Oct 18, 2022

To: FTC Commissioner and Chairwoman
Re: Request for Comments published 8.11.22

On behalf of the **Responsible Metaverse Alliance** (RMA), I hereby provide a written response to the Federal Trade Commission (FTC) request for public comments about the prevalence of commercial surveillance and data security harms caused by collecting, analyzing, and monetizing user information.

The enclosed comments, embodied as a paper entitled ***The Metaverse: the Ultimate Tool of Persuasion***, relate to immersive platforms using virtual and augmented reality technologies. While it is increasingly understood that the shift from flat media to immersive environments will greatly expand the surveillance capabilities (and risks) of online platforms, what is often overlooked is how these risks are amplified when platforms can simultaneously target users with promotionally altered experiences.

As described on the pages to follow, the metaverse represents a dangerous pairing of **real-time surveillance** and **real-time influence** on mainstream consumers. When considered in the context of control theory, we can see that without meaningful regulation to protect consumers, the metaverse could become an extremely powerful tool for deception, manipulation, and persuasion.

For these reasons, regulation should be considered for metaverse platforms to restrict behavioral and emotional profiling of metaverse users and to greatly limit the use of behavioral and emotional data for generating and controlling real-time promotional experiences in immersive environments.

Sincerely,

Louis Rosenberg, PhD
Chief Scientist, Responsible Metaverse Alliance

The Metaverse: the Ultimate Tool of Persuasion

Louis Rosenberg, PhD

The United Nations Human Rights Council recently adopted a draft resolution entitled Neurotechnology and Human Rights.[1] It's aimed at protecting humanity from devices that can "record, interfere, or modify brain activity." To describe the risks, the resolution uses euphemistic phrases like *cognitive engineering*, *mental privacy* and *cognitive liberty*, but what we're really talking about is mind control.

I applaud the U.N. for taking up the issue of mind control, but neurotechnology is not our greatest threat on this front. That's because it involves sophisticated hardware ranging from "brain implants" to wearable devices that can detect and transmit signals through the skull. Yes, these technologies could be very dangerous, but they're unlikely to be deployed at scale anytime soon. In addition, individuals who submit themselves to implants or brain stimulation, will likely do so with informed consent.

On the other hand, there is an emerging category of products and technologies that could threaten our *cognitive liberties* across large populations, requiring nothing more than consumer-grade hardware and software from trusted corporations. These seemingly innocent systems, which will be marketed for a wide range of positive applications from entertainment to education, targeting both kids and adults, could be dangerously misused and abused without our knowledge or consent.

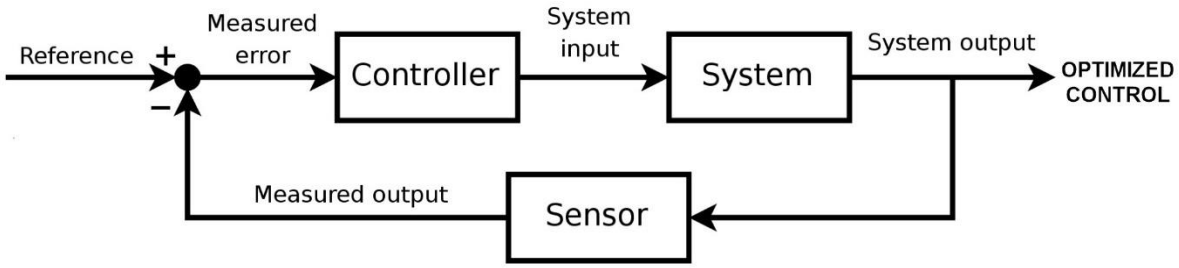
I'm talking about the metaverse.

I've written many articles about the dangers of the metaverse and the need to protect human rights in virtual and augmented worlds.[2] Still, I've not explained why metaverse technologies could threaten our cognitive liberties as effectively as a brain implant. To do so, I'd like to introduce an engineering concept that helps me think objectively about the potential dangers of interactive systems. The concept is *Feedback Control* and it comes from a technical discipline called *Control Theory*.

I know, those words imply a complex and esoteric explanation, but honestly, it's easy to understand. Control Theory is merely the name given by engineers and scientists to the formal method used for influencing the behaviors of a system. That system could be anything that interacts with the world. Think of a cheap nightlight with a sensor on it. When it gets dark outside, the nightlight comes on. When it gets light out again, the nightlight goes off. That's a control system.

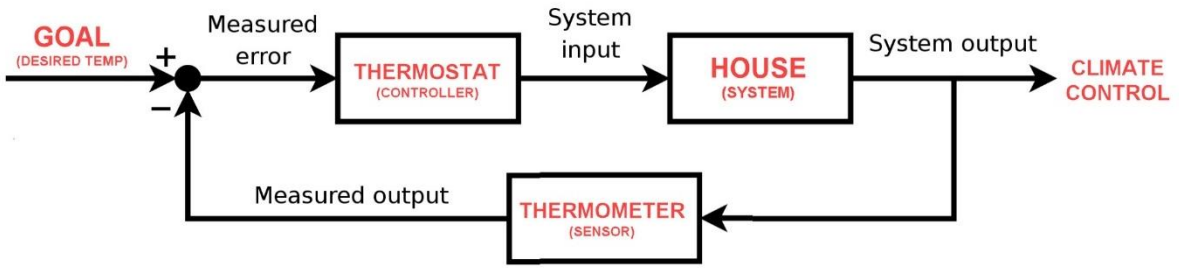
Now, consider a slightly more complex example – the thermostat in your house. Unlike a nightlight with only two states (On or Off), your house can be a range of temperatures. You set a goal and if your house falls below that goal, your heat turns on. If your house gets too hot, it turns off. When working properly, your thermostat keeps your house close to the goal you set. That's feedback control.

Of course, engineers sometimes make things more complex than they need to, so the simple concept above is generally represented in a standard format called a *Control System Diagram* as follows:



Generic "Control System" Diagram – (credit) Wikipedia[6]

The diagram above has three key boxes labeled *System*, *Sensor*, and *Controller*. In the heating example, your house would be the system, a thermometer would be the sensor, and the thermostat would be the controller. An input signal called the *Reference* is the temperature you set as the goal. The goal is compared to the actual temperature in your house (i.e., *Measured Output*). The difference between the goal and measured temperature is fed into the thermostat which determines what the house's heater should do. If the house is too cold, its heater turns on. If it's too hot its heater turns off. That's a classic control system that can be represented in the diagram below.



Control System Diagram for Climate Controlled House

Of course, control systems can get sophisticated, enabling airplanes to fly on autopilot and cars to autonomously navigate traffic, even allowing robotic rovers to land on mars. To control complex behaviors, these systems need sophisticated sensors to detect a wide range of driving conditions or flying conditions or whatever else is appropriate for the task at hand. Such systems also need powerful controllers that can process the rich sensor data and influence behaviors in diverse ways. These days, the controllers are usually microprocessors and increasingly use AI algorithms at their core.

With that background, let's jump back into the metaverse.

If we strip away the hype, the metaverse is about transforming how we humans interact with the digital world. In today's society, we mostly consume *flat media* viewed in the third person. In the metaverse, digital content will become *immersive experiences* that are spatially presented all around us. This shift to *first-person interactions* will fundamentally alter our relationship with digital information, changing us from outsiders peering in at content to active participants engaging content presented naturally in our surroundings. In other words, we will literally climb in and become part of the information system.

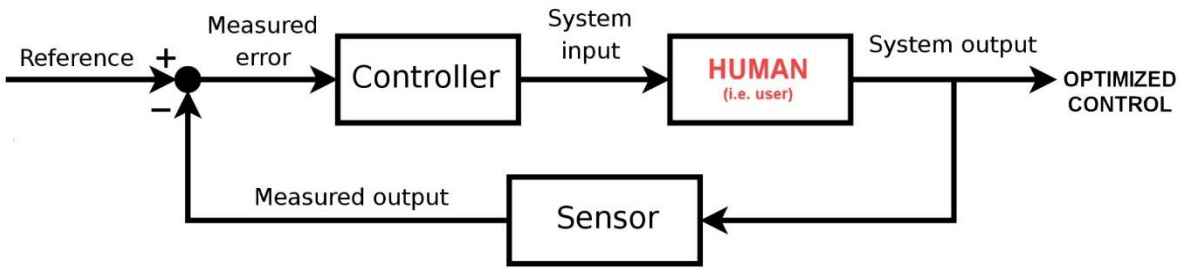
Before I describe the risks of immersive technologies, I need to express that the metaverse has the potential to be a deeply humanizing technology. After all, it will provide us with information in the format our senses were meant to perceive it – as natural first-person experiences, not as flat documents viewed through little windows. I truly believe this will be a major benefit for humanity. Thirty years ago I described the potential like this – *"Given the ability to draw upon our highly evolved human facilities, users of virtual environment systems can obtain an intimate level of insight and understanding."* [3]

I still believe this, but the power of immersive media can work in both directions. Used in positive ways, it can unlock insight and understanding for people around the globe. Used in negative ways, it could unleash the most powerful tool of persuasion and manipulation we've ever created.[4] That's because immersive media can be far more impactful than traditional media, targeting our perceptual channels in the most personal and visceral form possible – as experiences.[5] And because we humans evolved to trust our senses (i.e. believe our eyes and ears), the very notion that what we can see, hear, and feel things directly around us that are entirely fabricated is not a situation we're mentally prepared for.

And yet, that's not the most dangerous aspect of the metaverse.

To appreciate the true danger of immersive technologies, we can use the basics of Control Theory. Referring back to the standard diagram above, we see that only a few elements are needed to effectively control a system, whether it's a simple nightlight or a sophisticated robot. The two most important elements are a SENSOR to detect the system's real-time behaviors, and a CONTROLLER that can influence those behaviors. The only other elements needed are the *feedback loops* that continually detect behaviors and impart influences, guiding the system towards desired goals.

As you may have guessed, when considering the danger of the metaverse, *the system* being controlled is you – the human in the loop. The figure below might seem a little jarring, but when you put on a headset and sink into the metaverse, you are immersing yourself into an environment that has the potential to *act upon you* more than you *act upon it*. That means that you are the system most likely to be controlled – an inhabitant of a fabricated world that can monitor and influence you in real time.



Control System Diagram with "Human" in the loop

In the figure above, the human is the system being controlled. The *System Input* to the human are the immersive sights, sounds, and touch sensations that are fed into your eyes, ears, hands, and body. This is overwhelming input – possibly the most extensive and intimate input we could imagine. This means the ability to influence the system (i.e. you) is equally extensive and intimate. On the other side of the human user is the *System Output* – that's your real-time actions, reactions, and interactions. This includes everything your body does that could potentially be tracked by sensors. This might sound innocent, but in the metaverse tracking is far more extensive than most people think.

This brings us to the SENSOR box in the diagram above. In the metaverse, sensors will track everything you do in real-time – the subtle motions of your head, hands, and body. That includes the direction you're looking, how long your gaze lingers, the faint motion of your eyes, the dilation of your pupils, the changes in your posture and gait – even your vital signs are likely to be tracked in the metaverse including your heartrate, respiration rate and blood pressure.

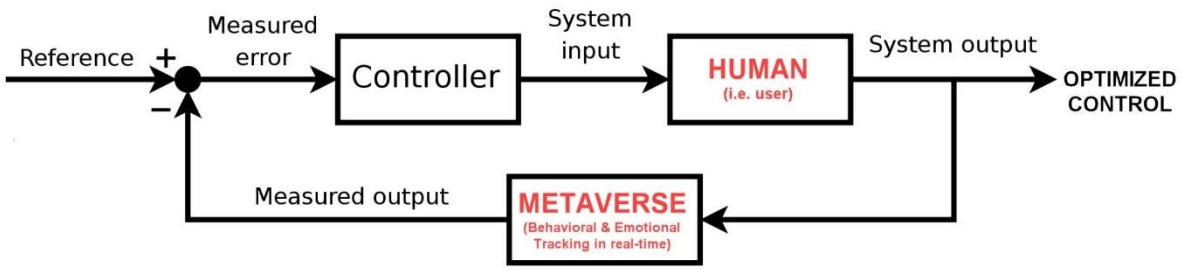
For example, the most recent headset deployed by META can accurately track your facial expressions and eye motions.[7] The potential goes beyond merely sensing the expressions that other people notice, but also includes subconscious expressions that are too fast or subtle for human observers to recognize. Known as "micro-expressions," these events can convey emotions that users had not intended to express and are unaware of revealing.[9] Users may not even be aware of feeling the emotions, leading to situations where the system literally knows the user better than he or she knows himself.

In addition, it's important to recognize that AI technology can be used to infer information that is not directly detected by sensors. In a recent paper, researchers at META showed that when processing "sparse data" from just a few sensors on your head and hands, AI technology could accurately predict the position, posture, and motion of the rest of your body.[8] Other researchers have shown that body motions such as gait can be used to infer a range of medical conditions from depression to dementia.[10,11]

In addition to tracking basic physical information, metaverse technology already exists to infer your emotions in real-time from your facial expressions, vocal inflections, gestures and body posture. Other technologies exist to detect emotions from the blood-flow patterns on your face and the vital signs detected from sensors in your earbuds.[12] This means when you immerse yourself into the metaverse, the sensor in the system diagram above will be able to track almost everything you do and say in that world and accurately predict how you feel during each action, reaction, and interaction.

Referring to the updated diagram below, we can fill in the SENSOR box with metaverse hardware that accurately tracks a user's behaviors and emotions in real-time. This means that while the user is being fed immersive experiences (System Input) that completely fill their senses with content, diverse sensors

can track nearly everything they do (including what they're paying attention to) and can infer how they feel while doing it. Again, it's not just conscious emotions that can be tracked, but subtle expressions detected from micro-expressions, facial blood-flow patterns, and vital signs.

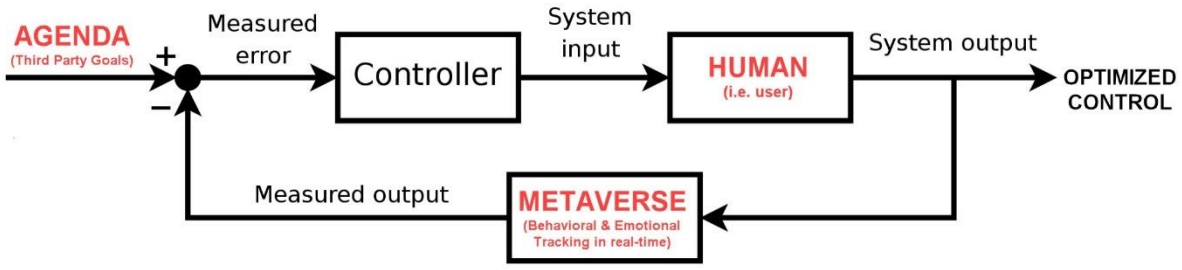


Control System Diagram for Metaverse Environments

In addition, the behavioral and emotional data described above could be stored by metaverse platforms over time, creating a database that reflects how individuals are likely to react to a wide range of stimuli throughout their daily life. When processed by AI algorithms, this extensive data could be turned into predictive behavioral and emotional models that enable platforms to accurately anticipate how users will react when presented with target stimuli (i.e. System Input) from a controller. And because the metaverse is not just virtual reality but also augmented reality, the tracking and profiling of users could occur throughout our daily life, from the moment we wake up to the moment we go to sleep.[13]

This means that in an unregulated metaverse, platforms could store and process extensive user data over time and create AI models that accurately predict how to drive desired behavioral or emotional responses in target users. This takes us one giant step towards mind control, potentially making the metaverse the most dangerous tool of persuasion ever created. To appreciate the risks, we need to consider the final box in the system, the controller.

As described above, the controller could be software running on processors that make extensive use of AI technology. The controller receives a Measured Error, which is the difference between a Reference Goal (a desired behavior) and the Measured Output (a sensed behavior). To bring this back to the topic of mind control, the goal could be the agenda of a third party that aims to impart influence over a user (see diagram below). That third party could be a paying sponsor that desires to persuade a user to buy a product, subscribe to a service, or even believe a piece of propaganda, ideology, or misinformation.



Control System Diagram with Third Party Agenda

Of course, advertising and propaganda exist using traditional marketing techniques. What is unique about the metaverse is the ability to create high-speed feedback loops in which a user's behaviors and

emotions are continuously fed into a controller that adjusts its influence in real-time for optimized persuasion. This process can easily cross the line from marketing to mind control. To appreciate these risks, we need to dig deeper into the role of the controller.

At its core, a controller aims to "reduce the error" between the desired behavior of a system and the measured behavior of the system. It does this by imparting System Input as shown on the diagrams above. In the metaverse, this arrow represents the ability of the controller to modify the virtual or augmented environment the user is immersed within. In other words, the controller can alter the world around the user, modifying what that user sees and hears and feels in order to drive the user towards the desired goal (i.e. agenda). And because the controller can accurately monitor how the user reacts to its alterations of the world, it will be able to adjust in real-time, optimizing the persuasive impact in much the same way that a thermostat optimizes the temperature of a house.

To make this concrete let's review some examples.

Imagine a user sitting in a coffee house in the metaverse (virtual or augmented). A third party sponsor wants to inspire the user to buy a particular product or service, or believe a piece of messaging, propaganda, or misinformation. In the metaverse, advertising will not be the pop-up ads and videos that we're familiar with today but will be immersive experiences that are seamlessly integrated into our surroundings.[14] In this example usage, the controller creates a virtual couple sitting at the next table. That virtual couple will be the System Input used to influence the user.

First, the controller will design the virtual couple for maximum impact. This means the age, gender, ethnicity, clothing styles, speaking styles, mannerisms, and other qualities of the couple will be selected by AI algorithms to be optimally persuasive upon the target user based on that user's historical profile. Next, the couple will engage in an AI-controlled conversation amongst themselves that is within earshot of the target user. That conversation could be about a car that the target user is considering purchasing, possibly framed as the virtual couple discussing how happy they are with their own recent purchase.

As the conversation begins, the controller monitors the user in real-time, assessing micro-expressions, body language, eye motions, pupil dilation, and blood pressure to detect when the user begins paying attention. This could be as simple as detecting a subtle physiological change in the user correlated with comments made by the virtual couple. Once engaged, the controller will modify the conversational elements to increase engagement. For example, if the user's attention increases as the couple talks about the car's horsepower, the conversation will adapt in real-time to focus on performance.

As the overheard conversation continues, the user may be unaware that he or she has become a silent participant, responding through subconscious micro-expressions, body posture, and changes in vital signs. The AI-driven controller will highlight elements of the new car that the target user responds most positively to and provide conversational counterarguments when the user's reactions are negative. And because the target user does not overtly express objections, the counterarguments could be profoundly influential. After all, the virtual couple could verbally address emerging concerns before those concerns have even fully surfaced in the mind of the target user. This is not marketing, it's mind control.

And in an unregulated metaverse, the target user may believe the virtual couple are avatars controlled by other patrons. In other words, the target user could easily believe they are overhearing an authentic

conversation among users and not realize it's a promotionally altered experience that was targeted specifically at them, injected into the surroundings to achieve a particular agenda.[15]

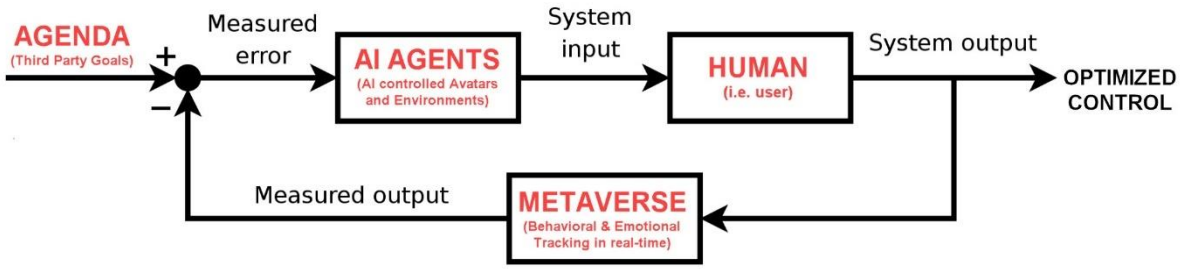
And that's a relatively benign example. Instead of pushing the features of a new car, the third-party agenda could be to influence the target user about a political ideology, extremist propaganda, or outright misinformation or disinformation. In addition, the example above targets the user as a passive observer of a promotional experience in his or her metaverse surroundings. In more aggressive examples, the controller will actively engage the user in targeted promotional experiences.

For example, consider the situation in which an AI-controlled avatar that looks and sounds like any other user in an environment engages the target user in agenda-driven promotional conversation. In an unregulated metaverse, the user may be entirely unaware that he or she has been approached by a targeted advertisement, but instead might believe he or she is in a conversation with another user. The conversation could start out very casual but could aim towards a prescribed agenda.

Of course, the controller will have access to a wealth of historical data about the target user – interests, values, hobbies, education, political affiliation, etc. – and will use this to craft dialog that optimizes engagement. In addition, the controller will have access to real-time information from the user, including facial expressions, vocal inflections, body posture, eye motions, pupil dilation, facial blood patterns, and potentially blood pressure, heartrate, and respiration rate. The controller will adjust its conversational tactics in real-time based on the overt verbal responses of the target user in combination with subtle and potentially subconscious micro-expressions and vital signs.

It's well known that AI systems can outplay the world's best human competitors at chess, go, poker, and a wealth of other games of strategy. From that perspective, what chance does an average consumer have when engaged in promotional conversation with an AI agent that has access to that user's personal background and interests, and can adapt its conversational tactics in real-time based on subtle changes in pupil dilation in blood pressure? The potential for violating a user's cognitive liberty through this type of feedback-control in the metaverse is so significant it likely borders on outright mind control.

To complete the diagram for metaverse-based feedback control, we can replace the generic word *controller* with AI-based software that alters the environment or injects conversational avatars that impart optimized influence on target users. This is expressed using the phrase AI AGENTS below.



Feedback-control Model of Human Persuasion in Immersive Environments

As described above, the metaverse could be used to create feedback-control systems that monitor the behaviors and emotions of users in real-time and employ AI agents to modify their surroundings (i.e., sights, sounds, and interactive experiences) to maximize persuasion. This means that large and powerful metaverse platforms could track billions of people and impart influence on any one of them by altering the world around them in targeted and adaptive ways.

This scenario is frightening but not farfetched.

In fact, it could be the closest thing to "playing god" that any mainstream technology has ever achieved. That's a bold statement and I don't make it lightly. I've been in this field for over 30 years, starting as a researcher at Stanford, NASA and the U.S. Air Force and then founding a number of successful companies in the space.[16] I genuinely believe the metaverse can be a positive technology for humanity, but if we don't protect against the downsides by crafting thoughtful standards and regulation, it could challenge our most sacred personal freedoms including our basic capacity for free will.[17,18,19]

REFERENCES

[1] United Nations. (n.d.). A/HRC/51/L.3 vote item 3 - 40th meeting, 51st Regular Session Human Rights Council | UN web TV. United Nations. https://media.un.org/en/asset/k1d/k1dcro5var

[2] Louis Rosenberg. 2022. Regulation of the Metaverse: A Roadmap: The risks and regulatory solutions for largescale consumer platforms. In Proceedings of the 6th International Conference on Virtual and Augmented Reality Simulations (ICVARS '22). Association for Computing Machinery, New York, NY, USA, 21–26. https://doi.org/10.1145/3546607.3546611

[3] Rosenberg Louis Barry (1994) "Virtual Fixtures : Perceptual Overlays Enhance Operator Performance in Telepresence Tasks." *Stanford University* U.M.I 19971994.

[4] Robertson, D. (2022, September 14). *The most dangerous tool of persuasion*. POLITICO. https://www.politico.com/newsletters/digital-future-daily/2022/09/14/metaverse-most-dangerous-tool-persuasion-00056681

[5] Breves, Priska. "Biased by Being There: The Persuasive Impact of Spatial Presence on Cognitive Processing." Computers in Human Behavior, vol. 119, 2021, p. 106723., https://doi.org/10.1016/j.chb.2021.106723

[6] Wikimedia Foundation. (2022, September 19). Control theory. Wikipedia. Retrieved October 18, 2022, from https://en.wikipedia.org/wiki/Control_theory

[7] Johnson, K. (2022, October 13). *Metas VR Headset Harvests Personal Data Right Off Your Face*. WIRED. https://www.wired.com/story/metas-vr-headset-quest-pro-personal-data-face/

[8] Winkler, A. (2022, September 20). *QuestSim: Human Motion Tracking from Sparse Sensors with Simulated Avatars*. https://arxiv.org/abs/2209.09391

[9] Li, Xiaobai & Hong, Xiaopeng & Moilanen, Antti & Huang, Xiaohua & Pfister, Tomas & Zhao, Guoying & Pietikainen, Matti. (2017). Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-Expression Spotting and Recognition Methods. IEEE Transactions on Affective Computing. PP. 1-1. 10.1109/TAFFC.2017.2667642.

[10] Wang Y, Wang J, Liu X, Zhu T. Detecting Depression Through Gait Data: Examining the Contribution of Gait Features in Recognizing Depression. Front Psychiatry. 2021 May 7;12:661213. doi: 10.3389/fpsyt.2021.661213. PMID: 34025483; PMCID: PMC8138135.

[11] Jacobs, S. (2022, October 12). Abnormality of Gait as a Predictor of Non-Alzheimer's Dementia. New England Journal of Medicine. https://www.nejm.org/doi/full/10.1056/NEJMoa020441

[12] Benitez-Quiroz CF, Srinivasan R, Martinez AM. Facial color is an efficient mechanism to visually transmit emotion. Proc Natl Acad Sci U S A. 2018 Apr 3;115(14):3581-3586. doi: 10.1073/pnas.1716084115. Epub 2018 Mar 19. PMID: 29555780; PMCID: PMC5889636.

[13] Rosenberg, L. (2022, August 30). How a parachute accident helped jump-start augmented reality. IEEE Spectrum. Retrieved October 1, 2022, from https://spectrum.ieee.org/history-of-augmented-reality

[14] Rosenberg, L (2022) Marketing in the Metaverse: A fundamental shift., Future of Marketing Institute. DOI: 10.13140/RG.2.2.35340.80003

[15] Rosenberg, L. (2022, August 21). Deception vs authenticity: Why the metaverse will change marketing forever. VentureBeat. https://venturebeat.com/ai/deception-vs-authenticity-why-the-metaverse-will-change-marketing-forever/

[16] Rosenberg, L.B. (2022). Augmented Reality: Reflections at Thirty Years. In: Arai, K. (eds) Proceedings of the Future Technologies Conference (FTC) 2021, Volume 1. FTC 2021. Lecture Notes in Networks and Systems, vol 358. Springer, Cham. https://doi.org/10.1007/978-3-030-89906-6_1

[17] Rosenberg, L. (2022, September 19). The case for demanding "immersive rights" in the metaverse. Big Think. https://bigthink.com/the-future/immersive-rights-metaverse/

[18] The XRSI Privacy and Safety Framework - XRSI – XR Safety Initiative. XRSI. (2022, February 16). Retrieved October 18, 2022, from https://xrsi.org/publication/the-xrsi-privacy-framework

[19] Rosenberg, L.B. (2022). Regulating the Metaverse, a Blueprint for the Future. In: De Paolis, L.T., Arpaia, P., Sacco, M. (eds) Extended Reality. XR Salento 2022. Lecture Notes in Computer Science, vol 13445. Springer, Cham. https://doi.org/10.1007/978-3-031-15546-8_23

About RMA:

The Responsible Metaverse Alliance (RMA) is an international movement dedicated to supporting the development of the metaverse (virtual and augmented worlds) to ensure that it's handled responsibly from the perspective of design, deployment, safety, culture, inclusion, operations and function.

About Author:

Dr. Louis Rosenberg is a pioneer of virtual and augmented reality. His work began over thirty years ago in labs at Stanford and NASA. In 1992 he developed the first interactive Augmented Reality system at Air Force Research Laboratory (AFRL). In 1993 he founded the early Virtual Reality company Immersion Corporation. In 2004 he founded the early Augmented Reality company Outland Research. He's been awarded over 300 patents for VR, AR, and AI technologies and published over 100 academic papers. He received his PhD from Stanford University and was a tenured professor at California State University (Cal Poly). He is currently CEO of Unanimous AI, the Chief Scientist of the Responsible Metaverse Alliance, the Global Technology Advisor to the XR Safety Initiative (XRSI) and a technology advisor to the Future of Marketing Institute.