Security and Privacy in the Metaverse: The Threat of the Digital Human

Lauren Buck
Trinity College Dublin
Dublin, Ireland
lauren.e.buck.12@gmail.com

Rachel McDonnell
Trinity College Dublin
Dublin, Ireland
ramcdonn@tcd.ie

ABSTRACT

Each year, researchers and technologists are bringing the vision of the Metaverse, which is predicted to be the future of the internet, closer to becoming a reality. People will spend most of their time in this space interacting face-to-face, so to speak, with highly customizable digital avatars that seamlessly convey precise non-verbal cues from the physical movements of the users themselves. This is an exciting prospect; however, there are many privacy and security concerns that arise from this new form of interaction. Precision motion tracking is required to drive high-fidelity animation, and this affords a mass of data that has never been available before. This data provides a wealth of physical and psychological information that can reveal medical conditions, mental disorders, personality, emotion, personal identity, and more. In this paper, we discuss some implications of the availability of this data, with a focus on the psychological manipulation and coercion capabilities made available by it.

CCS CONCEPTS

• **Security and privacy** → **Social aspects of security and privacy**; • **Human-centered computing** → *Collaborative and social computing theory, concepts and paradigms.*

KEYWORDS

virtual reality, personal data collection and use, shared virtual environments, avatars, agents, biometric data, machine learning, artificial intelligence

ACM Reference Format:

Lauren Buck and Rachel McDonnell. 2022. Security and Privacy in the Metaverse: The Threat of the Digital Human. In *Proceedings of CHI Conference on Human Factors in Computing Systems (CHI EA '22, Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality).* ACM, New York, NY, USA, 4 pages.

1 INTRODUCTION

In the present moment, we live in an age where personal data has been considered by some as more valuable than oil [2]. Tech companies deliberately design applications that are addictive to their users and source user data to create personalized ad experiences in order to generate revenue. Meta reported US$33bn in revenue in the 4th Quarter of 2021 alone [1], and ByteDance (TikTok) is currently valued around US$400bn as it reigns as the top grossing social media app of 2021 [6]. As individuals and governments grapple with the impact of personal data collection and how to protect individual users from big tech, small pockets of online scammers vie for the same data. Romance scams occur frequently online, where individuals are coalesced into fake relationships that often result in manipulation and theft, and infamous 'rug pull' scenarios involving cryptocurrencies have raked in over US$2.8bn in just the last year [16]. Protecting the individual has never been more important than ever as new technologies evolve and create new hosts of problems to resolve. In this paper, we consider some issues that are probable to become common, complex issues derived from the widespread adoption of virtual reality (VR) technology.

VR is a unique technological medium that is set apart from other media by one defining attribute: 3D interaction. VR environments emulate real world perceptions and allow users to move through space and interact with objects and people as they would naturally. This capability is driven by positional tracking, which calculates the precise position of a head-mounted display, controllers, and other trackers attached to the body, within Euclidean space. This tracking captures the motions of users and allows for the embodiment of bodily self-representations (i.e., avatars), which are an essential component of the VR experience. These capabilities alone introduce a myriad of privacy and security issues that require attention, especially considering the growing of body of knowledge about data extraction and psychological influence occurring due to the use of VR in experimental settings alone.

The existing catalog of literature that revolves around privacy and security in VR mainly focuses on motion tracking (gait analysis, eye tracking, general bodily motions), and there is increasing awareness of the fact that this data provides personally identifiable information. A shocking study carried out by Mark Miller and colleagues reports that five minutes of motion tracking data gathered only from a head-mounted display during a typical viewing task is enough to identify a user with 95% accuracy [25]. This type of data has the potential to reveal information about users that reflect mental state and medical status. Buck and Bodenheimer note that tracking a user's personal space representation can reveal social preferences and disorders like social anxiety [5]. Different types of body motions detected can predict levels of creativity [38] and learning [37], along with medical conditions such as autism, ADHD, and PTSD [25]. Diane Hosfelt, among others, have warned that the misuse and abuse of this data can produce life-altering consequences [15, 28].

However, the safety and privacy issues of VR that we as a community are cognizant of still remain a gray area. In this work, we hope to bring to light one pocket of issues directly related to the

CHI EA '22, Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality, April 29 - May 5, 2022, New Orleans, LA, USA

psychological manipulation of users by digital humans based on biometric data collection. There has been little publication, to our knowledge, on this issue.

2 THE DANGERS OF THE DIGITAL HUMAN

It is no secret that VR experiences can be psychologically compelling. Mel Slater and his colleagues have published an abundance of research that can attest to this. Men report feelings of empathy toward women after experiencing sexual harassment in a woman's body [27], the observation of a virtual body-double of oneself interacting with a crowd can reduce self-persecutory thoughts [13], and putatively stressful simulations can produce both physiological and psychological responses [22]. Why these simulations can be so impactful stems from the ability of VR users to embody virtual self-representations. The embodiment of a virtual character is a commonplace phenomenon in VR [20], and increases the perceived plausability of the simulation [32]. These graphical self-representations, or self-avatars, do not have to match the physicality of their users in order to elicit the sensation of embodiment [12], and virtual characters are not bound by physical constraints.

The importance of knowing that avatar appearance is mutable is in the detail that humans are naturally prone to making judgments based on physical appearance. It has been proven, even in early 2D games with elementary graphics, that the way users interact with one another has a lot to do with looks [10, 30]. It has been made clear that attractiveness dictates how VR users perceive themselves and others and the behaviors they choose to carry out. Avatars have already been posited as potential salespeople [17, 26], whose appearance can be persuasive enough to influence decision-making [9, 19], and when embodied can embolden users to engage in risk-taking behaviors [23]. VR gives us the flexibility to be digital chameleons, and we can connect the dots to understand that this ability in the hands of those with malicious intent will drive us toward a dystopian vision of the future we are all becoming increasingly familiar with.

This is where biometric data comes into play: the appearance of the computer or human-driven agents and avatars users are interacting with can be adapted to user preferences based on data that the user is unaware they have shared. Depending on the technology a system is using, eye tracking data can provide pupil dilation and gaze fixation data in response to visual stimuli [36], motion data can provide proxemic behavior and body language cues [5, 8], and physiological data like EEG and skin conductance can provide levels of emotional activation [33]. Additionally, facial tracking can also provide a window into emotional response [3]. These non-verbal cues can be fed into clever machine learning and artificial intelligence algorithms to create personalized, idealized interaction partners.

Besides outward appearance, both voice and motion can be manipulated to be appealing to users. Software can manipulate vocal tone, pitch and amplitude, which can allow users to change their voice from male to female and vice versa. Attractive voices that are smooth in texture and similar in pitch and timbre can be created easily via auditory morphing [4]. Vocal cloning software can mimic the sound of a particular person's voice. Motion data can give way to an expanse of physical and psychological information, which we have discussed in the introduction. Vocal expression in virtual characters has already been shown to impact social influence and attraction [34], and motion data is rich with cues that can be categorized into levels of attractiveness, which influence interaction behaviors [40].

These interaction partners may not only be designed to be attractive to users, but may understand users deeply on both a physical and psychological level. The personality of an agent could be adapted to be most likeable by the personality type the user is exhibiting [21]. A more advanced iteration of artificial intelligence or a human driving an avatar could perhaps detect and empathize with medical and mental conditions to create a sense of closeness and trust with a user. There are many applications of this psychological information that can take place with both benevolent and malicious intent.

Herein lies our danger. How far are we willing to take these technological capabilities? Generative adversarial networks (GANs) have already been leveraged to create human likenesses from scratch [18] and deepfakes generate video and audio to create scenes that have never happened in real life. Adapting agent and avatar appearance, personality, and interaction in order to sell products to users is not a far reach, as Amazon populates recommended items based on shopping habits, and social media sites generate personalised ads, by using machine learning techniques. Neither is it obscene to think avatars may be used for political duress, as some social media sites are notorious for serving politically polarizing content to users and manipulating emotions to increase engagement. Radical groups have been known to recruit impressionable, young people through online tactics. Additionally, Online gambling sites take advantage of those suffering with gambling addictions, and VR is already considered to promote high-risk gambling behaviors [29]. Additionally, virtual influencers are already materializing [35]. Could not a strategically placed agent persuade a user to engage in high-risk behavior? If we think about it, mental and physical traits extracted from biometric data could be exploited to coerce users into situations and behaviors they would otherwise refuse to engage in.

Particular attention needs to be paid to the potential for users to be manipulated not just by businesses and institutions, but by other individuals. It would be quite easy for a number of existing internet scams to spill into the Metaverse, and for their impact to be even more psychologically devastating because of the immersive aspect of VR. Extortion and bullying could put users in more personal, compromising situations, and concern comes with how children will be protected since they are particularly impressionable. Online predators will be handed a whole new toolkit of coercive measures with the availability of more natural interaction. Finally, cyberattacks will expose sensitive biometric data that could be sold on the dark web, which would be devastating to one's personal privacy.

Fortunately, there are many things that can be done to combat the misuse of biometric data before it begins, and we are all called to make positive contributions in this space. Power should be given to the user. People should be made aware of and educated on the implications of biometric data coupled with VR, and should be given the option to opt out of this type of data collection. Developers can implement cybersecurity protocols and can also choose to introduce noise to this type of data in order to generalize it and prevent it from revealing personally identifiable information. Finally, legislators

can introduce laws that prevent businesses and individuals from collecting and using this data with malicious intent.

Academics are called to make an impact through ongoing research to help understand and mitigate known and unknown psychological problems that will arise in the Metaverse. Potential research avenues include continuing to understand how users can be manipulated in advertisement scenarios [14, 24], what physical properties of agents and avatars are likely to have psychological influence over users [39], how risky behaviors translate from real to virtual scenarios [7], and the overall psychological impact of digital interaction in the Metaverse that will translate into the daily lives of users (think how augmented images affect self-esteem [11, 31]). There are many positive impacts that interaction with digital humans can have, and it is up to bring about an ethical iteration of the Metaverse.

3 CONCLUSIONS

Widespread adoption of the Metaverse comes with many unique threats to user privacy and security, some of which we have broached in this work with regard to digital humans. Biometric data reveals a host of personally identifiable information which can in turn be used to potentially manipulate users on a psychological level through the creation of avatars that are adaptable to user preferences. With respect to security and privacy issues, the VR community is in the midst of the Collingridge Dilemma; it is faced with the responsibility of understanding the potential risks that the Metaverse poses to the individual and mitigating those problems before it is too late. In the grand scheme of things, the digital human is something amazing and fearsome, and an aspect of VR that is not to be considered lightly.

ACKNOWLEDGMENTS

This research was funded by Science Foundation Ireland under the ADAPT Centre for Digital Content Technology (Grant No. 13/RC/2106_P2) and RADICal (Grant No. 19/FFP/6409).

REFERENCES

[1] 2021. Facebook Reports Third Quarter 2021 Results. https://investor.fb.com/investor-news/press-release-details/2022/Meta-Reports-Fourth-Quarter-and-Full-Year-2021-Results/default.aspx. Accessed: 2022-02-21.

[2] Anonymous. 2017. The world's most valuable resource is no longer oil, but data. *The Economist* (2017). https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data Accessed: 2022-02-01.

[3] Jeremy N Bailenson, Emmanuel D Pontikakis, Iris B Mauss, James J Gross, Maria E Jabon, Cendri AC Hutcherson, Clifford Nass, and Oliver John. 2008. Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International journal of human-computer studies* 66, 5 (2008), 303–317.

[4] Laetitia Bruckert, Patricia Bestelmeyer, Marianne Latinus, Julien Rouger, Ian Charest, Guillaume A Rousselet, Hideki Kawahara, and Pascal Belin. 2010. Vocal attractiveness increases by averaging. *Current biology* 20, 2 (2010), 116–120.

[5] Lauren E Buck and Bobby Bodenheimer. 2021. Privacy and Personal Space: Addressing Interactions and Interaction Data as a Privacy Concern. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 399–400.

[6] David Curry. 2022. Top Grossing Apps (2022). *Business of Apps* (2022). https://www.businessofapps.com/data/top-grossing-apps/ Accessed: 2022-02-01.

[7] Carla de Juan-Ripoll, José L Soler-Domínguez, Jaime Guixeres, Manuel Contero, Noemi Álvarez Gutiérrez, and Mariano Alcañiz. 2018. Virtual reality as a new approach for risk taking assessment. *Frontiers in psychology* (2018), 2532.

[8] Julius Fast. 1970. *Body language*. Vol. 82348. Simon and Schuster.

[9] Ylva Ferstl, Elena Kokkinara, and Rachel Mcdonnell. 2017. Facial features of non-player creatures can influence moral decisions in video games. *ACM Transactions on Applied Perception (TAP)* 15, 1 (2017), 1–12.

[10] Ylva Ferstl, Michael McKay, and Rachel McDonnell. 2021. Facial feature manipulation for trait portrayal in realistic and cartoon-rendered characters. *ACM Transactions on Applied Perception (TAP)* 18, 4 (2021), 1–8.

[11] Rebecca Fribourg, Etienne Peillard, and Rachel Mcdonnell. 2021. Mirror, Mirror on My Phone: Investigating Dimensions of Self-Face Perception Induced by Augmented Reality Filters. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 470–478.

[12] Mar Gonzalez-Franco and Tabitha C Peck. 2018. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI* 5 (2018), 74.

[13] Geoffrey Gorisse, Gizem Senel, Domna Banakou, Alejandro Beacco, Ramon Oliva, Daniel Freeman, and Mel Slater. 2021. Self-observation of a virtual body-double engaged in social interaction reduces persecutory thoughts. *Scientific reports* 11, 1 (2021), 1–13.

[14] Brittan Heller and Avi Bar-Zeev. 2021. The Problems with Immersive Advertising: In AR/VR, Nobody Knows You Are an Ad. *Journal of Online Trust and Safety* 1, 1 (2021).

[15] Diane Hosfelt. 2019. Making ethical decisions for the immersive web. *arXiv preprint arXiv:1905.06995* (2019).

[16] Stephanie Hughes. 2022. Rug-pull scams raked in over US$2.8 billion in crypto in 2021, report finds. *Financial Post* (2022). https://financialpost.com/fp-finance/cryptocurrency/rug-pull-scams-raked-in-over-us2-8-billion-in-crypto-in-2021-report-finds Accessed: 2022-02-01.

[17] Seung-A Annie Jin and Justin Bolebruch. 2009. Avatar-based advertising in Second Life: The role of presence and attractiveness of virtual spokespersons. *Journal of Interactive Advertising* 10, 1 (2009), 51–60.

[18] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4401–4410.

[19] Rabia Fatima Khan and Alistair Sutcliffe. 2014. Attractive agents are more persuasive. *International Journal of Human-Computer Interaction* 30, 2 (2014), 142–150.

[20] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (2012), 373–387.

[21] Tze Wei Liew and Su-Mae Tan. 2016. Virtual agents with personality: Adaptation of learner-agent personality in a virtual learning environment. In *2016 Eleventh International Conference on Digital Information Management (ICDIM)*. IEEE, 157–162.

[22] Marieke AG Martens, Angus Antley, Daniel Freeman, Mel Slater, Paul J Harrison, and Elizabeth M Tunbridge. 2019. It feels real: physiological responses to a stressful virtual reality environment and its impact on working memory. *Journal of Psychopharmacology* 33, 10 (2019), 1264–1273.

[23] Paul R Messinger, Xin Ge, Eleni Stroulia, Kelly Lyons, Kristen Smirnov, and Michael Bone. 2008. On the relationship between my avatar and myself. *Journal For Virtual Worlds Research* 1, 2 (2008).

[24] Abraham Hani Mhaidli and Florian Schaub. 2021. Identifying manipulative advertising techniques in xr through scenario construction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[25] Mark Roman Miller, Fernanda Herrera, Hanseul Jun, James A Landay, and Jeremy N Bailenson. 2020. Personal identifiability of user tracking data during observation of 360-degree VR video. *Scientific Reports* 10, 1 (2020), 1–10.

[26] Ian Mull, Jamie Wyss, Eunjung Moon, and Seung-Eun Lee. 2015. An exploratory study of using 3D avatars as online salespeople: The effect of avatar type on credibility, homophily, attractiveness and intention to interact. *Journal of Fashion Marketing and Management* (2015).

[27] Solène Neyret, Xavi Navarro, Alejandro Beacco, Ramon Oliva, Pierre Bourdin, Jose Valenzuela, Itxaso Barberia, and Mel Slater. 2020. An embodied perspective as a victim of sexual harassment in virtual reality reduces action conformity in a later milgram obedience scenario. *Scientific reports* 10, 1 (2020), 1–18.

[28] Mónika Nogel, Gábor Kovács, and György Wersényi. 2021. The Regulation of Digital Reality in Nutshell. In *12th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. 1–7.

[29] Sebastian Oberdörfer, David Schraudt, and Marc Erich Latoschik. 2022. Embodied Gambling-Investigating the Influence of Level of Embodiment, Avatar Appearance, and Virtual Environment Design on an Online VR Slot Machine. *Frontiers in Virtual Reality* (2022), 8.

[30] Connor P Principe and Judith H Langlois. 2013. Children and adults use attractiveness as a social cue in real people and avatars. *Journal of experimental child psychology* 115, 3 (2013), 590–597.

[31] Susruthi Rajanala, Mayra BC Maymone, and Neelam A Vashi. 2018. Selfies—living in the era of filtered photographs. *JAMA facial plastic surgery* 20, 6 (2018), 443–444.

[32] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3549–3557.

[33] Feng Tian, Minlei Hua, Wenrui Zhang, Yingjie Li, and Xiaoli Yang. 2021. Emotional arousal in 2D versus 3D virtual reality environments. *Plos one* 16, 9 (2021), e0256211.

[34] Ilaria Torre, Emma Carrigan, Katarina Domijan, Rachel McDonnell, and Naomi Harte. 2021. The Effect of Audio-Visual Smiles on Social Influence in a Cooperative Human–Agent Interaction Task. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 6 (2021), 1–38.

[35] Christopher Travers. 2022. *Virtual Humans.* https://www.virtualhumans.org/

[36] Joseph Tao-yi Wang. 2011. Pupil dilation and eye tracking. *A handbook of process tracing methods for decision research: A critical review and user's guide* (2011), 185–204.

[37] Andrea Stevenson Won, Jeremy N Bailenson, and Joris H Janssen. 2014. Automatic detection of nonverbal behavior predicts learning in dyadic interactions. *IEEE Transactions on Affective Computing* 5, 2 (2014), 112–125.

[38] Andrea Stevenson Won, Jeremy N Bailenson, Suzanne C Stathatos, and Wenqing Dai. 2014. Automatically detected nonverbal behavior predicts creativity in collaborating dyads. *Journal of Nonverbal Behavior* 38, 3 (2014), 389–408.

[39] Nick Yee and Jeremy Bailenson. 2007. The Proteus effect: The effect of transformed self-representation on behavior. *Human communication research* 33, 3 (2007), 271–290.

[40] Katja Zibrek, Benjamin Niay, Anne-Hélène Olivier, Ludovic Hoyet, Julien Pettré, and Rachel Mcdonnell. 2020. The effect of gender and attractiveness of motion on proximity in virtual reality. *ACM Transactions on Applied Perception (TAP)* 17, 4 (2020), 1–15.