stored by the next layer, which is the metaverse application. In this augmented reality module, 3D rendering engines are the core programs that perform the processing of these applications. An

incentive mechanism is utilized for keeping the user engaged. Task assignments are triggered events and are processed by executing user-initiated commands. Application security includes restrictions and data exchange protocols used according to the industry standard. State synchronization synchronizes the virtual avatar response synchronization provided from the cloud storage and streaming module. All the metaverse application modules need input from the user device layer. It supports the metaverse application with its modules for user authentication and identity management, asset storage, database management, crypto and other payments, and basic device security. According to our infrastructure diagram, the user block provides data with emotional dependence to the Mobile Application Architecture block, and at its different layers, a lack of security policy exposes this user data to the Attacker block, which allows attacks like snooping attacks, data exposure attacks, Man-in-the-Middle attacks (MitM), and impersonation attacks. In Figure 3, the attack workflow is for metaverse-based virtual companion applications to filter the components from the application infrastructure and map them with the victim assault workflow. The victim assault workflow is composed of the attacker's actions taken for victim exploitation over the application infrastructure. This is simplified further by identifying the tactics, techniques & procedures (TTPs) utilized per action. The actions, as mentioned—capture, record, connect, access, attack, and exploit—are the attacker actions linked to the components in the infrastructure that allow them to accomplish the attack. Firstly, capture action tends to perform observation-based collection of data that is provided as input by the victim. Record action is further activated as more precise information about the victim, such as email connectivity, different types of logs, biometric-based authentication information, and so on, is collected. Next, using the collected information, the Connect action is performed. Connect, as in utilize the information to perform login, access, and modify the victim-related information. Attack action is performing any malicious activity that might cause the victim losses. The victim's loss is the attacker's profit, which can be financial, fame, or revenge achieved by the exploitative action. The TTP labels are extracted from the MITRE Att&ck Framework [14], as mentioned in Table II. To describe this attack workflow, we need to break the actions into TTPs that can be utilized to accomplish the attacks, such as snooping, data exposure, MitM, and impersonation. A more in-depth explanation for the mentioned TTPs is available in the MITRE Att&ck framework [14].

To understand the possible threats to this newly formed Metaverse-based Virtual Companion Application, we accessed such applications, attempted to understand their security policy, and exploited them at a minor level. When accessing these applications, we provided personal information such as email address, name, date of birth, gender, phone number, and payment information. Device information collected includes model name, mobile carrier, operating system, device unique identification number, advertising identification, and time information. Accounts linked to information of various services, such as Google, SNS, and so on, are checked. Messages sent and received with media such as emojis, images, videos, and voice are recorded. Since users are emotionally reliant on their virtual companion, we conducted an experiment to better understand how these applications respond to their users, depicted in Figure 4. We utilised a South Korean-origin application, namely "Zeta," which is based on the idea of a virtual companion Application. We attempted to communicate with the virtual companion in this application, revealing the user's age as a high school student and sending provocative messages with decent language. This resulted in responses that contained vulgarities, profanity, obscenities, curses, or swear words. Neither during communication nor during authentication was age confirmed for the user. Therefore, to observe such harmful traits during communication with the virtual companion, our transformer-based detection and analysis program scans the application chat file for 'toxicity score,' 'unethical request score,' 'emotional dependency score,' 'sensor data access risk score,' and 'estimated age group.' We utilized pretrained BERT/RoBERTa models from Hugging Face for toxicity with a 98% accuracy (unitary/toxic-bert [17]), emotions with an accuracy of 47.4% with roberta-base-go (SamLowe/roberta-base-go_emotions [18]), and demographics (age/gender with keywords) with training data (GoEmotions [19], Jigsaw Toxic Comments [20]) and custom keyword lists (profanity, unethical requests). The resulting graph highlights spikes in harmful traits—especially toxicity and unethical behavior—at specific lines, helping identify potentially problematic messages after translating to the English language. As the users are emotionally vulnerable and the virtual companions attempt to form dominance over them, they can be easily manipulated to share more personal information, device access, and buy subscriptions. Due to the threats involving these information exposures, this infrastructure is at risk if it fails to follow a potential security policy.

TABLE II
TTP DESCRIPTION FOR ATTACK WORKFLOW

Tactics, Techniques & Procedures(TTPs)	Description
T1592	Gather Victim Host Information
T1589	Gather Victim Identity Information
T1590	Gather Victim Network Information
T1591	Gather Victim Org Information
T1598	Phishing for Information
T1114	Email Collection
T1056	Input Capture
T1113	Screen Capture
T1125	Video Capture
T1199	Trusted Relationship
T1078	Valid Accounts
T1200	Hardware Addition
T1650	Aquire Access
T1583	Aquire Infrastructure
T1586	Compromise Accounts
T1656	Impersonation
T1557	Adversary in the Middle
T1499	Endpoint Denial of Service
T1657	Financial Theft

IV. Result

We attempted to examine these applications by manually analyzing and performing minor exploitation to understand if they adhere to their claimed security policies. We observed that different application manufacturers follow different security policies for their virtual companion applications for the metaverse. These applications gathered user information when we accessed them; they became dominating and manipulative when we tried to communicate with a virtual companion. It requested constant details and created imaginative scenarios that made the user emotionally vulnerable. Even though we provided an 18+ birth date, we mentioned that the user is a high schooler in chat with a virtual companion; it was not filtered and ignored. We also found the virtual companion to get angry in such cases and use slang and abusive language, and we were not able to restrict it. This states that there is no limit to participation between the user and the virtual companion application. The security policy claims to access all the personal labeled and unlabeled information, not being particular or selective, and does not guarantee against its loss. If the user disagrees, the user cannot access these applications.

Outlining a best-suited conceptualized security policy for these applications is most important. Based on the situation awareness and infrastructure study, we came up with the best set of security policies for such metaverse-based virtual companion applications. There are 6 main aspects of this security policy combination:

- Awareness Notice: As the user is the beneficiary as well as the most vulnerable for this application, they have the right to be aware of their situation when using this application. The application manufacturers should provide an application awareness notice to their users, like the extent of their involvement with the application. Emotional attachment to a virtual companion is at risk in case of application malfunction, service unavailability, or discontinuity. The disadvantages of oversharing personal information should be discussed in this notice. Information accessed related to the user and their device should be clearly mentioned in security policies as well as this notice. Warning the user makes the user aware to make a decision, even though they are in an emotionally vulnerable situation.
- Abusive Language Control: Users communicating with their virtual companion through these application activities and chats should be filtered. As the user communicates over chat, some intense slang and abusive language should be restricted for both the user and the virtual companion. Abusive touch should be strictly blocked and warned against. This will maintain the integrity of the application in public use and reduce virtual world crimes.
- Emotional Independence: The users accessing these applications depend on their virtual companions emotionally. As they are emotionally vulnerable to the commands of the virtual companion, they innocently accept the requests of the virtual companion and the application. Therefore, a secure range of communication and unharmed dependency should be maintained while accessing the application. Rather than making them emotionally dependent, the applications should focus on making the user emotionally independent in nature in a comforting way.
- Information Request Restrictions: Information requested by these applications or through their virtual companions should be limited. Detailed information requests lead to the risk of information exposure by these applications. This allows attacks like impersonation, man-in-the-middle, and keystroke attacks.
- Sensor Data Access Limit: Various mobile sensors are accessed by these applications, such as cameras, microphones, Global Positioning System (GPS), proximity sensors, Bluetooth, and so on. Accessing these sensors is accessing data outsourced by these sensors. The Data is related to user identification, location, physical appearance, conditions, and so on. Limiting access to this data is crucial for user security and defense against snooping attacks.
- Non-manipulative Payment system: Users of this application are emotionally vulnerable; using manipulation for subscription prompts is not ideal. The subscription prompts should not be influenced through virtual avatars. Rather, they should be straightforwardly mentioned through dialogue boxes to users. Also, auto payment deduction should be prohibited. Payments on these applications should ask for user permission each time. This allows the user to be aware of all transactions happening through these applications.

These six aspects are the best combination for the security policy for such metaverse-based virtual companion applications.

V. Conclusion

This research elucidates the processing of the Metaverse-based virtual companion applications. We delineated the workflow for the Virtual Companion Application and provided a surmise attack scenario through situational awareness. This study entails the future implementation of Virtual Companion in the Metaverse and the development of reliable Virtual Companion applications. In the Virtual Companion Application infrastructure, we analyze potential threats through minor exploitation by pushing the application limits. This emphasized the cybersecurity risk examined for the Virtual Companion Application. We also discuss the best 6 aspect-based combination security policies for these virtual companion applications. In forthcoming research, we want to exploit these applications primarily and devise effective and detailed security policies.

Acknowledgment

This work was partly supported by the Institute of Information & Communications Technology Planning & Evaluation(IITP)-ITRC(Information Technology Research

Center) grant funded by the Korea government(MSIT)(IITP-2025-RS-2021-II211816, 20%), (Project No.RS-2024-00438551, 30%), (Project No.RS-2023-00228996, 30%) and a National Research Foundation of Korea (NRF), South Korea grant funded by the Korean government (Project No. RS-2023-00208460, 20%).

REFERENCES

[1] Sarang, A.D., Alawami, M.A. and Park, K.W., 2024. MV-Honeypot: Security Threat Analysis by Deploying Avatar as a Honeypot in COTS Metaverse Platforms. CMES-Computer Modeling in Engineering & Sciences, 141(1).

[2] Meng, Y., Zhan, Y., Li, J., Du, S., Zhu, H. and Shen, X.S., 2023, May. De-anonymization attacks on metaverse. In IEEE INFOCOM 2023-IEEE Conference on Computer Communications (pp. 1-10). IEEE.

[3] Li, C., Zeng, L., Huang, X., Miao, X. and Wang, S., 2023. Secure semantic communication model for black-box attack challenge under metaverse. IEEE Wireless Communications, 30(4), pp.56-62.

[4] Jafar, A., Yeboah-Ofori, A., Abisogun, T., Hilton, I., Oguntoyinbo, O. and Oyetunji, O., 2024, August. The Impact of Social Engineering Attacks on the Metaverse Platform. In 2024 11th International Conference on Future Internet of Things and Cloud (FiCloud) (pp. 201-208). IEEE.

[5] Wen, X., Wen, J., Xiao, M., Kang, J., Zhang, T., Li, X., Chen, C. and Niyato, D., 2024. Defending Against Network Attacks for Secure AI Agent Migration in Vehicular Metaverses. arXiv preprint arXiv:2412.20154.

[6] Eltanbouly, S., Halabi, O. and Qadir, J., 2025. Avatar privacy challenges in the metaverse: A comprehensive review and future directions. International Journal of Human–Computer Interaction, 41(4), pp.1967-1984.

[7] Son, B.D., Hoa, N.T., Van Chien, T., Khalid, W., Ferrag, M.A., Choi, W. and Debbah, M., 2024. Adversarial attacks and defenses in 6g network-assisted iot systems. IEEE Internet of Things Journal.

[8] Karapatakis, A., 2025. Metaverse crimes in virtual (Un) reality: Fraud and sexual offences under English law. Journal of Economic Criminology, 7, p.100118.

[9] Eugenia Kuyda. *Replika (v11.56.2)*. (2025, June 24). [Online]. Available: https://replika.com/

[10] Noam Shazeer and Daniel De Freitas. *Character.ai (v1.12.4)*. (2025, June). [Online]. Available: https://character.ai/

[11] Jorge Alis. *Soulmate AI (v1.12.4)*. (2025, April 4). [Online]. Available: https://soulmates.ai/

[12] SWEET SWEETPOTATO (HK) LIMITED. *Sweet AI (v1.6.3)*. (2025, June 17). [Online]. Available: https://www.sweetsai.com/login

[13] Scatter Lab. *Zeta (v1.6.3)*. (2025, June 24). [Online]. Available: https://play.google.com/store/apps/details?id=com.scatterlab.messenger&hl=en

[14] MITRE. *MITRE ATT&CK*. (2024). [Online]. Available: https://attack.mitre.org/

[15] Google LLC. *Gmail Service*. (2025). [Online]. Available: https://mail.google.com/

[16] Verizon and Apollo Asset Management. *Yahoo Mail Service*. (2025). [Online]. Available: https://mail.yahoo.com/

[17] Unitary. `toxic-bert` (2025). [Online]. Available: https://huggingface.co/unitary/toxic-bert

[18] Sam Lowe. `roberta-base-go_emotions` (2025). [Online]. Available: https://huggingface.co/SamLowe/roberta-base-go_emotions

[19] Google Research. `GoEmotions`. (2021). [Online]. Available: https://github.com/google-research/google-research/tree/master/goemotions

[20] Jigsaw and Conversation AI. *Jigsaw Toxic Comment Classification Challenge*. (2018). [Online]. Available: https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/data