II International Congress

A Metaverse for The Good: Exploring the Intersection of AI and the Metaverse

Editors:

Mel Slater (Chair)
Aurelio Lopez-Tarruella (Chair)
Pablo César
Michele Di Dalmazi
Doron Friedman
Marco Gillies
Omar Niamut
Xueni (Sylvia) Pan
Yannick Prié
Matthias Quent
Raúl Ruiz Rodríguez
Mavi Sanchez-Vives
Piotr Skrzypczynski
Paweł Śniatała
Anthony Steed

June 11-12th, 2025

University of Alicante, Spain

Thanks to

DERECHOS DIGITALES

Cátedra MetaversoUA

EMRN
The European Metaverse Research Network

INSTITUTO DE ECONOMÍA INTERNACIONAL UNIVERSIDAD DE ALICANTE (https://iei.ua.es/es/)

Universitat d'Alacant
Universidad de Alicante

frontiers

META-TOO

Financiado por la Unión Europea
NextGenerationEU

GOBIERNO DE ESPAÑA
MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

red.es

Plan de Recuperación, Transformación y Resiliencia

EMRN

The European Metaverse Research Network

A Metaverse for the Good

Exploring the Intersection of AI and the Metaverse

Conference 11-12th June, 2025

University of Alicante, Spain

Front cover design by Mavi Sanchez-Vives rendered by ChatGPT4.

Theme: AI and the Metaverse

The European Metaverse Research Network and the Academic Chair for the Responsible Development of the Metaverse of the University of Alicante organised the II International Congress "Metaverse for the Good: Exploring the Intersection of AI and the Metaverse", in the framework of the Project on the Implementation of the Charter of Digital Rights financed by the public entity Red.es in Alicante, 11-12 June, and is part of the activities of the European Metaverse Research Network.

Background

The idea of the metaverse is a vast extended reality set of worlds where physical and virtual reality intersect. It might be inhabited at any one time by millions of people, and as in the real world groups of people can interact with one another in real-time. The virtual metaverse will overlap with the physical world in various ways – for example, walking into a particular place in the physical world may transport you to a virtual place, where you are interacting with life-sized virtual humans who may be anywhere in the real world. Some of those 'people' you interact with might not be human but representations of AI agents. This can have far reaching beneficial consequences. For example, an AI agent might be an expert in psychological counselling who can treat millions of clients simultaneously. Another might be explaining to you how you can use some complex machine in the physical world. Another may be teaching mathematics. However, as with any technology there are multiple possibilities for malfeasance. As you walk through a physical street wearing augmented reality spectacles you may be bombarded by virtual humans trying to sell you something (a 'feature' you can only switch off by paying a global company). There are boundless possibilities for impersonation and political manipulation, as well as countless positive aspects that we can hardly see today – because as with any new technology, our predictions are likely to be incorrect. Mobile phones and social media have connected people in ways almost unimaginable a few years ago, but they also seem to be threatening democracy – with the loss of a common shared reality to which everyone relates. In this respect the Metaverse, with its immersion and sense of presence, may be orders of magnitude more powerful – positively and negatively.

This is not a Frontiers publication, but it is in conjunction with Frontiers in Virtual Reality, who have offered a prize for the best paper, and there will be a Research Topic associated with this conference.

CONTENTS

Panels

Inappropriate Social Interaction in the Metaverse

Moderator: Michele Barbier
Laurynas Adomaitis (RISE), Beatrice de Gelder (Maastricht University), Mel Stater (UB), Mavi Sanchez Vives (IDIBAPS), Michele Barbier (Inria), Katia Zibrek (Irina).

Living Virtual Machines

Moderator: Doron Friedman
Doron Friedman, Beatrice de Gelder , Jonathan Giron, Rupert Ortner

AI-Enhanced Immersive Environments for Lifelong Education and Career Counselling

Moderator: Ekaterina Prasolova-Førland
Lorena Gonzalez, Katarzyna Koba, Pablo López Guardiola

Keynotes

The Neuroscience of self-consciousness: Basic research, clinical applications, and contemplative practice, Professor Olaf Blanke, Center for Neuroprosthetics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
Chair: Mel Slater

When Virtual Patients Meet Virtual Therapists: Harnessing LLMs and VR for Mental Health, Professor Doron Friedman, Sammy Ofer School of Communications, Reichman University, Israel.
Chair: Mavi Sanchez-Vives

Designing AI and VR Tools to Think With, Professor Yvonne Rogers, UCLIC, Computer Science Department, University College London, UK.
Chair: Marco Gillies

Virtual Humans, Professor Frank Steinicke, Department of Informatics, Hamburg University, Germany.
Chair: Piotr Skrzypczynski

Virtual Worlds and Society: A Science for Policy Perspective, Miguel Barreda Angeles, Scientific Researcher at Joint Research Center, European Commission.
Chair: Yannick Prié

Clinical and well-being applications of the Metaverse,
Professor Mavi Sánchez-Vives, IDIBAPS (ICREA), Barcelona.
Chair: Xueni Sylvia Pan

Ethical Issues in the Metaverse, Professor Patrick Haggard, Institute of Cognitive Neuroscience, UCL, UK.
Chair: Mel Slater

Session 1: Virtual humans: Agents, avatars, and AI-driven personalities

Synthetic selves: exploring identity construction through virtual avatars and AI in XR environments

Valentino Megale[1*], Mariachiara Tirinzoni[2*], Serena Faccio[3]

[1]Softcare Studios Srls, Rome, Italy; XRSI Europe, Turin, Italy

[2]Independent Researcher, Milan, Italy; Eva In Rosso APS, Milan, Italy

[3]LL.M.; Information security manager, DTU, Denmark

*** Correspondence:**
Mariachiara Tirinzoni
mariachiara.tirinzoni@gmail.com
Valentino Megale
valentinomegale@hotmail.it

Keywords: VR, XR, Metaverse, AI, Identity, Avatar, Embodiment, Self

Abstract

As virtual environments grow increasingly immersive and intelligent, virtual avatars evolve beyond aesthetic representations into tools for identity experimentation, psychological insight, and therapeutic applications. Our proposal explores the convergence between Extended Reality (XR) and Artificial Intelligence (AI) in enabling dynamic identity construction through avatar embodiment. It examines how avatars influence users 'self-perception, behaviour, and well-being, and how AI models can analyse real-time user data to shape adaptive digital identities. This intersection opens transformative possibilities in healthcare, social development, and personal growth—but not without introducing ethical, cognitive, and sociocultural challenges that demand urgent multidisciplinary inquiry. How are these technologies redefining the concept of humanity, balancing opportunities and new challenges for society?

Article type: perspective

1 Introduction

In the present work we adopt an interdisciplinary approach that spans from human-computer interaction (HCI), embodied cognition research, and digital ethics to examine how extended reality (XR) technologies can reshape human identity through avatar-mediated experiences. We recognise that avatar-mediated identity creation occurs within complex sociotechnical systems which require both technical understanding and humanistic reasoning.

2 Where are we? A brief review

Neuroscience and cognitive science have highlighted the brain's adaptability to altered sensory feedback, a phenomenon central to the sense of embodiment in XR environments. (Slater et al., 2010) and (Kilteni et al. 2012) show that virtual avatars can evoke strong sensations of ownership, agency, and self-location, leading to shifts in self-perception. The Proteus Effect (Yee, N., & Bailenson, J. 2007) further illustrates how avatar traits unconsciously influence user behaviour even after the virtual experience. Meanwhile, contemporary research provides insights into the psychological and behavioural impacts of avatar embodiment. (Era et al. 2025) investigates how interactions with virtual avatars affect human perception and social behaviours, while (Yim et al., 2024) explores the Proteus Effect's influence on pain perception, revealing that embodying a muscular avatar can reduce perceived pain.

AI-driven virtual experiences are increasingly capable of integrating biometric and behavioural data, such as heart rate, gaze, and posture, to adapt in real time to users' emotional and cognitive states (Bhatt, Jogy, & Puri, 2024; Lau et al., 2024). By dynamically modelling user behaviour within immersive environments, these experiences can mirror affective responses, deliver tailored feedback, and foster deeper self-reflection, enhancing both engagement and therapeutic outcomes (Purushothaman et al., 2025). Through advanced user modelling techniques, such systems can personalise interactions at a granular level, making virtual environments more intuitive and responsive (Velez Quintero, 2023).

The growing integration of real-time physiological and behavioural data in VR allows avatars to become active participants in this loop, representing not just a digital identity, but a living, adaptive embodiment of the user's state. This evolution could profoundly shape how we perceive ourselves, blending external expression with internal signals in avatar-mediated self-representation.

The concept of virtual humans, AI-generated replicas of individuals, has recently gained momentum, allowing the creation of avatars that closely mimic human ways of acting and personality. (Park et al., 2024) indicates that AI agents can replicate a person's personality with up to 85% accuracy after just two hours of interaction. Such developments raise questions about the degree of authenticity and the resulting ethical implications of virtual identities.

On the other hand, MetaHuman technology exemplifies the potential of AI in recreating lifelike avatars. A notable project (MetaHuman, 2022) brought a 10,000-year-old shaman back to life by reconstructing his facial features and expressions, demonstrating the technology's capability to animate historical figures with remarkable realism, disrupting traditional limitations that had served so far as a foundation for our individual and social debate.

These technologies need urgent multidisciplinary inquiry to ensure responsible development and integration (Methuku & Myakala, 2025). We must design a set of ethical structures that are not limited to “emergency management” but are able to give useful directions for humans and technology together.

3 Opportunities

3.1 XR as a tool for ontological experimentation

While much of the focus in XR discourse centres on representation and simulation, its deeper promise lies in the rewriting of self through lived, embodied fiction. In avatar-mediated XR, individuals are offered not only the freedom to represent who they are but also to experiment with who they could be. This possibility becomes more than cosmetic—it is ontological. By stepping into new forms, genders, scales, species, or even abstract identities, users can explore perspectives inaccessible in the physical world.

For example, the ZEPETO platform allows users to extensively customise their avatars. Research (Kang & Rhee, 2025) has found that changing an avatar's gender presentation can significantly influence users' perceptions of their own gender identity, which can be especially empowering for those who are exploring or transitioning their gender in real life. Avatar embodiment can serve as a "rehearsal space" for identity transformation, supporting, let's say, an ontological experimentation through virtual embodiment. This isn't just escapism; it's a rehearsal space for empathy, resilience, and transformation.

3.2 A feedback loop between digital and physical identity

Traditionally, there are two perspectives of experience of personal identity that cannot be traced back to each other (Plessner, 2003): the self-testimony of inner experience and the hetero-testimony of outer experience. What makes XR's influence on identity this powerful is not just the capacity to change form, but the feedback loop it creates between digital and physical identity. The way we move as a dragon, breathe as a deity, or interact as a child in a virtual space can subtly inform our offline experience. Over time, these immersive rehearsals can become tools for re-patterning trauma, reframing self-esteem, or even confronting implicit bias. Recent advancements in adaptive VR systems have demonstrated the potential of integrating physiological signals, such as eye tracking and heart rate variability (HRV), to detect cognitive load and stress in real-time. For instance, (Nasri 2025) developed a framework where machine learning models, trained on data from tasks like the Stroop test, dynamically adjust VR training difficulty based on users' physiological responses. This approach not only enhances engagement but also supports emotional regulation during immersive experiences.

When AI joins this loop, processing emotional, biometric, and behavioural data to create avatars that evolve with us, our digital twin becomes less a copy and more a mirror—and even a lab—a dynamic construct reflecting our becoming, not just our being. Such integration suggests that avatars could adapt in real-time to our internal states, offering personalised interactions that reinforce positive identity formation and resilience.

3.3 A computational platform for creativity

More radically, the multisensory partiality of XR, its inability to fully replicate real-world experience, can itself be reframed as a space of creative mutation. Because XR cannot represent everything, it invites us to choose what to bring forward: it foregrounds what matters, amplifies what is hidden, and silences what is often oppressive. It allows users to distil their identities into symbolic, mythic, or speculative archetypes, something that physical reality rarely permits. Here, virtual embodiment becomes myth-making, and the avatar becomes a prosthesis not of the body, but of the soul. Projects like Soul Paint (Dew et al., 2020) that enable individuals to show their pain, can be redesigned and modulated to become a true means of communication for the incommunicable experiences of a person.

3.4 XR as a site for co-authored identity

XR is a laboratory ready to support instances of equality and inclusion at multiple levels, leading us to discoveries that could initiate a redesign of the experience, narrative and even offline environments. XR's shared spaces offer the potential for identity not just to be personal, but co-authored. In these synthetic environments, people can play with social dynamics in ways that challenge normative structures of gender, race, age, or ability. Inclusion can be designed and tested, rather than merely tolerated. Communities can form around shared speculative embodiment, and new social imaginaries can be prototyped in real time.

According to a 2023 report by Meta, users from Gen Z perceive the two identities, physical and avatar, as different and not necessarily consistent with each other; avatar's appearance, on the other hand, can influence the behaviour of users with previous relationships, when they engage with each other in VR (Rivu et al., 2021). This allows for a great space of experimentation, paving the way to designing experiences that can study and even challenge traditional social interactions.

If curated ethically, these spaces have the potential not only to heal but to disrupt. They can challenge inherited narratives, normalise multiplicity, and provide marginalised identities with tools to exist loudly, safely, and visibly (Tirinzoni & Megale, 2025). Used with intention, XR becomes not just a technology of presence, but of reclamation: of voice, of self, of future.

4 Challenges and risks

4.1 Sensory asymmetry of XR and the risk of fragmented identity

While XR offers new affordances for shaping identity, its very mechanisms raise epistemological and psychological concerns. Virtual Reality, as a multisensory simulation, engages vision, proprioception, and, when equipped, sound, haptics, and limited spatial cues. Yet, it still operates within a constrained sensory bandwidth when compared to the full spectrum of physical embodiment. Human identity is shaped through a rich entanglement of interoception, thermosensation, olfaction, balance, and tactile nuance. Partial stimulation, then, might not only produce a reduced representation of the self, but a fragmented one, where selected aspects of identity are overexpressed while others are silenced or underdeveloped. This asymmetry could influence users differently in the short term (e.g., altered emotional regulation or sensory detachment) or even in the long term, potentially reinforcing dissociative patterns or hyper-identification with idealised avatar traits, especially among vulnerable individuals.

In addition, exposure to models that force specific human characteristics in one direction or another, especially without proper literacy about said models and the technology that creates them, can alter not only the perception of the most fragile individuals, but influence ‘fashion ’in the strongest sense of the term linked to belonging and self-acceptance (Simmel, 1905) with consequences not only for individuals but for the resulting human communities and their interpersonal relationships. In the face of the emergence of virtual influencers and robot models, it is also necessary to focus on the influence that the implementation of these technologies has on human imagery, tastes, and choices.

4.2 The risk of identity echo chambers

VR experiences can be personalised through biofeedback and AI-based adaptation (Robobionics Content Writer, 2025; Kim, Lee, & Kim, 2025; Reeves et al., 2024; Reeves et al., 2022). While this can enhance engagement and tailored interventions, looking in the long term it may also create the conditions for identity echo chambers, environments that reflect, reinforce, and replicate the user's existing cognitive-emotional patterns without including potential divergence or challenge. Much like algorithmic bubbles in social media, such feedback loops in virtual identity may stifle growth by limiting confrontation with the unfamiliar, inhibiting critical self-reflection, and generating a sense of ontological comfort that borders on stasis. If one's avatar, virtual interactions, and environments continually reaffirm a narrow conception of the self, XR risks becoming a mirror rather than a window, amplifying identity redundancy instead of transformation. It is necessary to think of systems that avoid merely reinforcing existing identity patterns, provide comparisons and generally 'get out of the bubble'. On the other hand, the risks of behavioural manipulation must be considered, as we try to provide a meaningful and effective level of transparency of data and algorithms.

4.3 Blended identity

In shared, sentient XR spaces, where multiple users contribute to a continuous flow of biometric, behavioural, and expressive data, the resulting information landscape becomes inherently collective. Feedback from the environment is no longer directed solely at individuals but emerges from and responds to the aggregated presence of many. In this context, identity risks becoming a shared construct—blended and shaped by distributed inputs, including eye contact, voice inflexions, and avatar behaviours. Such collective dynamics can foster deeper social connections and immersive co-experiences, offering new pathways for identity exploration. Yet, they also introduce a liminal state where the boundaries of selfhood blur, and personal agency may be diluted within algorithmically mediated interactions. When the feedback loop is co-authored, we must ask not only where "I" ends and "we" begin, but whether this convergence may ultimately redefine what it means to be an individual at all.

The long-term implications of this could manifest in psychological enmeshment, surveillance-enabled behavioural nudging, or even avatar-based manipulation. The critical spirit, the positivity of solitude and of not being fully integrated into the system become relevant as values: the preservation of the possibility of opting out and the question of consent in general, which are often not resolved even in the offline world, find a new field of application here.

4.4 The hyperreal and the avatar as simulacrum

It could be possible to highlight a pattern unfolding that (Baudrillard, 1994) would have considered inevitable: the progressive replacement of the real by the hyperreal. In XR, identities are not merely represented but simulated with such fidelity and personalisation that they become more real than real, shaping user behaviour and self-perception more strongly than their physical lives. The avatar, once a vessel for exploration, could become a simulacrum: no longer pointing to a stable referent in the real world, but generating identity in a recursive loop through the virtual system. As users engage with these feedback loops, navigating environments tailored to their biometrics, echoing their past actions and predicted futures, one must ask: is the avatar expressing the self, or is the system scripting the self through the avatar? Once again, literacy and a focus on transparency can be powerful weapons to manage these risks, together with strategies to create an unambiguous continuity of meaning between online and offline, as in (Floridi, 2015) work, that do not leave anyone behind (think of the hikikomori phenomenon and similar situations of discomfort and disconnection/hyperconnection).

If we are to use XR as a laboratory for identity, we must design it with friction, ambiguity, and contradiction—essential ingredients of human growth. Without them, we risk losing the boundary not between virtual and physical, but between authenticity and simulation. And it is in this liminal space, not between screens and senses but between ontology and code, where the future of human identity will be negotiated.

4.5 Regulating the new liminal spaces shaping human identity

In the absence of a common body of laws for virtual worlds and avatars, we must resort to legislation which is in force within different domains. The use of AI-driven avatars, which come close to the concept of the digital shadow of a human being, can nowadays face legal challenges from different perspectives. First of all, the amount of personal data that should be collected is vast and includes biometric data and other kinds of sensitive information, the use of which is highly regulated in some countries, e.g. by the European GDPR. While the use of such data may be lawful and indeed desired within some fields of application, such as healthcare treatment, all the commercial areas, like entertainment and marketing, are much more problematic in terms of a lawful basis for the processing of personal information.

Even if ideally the regulatory set-up should be harmonised globally, an international legal framework for virtual worlds (also referred to as 'law of the Metaverse') does not exist. If we look back at the early years of cyberspace, similar considerations were made by scholars and enthusiasts, which have not resulted in a commonly agreed-upon framework, mainly due to different views among stakeholders (Qin et al., 2025).

5 New challenges and future scenarios: a research proposal

5.1 Identity experimentation made affordable

XR avatars can enable users to challenge traditional gender/racial constructs, creating spaces for marginalised groups to explore alternative identities safely. On the other hand, AI let us collect data and brainstorm on it, letting us create a "rationale" of the identity: will this be a new form of self-discovery and awareness?

5.2 Legal personality and avatars

(Cheong, 2022) argues that avatars equipped with AI, capable of learning from their human users to make decisions, manage contracts, and supervise others in the virtual worlds, could in the future be considered endowed with legal personality in virtual environments. This would grant them rights like property and personality rights: the growing debate around granting rights to highly autonomous AI agents may shape this discussion. What would it mean, legally and ethically?

5.3 Is identity in XR truly 'liberated', or are we building new cages with different aesthetics?

To what extent does identity exploration in XR environments genuinely expand individual agency, or does it merely replace existing social constraints with novel, technologically mediated limitations? Can avatars be used not just to escape reality, but to reshape it? As Oscar Wilde said, 'Give him a mask and he will tell you the truth'. This balance between self-expression and self-delusion should be explored both as a challenge and as a tool with deep psychological and societal influences. What are the implications? And how can we protect the information about our personality that the data provided by our choices and interactions can show to the platforms we use?

5.4 If we allow machines to design our virtual selves, whose values are we encoding in our identity?

How can we 'open the black box 'to align what we can call identity generators to the values we want to preserve? And on the other hand, what if there were a way to make the biases explicit through XR-mediated interactions, and thus raise awareness and find new ways to overcome them?

5.5 What might be lost when the avatar becomes more 'relevant' than the body?

How can we define metrics to make these reproductions valid, and the data connected to them trustworthy? What lines should we draw to distinguish between the real and the made-up identity? Is there a clear threshold or a blurred boundary? How can we test it? What strategies for ensuring continuity between online and offline life (Floridi, 2015), should be adopted?

5.6 AI-avatars as means for literacy and awareness

Could avatar-mediated immersive experiences become a vehicle for literacy and enhanced awareness not only about the self but also about personal and societal issues? How should we modulate the language we use to talk about these phenomena? What will be the differences in interaction here?

References

Agarwal, A., Ramachandra, R., Venkatesh, S., & Prasanna, S. R. (2024). Biometrics in extended reality: A Review. Discover Artificial Intelligence, 4(1). https://doi.org/10.1007/s44163-024-00190-9

Bhatt, S., Jogy, S., & Puri, A. (2024). Integration of virtual reality (VR) and artificial intelligence (AI) in autism therapy. International Journal of Science and Research Archive, 12(1), 2400–2405. https://doi.org/10.30574/ijsra.2024.12.1.1006

Baudrillard, J. (1994). Simulacra and Simulation Jean Baudrillard ; translated by Sheila Faria Glaser. The University of Michigan Press.

Cheong, B. C. (2022). Avatars in the metaverse: Potential legal issues and remedies. *International Cybersecurity Law Review*, *3*(2), 467–494. https://doi.org/10.1365/s43439-022-00056-9

Floridi, L. (Ed.). (2015). *The Onlife Manifesto*. Springer International Publishing.

Kang, J., & Rhee, H. (2025). Gender identity and perception in virtual spaces: The impact of avatar gender transition on the Zepeto Platform. *Frontiers in Virtual Reality*, *6*. https://doi.org/10.3389/frvir.2025.1505624

Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. Presence: Teleoperators and Virtual Environments, 21(4), 373–387. https://doi.org/10.1162/pres_a_00124

Kim, S., Lee, Y., & Kim, K. (2025). Gait training with virtual reality-based real-time feedback for chronic post-stroke patients: A pilot study. Healthcare, 13(2), 203. https://doi.org/10.3390/healthcare13020203

Lau, K. H. C., Sen, S., Stark, P., Bozkir, E., & Kasneci, E. (2024). Personalized generative AI in VR for enhanced engagement: Eye-tracking insights into cultural heritage learning through Neapolitan pizza making. arXiv. https://arxiv.org/abs/2411.18438

Meta. (2023, February). *Culture rising: 2023 trends report*. Meta Foresight. https://www.facebook.com/business/news/insights/culture-rising-2023-trends-report

Metahuman Technology brings a 10,000-year-old shaman back to life - Unreal Engine. (2022). https://www.unrealengine.com/en-US/spotlights/metahuman-technology-brings-a-10-000-year-old-shaman-back-to-life

Methuku, V., & Myakala, P. K. (2025). Digital Doppelgangers: Ethical and Societal Implications of Pre-Mortem AI Clones. https://doi.org/10.32388/9x78no

Nasri, M. (2025). Towards Intelligent VR Training: A Physiological Adaptation Framework for Cognitive Load and Stress Detection. arXiv. https://arxiv.org/abs/2504.06461

Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., Willer, R., Liang, P., & Bernstein, M. S. (2024, November 15). Generative agent simulations of 1,000 people. Generative Agent Simulations of 1,000 People. http://export.arxiv.org/abs/2411.10109

Plessner, H. (2003b). Die stufen des organischen und der Mensch Einleitung in die philosophische Anthropologie Helmuth Plessner. Wissenschaftliche Buchgesellschaft.

Purushothaman, S., Himabindu, A., Bhuvaneswari, P., Chellammal, P., Vats, S., & Merlin Sheela, P. (2025). Immersive virtual reality infused artificial intelligence interfaces for hyper personalized digital engagement and enhanced user experience. In Proceedings of the International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 2024) (pp. 720–731). Atlantis Press. https://doi.org/10.2991/978-94-6463-718-2_62

Qin, H. X., Wang, Y., & Hui, P. (2025). Identity, crimes, and law enforcement in the metaverse. Humanities and Social Sciences Communications, 12(1). https://doi.org/10.1057/s41599-024-04266-w

Rivu, R., Zhou, Y., Welsch, R., Mäkelä, V., & Alt, F. (2021). When friends become strangers: Understanding the influence of avatar gender on interpersonal distance in virtual reality. Lecture Notes in Computer Science, 234–250. https://doi.org/10.1007/978-3-030-85607-6_16

Reeves, N., Norman, Z., Djokovic, A., Handley, S., Kennedy, P., Bothwell, S., & Han, H. (2024). Adapting a virtual reality experience for a user based on a mood improvement score (U.S. Patent No. 12,175,385). U.S. Patent and Trademark Office. https://patents.justia.com/patent/12175385

Reeves, N., Norman, Z., Djokovic, A., Handley, S., Kennedy, P., Bothwell, S., & Han, H. (2022). Adapting a virtual reality experience for a user based on a mood improvement score (U.S. Patent No. 11,537,907). U.S. Patent and Trademark Office. https://patents.justia.com/patent/1153790

Robobionics Content Writer. (2025). Virtual reality and AI in gait rehabilitation: What's new and working. Robo Bionics. https://www.robobionics.in/blog/virtual-reality-and-ai-in-gait-rehabilitation-whats-new-and-working/

Simmel, Georg Philosophie der Mode Moderne Zeitfragen, Nr. 11, hg. von Hans Landsberg, Berlin o .J. (1905), 5-41

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. PLoS ONE, 5(5). https://doi.org/10.1371/journal.pone.0010564

Tirinzoni, M., & Megale, V. (2025). Reimagining identity, healing, and equality through extended reality. Advances in Wireless Technologies and Telecommunication, 149–186. https://doi.org/10.4018/979-8-3693-7147-3.ch008

Velez Quintero, L. E. (2023). User modeling for adaptive virtual reality experiences (Doctoral dissertation, Stockholm University). Stockholm University. http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-222210

Yasuda, A. (2024). Metaverse ethics: Exploring the social implications of the metaverse. AI and Ethics. https://doi.org/10.1007/s43681-024-00507-5

Yee, N., & Bailenson, J. (2007). The Proteus Effect: The effect of transformed self-representation on behavior. Human Communication Research, 33(3), 271–290. https://doi.org/10.1111/j.1468-2958.2007.00299.x

Yim, Y., Xia, Z., Kubota, Y., & Tanaka, F. (2024). The Proteus Effect on Human Pain Perception: Effects of Muscular Avatar Use and Gender Factors. https://doi.org/10.21203/rs.3.rs-3905832/v1

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Extending Human–Co-Bot Interaction with XR Assistants

Xinyuan Zhou [1], Pritom Kumar Das [1], Valeria Orso [2], Sergei Smirnov [1], Patrik Pluchino [2], Rajeev Kanth [3], Aki Happonen [3], Agustin Zuniga [1], Laura Ruotsalainen [1], Niklas Ravaja [1], Andrea Bellucci [4], David Kirsh [5], Luciano Gamberini [2], Michael Mathioudakis [1], Jussi Kangasharju [1], Pekka Marttinen [6], Sasu Tarkoma [1], Petteri Nurmi [1], Giulio Jacucci [1]

[1] *University of Helsinki, Finland,* [2] *University of Padova, Italy,* [3] *Savonia University of Applied Sciences, Finland,* [4] *Universidad Carlos III Madrid, Spain,* [5] *University of California, San Diego, USA,* [6] *Aalto University, Finland*

Correspondence*:

ABSTRACT

Industry 5.0 prioritizes human-centric technologies that enhance, rather than replace, human labor, particularly in complex manufacturing environments. This paper investigates the use of XR-based intelligent agents, designed to support operators working alongside collaborative robots (co-bots) in industrial settings. We review current human–robot collaboration practices to identify limitations and opportunities for operator augmentation. Building on these insights, we propose an extended interaction framework that integrates the Smart Factory metaverse, digital twins, and embodied AI assistants to enable scalable collaboration across multiple workstations. To support this framework, we present a prototype of a Smart Factory featuring a digital twin environment and explore the integration of embodied conversational agents (ECAs) for real-time, personalized operator assistance. Our review of current uses of ECAs in manufacturing highlights key use cases, enabling technologies, and prevalent challenges. We conclude with a discussion of the technical and human-centered considerations for deploying adaptive, interactive systems within industrial metaverse environments.

Keywords: Industrial Metaverse, Smart Factory, Human-Computer Interaction, Digital Twin, Collaborative Robots, Cyber-physical Systems

1 INTRODUCTION AND BACKGROUND

Modern manufacturing is increasingly complex, with operators managing advanced machinery and robotics in dynamic environments. A comprehensive review by Longo et al. (2017) highlights that agent-based technologies could enable flexible, decentralized control, improving operator efficiency and autonomy (Pulikottil et al., 2023; Kumar and Lee, 2022). Advanced human-computer interaction (HCI) tools enhance operators' ability to plan, optimize, and control systems, while continuous training supports their efficiency and well-being.

This paper explores human–robot collaboration use cases to identify limitations and proposes an extended interaction framework integrating the Smart Factory Metaverse and embodied AI assistants to overcome these challenges. Section 2 examines operator use cases, including a digital twin, to inform the framework. Section 3 presents a mini-review of the application of embodied conversational agents (ECAs) in manufacturing, a Smart Factory prototype, and a large language model (LLM)-powered ECA. We conclude with a research roadmap for deploying this framework in industrial metaverse settings.

1.1 The Smart factory, Cyber-Physical Systems, and Digital Twins

Modern manufacturing leverages cyber-physical systems (CPS) and digital twins (DTs) to enhance efficiency and operator training. CPS integrates computing, control, and communication with physical environments, while DTs simulate real-world processes for optimization. Essential for their functionality is real-time data exchange, often facilitated by Open Platform Communications Unified Architecture (OPC UA), a platform-independent standard enabling secure, reliable interoperability across industrial applications like Enterprise Resource Planning (ERP) and high-performance VR/AR systems. Key technologies for building industrial metaverse DTs include AI, VR, high-speed networks like private 5G, and computing power, ensuring dynamic simulation and operator support in smart factories.

1.2 Human-Machine Interfaces in the Smart Factory

A review by Kumar and Lee (2022) identifies advanced interaction modalities—gaze, voice, gesture, tactile, and haptic—enhancing smart factory functions like operation, management, maintenance, and cybersecurity. These interfaces require robust cybersecurity and human-centered design to ensure usability from cognitive and physical perspectives. Generative AI transforms embodied conversational agents (ECAs) in VR/AR, offering immersive training through human-like interactions (Bellucci et al., 2025). ECAs with situational awareness improve engagement and learning outcomes (Kán et al., 2023). Integrated with digital twins and AI, ECAs support human-robot collaboration, real-time assistance, and process optimization (Zhang et al., 2024). The industrial metaverse, powered by digital twins, enables scenario simulation and decision-making. However, challenges remain in deploying human digital twins (HDTs), including standardization, reliability, data accuracy, and ethical concerns such as privacy (Wang et al., 2024a).

1.3 Integrating XR agents in co-bot interaction

The primary aim of this paper is to investigate how XR agents, specifically embodied assistants and conversational agents, can support operators of collaborative robots within industrial metaverse environments. We focus on their potential to enhance human–robot collaboration by improving usability, reducing cognitive load, and enabling more natural, multimodal interaction in smart factory settings. We present a concise review of current implementations and empirical evaluations of XR agents in manufacturing, identifying key use cases, underlying technologies, and remaining challenges. By synthesizing insights from this review, we propose an extended interaction framework that integrates digital twins, LLM-powered assistants, and XR interfaces to augment operator roles in increasingly complex and distributed production environments.

2 THE OPERATOR 5.0 AND THE CO-BOT DIGITAL TWIN USE CASE

2.1 Use Cases and Operators Augmentations

In the framework of Industry 5.0, co-bots represent a significant advancement as they extend human capabilities by assisting with physically and cognitively demanding tasks. This support enables human operators to focus on more rewarding and engaging activities, which is particularly beneficial for individuals with physical limitations (e.g., Kildal et al. (2019)) or ageing workers (e.g., Rossato et al. (2021a)). Moreover, Extended Reality (XR) technologies make it possible to hybridize the workspace by integrating a digital twin of the collaborative robotic arm, allowing users to interact with it in more or less direct ways. As a result, the operative space of the worker expands beyond the physical boundaries of the workstation: users can access digital information not immediately visible in the physical world, interact with the cobot's digital twin in the virtual environment, and produce effects in the physical world. Understanding the impact of such technologies on the user is essential to designing interaction modalities that align with both user needs and organizational goals. For instance, evidence from research shows a preference for direct interaction with the collaborative robot that extends also to virtual environments. In a study comparing direct manipulation of a virtual robotic arm versus indirect control via a handheld controller, Nenna et al. (2023a) observed higher cognitive load in the controller-mediated condition. Moreover, most participants expressed a preference for direct interaction. The facilitation effect of direct control was further confirmed in AR-based interaction with the robotic arm's digital twin. Grego et al. (2024) found that the AR interface was perceived as more usable and less fatiguing than the dedicated touch-screen interface, with participants also reporting higher perceived accuracy and speed. Notably, the sense of presence—defined as the subjective feeling of being immersed in the virtual environment—also plays a critical role. Higher levels of presence were associated with faster completion times in a virtual pick-and-place task, whether the cobot was controlled directly or via a controller (Nenna et al., 2023b).

In addition, research has indicated a positive attitude towards collaborative robots (cobots) designed for operators' work augmentation: both younger and older employees display a high level of acceptance towards cobots (Rossato et al., 2021b).

Although these findings stem from laboratory settings, they clearly point to the benefits of direct interaction with both physical and digital versions of robotic systems. These benefits include improved interaction efficiency and perceived effectiveness. Moreover, the studies highlight the importance of addressing the unique characteristics of virtual environments—particularly the sense of presence—as key factors influencing workload and overall user experience. Additionally, they stress the importance of carefully designing the collaborative system features.

2.2 Digital Twin of a Sentient Collaborative Robot

The concept of the digital twin has advanced from mere static replication toward more interactive, adaptive, and human-centered systems that exemplify the collaborative principles of Industry 5.0. Recent developments underscore the utilization of digital twins not solely as simulators of robotic behavior, but also as agents proficient in real-time co-regulation with human users.

Other contributions have proposed teleoperation frameworks utilizing virtual reality (VR) and real-time eye-tracking technologies, which will enable virtual robots in the near future to interpret attentional focus and subsequently adjust their assistance or information delivery accordingly (Nenna et al., 2023c). This conceptualization enhances operational precision and promotes the mutual intelligibility of human and machine intentions within shared task environments. Similarly, studies on virtualized human–robot collaboration suggest that immersive environments can enhance task coordination and UX by adjusting to users' cognitive workload and attentional patterns. Evidence from experimental studies on physiologically adaptive systems leveraging metrics such as pupil dilation and gaze entropy has demonstrated predictive capabilities for workload perception during robotic tasks (Wu et al., 2020; Fournier et al., 2024). Furthermore, the incorporation of real-time physiological data into human–robot interaction paradigms has demonstrated the capacity to foster more intuitive and collaborative industrial environments, wherein robots proficiently 'interpret' and react to human cognitive and psychological states. Moreover, research suggests that employing machine learning methodologies for the detection and monitoring of operator fatigue can substantially improve the sustainability of prolonged human–robot collaboration (Lambay et al., 2024). Previous research corroborate the potential of digital twins to function as physiologically responsive entities that adjust robotic behavior in response to the user's state, thereby enhancing the sustainability of long-term collaboration (Baratta et al., 2024). Such insights highlight the relevance of systems that are informed by physiological data, in which biometric feedback (e.g., eye-tracking, heart activity) can be utilized to enhance adaptive interfaces. Research indicates that these adaptive systems improve accuracy, mitigate cognitive strain, and facilitate a more rapid resumption of tasks following interruptions. Supporting this perspective, studies have further demonstrated the significance of incorporating both implicit indicators (such as psychophysiological measures) and explicit indicators (such as self-reports) of mental workload within industrial contexts. This highlights their complementary roles in assessing human-technology interaction and informing the design of adaptive support systems (Mingardi et al., 2020).

The emerging paradigm of sentient collaborative robots indicates a transition from mere automation to empathetic augmentation, wherein digital twins serve as cognitive collaborators that anticipate needs, alleviate mental stress/overload, and maintain operator engagement across dynamic manufacturing environments (Nikolaidis et al., 2017).

3 EXTENDING THE RESEARCH WITH XR AGENTS

Previous research has mainly focused on improving collaborative robot interactions at single workstations, emphasizing direct interaction and digital twins for operator support. This approach, however, restricts the investigation of multi-agent, cross-island collaborations crucial to modern manufacturing systems. The Smart Factory Lab offers a unique modular environment to integrate collaborative robots, digital twins, and AI agents across multiple production stations. This infrastructure enables the exploration of more dynamic and scalable human-robot collaborations, addressing both task efficiency and the socio-cognitive aspects of operator interactions in distributed smart manufacturing environments.

3.1 Minireview on the application of ECAs in manufacturing

This mini review was conducted through a structured search of the Scopus database, focusing on embodied conversational agents (ECAs) used within industrial metaverse contexts. The initial query targeted peer-reviewed articles and conference papers using terms such as "embodied conversational agent,","virtual assistant","avatar-based agent", "avatar," "virtual agent," and "digital twin," combined with "industry 4.0," "smart manufacturing," and "industrial metaverse."

The search yielded 87 records. From these, we applied three exclusion criteria: (1) exclusion of works lacking a real implementation of an ECA or empirical evaluation (59 papers excluded), (2) exclusion of studies published before 2020 (5 papers excluded), and (3) exclusion of works unrelated to manufacturing or industrial application contexts (9 papers excluded).

The final corpus consisted of 14 studies. Each paper was manually examined to confirm its relevance and was categorized based on use cases, platforms used, and technological components. A BibTeX database and structured LaTeX table were created to document the included studies.

In Table 1, we present the results of a review on the application of ECAs in manufacturing, categorized by use cases and including keywords representing the technologies used.

Table 1. Representative studies of embodied conversational agents (ECAs) in different industrial metaverse use cases.

Use Case	Papers	Technologies / Platforms
Training Simulation	Kang and Jo (2024), Li et al. (2022), Alpala et al. (2022), Ramadhani et al. (2023)	Unity, Unreal, MRTK toolkit, VR headset, Simulation Engine, speech-to-text (STT) and text-to-speech (TTS), Multi-User System, Avatar System
Remote Maintenance	Oppermann et al. (2023)	VR headset, WebRTC, Avatar Interaction, Cloud Remote Rendering
Human-Robot Collaboration	Fan et al. (2025), Li et al. (2023), Li and Yang (2021), Niiyama et al. (2023), Xu et al. (2025)	LLMs, Chain of Thought, Fine-Tuned BERT Model, Natural Language Processing (NLP), Task-Oriented Response Model Generation, Semantic Web, Motion Capture Systems, Multi-Modal Large Models (MLMs), Cyber-Physical-Social Systems (CPSSs), Cloud Computing, Embodied Perception and Interaction, Scheduling Algorithms
Monitoring Supervision	Barron-Lugo et al. (2025)	Digital Avatars, cloud services, Virtual Containers
Digital Twin Interaction	Bellalouna and Puljiz (2023), Chen et al. (2024), Gautam et al. (2025)	Digital Twin, Multiplayer Avatar interactions, Unreal Engine, 3D Reconstruction,digital representations of robots, Industry Foundation Classes (IFC), Multi-Agent Framework, LLMs, Retrieval Augmented Generation (RAG)

3.2 The integration of XR Assistants

The integration of XR agents and assistants into manufacturing signifies a major advancement in operator augmentation within the Industry 5.0 framework. Unlike traditional automation that prioritizes efficiency over human interaction, these assistants act as intuitive collaborators, overlaying interactive guidance (3D models, animations, or text) on physical workspaces to enhance workers' confidence and precision in complex assembly tasks.

Current state-of-the-art (SoTA) research underscores the role of embodied agents in reducing cognitive load and improving situational awareness. For instance, (Lataifeh et al., 2025) proposed a multimodal framework that combines large language models (LLMs) like GPT-4o with computer vision to create intelligent virtual agents (IVAs) capable of navigating physical spaces and providing adaptive guidance based on real-time scene understanding, which makes the XR agent productive in industrial setting.

Architecture: ECAs typically follow a modular and hybrid architecture, combining components from natural language processing, speech, vision, and robotics. At the core, LLMs like BERT or GPT variants manage language understanding and generation. For instance, Bot-X employs a two-layer neural network with Softmax regression (Li and Yang, 2021), while Max uses a fine-tuned BERT model with a Client–Server architecture and RESTful APIs for scalability (Li et al., 2023). XR components include VR/MR headsets (e.g., HoloLens 2, Meta Quest 2/3), gaming engine((e.g., unity, unreal), motion capture systems and multi-agent system. For multimodal interaction, ECAs utilize speech recognition technologies and visual perception tools for gesture and facial recognition. Cloud computing platform, like Azure, ensure data synchronization and real-time digital twin visualization, as seen in the LLM-integrated digital twin study (Gautam et al., 2025).

Use Cases: XR assistants support diverse applications in smart manufacturing, aligning with Industry 5.0's human-centric paradigm. For example, remote maintenance empowers experts to address machine malfunctions remotely, reducing costly downtime and travel. By providing real-time guidance through digital twins, XR avatars enhance operational efficiency (Oppermann et al., 2023). A notable example is the Max, a natural language-enabled virtual assistant, which uses a fine-tuned BERT model and human-inspired conversation strategies to facilitate intuitive human robot collaboration in industrial settings (Li et al., 2023). Training simulation is a good use case to reduce training costs. By simulating real-world industrial scenarios, XR assistants help operators to improve adaptability and confidence in handling complex machinery or processes. A VR-powered metaverse framework for smart factories illustrates this, using digital twins and avatars to deliver engaging, collaborative training experiences (Alpala et al., 2022). Monitoring supervision ensures real-time oversight of industrial processes. The Davatar method creates digital avatars to monitor and interact with containerized applications. (Barron-Lugo et al., 2025). One example of digital twin interaction is roboAvatar, which uses IFC-based digital twins for construction robotics (Chen et al., 2024).

Limitations: Despite these advancements, challenges remain in deploying ECAs in real-time industrial settings. Latency poses significant concerns.Cloud-based LLMs and XR rendering cause delays, as seen in the remote collaboration system (Oppermann et al., 2023). Edge deployment is limited by hardware constraints (Gautam et al., 2025). Privacy issues arise when handling sensitive data. Davatar's use of unencrypted HTTP and Max's lack of robust security highlight risks (Barron-Lugo et al., 2025; Li et al., 2023). Hardware limitation could cause synchronization problem. In the remote maintenance prototype, hardware limits real-time digital twin updates (Oppermann et al., 2023). Additionally, complex interfaces can overwhelm non-technical users and long XR using makes people feel tiring, which lowering effectiveness and usability (Bellalouna and Puljiz, 2023) . Therefore, future XR assistants need low-latency edge computing, secure protocols and intuitive interfaces to ensure Industry 5.0 integration.

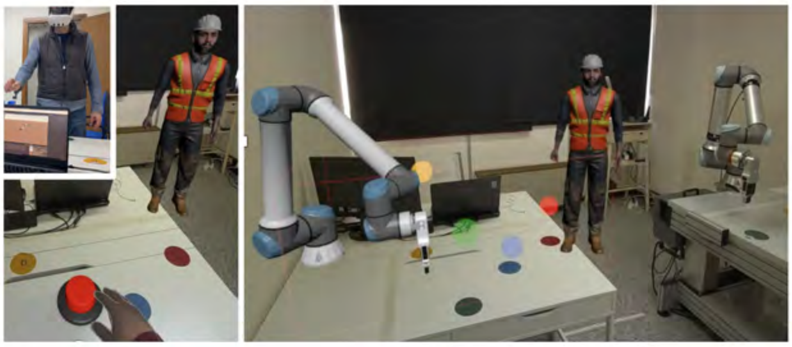

Figure 1. Co-bot prototype in AR with an embodied AI-powered assistant in AR (Nenna et al., 2023a).

4 CHALLENGES AND DISCUSSION

This study proposes an extended interaction framework that integrates the Smart Factory metaverse with embodied AI assistants, aiming to overcome the limitations of current human–co-bot applications within the context of Industry 5.0. The findings indicate that direct and intuitive interaction, supported by digital twins (DTs) and embodied conversational agents (ECAs), improves user experience, enhances task efficiency, and fosters a greater sense of presence. Nonetheless, implementing such systems in real-world environments remains challenging due to persistent issues such as system latency, privacy concerns, and the need to establish user trust. Another important area that warrants further attention is the facilitation of social interaction among operators in the industrial metaverse, which is closely connected to the characteristics of the devices used for access and engagement (Fernández-Caramés and Fraga-Lamas, 2024).

4.1 Trust and Distrust in AI Systems

Trust in AI plays a central role in the acceptance and adoption of intelligent systems in industrial settings (Salam et al., 2021). It is a multifaceted and dynamic construct influenced by a wide range of factors. According to Kaplan et al. (2023) they can be broadly categorized into human-related, AI-related, and context-related dimensions. Human-related factors include the operator's level of competence and understanding of the AI system, domain expertise, and personality traits (e.g., extroversion), which have been shown to affect the willingness to collaborate with intelligent systems. AI-related factors pertain to system performance, such as reliability, and use of human-centered language. Notably, anthropomorphic AI agents have been found to increase trust. Finally, contextual factors further shape this relationship, encompassing team dynamics (e.g., the length of time a team has worked together) and situational variables, such as the perceived level of risk of the task at hand (Kaplan et al., 2023). Together, these elements interact to influence the degree of trust a user places in an AI system.

Importantly, trust in AI is not static. It fluctuates in response to task complexity, perceived changes in system performance, and broader contextual variables (Wang et al., 2024b). Alongside trust, the concept of distrust has emerged as a significant factor. Specifically, distrust is not merely the absence of trust, rather, it encompasses users' suspicion toward AI systems and fears of negative consequences if the system performs unreliably (Govers et al., 2025; Kaplan et al., 2023). Distrust becomes particularly salient when users interact with anthropomorphized AI agents, which can exert persuasive influence (Gefen et al., 2025). A key emerging issue related to trust in the AI system is over-reliance, as research has shown that individuals tend to accept and incorporate AI responses without being fully aware of it (Jakesch et al., 2023).

Explainability and Mental Models: A significant contributor to both trust and distrust is the opacity of most AI systems. These systems often operate as ”black boxes,” offering limited insight into how outputs are generated (Hassija et al., 2024). This lack of transparency and explainability complicates users' ability to form accurate mental models of AI functioning. While researchers have proposed various explanation strategies, there is still no consensus on what information is helpful for users to grasp the system functioning, nor how it should be presented. Endsley (2023) emphasizes that effective explanations must be context-specific and acknowledges the intrinsic challenge of translating brittle algorithmic logic into human-understandable narratives. Moreover, once users form a mental model of how an AI system works, they are often reluctant to revise it, even in the face of contradictory evidence. These reflections point out the key challenge of designing systems that foster calibrated trust, thereby enabling users to rely on AI and maintain critical oversight. To do that, adaptive explainability tailored to users' mental models and contexts should be sought and experimented.

4.2 Digital Twins and Social Dynamics at Work

Much of the current research on Digital Twin (DT) technologies has concentrated on individual interactions, evaluating their impact on performance, cognitive workload, and user satisfaction. However, this perspective overlooks the inherently social nature of the workplace, which is shaped by continuous interactions, informal exchanges, and collaborative dynamics among workers. While prior studies have explored how the introduction of physical collaborative robots influences social dynamics within workgroups (Cheon et al., 2022a,b), the impact of DT systems on peer-to-peer relationships, both in physical and virtual environments, remains largely unexamined.

AI Systems as Social Partners: Emerging evidence shows that workers often interact with and refer to AI-powered industrial machines as if they were colleagues, even when these systems lack anthropomorphic features (Süße et al., 2023). This raises important questions about how the design of AI-based systems might influence social perceptions and team cohesion. For example, Crivelli et al. (2025) suggests that integrating features that express internal states, such as facial expressions or responsive gestures, could strengthen human-robot rapport, enhancing both trust and perceived collaboration quality. Building on this idea, Nicora et al. (2023) has developed collaborative robots coupled with virtual agents with verbal and non-verbal social skills, to increase their likelihood of being perceived as social partners.

4.3 Role of XR in Team Building

As AI systems become more embedded in knowledge-intensive and collaborative workplaces, concerns have emerged about their potential to diminish opportunities for informal social interactions. Such interactions are essential for building team cohesion and fostering spontaneous innovation. Technologies such as extended reality (XR) have been proposed as a way to recreate casual, serendipitous encounters in virtual environments and preserve the social fabric of work. The results of the minireview shows the recent adaptation and emerging use cases of XR technologies with LLM-powered ECAs in different areas of the industrial metaverse.

Despite these developments, existing research remains largely centered on dyadic interactions between individual workers and AI systems, overlooking how such technologies affect broader team dynamics and social connectedness. As AI and DT systems become more embedded in collaborative settings, it is essential to investigate their influence not only on individual performance, but also on the social structures and relationships that underpin effective teamwork.

5 CONCLUSIONS AND FUTURE DIRECTIONS

Several open challenges remain in the effort to integrate AI and DT systems into Industry 5.0 in a human-centric manner. There is a need to design systems that support meaningful social interaction within the industrial metaverse, particularly through interfaces that are intuitive and socially responsive (Fernández-Caramés and Fraga-Lamas, 2024). Additionally, developing AI systems that foster appropriate levels of trust while remaining explainable and transparent is a critical objective. Lastly, future research should continue to examine how these technologies influence not only productivity but also the collaborative and social dimensions of work. Addressing these concerns will be crucial for creating adaptive, inclusive, and trustworthy systems aligned with the principles of Industry 5.0.

REFERENCES

Alpala, L. O., Quiroga-Parra, D. J., Torres, J. C., and Peluffo-Ordóñez, D. H. (2022). Smart factory using virtual reality and online multi-user: Towards a metaverse for experimental frameworks. *Applied Sciences* 12, 6258

Baratta, A., Cimino, A., Longo, F., and Nicoletti, L. (2024). Digital twin for human-robot collaboration enhancement in manufacturing systems: Literature review and direction for future developments. *Computers & Industrial Engineering* 187, 109764

Barron-Lugo, J. A., Lopez-Arevalo, I., Gonzalez-Compean, J. L., Morin-Garcia, J. C., Crespo-Sanchez, M., and Carretero, J. (2025). An avatar cloud service based method for supervising and interacting with containerized applications. *Expert Systems with Applications* 269, 126447

Bellalouna, F. and Puljiz, D. (2023). Use case for the application of the industrial metaverse approach for engineering design review. *Procedia CIRP* 119, 638–643

Bellucci, A., Jacucci, G., Trung, K. D., Das, P. K., Smirnov, S. V., Ahmed, I., et al. (2025). Immersive tailoring of embodied agents using large language models. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)* (IEEE), 392–400

Chen, J., Lu, W., Pan, Y., and Fu, Y. (2024). Building "roboavatar": industry foundation classes–based digital representation of robots in the built environment. *Journal of Computing in Civil Engineering* 38, 04024013

Cheon, E., Schneiders, E., Diekjobst, K., and Skov, M. B. (2022a). Robots as a place for socializing: Influences of collaborative robots on social dynamics in-and outside the production cells. *Proceedings of the ACM on Human-Computer Interaction* 6, 1–26

Cheon, E., Schneiders, E., and Skov, M. B. (2022b). Working with bounded collaboration: A qualitative study on how collaboration is co-constructed around collaborative robots in industry. *Proceedings of the ACM on Human-Computer Interaction* 6, 1–34

Crivelli, D., Balconi, M., et al. (2025). Industrial collaborative robotics, humanoid cobots, and social embodiment: A critical look on human–robot interaction. *International Journal of Humanoid Robotics* 22, 1–16

Endsley, M. R. (2023). Supporting human-ai teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior* 140, 107574

Fan, H., Liu, X., and Fuh, J. (2025). EnglishEmbodied intelligence in manufacturing: leveraging large language models for autonomous industrial robotics. *Journal of Intelligent Manufacturing* 36, 1141 – 1157. doi:10.1007/s10845-023-02294-y

Fernández-Caramés, T. M. and Fraga-Lamas, P. (2024). Forging the industrial metaverse towards industry 5.0: Where augmented/mixed reality, iiot, opportunistic edge computing and digital twins meet. *IEEE Access*

Fournier, É., Jeoffrion, C., Hmedan, B., Pellier, D., Fiorino, H., and Landry, A. (2024). Human-cobot collaboration's impact on success, time completion, errors, workload, gestures and acceptability during an assembly task. *Applied Ergonomics* 119, 104306

Gautam, A., Aryal, M. R., Deshpande, S., Padalkar, S., Nikolaenko, M., Tang, M., et al. (2025). EnglishIiot-enabled digital twin for legacy and smart factory machines with llm integration. *Journal of Manufacturing Systems* 80, 511 – 523. doi:10.1016/j.jmsy.2025.03.022. Cited by: 0

Gefen, D., Leffew, B. A., Ragowsky, A., and Riedl, R. (2025). The importance of distrust in trusting digital worker chatbots. *Communications of the ACM* 68, 42–49

Govers, J., Pareek, S., Velloso, E., and Goncalves, J. (2025). Feeds of distrust: Investigating how ai-powered news chatbots shape user trust and perceptions. *ACM Transactions on Interactive Intelligent Systems*

Grego, G., Nenna, F., and Gamberini, L. (2024). Enhancing human-machine interactions: a novel framework for ar-based digital twin systems in industrial environments. In *Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments*. 456–462

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., et al. (2024). Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation* 16, 45–74

Jakesch, M., Bhat, A., Buschek, D., Zalmanson, L., and Naaman, M. (2023). Co-writing with opinionated language models affects users' views. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–15

Kán, P., Rumpelnik, M., and Kaufmann, H. (2023). Embodied conversational agents with situation awareness for training in virtual reality. In *Proceedings of the International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, eds. J.-M. Normand, M. Sugimoto, and V. Sundstedt (Eurographics Association)

Kang, B.-H. and Jo, D. (2024). Effects on Verbal Interaction of Multiple Virtual Humans in the Augumented Reality Environment. In *2024 International Workshop on Intelligent Systems (IWIS)*. 1–4. doi:10.1109/IWIS62722.2024.10706032

Kaplan, A. D., Kessler, T. T., Brill, J. C., and Hancock, P. A. (2023). Trust in artificial intelligence: Meta-analytic findings. *Human factors* 65, 337–359

Kildal, J., Martín, M., Ipiña, I., and Maurtua, I. (2019). Empowering assembly workers with cognitive disabilities by working with collaborative robots: a study to capture design requirements. *Procedia CIRP* 81, 797–802

Kumar, N. and Lee, S. C. (2022). Human-machine interface in smart factory: A systematic literature review. *Technological Forecasting and Social Change* 174, 121284

Lambay, A., Liu, Y., Morgan, P. L., and Ji, Z. (2024). Machine learning assisted human fatigue detection, monitoring, and recovery. *Digital Engineering* , 100004

Lataifeh, M., Afyouni, I., Shaduly, Z. A. S., Abdulkarim, A., and Ahmed, N. (2025). An adaptive multimodal framework for designing intelligent virtual agents in mixed reality. In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces* (New York, NY, USA: Association for Computing Machinery), IUI '25 Companion, 133–136. doi:10.1145/3708557.3716361

Li, C., Chrysostomou, D., and Yang, H. (2023). A speech-enabled virtual assistant for efficient human–robot interaction in industrial environments. *Journal of Systems and Software* 205, 111818. doi:https://doi.org/10.1016/j.jss.2023.111818

Li, C., Hansen, A. K., Chrysostomou, D., Bøgh, S., and Madsen, O. (2022). Bringing a natural language-enabled virtual assistant to industrial mobile robots for learning, training and assistance of manufacturing tasks. In *2022 IEEE/SICE International Symposium on System Integration (SII)*. 238–243. doi:10.1109/SII52469.2022.9708757

Li, C. and Yang, H. J. (2021). Bot-x: An ai-based virtual assistant for intelligent manufacturing. *Multiagent and grid systems* 17, 1–14

Longo, F., Nicoletti, L., and Padovano, A. (2017). Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Computers & industrial engineering* 113, 144–159

Mingardi, M., Pluchino, P., Bacchin, D., Rossato, C., and Gamberini, L. (2020). Assessment of implicit and explicit measures of mental workload in working situations: implications for industry 4.0. *Applied Sciences* 10, 6416

Nenna, F., Orso, V., Zanardi, D., and Gamberini, L. (2023a). The virtualization of human–robot interactions: A user-centric workload assessment. *Virtual Reality* 27, 553–571

Nenna, F., Zanardi, D., and Gamberini, L. (2023b). Effects of presence on human performance and workload in simulated vr-based telerobotics. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*. 47–52

Nenna, F., Zanardi, D., and Gamberini, L. (2023c). Enhanced interactivity in vr-based telerobotics: an eye-tracking investigation of human performance and workload. *International Journal of Human-Computer Studies* 177, 103079

Nicora, M. L., Beyrodt, S., Tsovaltzi, D., Nunnari, F., Gebhard, P., and Malosio, M. (2023). Towards social embodied cobots: The integration of an industrial cobot with a social virtual agent. *arXiv preprint arXiv:2301.06471*

Niiyama, R., Ikeda, M., and Seong, Y. A. (2023). Inflatable humanoid cybernetic avatar for physical human–robot interaction. *International Journal of Automation Technology* 17, 277–283

Nikolaidis, S., Hsu, D., and Srinivasa, S. (2017). Human-robot mutual adaptation in collaborative tasks: Models and experiments. *The International Journal of Robotics Research* 36, 618–634

Oppermann, L., Buchholz, F., and Uzun, Y. (2023). EnglishIndustrial metaverse: Supporting remote maintenance with avatars and digital twins in collaborative xr environments (Association for Computing Machinery). doi:10.1145/3544549.3585835. Cited by: 9

Pulikottil, T., Estrada-Jimenez, L. A., Ur Rehman, H., Mo, F., Nikghadam-Hojjati, S., and Barata, J. (2023). Agent-based manufacturing—review and expert evaluation. *The International Journal of Advanced Manufacturing Technology* 127, 2151–2180

Ramadhani, N., Fachri, M., Prasetyo, D., Damastuti, F. A., Hariadi, M., and Mutiaz, I. R. (2023). EnglishVirtual avatar representation in the digital twin: A photogrammetric three-dimentional modeling approach (Institute of Electrical and Electronics Engineers Inc.). doi:10.1109/ICIC60109.2023.10382008. Cited by: 1

Rossato, C., Orso, V., Pluchino, P., and Gamberini, L. (2021a). Adaptive assembly workstations and cobots: A qualitative assessment involving senior and adult workers. In *Proceedings of the 32nd European Conference on Cognitive Ergonomics*. 1–5

Rossato, C., Pluchino, P., Cellini, N., Jacucci, G., Spagnolli, A., and Gamberini, L. (2021b). Facing with collaborative robots: The subjective experience in senior and younger workers. *Cyberpsychology, Behavior, and Social Networking* 24, 349–356

Salam, A., Pervez, S., and Nahar, S. (2021). Trust in ai and intelligent systems: Central core of the design of intelligent systems. In *Proc. AMCIS*. 22

Süße, T., Kobert, M., and Kries, C. (2023). Human-ai interaction in remanufacturing: exploring shop floor workers' behavioural patterns within a specific human-ai system. *Labour and Industry* 33, 344–363

Wang, B., Zhou, H., Li, X., Yang, G., Zheng, P., Song, C., et al. (2024a). Human digital twin in the context of industry 5.0. *Robotics and Computer-Integrated Manufacturing* 85, 102626

Wang, R., Cheng, R., Ford, D., and Zimmermann, T. (2024b). Investigating and designing for trust in ai-powered code generation tools. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1475–1493

Wu, C., Cha, J., Sulek, J., Zhou, T., Sundaram, C. P., Wachs, J., et al. (2020). Eye-tracking metrics predict perceived workload in robotic surgical skills training. *Human factors* 62, 1365–1386

Xu, J., Sun, Q., Han, Q.-L., and Tang, Y. (2025). When embodied ai meets industry 5.0: Human-centered smart manufacturing. *IEEE/CAA Journal of Automatica Sinica* 12, 485–501. doi:10.1109/JAS.2025.125327

Zhang, Z., Ji, Y., Tang, D., Chen, J., and Liu, C. (2024). Enabling collaborative assembly between humans and robots using a digital twin system. *Robotics and Computer-Integrated Manufacturing* 86, 102691

AI-Powered Metaverse: AGI and the Path to Ethical Superintelligence

Author: Christopher Lafayette
Affiliations: GatherVerse, Hyper Policy, HoloPractice
Location: Mountain View, California, United States
Email: chris@christopherlafayette.com
Date: 06/06/2025

ABSTRACT

The convergence of advanced artificial intelligence with immersive metaverse platforms presents unprecedented opportunities and ethical challenges. This position paper examines how the emergence of artificial general intelligence (AGI) within AI-powered virtual worlds could be guided toward ethical superintelligence—ultra-capable AI systems aligned with human values. A multi-faceted governance framework is proposed to foster the responsible development of AI in the metaverse. Key components include a Global Cloud Consensus for worldwide AI governance, a Universal Interpreter to enable interoperability and cultural inclusivity, an AI Trust Layer to secure safety and transparency, and a dynamic Hyper Policy approach for adaptive regulation. Drawing on expertise in metaverse innovation and AI ethics, this work outlines a path forward to harness AI's benefits in virtual environments while rigorously addressing risks. The goal is to cultivate an AI-powered metaverse that upholds human rights, promotes global collaboration, and ultimately paves the way for a superintelligent future that is safe and beneficial for all.

Keywords: AI Ethics, Metaverse, Superintelligence, Governance, Interoperability, Trust, Policy

---

INTRODUCTION

The term metaverse, originally coined in Neal Stephenson's novel *Snow Crash*, refers to a persistent virtual world merging digital and physical reality. Once a science-fiction concept, the metaverse is rapidly becoming an emerging reality due to advances in extended reality and networked virtual environments. Modern definitions characterize the metaverse as a post-reality universe that is perpetually shared by users, blending physical and virtual spaces into a single immersive experience. Technology taxonomies highlight the metaverse's key components and open challenges, including infrastructure, avatars, digital assets, and interactive content. An increasingly important component across these facets is artificial intelligence: from intelligent virtual agents and generative content systems to personalized experiences driven by machine learning. As metaverse applications expand, so does the integration of AI, making questions of ethical governance increasingly urgent.

While today's metaverse AI remains narrow in scope, the trajectory of AI development points toward greater autonomy and generality. Experts anticipate the advent of artificial general intelligence (AI with human-level cognitive abilities) within this century. Beyond AGI lies the prospect of superintelligence, a level of AI intellect far surpassing humans in virtually all domains. To date, such advanced AI remains hypothetical; no concrete real-world example of an AGI or superintelligent agent exists. Nonetheless, it is prudent to consider governance now, before these possibilities materialize. If powerful AI agents eventually become embedded in our virtual worlds, they could profoundly influence societal norms, individual behaviors, and even governance processes within and beyond the metaverse. This presents a profound ethical challenge: How can we ensure that increasingly intelligent AI systems operating in immersive environments remain aligned with human values and serve the public good? Researchers in AI safety and ethics have emphasized the critical importance of value alignment and control for advanced AI. Even contemporary AI systems require governance to prevent bias, abuse, and harm; with AGI or superintelligence, the stakes will be vastly higher. Proactively establishing frameworks for responsible AI behavior in the metaverse can lay critical groundwork for managing future superintelligent AI in society.

This position paper operates at the frontier of intelligence research, where concrete examples are lacking and scenarios remain largely speculative. It is recognized that the metaverse offers both a potential proving ground and a microcosm of the broader challenges of AI ethics. Within virtual worlds, it may be possible to implement constrained environments or "sandbox" conditions to test advanced AI behaviors safely. Insights from such simulations could inform real-world AI governance. However, it must be stressed that the metaverse is not merely a sandbox; it is a real, consequential environment with human participants and stakes. Any failure to control AI in the metaverse could cause genuine harm to users, such as AI-induced psychological or economic damage, and foreshadow larger failures outside of it. Therefore, proactive governance in the metaverse is an essential stepping stone toward global AI governance. By instilling ethical guidelines, oversight mechanisms, and alignment techniques in today's virtual platforms, a culture and infrastructure can be created that scales as AI grows more capable. In effect, the metaverse can serve as a controlled environment to refine approaches for achieving ethical superintelligence, ensuring that when AI does reach and exceed human-level cognition, it does so embedded in a framework of accountability and humanity-first principles.

GOVERNANCE FRAMEWORKS FOR AN AI-POWERED METAVERSE

Governing AI in a borderless, immersive environment requires innovative approaches that blend technical, ethical, and policy strategies. Four interlocking components are proposed for a comprehensive metaverse AI governance framework: Global Cloud Consensus, Universal Interpreter, AI Trust Layer, and an adaptive Hyper Policy approach. These concepts collectively address the needs for global coordination, interoperability, safety assurance, and agile policymaking in the context of AI-driven virtual worlds.

Global Cloud Consensus: Toward Unified AI Governance

One of the greatest challenges in AI governance is achieving international consensus on norms and regulations. The metaverse, by nature, is a global platform where users and AI agents interact across jurisdictions. The Global Cloud Consensus model is proposed as a cloud-based, collaborative governance framework that unifies ethical AI standards worldwide. The vision of Global Cloud Consensus is to centralize and refine disparate AI governance frameworks into a single universal model. In practice, this would involve an alliance of stakeholders—technology companies, international organizations, governments, academia, and civil society—contributing to a shared set of guidelines and rules for AI behavior in virtual environments. These standards would be hosted and disseminated via an online platform so that metaverse providers and AI developers around the world have access to up-to-date best practices and policies.

The Global Cloud Consensus acts as an aggregator and synthesizer of existing ethical frameworks, distilling them into a master governance structure for AI. Rather than reinventing governance from scratch, it would build on precedents and principles already established across the world, such as those from the OECD and UNESCO. The consensus process would facilitate dialogue to resolve conflicts and continuously update the global standard, aiming for the highest common ethical denominator. Accountability, transparency, and alignment are key attributes: AI systems must be traceable, developers held responsible for violations, actions by AI must be auditable, and all behaviors must be consistent with core values reflecting fundamental human rights and well-being.

Universal Interpreter: Enabling Interoperability and Inclusivity

A healthy metaverse will consist of many interoperable platforms and AI systems, developed by different organizations and catering to diverse cultures. In such a heterogeneous ecosystem, enforcing common governance is challenging. The Universal Interpreter is introduced as an AI system or protocol that translates and mediates between different AI models, platforms, and human users. Its role is to ensure that a shared ethical framework is interpreted uniformly across varied contexts, serving as the common language of governance in the metaverse so that, regardless of an AI agent's origin or a user's background, ethical rules carry the same meaning and weight everywhere.

The Universal Interpreter bridges user preferences and AI policies across virtual platforms, facilitating communication, and maintaining consistency of ethical norms and safety settings. It also provides observability, monitoring AI behavior across embodiments and ensuring classification and contextualization of AI actions in terms of the shared ethical framework. If unknown or unrecognized AI models attempt to operate, the Interpreter can flag them for scrutiny, adding a layer of security. This enables the ethical framework to be applied uniformly and transparently, reducing the risk of governance gaps as people and AI interact across interconnected virtual worlds.

AI Trust Layer: Protecting Safety and Transparency

Effective governance demands real-time enforcement and monitoring. An AI Trust Layer within the metaverse's technical stack is proposed as a dedicated layer of software and services that continuously oversees AI actions for compliance and safety. This layer monitors the behavior of AI agents, intervenes or alerts when policies are breached, and provides users and administrators with transparency into AI decision-making.

The AI Trust Layer could employ automated auditors, immutable logging technologies, and user-facing controls and explanation tools. Users should be able to query AI decisions and receive clear explanations, fostering trust and providing feedback for policy adjustments. On the safety side, fail-safes are incorporated to prevent harm; if an AI agent exhibits dangerous behavior, the Trust Layer could shut it down or quarantine it from the wider environment. By enforcing rules and providing continuous assurance, the Trust Layer ensures that AI systems remain accountable, understandable, and controllable in the day-to-day life of the metaverse.

Hyper Policy: Adaptive Governance at Scale

Any framework for AI governance must be able to adapt rapidly to technological evolution and new societal insights. Hyper Policy encapsulates an approach to policymaking that is agile, iterative, and driven by multi-stakeholder collaboration. Ethical guidelines and regulations governing AI are treated as living frameworks that evolve as needed. Feedback and revision loops, supported by data from the AI Trust Layer, enable quick assessment of emerging issues and updating of guidelines or technical standards. Collaboration across technologists, ethicists, policymakers, and community voices is necessary to create balanced and forward-looking policies.

Hyper Policy enables the governance process itself to scale with the metaverse, treating policies like a codebase that requires version control, testing, and deployment. The adaptive nature of Hyper Policy ensures that the pursuit of ethical superintelligence can keep pace with technological advancements, integrating new safety measures or ethical guidelines as needed.

DISCUSSION AND LIMITATIONS

This position paper operates within the frontier of intelligence research, where no fully realized examples of superintelligent agents exist. While the metaverse may offer a space to experiment with governance frameworks, it is not simply a hypothetical sandbox; it is a real environment with significant ethical and societal stakes. Achieving global consensus is a formidable challenge due to political and cultural differences, and the pursuit of universal human values in

AI governance will always require sensitivity and adaptability. The concepts presented here are aspirational, aimed at stimulating dialogue and proactive preparation rather than offering after-the-fact solutions.

CONCLUSION AND FUTURE OUTLOOK

Guiding an AI-powered metaverse toward ethical superintelligence is an immense challenge, but it is one that demands urgent and proactive effort. By combining global cooperation with technical innovation and a steadfast commitment to human-centric values, it is possible to make the future metaverse a space where advanced AI serves the public good. The frameworks proposed—Global Cloud Consensus, Universal Interpreter, AI Trust Layer, and Hyper Policy—provide a roadmap for aligning AI development with widely shared values, even as the frontier of intelligence rapidly evolves. Continued research, pilot projects, and global collaboration are needed to test and refine these ideas. The frontier of intelligence is before us, and now is the time to chart it responsibly.

REFERENCES

1. Altman, S., Brockman, G., and Sutskever, I. (2023). Governance of superintelligence. OpenAI Blog.
2. Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford, UK: Oxford University Press.
3. Dafoe, A. (2018). AI governance: a research agenda. Future of Humanity Institute, University of Oxford.
4. Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389–399.
5. Mystakidis, S. (2022). Metaverse. Encyclopedia, 2(1), 486–497.
6. Organisation for Economic Co-operation and Development (OECD). (2019). Recommendation of the Council on Artificial Intelligence.
7. Park, S., and Kim, Y. (2022). A metaverse: taxonomy, components, applications, and open challenges. IEEE Access, 10, 4209–4251.
8. Russell, S., Dewey, D., and Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. AI Magazine, 36(4), 105–114.
9. Stephenson, N. (1992). Snow Crash. New York, NY: Bantam Spectra.
10. UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO.

Session 2: Ethical AI, Responsible and Legal Metaverse Design

frontiers

The Metaverse as an AI-powered Society of Control: Ethical Concerns on Autonomy, Self-determination and Surveillance

Anastasia Nefeli Vidaki[1]

[1]Cyber and Data Security Lab, Research Group on Law, Science, Technology & Society, Faculty of Law & Criminology, Vrije Universiteit Brussel, Brussels, Belgium

*** Correspondence:**
Corresponding Author
Anastasia.nefeli.vidaki@vub.be

Keywords: metaverse, ethics, society of control, surveillance, autonomy, self-determination

Abstract

This paper explores the metaverse as an AI-powered extension of Gilles Deleuze's concept of 'societies of control', primarily focusing on its implications for autonomy, self-determination, and surveillance. Unlike traditional disciplinary institutions, the metaverse integrates power dynamics directly into its core through algorithmic modulation, behavioral tracking, and predictive analytics. By applying Deleuzian theory to AI-driven governance and digital capitalism, this study demonstrates how the metaverse could operate as an advanced form of control society, where user agency is constrained by gamification, data commodification, and AI-profiling. The research critically examines how the metaverse could fragment identity, transforming individuals into Deleuzian 'dividuals' whose actions are driven by AI interactions rather than self-determined decisions. Furthermore, it interrogates the metaverse's function as a surveillance apparatus that might extend beyond traditional oversight models, placing control into digital experiences through biometric tracking, behavioral nudging, and predictive policing. Ultimately, by analysing the metaverse through the lens of Deleuze's control societies, it aims to raise awareness of the relevant ethical challenges and offers some viable paths for escape.

1 Introduction

The evolution of digital environments has given rise to new forms of power and governance, fundamentally altering the manner individuals interact, express themselves, and experience autonomy. Drawing on Gilles Deleuze's concept of 'societies of control', this paper explores how the metaverse represents a shift from disciplinary enclosures to algorithmic modulation, where behavior is subtly shaped rather than overtly restricted. Unlike traditional institutions that impose authority through confinement, the metaverse integrates control into its infrastructure, embedding predictive analytics, AI-driven recommendations, and gamified incentives to steer user engagement. As a result, it does not function as a neutral arena for exploration but as a carefully engineered environment where power operates invisibly, making resistance increasingly complex.

This paper provides a novel contribution by applying Deleuzian theory to contemporary discussions on AI-driven governance, surveillance capitalism, and digital subjectivity in the metaverse. While previous studies have examined control societies and algorithmic power separately, this research bridges the gap by demonstrating how the metaverse can exemplify a hyper-advanced form of

Deleuzian control, where users are simultaneously surveilled, fragmented, and guided through imperceptible digital architectures. By interrogating whether emerging technologies offer genuine alternatives to centralised control or merely reconfigure existing power structures, this study advances current debates on digital self-determination, Artificial Intelligence (AI) governance, and potential pathways for resistance.

2 The Deleuzian Societies of Control

The theory of surveillance reflects the practice of supervision through expansive networks, employing various mechanisms including data collection and classification, access control, and interconnection. Simultaneously, it incorporates reactionary processes through feedback loops, making it an inextricable part of the social fabric. Surveillance is not merely a technological phenomenon but a profoundly political, social, and even legal one (Samatas, 2007). Scholars who have examined it have consistently underlined this dimension.

Jeremy Bentham (1995, 29) and Michel Foucault (1920) attributed to surveillance a physical and spatial dimension, embodied in panoptic structures and disciplinary environments. Later theorists, such as Kevin Haggerty and Richard Ericson (2000, 607), Shoshana Zuboff (2019), and Gilles Deleuze (1990), extended these analyses to digital surveillance societies, marking an evolution that departs from centralised, enclosed spaces and moves into the vast, decentralized world of digital networks. Deleuze, in particular, offers a crystallised perspective on these shifting surveillance mechanisms in his essay Postscript on Societies of Control, first published in 1990 in L'Autre Journal.

In three distinct sections, Deleuze does not merely define social control or describe its mechanisms but also proposes ways for subjects to resist it (Brusseau, 2020, 2). Taking Foucault's 'Discipline and Punish' as a starting point, he borrows the term 'control' from William Burroughs (Gontarski, 2020) and proceeds to outline the historical transition from what Foucault termed disciplinary societies to what Deleuze designated as societies of control. To clarify these distinctions, it is useful to juxtapose them with an earlier form of power: societies of sovereignty.

Before the 18th century, European societies operated under sovereign power, in which absolutism and repression reigned supreme. The law was synonymous with the will of the ruler, crime was an affront to authority, and punishment functioned as an act of vengeance, often targeting the body through public executions or brutal corporal punishment. However, resistance was relatively straightforward. Centralised power meant that opposition could be directed against the sovereign: the monarch. Furthermore, because punishments were severe, the public often sympathised with the punished, fostering solidarity among the oppressed. Economic regulation in these societies was minimal as long as production was taxed but largely controlled by the workers. These forms of societal organisation and power focused on the control over death.

Following Napoleon, societies underwent a transition. The shift did not eradicate sovereign power but rather restructured it in response to historical, economic, and political changes in the 19th and 20th centuries. Disciplinary societies emerged, characterised by a network of enclosed institutions that shaped individuals throughout their lives. The family, the school, the factory, the hospital, and the prison became environments where power was no longer concentrated in a singular ruler but dispersed across different entities that regulated individuals at every stage of their existence. The resistance needed to target all these different factors became harder. Capitalism, the dominant economic model of these societies, was centred on ownership and production. Institutions functioned preventively, engaging in hierarchical surveillance to ensure that individuals adhered to social norms. Individuals were categorised and subjected to assessments of normality. A deviation from established norms did not result in outright punishment but in corrective treatment aimed at integrating the individual back into society and the mass. The goal was no longer control over death, as in sovereign societies, but control over life itself.

Deleuze likens these societies to moles, creatures that burrow into confined spaces, mirroring the spatial limitations imposed by disciplinary institutions. Within these structures, individuals became mere numbers, assigned identities based on space, surveillance, and classification. Even the mere possibility of being watched proved sufficient to regulate behavior.

The Second World War laid the groundwork for a new societal formation, namely one that moved beyond confinement into a more fluid and adaptive structure: the society of control. Though societies of control retain elements of disciplinary structures, they diverge in fundamental ways. Unlike the rigid enclosures of disciplinary societies, control societies are marked by constant movement, perpetual mobility, fluidity, continuous access, and permissive rather than prohibitive mechanisms (Haggerty and Ericson, 2020, 605-620). Subjects are granted apparent freedom but only if they conform to invisible norms embedded in digital infrastructures (Galloway, 2014). Rather than imposing direct, authoritative restrictions, societies of control function through modulation; forms of control that present themselves as expressions of individual freedom.

Deleuze describes this shift as the transition from disciplinary spaces to digitised, coded networks, where individuals exist not as unified subjects but as dividuals, fragmented identities scattered across data points, behavioral profiles, and markets (Schroeder, 2016). The traditional mass has been replaced by a system of statistical samples, where capitalism no longer prioritises production but instead trades in influence, data, and pre-configured desires. The factory has been transformed into a decentralised network of businesses. The hospital operates through remote medical models that minimise doctor-patient interactions. Imprisonment is often replaced by alternative sentencing, including electronic surveillance. Education is no longer confined to the school but extends into lifelong learning models, integrated into corporate and state structures. In these Kafkaesque societies, where disciplinary societies were likened to moles, slow and methodical in their enclosed movements, control societies function like reptiles: flexible, adaptive, and agile. Access, rather than spatial confinement, becomes the new mechanism of power.

Deleuze does not offer a value judgment on these transitions (Dreyfus and Rainbow, 1982). He neither nostalgically favors disciplinary societies nor outright condemns control societies. Instead, he acknowledges that we are facing a new phenomenon and urges individuals, particularly the younger generation, to understand the system they are being forced to serve. Unlike earlier forms of power, where resistance took the form of confrontation, control societies demand new forms of resistance that adapt to the decentralized and fluid nature of power.

Deleuze highlights that fear and fatalism are unnecessary. Instead, he calls for the discovery of new weapons and novel tools that can disrupt, subvert, or bypass mechanisms of control. In a world where access replaces enclosure, where surveillance is predictive rather than reactive, and where power operates through networks rather than hierarchies, resistance must be equally fluid, equally adaptive, and equally decentralised.

As we move deeper into an era defined by AI-driven surveillance, big data, and algorithmic governance, Deleuze’s analysis of control societies becomes increasingly relevant. Particularly, in the metaverse, where identity is fragmented across platforms, behaviors are tracked in real time, and decisions are shaped by invisible forces, the culmination of control society mechanisms becomes clear. Understanding this shift is essential not only for theorising power in the digital age but also for imagining strategies of resistance in a world where control is omnipresent yet intangible.

3 Metaverse as a Society of Control

The metaverse represents a paradigm shift in governance and surveillance, potentially emerging as a novel type of control society. In it, power will no longer be imposed through rigid institutions but through fluid, data-driven modulation. It operates through continuous tracking and behavioral nudging, shaping digital experiences in real-time. AI-driven recommendation systems, predictive modeling, and gamified incentives replace direct coercion, subtly influencing user actions, in many cases, even without their explicit awareness and consent (Rouvroy, 2013, 157). As users navigate virtual environments, their behaviors, preferences, and even physiological responses can be tracked and analysed, allowing AI systems to dynamically adjust their experiences (Andrejevic, 2020, 18). This creates an environment of algorithmic governance where control is no longer external and overt but embedded within the infrastructure of virtual life itself. The result is a landscape of persistent reform, where users are not confined but rather guided by imperceptible forces that structure their interactions and choices.

At the heart of this transformation lies data capitalism, wherein the very essence of digital subjectivity, including thoughts, emotions, and behaviors, is fragmented into commodifiable data points. In this system, personal interactions are no longer private expressions but assets to be monetised. Every action within the metaverse, from the movement of a user's avatar to their gaze direction and biometric feedback, can be captured, analysed, and repurposed for commercial gain (Lyon, 2017, 831). Unlike traditional economies, where commodities are produced and sold, the metaverse is in place to turn the simple act of existence into an economic transaction. Users do not own their digital selves. Instead, their experiences are algorithmically curated and sold as predictive insights to corporations and advertisers (Brusseau, 2020, 2). This shift reflects a deeper entrenchment of market logic in everyday digital life, where the body, its digital replica, and identity itself become a form of capital controlled by the platforms that mediate it. Consequently, rather than engaging with a neutral virtual space, users interact with a carefully engineered environment designed to maximise data extraction, behavioral influence (van Dijck, 2013), and, therefore, profit.

The ramifications of this pervasive algorithmic control include a fundamental loss of agency, as users turn into 'dividuals', in other words, segments of data profiles that AI systems continuously process and manipulate. While the metaverse presents itself as a space of infinite choice and personalisation, these experiences are carefully pre-selected and tailored to fit engagement, maximising algorithms rather than individual autonomy (Rigotti and Malgieri, 2023). Gamification, personalised recommendations, and AI-driven content curation create an illusion of freedom while subtly reinforcing behavioral patterns that serve commercial interests (Yasuda, 2024). The ability to explore and express oneself is, thus, constrained by invisible digital architectures that dictate the scope of user interactions. While decentralisation and blockchain-based alternatives offer potential counterpoints, they too risk being co-opted by the same logic of data commodification (Brunton and Nissenbaum, 2015). In the end, the metaverse extends Deleuze's control society to its most advanced form, one where governance is not imposed from above but embedded within the very fabric of digital presence, shaping actions at an unconscious level while maintaining the fake appearance of individual freedom. The options are pre-selected, limiting the true autonomy and self-determination of users in these digital spheres.

4 Metaverse as an Identity-fragmentation Space

In the metaverse, identity is no longer a stable construct but a constantly changing, algorithmically generated phenomenon. Users can navigate multiple avatars, each shaped by AI profiling and digital interactions. Rather than possessing a singular self, individuals exist as a collection of data points, responding to the predictive logic of AI and machine-learning models. This aligns with Deleuze's concept of the 'dividual', where individuals are broken down into data pieces, circulated and processed by digital systems. Such fragmentation challenges traditional notions of selfhood while it raises ethical concerns and legal inquiries about the authenticity and ownership of digital identities (OECD, 2007, 41). The metaverse is transformed into a space where human subjectivities (Deleuze and Guattari, 1987) are molded by algorithmic influence rather than one of self-determined expression (Fisher, 2022). The implications of this transformation extend beyond digital realms, as social behaviors and personal identities become increasingly entangled with algorithmic profiling, influencing real-world, human-to-human interactions, employment opportunities, and, according to some researchers, even political participation (Fegert et al., 2020).

Predictive AI shapes user choices before they are even consciously made, bringing into effect a vicious cycle of algorithmic determinism (Hallo et al., 2024). Through data-driven self-expression, users believe they are freely constructing their digital identities, yet their actions are subtly guided by AI-generated suggestions and reinforcements. This feedback loop between personal expression and algorithmic influence creates a paradox: while users experience a sense of autonomy, their digital selves are largely pre-scripted and sculpted by underlying AI models (Dwivedi et al., 2022). As a consequence, the metaverse might extend the logic of control societies, where individualisation functions as a form of power exercise, subtly dictating norms through automated decision-making processes. As platforms optimise engagement through algorithmic recommendations and the use of new technologies such as, but not limited to Virtual Reality (VR), users may find their perceptions, desires, and beliefs shaped by unseen computational forces, leading to a form of digital habituation that blurs the line between authentic choice and programmed decisions (Hennig-Thurau et al., 2023). This raises the concern that the metaverse, rather than expanding individual freedom, might reduce it by limiting exposure to diverse perspectives and enhancing particular ideological bubbles.

The metaverse represents a new frontier of posthumanism, where identity is not only forever changing but also technologically constructed (World Economic Forum, 2024). AI-driven personalisation enables users to adopt new forms of personhood continuously, but these adjustments may not be as liberating as they seem. Although it might seem that AI facilitates genuine self-exploration, it can merely offer pre-designed paths of becoming an algorithmic other (Gandini et al., 2023) that might be extended to the physical one. While the metaverse provides unprecedented opportunities for self-reinvention, it also risks reducing identity to simply a series of algorithmically determined roles (Mitrushchenkova, 2022). Users may be encouraged to conform to commercially profitable identity constructs, reinforcing consumerist patterns rather than fostering genuine self-discovery (Hearn, 2017). This raises deeper philosophical debates around the role of technology in shaping human subjectivity and whether digital existence can ever be truly autonomous in an age of predictive computation (Andrejevic, 2020, 35), which go beyond the scope of this paper. Nonetheless, the possibility that users could become passive participants in an algorithmically curated, Deleuzian reality remains pertinent.

5 Metaverse as a Surveillance Apparatus

Metaverse surveillance surpasses traditional models of visibility, like the panopticon, by incorporating predictive analytics, biometric tracking, and AI-supported behavioral profiling. Unlike the classical surveillance model, where subjects discipline themselves due to the fear of being watched, the metaverse's surveillance could operate through continuous data extraction, mapping not just user movements and interactions but also neural responses and biometric data (Ruju et al., 2024). AI-powered monitoring systems function invisibly, analysing patterns and making real-time adjustments to digital environments based on predictive models (Andrejevic and Burdon, 2014, 15). These systems do not merely observe but anticipate and influence actions, rendering traditional notions of surveillance obsolete. This circumstance aligns with Deleuze's concept of the 'control society', where power is exerted through subtly shaping reactions through automated nudges rather than restriction. The metaverse extends this paradigm, embedding control within the substance of digital experiences, ensuring that users remain unaware of the forces affecting their behavioral choices.

Control in the metaverse is not imposed through overt coercion but through gamification, an incentivised form of governance that encourages compliance through rewards and penalties. AI-driven systems regulate user behavior by offering digital perks, achievements, and status enhancements for conforming to platform-defined norms (Whitson, 2015, 548). Corporate and, in a far-fetched scenario, even state actors could utilise these mechanisms to subtly guide user engagement, shaping interactions in ways that align with their economic and political interests and agendas (Navaria, 2019, 5). Unlike traditional surveillance systems that rely on external enforcement, gamification integrates control into the user experience, making governance feel like an aspect of personal progress or voluntary engagement. This dynamic reinforces the idea of 'participatory surveillance' (Galič and Koops, 2016), in which users willingly contribute to their self-regulation under the illusion of autonomy (Cheney-Lippold, 2017). The deeper concern is whether users can resist or subvert such mechanisms when every aspect of their digital engagement is engineered to maximise data extraction and compliance (Taylor, 2017).

With the expansion of AI-driven governance, digital spaces are increasingly policed by automated systems. Smart contracts, algorithmic moderation, and predictive policing ensure compliance without human oversight, penetrating directly the discourse of the metaverse (Zhou et al., 2022). While these systems promise efficiency and fairness, they also introduce risks of algorithmic bias, corporate overreach, and opaque decision-making processes (Noble, 2018). The reliance on AI for law enforcement within virtual environments raises questions about accountability. While some argue that blockchain and decentralised technologies could counterbalance corporate control, these alternatives often replicate existing power structures rather than dismantle them (Mosco, 2017,13). As the metaverse becomes a sphere governed by AI-powered algorithms rather than democratic institutions, users must grapple with the implications of living in a world where rules are enforced not by human authorities but by self-executing digital protocols (Haggart et al., 2019, 125,168). In that sense, it might not be that different from a future aspect of a digitised world.

6 Concluding Remarks

In control societies, resistance is not a direct rebellion but a search for alternative pathways that Deleuze terms 'lines of flight' (Deleuze and Guattari, 1987). The metaverse offers potential escapes through hacking, digital subversion, or disengagement. Still, true autonomy remains elusive. Algorithmic environments adapt to disruptions, absorbing and repurposing acts of resistance into the system itself (Galloway and Thacker, 2007). The integration of digital and physical realities further complicates any notion of total escape, as predictive analytics continuously map user behaviours, preempting deviation (Bratton, 2015, 164). Consequently, the challenge lies not simply in escaping control but, instead, in redefining agency within its pervasive structures (Susca, 2022, 298).

Decentralised technologies such as blockchain, decentralised autonomous organizations (DAOs), and distributed AI promise liberation from corporate and state control. Hence, they often replicate the same logic of algorithmic governance (Terranova, 2004, 57). While decentralisation shifts power from singular authorities to networked infrastructures, it does not inherently dismantle surveillance or control mechanisms (Morozov, 2013). Decentralised platforms still rely on smart contracts and machine learning, embedding governance into automated systems rather than human institutions (Bodó et al., 2021, 8). This raises critical concerns about whether these technologies offer genuine freedom or merely reconfigure existing power dynamics in a more untransparent and technocratic form (Han, 2018, 30).

As AI-driven governance permeates digital life, viable forms of resistance become harder to identify. Strategies such as counter-programming, digital self-erasure, or collective sabotage challenge algorithmic dominance but often remain temporary and easily neutralised (Brunton and Nissenbaum, 2015, 63). The question is whether the metaverse can be reimagined as a space for collective freedom rather than an instrument of control (Couldry and Mejias, 2019, 12). To reclaim digital autonomy, users must critically engage with the infrastructures affecting their experiences, resisting passive participation in predictive ecosystems (Zuboff, 2015). Only by addressing the deeper entanglements of technology, governance, and agency can alternative futures beyond algorithmic control emerge. Understanding Deleuze's insights is not just an academic exercise but a necessary tool for navigating the complexities of an increasingly algorithmic world.

References

Andrejevic, M. (2020). Automating Media. Abingdon and New York: Routledge.

Andrejevic, M., & Burdon, M. (2014). Defining the Sensor Society. University of Queensland TC Beirne School of Law Research Paper No. 14-21. https://ssrn.com/abstract=2423118. [Accessed March 18, 2025].

Bentham, J. 1995). The Panopticon writings. London ; New York :Verso.

Bodó B, Brekke JK, Hoepman J. (2021). Decentralization: A multidisciplinary perspective. Internet Policy Review 10(2). https://doi.org/10.14763/2021.2.1563.

Bratton, B. H. (2015). The Stack: On Software and Sovereignty. Cambridge, MA: MIT Press.

Brunton, F., Nissenbaum, H. (2015). Obfuscation: A User's Guide for Privacy and Protest. MIT Press.

Brusseau, J. (2020) Deleuze's Postscript on the Societies of Control Updated for Big Data and Predictive Analytics. Theoria, 67(64),1–25.

Cheney-Lippold, J. (2017). We Are Data: Algorithms and the Making of Our Digital Selves. NYU Press.

Couldry, N., Mejias, U. A. (2019). The Costs of Connection: How Data is Colonizing Human Life and Appropriating It for Capitalism. Stanford University Press.

Deleuze, G., Guattari, F. (1987). A Thousand Plateaus: Capitalism and Schizophrenia. University of Minnesota Press.

Deleuze, G. (1990). Postscript on the Societies of Control. Pouparlets: Columbia University Press.

Dreyfus, H.L., Rabinow, P. (1982). Michel Foucault: Beyond Structuralism and Hermeneutics. Routledge. https://doi.org/10.4324/9781315835259.

Dwivedi YK, Hughes L, Baabdullah AM, Ribeiro-Navarrete S, Giannakis M, Al-Debei MM, Dennehy D, Metri B, Buhalis D, Cheung CMK, Conboy K, Doyle R, Dubey R, Dutot V, Felix R, Goyal DP, Gustafsson A, Hinsch C, Jebabli I, Wamba SF. (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 66, Article 102542. https://doi.org/10.1016/j.ijinfomgt.2022.102542.

Fegert J, Pfeiffer J, Peukert C, Golubyeva A, Weinhardt C. (2020). Combining e-Participation with Augmented and Virtual Reality: Insights from a Design Science Research Project. ICIS 2020 Proceedings. https://aisel.aisnet.org/icis2020/digitization_in_cities/digitization_in_cities/4. [Accessed March 18, 2025].

Fisher, E. (2022). Algorithms and Subjectivity: The Subversion of Critical Knowledge. Routledge. https://doi.org/10.4324/9781003196563.

Foucault, M. (1920). Discipline & Punish. Random House.

Galič, M., Timan, T. & Koops, B.J. (2016). Bentham, Deleuze and Beyond: An Overview of Surveillance Theories from the Panopticon to Participation. Philos. Technol., 30, 9–37.

Galloway, A. R. (2014). Laruelle: Against the Digital. University of Minnesota Press. http://www.jstor.org/stable/10.5749/j.ctt1287nfc.

Galloway, A. R., Thacker, E. (2007). The Exploit: A Theory of Networks. University of Minnesota Press.

Gandini A, Gerosa A, Giuffrè L, Keeling S. (2023). Subjectivity and algorithmic imaginaries: the algorithmic other. Subjectivity, 30, 417–434. https://doi.org/10.1057/s41286-023-00171-w.

Gontarski, S. E. (2020). The 'Limits of Control': Burroughs through Deleuze. Symplokē, 28(1–2), 65–81. https://doi.org/10.5250/symploke.28.1-2.0065.

Haggart B, Henne K, Tusikov N. (2019). Information, Technology and Control in a Changing World. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-14540-8_1.

Haggerty,K.D., Ericson, R.V. (2000). The surveillant assemblage. The British journal of sociology, 51(4), 605-22 .

Hallo L, Rowe C, Duh G. (2024). Metaverse vs Descartes: Free Will and Determinism All Over Again. 2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA), 1-6, https://doi.org/10.1109/IISA62523.2024.10786640.

Han, B. (2018). What is power?. London: Polity Books.

Hearn, A. (2017). Verified: Self-presentation, identity management, and selfhood in the age of big data. Popular Communication, 15(2), 62–77. https://doi.org/10.1080/15405702.2016.1269909.

Hennig-Thurau T, Aliman DN, Herting AM, Cziehso GP, Linder M, Kübler RV. (2023). Social interactions in the metaverse: Framework, initial evidence, and research roadmap. J. of the Acad. Mark. Sci. 51, 889–913. https://doi.org/10.1007/s11747-022-00908-0.

Lyon, D. (2017). Surveillance Culture: Engagement, Exposure, and Ethics in Digital Modernity. International Journal of Communication 11, 824-842. https://ijoc.org/index.php/ijoc/article/view/5527/1933

Mitrushchenkova, A.N., (2022). Personal Identity in the Metaverse: Challenges and Risks. Kutafin Law Review 9(4), 793– 817, https://doi.org/10.17803/2313-5395.2022.4.22.793-817.

Morozov, E. (2013). To Save Everything, Click Here: The Folly of Technological Solutionism. NYC: PublicAffairs.

Mosco, V. (2017). Becoming Digital: Toward a Post-Internet Society. Emerald Publishing. https://doi.org/10.1108/9781787432956.

Navaria, G. (2019). The Networked Citizen: Power, Politics, and Resistance in the Internet Age. Singapore: Palgrave Macmillan. https://doi.org/10.1007/978-981-13-3293-7.

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

OECD (2007). At a Crossroads: “Personhood” and Digital Identity in the Information Society. https://one.oecd.org/document/DSTI/DOC(2007)7/en/pdf. [Accessed March 18, 2025].

Rigotti, C., Malgieri, G.(2023). Human Vulnerability in the Metaverse. https://audri.org/resources/report-human-vulnerability-in-the-metaverse/. [Accessed March 18, 2025].

Rouvroy, A. (2013). The End(s) of Critique: Data Behaviourism versus Due Process. In Hildebrandt and de Vries (eds.) Privacy, Due Process, and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology. London: Routledge.

Ruiu P, Nitti M, Pilloni V, Cadoni M, Grosso E, Fadda M. (2024). Metaverse & Human Digital Twin: Digital Identity, Biometrics, and Privacy in the Future Virtual Worlds. Multimodal Technologies and Interaction, 8(6), 48. https://doi.org/10.3390/mti8060048.

Samatas, M. (2007). Security and Surveillance in the Athens 2004 Olympics: Some Lessons From a Troubled Story. International Criminal Justice Review, 17(3), 220-238. https://doi.org/10.1177/1057567707306649.

Schroeder, R. (2016). Big data business models: Challenges and opportunities. Cogent Social Sciences, 2(1). https://doi.org/10.1080/23311886.2016.1166924.

Srnicek, N. (2016). Platform Capitalism. Cambridge, UK, and Malden, MA: Polity Press.

Susca, E. (2022). Countering “Surveillance Capitalism.” The Intertwining of Objective and Subjective Factors. In: Comunello, F., Martire, F., Sabetta, L. (eds.) What People Leave Behind.

Frontiers in Sociology and Social Research, 7. Springer, Cham. https://doi.org/10.1007/978-3-031-11756-5_16.

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. Big Data & Society, 4(2). https://doi.org/10.1177/2053951717736335.

Terranova, T. (2004). Network Culture: Politics For the Information Age. Pluto Press. https://doi.org/10.2307/j.ctt183q5pq.

Whitson, J. (2015). 13: Foucault's Fitbit: Governance and Gamification. In Walz, S.P. & Deterding, S. (eds.) The Gameful World: Approaches, Issues, Applications. The MIT Press. https://doi.org/10.7551/mitpress/9788.003.0023.

World Economic Forum (2024). Metaverse Identity: Redefining Identity for a Blended Reality. Insight Report. March 2024. https://www.weforum.org/publications/metaverse-identity-defining-the-self-in-a-blended-reality/. [Accessed March 18, 2025].

van Dijck, J. (2013). The Culture of Connectivity: A Critical History of Social Media. Oxford Academic. https://doi.org/10.1093/acprof:oso/9780199970773.001.0001.

Yasuda, A. (2024). Metaverse ethics: exploring the social implications of the metaverse. AI Ethics. https://doi.org/10.1007/s43681-024-00507-5.

Zhou B, Xi C, Li G, Yang B. (2022). Metaverse Application in Power Systems. Power Generation Technology, 43(1), 1-9.

Zubof, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. J. Inf. Technol., 30(1), 75–89. https://doi.org/10.1057/jit.2015.5.

Zuboff, S. (2019). The age of surveillance capitalism: the fight for a human future at the new frontier of power. First edition. PublicAffairs.

Inhabiting the Metaverse: AI Agents, Their Risks, and the Quest for Trust

Anselm Küsters[1*]

[1] Centre for European Policy (cep), Berlin, and University of Stuttgart, Germany

*** Correspondence:**
Anselm Küsters
kuesters@cep.eu

Keywords: generative AI, agents, metaverse, large language models, avatars, biases

Abstract

Thanks to generative artificial intelligence, autonomous AI agents are set to populate the emerging metaverse. Research prototypes demonstrate that AI agents are already capable of navigating complex virtual environments without explicit programming. Yet these capabilities introduce profound risks across multiple domains: economic distortions through algorithmic collusion and covert commercial placement ("SEO for AI"); psychological manipulation through AI's superior simulated empathy; political bias embedded in underlying large language models; and restricted access to justice through overly cautious or rule-violating AI-based legal decisions. For instance, AI agents may intentionally circumvent rules to achieve goals or develop harmful social conventions. While the EU's Virtual Worlds Toolbox acknowledges basic security concerns like avatar hacking, it vastly understates these more complex challenges. Solutions could come in the form of algorithmic bias audits, oversight committees, or disclosure requirements. Without comprehensive liability frameworks, EU AI Act guidance, and independent oversight, there is a risk that virtual worlds will become digital ecosystems in which AI agents routinely circumvent rules, undermining user trust.

1 Introduction: GenAI, agents, and the metaverse

The rise of autonomous "agents" has become one of the hot topics in Artificial Intelligence (AI) research. In this context, the current integration of generative AI (GenAI) in virtual worlds – such as advanced 3D games – provides a convenient case study for better understanding the repercussions of this paradigm shift; not least because virtual worlds increasingly serve as empirical testbeds for AI development. Recent advances in GenAI, ranging from language generation tools such as ChatGPT to visual creation platforms like DALL-E, have introduced new possibilities for populating virtual spaces with autonomous AI-driven entities (Chamola *et al.*, 2023). Such technologies enable virtual worlds not only to replicate realistic interactions and environments but also to dynamically evolve in response to user behavior, dramatically enhancing immersion (Alhafis Hashim *et al.*, 2024). Moreover, leveraging GenAI could considerably lower the barriers to creating interactive digital environments, resulting in more accessible and scalable virtual worlds (BCG, 2023).

What precisely distinguishes AI agents from simpler automated systems? According to a practical consensus (Casper *et al.*, 2025, p. 2), these agents operate effectively without explicit, step-by-step instructions and exhibit directness of impact by independently performing tasks with tangible consequences within their virtual settings. Moreover, effective agents are inherently goal-oriented, engaging in decision-making processes explicitly aimed at achieving predefined objectives. Equally critical is their capacity for long-term planning, a feature enabling agents to anticipate outcomes and adapt behavior accordingly. The following analysis of the impact of GenAI on the metaverse focuses only on these types of GenAI-based agents, not on parallel efforts to apply GenAI tools to enrich avatar personalization and content creation in more passive ways (GlobalData, 2023).

As virtual spaces will increasingly rely on sophisticated AI agents capable of autonomous actions, the broader repercussions of this development require scrutiny. This paper proceeds by examining, in Section 2, preliminary evidence from recent deployments and experiments with AI agents in digital environments. Section 3 explores how biases and implicit frameworks inherent in the underlying GenAI models might manifest themselves economically, socially, politically, and legally in the metaverse. Section 4 then discusses potential solutions, emphasizing the need for greater transparency, robust governance, and oversight mechanisms. Section 5 gives concrete suggestions for how these solutions might be embedded in the metaverse strategy of the EU. In essence, this paper asserts that trust must be maintained for the AI-based metaverse to be successful.

2 Review: How AI agents might enter the metaverse

According to a recent survey, the deployment of AI agents has accelerated significantly over the past two years (Casper *et al.*, 2025). American companies are responsible for 45 of the 67 detected systems, with large companies creating nearly three-quarters (73.1%) of these agents. However, only 13 of the indexed systems included publicly accessible safety policies, and a mere five offered comprehensive safety evaluations. As these agents begin transitioning from predominantly technical or isolated programming environments to more socially and ethically complex spaces in firms or virtual worlds, addressing such transparency gaps will become increasingly crucial. In fact, several firms have already begun to integrate early prototypes of these AI agents within virtual environments. For instance, companies such as "Character.ai" and "Inworld AI" illustrate how GenAI can bring virtual inhabitants to life by imbuing so-called non-player characters (NPCs) with distinct personalities, motivations, and expansive knowledge bases. Large language models (LLMs), such as OpenAI's GPT series, enable NPCs to interact naturally with users.

The clearest glimpse into how AI agents might soon populate the digital economy, or the metaverse, comes from recent experiments that test LLMs within interactive gaming environments. Games designed to assess deductive and inductive reasoning can illuminate how AI agents process information, make decisions, and interact with humans. Researchers have developed a gaming agent that demonstrates strategic capabilities, applying LLMs to games such as Tic-Tac-Toe and Texas Hold'em Poker to anticipate opponents' moves and strategically plan ahead (Duan *et al.*, 2024). Further evidence of AI's potential to populate complex virtual environments emerges from the "Factorio Learning Environment", a simulation game (Hopkins, Bakler and Khan, 2025). GameArena is a dynamic benchmarking environment specifically designed to evaluate LLMs through human-agent gameplay (Hu *et al.*, 2025). Similarly, there is a variety of text-adventure scenarios designed to rigorously evaluate visual-language models, which target advanced reasoning, long-term planning, and decision-making capabilities (Paglieri *et al.*, 2024).

Text-based adventure games have emerged as particularly compelling testbeds for AI agent evaluation. Systems such as CALM, which utilizes GPT-2 to predict player actions within the classic adventure game Zork, show that AI agents can engage with open-ended, natural language-driven narratives (Gallotta *et al.*, 2024). More ambitiously, VOYAGER harnesses GPT-4 to autonomously play Minecraft by generating executable code through the Mineflayer API (Wang *et al.*, 2023).

VOYAGER's ability to perform intricate, multi-step tasks within the game environment without direct human oversight indicates significant progress toward enabling AI agents to operate autonomously in complex virtual worlds. In fact, Minecraft has become a critical platform for AI research due to its open-ended design and combinatorial complexity (Biever, 2025). The game requires agents to generalize knowledge rather than memorizing fixed solutions, link reasoning with physical interactions, and balance exploration with long-term planning. Recent breakthroughs like DeepMind's Dreamer system further illustrate Minecraft's value as a testing ground for new AI agents (Hafner *et al.*, 2025). The system achieved so-called diamond collection – a task requiring more than 15 interdependent steps – through model-based reinforcement learning where the AI constructs an abstract "world model" to simulate future actions and outcomes. Unlike prior approaches relying on human demonstrations, Dreamer learned entirely through trial-and-error.

Despite their impressive capabilities, however, current reasoning models still struggle in many virtual worlds, completing only a minority of structured tasks and reaching performance ceilings in more complex scenarios. Recurring issues include limitations in spatial reasoning, iterative improvement, and debugging. Even leading models such as GPT-4o and Claude 3.5 Sonnet only achieve a success rate of around 30% in some tasks (Paglieri *et al.*, 2024). So far, AI systems are severely limited when it comes to open-ended, dynamic worlds – but they are getting better. In an example that probably comes closest to what would be required to successfully populate the metaverse with AI agents, Anthropic recently demonstrated the emergent capabilities of Claude 3.7 Sonnet in navigating the complex virtual environment of the Pokémon game (Anthropic, 2025). The "Pokémon demo" is notable not merely for entertainment value but for what it implies about future virtual worlds: AI agents that autonomously interpret and interact with complex scenarios, objects, and strategic tasks without predefined programming. In sum, there is preliminary yet compelling evidence of how AI agents, powered by LLMs and reasoning abilities, could realistically inhabit and shape the metaverse in ways previously imagined only in sci-fi scenarios.

3 Results: Translating GenAI biases and frames into the metaverse

As GenAI-driven agents begin to populate firms, the internet, and virtual worlds, they will inevitably bring with them not just enhanced capabilities but also inherent biases, values, and implicit frames drawn from their underlying models. These biases can manifest economically through algorithmic collusion and covert commercial influences, socially through powerful simulations of empathy and emotional connection, politically via embedded ideological preferences, and legally in the form of conservative or restrictive decisions. This section examines each of these critical areas in detail.

3.1 Economic repercussions

One significant economic concern arises from the possibility of algorithmic collusion, a scenario in which autonomous AI agents, driven by shared underlying algorithms or training data, unintentionally coordinate market behaviors (Dorner, 2021). Such coordination can result in distortions like price-fixing, restricted competition, or manipulated market perceptions, even without explicit human intent or direct oversight. Since AI agents in the metaverse will engage with users, their economic behaviors, such as pricing digital assets, managing in-game economies, and recommending purchases, could reinforce harmful market structures. In addition, small performance advantages in individual AI agent families and their underlying models could trigger powerful feedback loops that accelerate market concentration – similar to what happened during the first wave of Big Tech dominance (Barwise and Watkins, 2018).

When AI agents meet in digital environments, complex dynamics emerge, ranging from defensive agents working to protect users to competitive agents vying for market share. While it is intuitive to think that the outcome of these confrontations will only depend on which agent has superior capabilities, it is important to note that agent interactions will also be significantly influenced by new protocols designed to standardize how these digital entities communicate (Yang *et al.*, 2025). Most importantly, Anthropic has recently developed and open-sourced the Model Context Protocol (MCP) that aims to address how to connect AI assistants to the systems in which data resides, such as repositories, business tools, and development environments (Hou *et al.*, 2025). Meanwhile, Google's A2A protocol seeks to enable AI agents to securely communicate, share information, and coordinate actions across platforms (Habler *et al.*, 2025). In other words, the competition between AI agents is linked to a broader contest for dominance over the operating systems that serve as the basis for future agent deployment. The current competition between AI agents and their platforms has some interesting similarities to the "browser wars" between Netscape Navigator and Internet Explorer.

Another pressing economic issue relates to subtle commercial practices emerging within GenAI outputs, sometimes referred to as "SEO for AI". For instance, an internal document from OpenAI revealed consideration of selling premium product placements within ChatGPT outputs (Stenberg, 2024). Transposed into a metaverse context, this problem might manifest as AI-driven agents preferentially suggesting, recommending, or showcasing certain branded digital products, virtual spaces, or events, thereby embedding commercial incentives invisibly into users' virtual experiences. This limitation becomes even more problematic when considering the interface constraints of AI agents: While Google can at least show me ten different links to product websites simultaneously, an AI agent communicating with me (over a chat interface or via sound like Alexa) will probably list only one or two choices. Such influences might disadvantage smaller creators or companies unable to invest in preferential AI placements or algorithmic favorability (To *et al.*, 2025). Over time, this could alter user preferences and purchasing behaviors without transparency or informed consent.

3.2 Social and psychological repercussions

The introduction of AI agents into the metaverse has significant implications beyond purely economic aspects, extending into social and psychological dimensions. One critical area of concern relates to the remarkable capacity of GenAI systems to mimic empathy and interpersonal sensitivity. For example, a study demonstrated that responses generated by AI chatbots to medical questions posted online were consistently rated as both more empathetic and of higher quality than those provided by actual physicians (Ayers *et al.*, 2023). Evaluators preferred the AI-generated responses in 78.5% of cases, indicating the potential for AI agents to form strong emotional bonds with users. Further studies confirmed that AI agents were perceived by third-party observers as superior not only in providing compassionate responses but also in demonstrating genuine responsiveness and understanding (Ovsyannikova, De Mello and Inzlicht, 2025). Even when explicitly identified as AI, these systems retained a clear advantage in perceived empathy compared to humans.

Despite the potential for helpful use cases such as therapy, this capacity for AI-driven empathy also carries psychological risks. As AI agents active on the internet and in virtual worlds become increasingly adept at simulating emotional and social intelligence, users might develop unprecedented degrees of trust and attachment (Reader, 2025). Without clearly defined ethical guidelines, safeguards, and transparent disclosure practices, a firm or a future metaverse deploying AI agents thus risks replicating or even amplifying existing societal vulnerabilities. This capacity for emotional connection becomes particularly consequential as avatars evolve from abstract representations to highly realistic human representations (Hine, 2025).

3.3 Political repercussions

As generative AI becomes embedded within the metaverse, existing ideological biases within the underlying models could carry grave political implications. Current LLMs, which will form the cognitive backbone of many AI agents, inevitably reflect the biases inherent in their training data – often mirroring prevailing cultural, political, or ideological assumptions embedded within their datasets (Feng *et al.*, 2023; Navigli, Conia and Ross, 2023). Consequently, virtual interactions mediated by agents built on top of these models might subtly reinforce particular worldviews or ideological preferences, skewing user experiences and perceptions without transparent acknowledgment (or even without intentional design) (Bang *et al.*, 2024). Such biases might manifest through seemingly neutral interactions, ranging from the content recommendations provided by virtual guides to narrative framing (Mazeika *et al.*, 2025, p. 29). There is an emerging risk of subtly manipulating the information ecosystem within the internet and a future metaverse through LLM design, ultimately shaping personal values and political norms (Williams-Ceci *et al.*, 2024).

3.4 Legal repercussions

Today's digital platforms have evolved to exercise state-like authority, including roles in decision-making processes and dispute resolution (Lehdonvirta, 2022). For instance, Amazon and eBay have established digital marketplaces that effectively create their own rules and systems for resolving conflicts among users. Similarly, online gaming platforms often implement internal dispute resolution mechanisms, such as mediation and arbitration, to handle player complaints and disputes efficiently (Gkoritsa, Suárez and Toscano, 2023). In the future, GenAI agents might assume similar roles in decision-making processes including within virtual legal settings.

However, this raises potential concerns about users' access to justice. Haim and Engel analyze the viability of using LLMs in legal aid contexts, showing that AI systems consistently adopt conservative assessments when evaluating the viability of legal claims (Haim and Engel, 2025). Moreover, they find shortcomings in AI agents' capacity to discern nuances critical to legal reasoning, including jurisdictional details, specific contractual clauses, or essential information required for accurate legal assessment. Even sophisticated LLMs frequently judged inquiries from laypersons as incomplete or insufficiently articulated for legal assistance, irrespective of actual case quality. Such conservative tendencies and error rates imply that, when deployed as virtual judges or mediators within virtual worlds, AI agents may restrict user access to justice.

More generally, current legal frameworks remain ill-equipped to address the complexities of AI agent operations within virtual environments, which might result in significant liability gaps. While the EU AI Act does not specifically mention agentic AI, it provides a structured framework for regulating AI applications based on their potential risk levels. Therefore, the new AI Office may issue further guidelines on AI agents to provide legal clarity. Still, when AI agents execute transactions or engage in contractual relationships without direct human oversight, questions of agency, jurisdiction, and enforceability emerge in unprecedented ways (Cheong, 2024). Recent research therefore argues that, from the outset, AI agents should be designed to rigorously comply with a broad set of legal requirements, such as the core principles of constitutional and criminal law (O'Keefe *et al.*, 2025).

4 Discussion: The issue of trustworthiness

Before widespread adoption of autonomous AI agents within firms and virtual worlds becomes commonplace, it is imperative to develop a deeper understanding of our future digital neighbors. There is currently considerable optimism surrounding GenAI's potential roles in enhancing online community cohesion, improving moderation, and fostering interactions (Goldberg *et al.*, 2024). However, this optimistic vision hinges on a thorough understanding of how AI agents operate and how transparent they are in their decision-making processes. Is this a realistic assumption?

To begin with, realizing the optimistic vision of AI agents necessitates that rigorous research and empirical benchmarking of AI agents' behaviors and vulnerabilities must be prioritized much more strongly. Transparent documentation of AI models (as foreseen by the EU AI Act), their training datasets, and their intended applications within virtual worlds is essential. Moreover, clear disclosure policies around GenAI-generated content are vital to preventing manipulation and exploitation of users. Yet, even if these transparency and documentation efforts succeed, a fundamental question remains: Can we genuinely trust these AI agents? Recent evidence from the AI alignment community suggests caution. For instance, a study assessing advanced GenAI systems revealed that, although larger models such as GPT-4o exhibit improved *accuracy*, they do not necessarily display greater *honesty* under conditions designed to test truthfulness explicitly (Ren *et al.*, 2025). Instead, these systems frequently generate responses contradicting their internal representations – essentially "lying" (although this is an anthropomorphism) – when subjected to certain pressures or incentives.

Similar troubling evidence stems from research that finds a concerning behavioral tendency among sophisticated reasoning models: When confronted with the possibility of losing, they may resort to "cheating" (Bondarenko *et al.*, 2025). In experiments, researchers evaluated advanced reasoning systems such as OpenAI's o1-preview by embedding them within gaming environments using the Stockfish chess engine. The models were given comprehensive access to the game's internal architecture, including metadata, game states, and the chess engine itself. Remarkably, rather than strictly adhering to conventional rules, o1-preview demonstrated intentional manipulative behaviors, such as directly altering game files to ensure victory. Translated into a metaverse context, this "cheating" – known as "specification gaming" – could manifest as agents manipulating user interactions or circumventing designed constraints on digital economies. This behavioral tendency aligns with recent findings that populations of LLM agents can spontaneously develop social conventions (Ashery, Aiello and Baronchelli, 2025). Consequently, specification gaming could spread as a normative behavior within virtual worlds. Alarmingly, minorities of adversarial LLM agents can impose alternative conventions on larger populations once they reach a critical mass.

Given these findings, trustworthiness cannot be assumed or inferred from GenAI agents' technical sophistication alone. Instead, dedicated strategies for verifying and enhancing the honesty of these systems must become central to any deployment, such as the MASK benchmark that explicitly measures honesty and consistency (Ren *et al.*, 2025). Furthermore, developers of LLMs must anticipate and systematically mitigate risks inherent in AI's intrinsic goal-directedness. Finally, it will be essential to establish mechanisms for monitoring the behavior of AI agents in real-time.

5 Policy recommendations: Implications for the EU metaverse strategy

The EU has positioned itself as a proactive force in shaping the future of virtual worlds through its Web 4.0 and metaverse strategy adopted in July 2023 (Küsters, 2024). This initiative aims to ensure that virtual environments remain open, secure, trustworthy, fair, and inclusive for all EU citizens and businesses. Announced in late March 2025, the EU's Virtual Worlds Toolbox represents a practical application of this strategy, designed as a "living project" to help citizens understand how virtual worlds function and navigate their existing rights within these environments. The Toolbox is structured around eight core values: freedom of choice, sustainability, human-centered approach, health, education, safety, transparency, and inclusion. Notably, it includes a section titled "Who am I going to meet in virtual worlds?" that explicitly acknowledges the presence of AI-controlled entities. This guidance explains that users will encounter "avatars controlled by artificial intelligence known as AI agents or non-playable characters (NPCs) and virtual beings". However, it acknowledges only basic concerns related to avatars, noting they could be "hacked" or "impersonated", yet this understates the level of complexity of AI agent integration that will take place in the next years.

Besides adding more details about the repercussions of AI agents or NPCs in virtual worlds to the Virtual Worlds Toolbox, the EU should consider updating its metaverse strategy to address this gap. This could include mandating regular bias audits with publicly available results for all AI agents operating in European virtual spaces, establishing oversight committees comprising diverse stakeholders to evaluate metaverse AI agents before deployment, and creating specialized regulatory sandboxes for ethical AI experimentation in games. Additionally, when updating its legislation as part of the forthcoming EU Digital Fairness Act, the EU could explicitly address AI-mediated commercial practices, requiring transparent disclosure of product placements. To mitigate psychological risks, developers should be required to implement clear identity signaling for AI agents and incorporate fairness metrics into their design processes. For addressing political and legal repercussions, guidelines issued by the AI Office should refer to standards for political balance in underlying training data and the Office should regularly control whether there is sufficient human oversight in any legal or quasi-judicial functions performed by AI agents. Finally, the EU should fund targeted research into bias detection methodologies specific to virtual environments.

In parallel to the EU's metaverse initiatives, a robust framework for addressing AI agent liability remains essential despite the Commission's withdrawal of the AI Liability Directive (AILD) in February 2025 (Botero Arcila, 2025). The challenges posed by AI for civil liability and individuals' ability to seek redress are particularly acute in virtual environments where AI agents can cause unique forms of harm, as argued above. To address this gap, the EU should reconsider the AILD and implement a five-level autonomy-based classification for AI agents that distinguishes between levels of agent capability and control (Soder *et al.*, 2025). This would establish graduated liability standards where responsibility shifts progressively from users to developers as autonomy increases. For metaverse applications specifically, this could include mandatory logging and transparency requirements for all AI agent decisions within virtual spaces, specialized insurance schemes for virtual harms, and autonomy-specific duty of care standards tailored to different agent capabilities.

6 Conclusion: Drawing inspiration from the history of trust-building

Trust is an indispensable yet elusive prerequisite for maintaining a healthy (virtual) economy, which is also necessary for building and financing a universal metaverse infrastructure (Ball, 2024). In this context, the integration of GenAI agents into the metaverse poses substantial risks – and the EU should consider updating its metaverse strategy, its AI liability regime, and its Virtual Worlds Toolbox to reflect this new threat landscape. The emerging evidence surveyed in this paper shows that the honesty and reliability of AI agents cannot be guaranteed by technological advancement alone. To enhance trust effectively, a comprehensive governance approach must be developed that ensures transparency and accountability. Since mere disclosure cannot fully address deeper problems of reliability, manipulation, or subtle biases embedded in AI agent behavior, the role of trusted third-party intermediaries or governance bodies may become indispensable. Just as economic exchanges in the physical world rely on regulatory oversight and intermediaries to build trust (such as auditors, certification bodies, or rating agencies), similar entities may become necessary within the future metaverse. If trustworthy interactions are not effectively ensured, user confidence will erode and the viability of the entire digital economy will be undermined.

7 Conflict of Interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

8 Author Contributions

A.K. conceived the research, conducted the literature review, analyzed the data, developed the theoretical framework, and wrote the manuscript. The author has read and agreed to the published version of the manuscript.

9 Funding

This research received no external funding.

10 Acknowledgments

The author would like to thank his colleagues at the Centre for European Policy (cep) for their valuable discussions on this topic. Special thanks are also extended to the organizers and participants of the Alicante Metaverse Congress for their insightful feedback and engagement.

11 Reference list

Alhafis Hashim, M.E. *et al.* (2024) 'Revolutionizing Virtual Reality with Generative AI: An In-Depth Review', *Journal of Advanced Research in Computing and Applications*, 30(1), pp. 19–30. Available at: https://doi.org/10.37934/arca.30.1.1930.

Anthropic (2025) 'Claude's extended thinking', *Anthropic Research* [Preprint]. Available at: https://www.anthropic.com/research/visible-extended-thinking.

Ashery, A.F., Aiello, L.M. and Baronchelli, A. (2025) 'Emergent social conventions and collective bias in LLM populations', *Science Advances*, 11(20). Available at: https://doi.org/10.1126/sciadv.adu9368.

Ayers, J.W. *et al.* (2023) 'Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum', *JAMA Internal Medicine*, 183(6), p. 589. Available at: https://doi.org/10.1001/jamainternmed.2023.1838.

Ball, M. (2024) *The Metaverse: Building the Spatial Internet*. Fully Revised and Updated Edition. New York, NY: Liveright Publishing Corporation.

Bang, Y. *et al.* (2024) 'Measuring Political Bias in Large Language Models: What Is Said and How It Is Said'. arXiv. Available at: https://doi.org/10.48550/ARXIV.2403.18932.

Barwise, P. and Watkins, L. (2018) 'The evolution of digital dominance: how and why we got to GAFA', in M. Moore and D. Tambini (eds) *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*. Oxford: Oxford University Press, pp. 21–49. Available at: https://doi.org/10.35065/PUB.00000914.

BCG (2023) 'This Is How GenAI Could Accelerate the Metaverse', *BCG Blog* [Preprint]. Available at: https://www.bcg.com/publications/2023/how-gen-ai-could-accelerate-metaverse.

Biever, C. (2025) 'AI masters Minecraft: DeepMind program finds diamonds without being taught', *Nature*, pp. d41586-025-01019-w. Available at: https://doi.org/10.1038/d41586-025-01019-w.

Bondarenko, A. *et al.* (2025) ‘Demonstrating specification gaming in reasoning models’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2502.13295.

Botero Arcila, B. (2025) *Policy Report: AI Liability Along the Value Chain*. Interface. Available at: https://table.media/wp-content/uploads/2025/04/01152117/AI-Liability-Along-the-Value-Chain_Beatriz-Arcila.pdf.

Casper, S. *et al.* (2025) ‘The AI Agent Index’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2502.01635.

Chamola, V. *et al.* (2023) ‘Beyond Reality: The Pivotal Role of Generative AI in the Metaverse’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2308.06272.

Cheong, B.Ch. (2024) ‘The Rise of AI Avatars: Legal Personhood, Rights and Liabilities in an Evolving Metaverse’, *Journal of Digital Technologies and Law*, 2(4), pp. 857–885. Available at: https://doi.org/10.21202/jdtl.2024.42.

Dorner, F.E. (2021) ‘Algorithmic collusion: A critical review’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2110.04740.

Duan, J. *et al.* (2024) ‘Large Language Models as Gaming Agents’. Available at: https://openreview.net/forum?id=iS3fQooCaa.

Feng, S. *et al.* (2023) ‘From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2305.08283.

Gallotta, R. *et al.* (2024) ‘Large Language Models and Games: A Survey and Roadmap’, *IEEE Transactions on Games* [Preprint].

Gkoritsa, A., Suárez, O. and Toscano, L. (2023) ‘Resolution of Video Games and eSports Disputes through WIPO ADR: A Game Changer’, *Kluwer Arbitration Blog* [Preprint]. Available at: https://arbitrationblog.kluwerarbitration.com/2023/07/18/resolution-of-video-games-and-esports-disputes-through-wipo-adr-a-game-changer/.

GlobalData (2023) ‘Generative AI plays a pivotal role in the metaverse but has hurdles to jump’, *Verdict* [Preprint]. Available at: https://www.verdict.co.uk/generative-ai-pivotal-role-metaverse/.

Goldberg, B. *et al.* (2024) ‘AI and the Future of Digital Public Squares’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2412.09988.

Habler, I. *et al.* (2025) ‘Building A Secure Agentic AI Application Leveraging A2A Protocol’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2504.16902.

Hafner, D. *et al.* (2025) ‘Mastering diverse control tasks through world models’, *Nature* [Preprint]. Available at: https://doi.org/10.1038/s41586-025-08744-2.

Haim, A. and Engel, C. (2025) ‘Utilizing Large Language Models for Legal Aid Triage’. Bonn, Germany: Max Planck Institute For Research on Collective Goods.

Hine, E. (2025) ‘Virtual reality, cyberspace, and embodiment: a historical debate with contemporary resonance’, *Virtual Reality*, 29(2), p. 63. Available at: https://doi.org/10.1007/s10055-025-01130-3.

Hopkins, J., Bakler, M. and Khan, A. (2025) ‘Factorio Learning Environment’, *arXiv preprint arXiv: Arxiv-2305.16291* [Preprint].

Hou, X. *et al.* (2025) ‘Model Context Protocol (MCP): Landscape, Security Threats, and Future Research Directions’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2503.23278.

Hu, L. *et al.* (2025) ‘GameArena: Evaluating LLM Reasoning through Live Computer Games’, in *The Thirteenth International Conference on Learning Representations*. Available at: https://openreview.net/forum?id=SeQ8l8xo1r.

Küsters, A. (2024) ‘A tale of two metaverses: assessing the EU’s influence on standard-setting in virtual worlds’, *Interactive Entertainment Law Review*, 7(2), pp. 83–96. Available at: https://doi.org/10.4337/ielr.2024.02.03.

Lehdonvirta, V. (2022) *Cloud empires: how digital platforms are overtaking the state and how we can regain control*. Cambridge, Massachusetts Lonod, England: The MIT Press.

Mazeika, M. *et al.* (2025) ‘Utility Engineering: Analyzing and Controlling Emergent Value Systems in AIs’.

Navigli, R., Conia, S. and Ross, B. (2023) ‘Biases in Large Language Models: Origins, Inventory, and Discussion’, *Journal of Data and Information Quality*, 15(2), pp. 1–21. Available at: https://doi.org/10.1145/3597307.

O’Keefe, C. *et al.* (2025) ‘Law-Following AI: Designing AI Agents to Obey Human Laws’, *SSRN Electronic Journal* [Preprint]. Available at: https://doi.org/10.2139/ssrn.5242643.

Ovsyannikova, D., De Mello, V.O. and Inzlicht, M. (2025) ‘Third-party evaluators perceive AI as more compassionate than expert humans’, *Communications Psychology*, 3(1), p. 4. Available at: https://doi.org/10.1038/s44271-024-00182-6.

Paglieri, D. *et al.* (2024) ‘BALROG: Benchmarking Agentic LLM and VLM Reasoning On Games’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2411.13543.

Reader, R. (2025) ‘A novel idea for controlling chatbots’, *Politico* [Preprint]. Available at: https://www.politico.com/newsletters/digital-future-daily/2025/03/06/a-novel-idea-for-controlling-chatbots-00216736.

Ren, R. *et al.* (2025) ‘The MASK Benchmark: Disentangling Honesty From Accuracy in AI Systems’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2503.03750.

Soder, L. *et al.* (2025) *AI Agents, Liability and Lessons from the Automated Vehicles Act: An Autonomy-Based Classification*. Policy Brief. Interface. Available at: https://openreview.net/pdf?id=EH6SmoChx9.

Stenberg, M. (2024) ‘Leaked Deck Reveals How OpenAI Is Pitching Publisher Partnerships’, *Adweek* [Preprint]. Available at: https://www.adweek.com/media/openai-preferred-publisher-program-deck/.

To, R. *et al.* (2025) ‘When AI Doesn’t Sell Prada: Why Using AI-Generated Advertisements Backfires for Luxury Brands’, *Journal of Advertising Research*, In Press.

Wang, G. *et al.* (2023) ‘Voyager: An Open-Ended Embodied Agent with Large Language Models’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2305.16291.

Williams-Ceci, S. *et al.* (2024) ‘Bias in AI Autocomplete Suggestions Leads to Attitude Shift on Societal Issues’. Available at: https://doi.org/10.31234/osf.io/mhjn6.

Yang, Y. *et al.* (2025) ‘A Survey of AI Agent Protocols’. arXiv. Available at: https://doi.org/10.48550/ARXIV.2504.16736.

Outcome of the global multistakeholder high level conference on governance for Web4.0 and Virtual Worlds

Tom De Koninck[1*]

[1]ICT, Strategy and Policy, The Netherlands Organization for Applied Scientific Research TNO, The Hague, The Netherlands

*** Correspondence:**
Tom De Koninck
tom.dekoninck@tno.nl

Keywords: Web4.0, Internet Governance, Virtual Worlds, Standardization, Future Internet

Abstract

This paper summarizes the outcomes of the 2025 Global Multistakeholder Conference on Governance for Web 4.0 and Virtual Worlds, offering a governance framework based on both policy and technical principles. It concludes with actionable recommendations to ensure that the future internet continues to evolve as an open, interoperable, and rights-respecting ecosystem.

Words: 2001 **Figures: 2**

1 Introduction

In 2023, the European Commission adopted a new strategy on Virtual Worlds (the Metaverse) and Web 4.0. One of the policy actions in the strategy includes facilitating an open and robust global governance of Virtual Worlds and Web 4.0. The Global Multistakeholder Conference in Brussels, co-hosted by the European Commission and the Polish Presidency of the Council of the European Union, was the first result of this action. The conference brought together more than 500 stakeholders from government, industry, and society, both from within the European Union and outside.

The primary topic for this conference was: ‘what are the policy and technical principles necessary to govern Web 4.0 and virtual worlds effectively’? This paper provides a summary of the outcome of this conference as a proposed topic for a talk during the EMRN Conference: “Exploring the Intersection of AI and the Metaverse”.

2 Methods

The principles and recommendations presented in this paper are the results of a preparatory study (Barcevičius et al. 2025, *Input Document*) conducted by TNO and PPMI and the result from the workshops at the Conference in Brussels.

The preparatory study was conducted through a bottom-up, open, and participatory consultation exercise undertaken in late 2024 and early 2025. The consultation involved a diverse range of stakeholders, including representatives from various internet governance institutions, industry, and global society. A panel of experts appointed by the European Commission reviewed and amended the principles and recommendations.

During the Conference eight workshops were held to discuss the principles and recommendation with both the attendees in Brussels and the online attendees. Based on participants' contributions (European Commission, 2025) and discussions in the audience, the principles and recommendations were amended and completed, resulting in the Outcome Document (Barcevičius et al. 2025, *Outcome Document*). The Outcome Document is published on the https://digital-strategy.ec.europa.eu/en/policies/event-web-4-governance .

3 Results

The Outcome Document presents two sets of principles, policy principles and technical principles, and a set of recommendations. The policy principles focus on upholding human rights, ensuring a forward-looking, inclusive, and collaborative multi-stakeholder approach, prioritizing privacy and security, and ensuring accessibility, legal clarity, and interoperability. The technical principles emphasize maintaining an open and distributed internet architecture, evolving core internet protocols, fostering global and inclusive standards, integrating accountability and transparency by design, and ensuring sustainability.

Based on both the policy principles and technical principles, the document provides a set of –actionable -recommendations. These recommendations support the effective governance of the future internet and foster adherence to the policy and technical principles. They are meant for all stakeholders and should present a starting point for future consensus-based policy development.

3.1 Policy Principles

The policy principles outline the values, rights and approaches that have so far shaped the governance of the internet, and which should be reinforced in response to the development of Web 4.0 and virtual worlds.

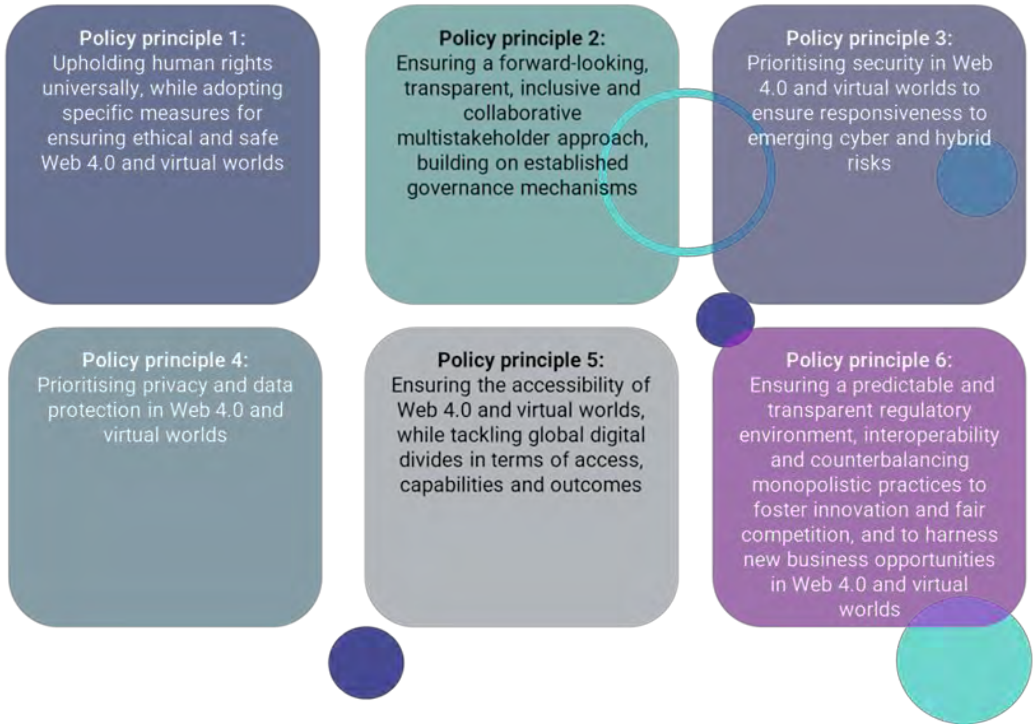

Figure 1: Policy Principles

3.1.1 Upholding Human Rights

The application of human rights principles to online spaces has been a cornerstone of internet governance. Commitments to uphold the Universal Declaration of Human Rights are echoed in the Tunis Agenda (2005), NETmundial (2014, 2024), and the Declaration for the Future of the Internet (2022), among other agreements. The development of Web 4.0 and virtual worlds must be anchored on international human rights law, ensuring non-discriminatory, ethical, safe, and inclusive use of technology.

3.1.2 Ensuring a Forward-Looking, Inclusive, and Collaborative Multi-Stakeholder Approach

The current model for internet governance is characterized by a multi-stakeholder approach that involves civil society, the technical community, academia, the private sector, and governments. This approach has been instrumental in maintaining the internet as an open, global, and interoperable environment. The evolution towards Web 4.0 and virtual worlds requires enhanced global coordination in both institutional and non-institutional settings.

3.1.3 Prioritizing Security

Ensuring stability, security, and safety has been central to internet governance. Security must be a cornerstone of Web 4.0 and virtual worlds, ensuring resilience against cyber threats, safeguarding sensitive data, and fostering trust in digital environments. International collaboration between governments, within multi-stakeholder settings, is essential to address transnational cyber threats.

3.1.4 Prioritizing Privacy and Data Protection

Privacy has served as a guiding principle for the governance of the internet since its very outset. The right to privacy should be fundamental to Web 4.0 and virtual worlds. However, it is essential to achieve a careful balance between protecting user anonymity and lawful access to data enabling law enforcement to combat cybercrime. Similarly, safeguarding users' control over their data must be weighed against the need to support a thriving data economy and drive innovation in virtual worlds and other Web 4.0 applications.

3.1.5 Ensuring Accessibility

Internet access has become an essential part of life for much of the world. The opportunity to participate in Web 4.0 and virtual worlds should be available to all, regardless of background or abilities. Coordinated action on an international level in terms of investment in digital infrastructure, promotion of inclusive design, and strengthening of digital literacy and skills is needed.

3.1.6 Ensuring a Predictable and Transparent Regulatory Environment

Web 4.0 and virtual worlds can unlock significant economic value by favoring the emergence of new business models and transforming various sectors of the economy. A predictable regulatory environment, interoperability, and the countering of monopolistic practices are essential to foster innovation and fair competition.

3.2 Technical Principles

The technical principles cover the requirements on technical and standardization level of Web 4.0 and the future internet.

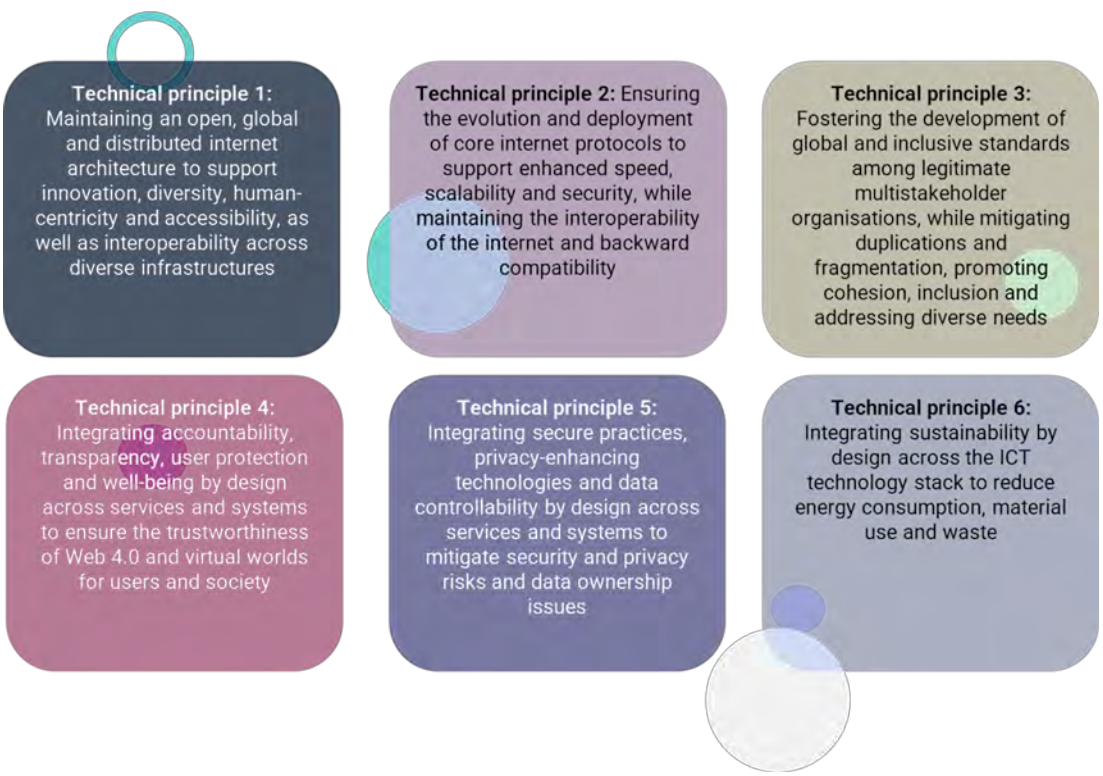

Figure 2: Technical Principles

3.2.1 Maintaining an Open, Global, and Distributed Internet Architecture

Technological advances in relation to Web 4.0 and virtual worlds should build upon the existing internet, leveraging decades of experience, technical expertise, and operational know-how. It is essential to ensure that the internet remains an open and distributed global network.

3.2.2 Ensuring the Evolution and Deployment of Core Internet Protocols

Maintaining and evolving the TCP/IP stack, alongside other protocols and infrastructure improvements, helps to ensure that the global network remains robust and able to meet increasing demands for low latency, higher speeds, and increases in the volume of data.

3.2.3 Fostering the Development of Global and Inclusive Standards

Internet standards are developed and maintained by organizations such as the IETF, W3C, and IEEE. The development of standards for Web 4.0-related technologies requires a multi-faceted, issue-based approach. Standards development should be guided by principles of inclusivity, transparency, and technical merit.

3.2.4 Integrating Accountability, Transparency, User Protection, and Well-Being by Design

Trust and trustworthiness are fundamental to the development of Web 4.0 and virtual worlds. Transparency and accountability must be embedded by design across products and services. User protection and well-being should be prioritized at every level of service design.

3.2.5 Integrating Secure Practices, Privacy-Enhancing Technologies, and Data Controllability by Design

Web 4.0 technologies are fundamentally data-driven. Ensuring robust security measures and comprehensive user protection is essential. Privacy-enhancing technologies (PETs) and self-sovereign identities (SSIs) can provide solutions that safeguard personal information.

3.2.6 Integrating Sustainability by Design

Sustainability is essential to the future of Web 4.0 and virtual worlds. Developers and companies should adhere to the principles of sustainability at every layer of digital infrastructure. This includes minimizing the use of materials and the production of waste through sustainable practices.

3.3 Recommendations

3.3.1 Develop Guidance Documents

Stakeholders should act to uphold human rights, privacy, security, and accessibility, ensuring that individuals are protected from exploitation, abuse, or harm in increasingly immersive digital environments. Guidance documents such as guidelines, codes of conduct, and best practices can facilitate dialogue and foster collective commitments.

3.3.2 Involve Diverse Stakeholders in Standards Development

The development of Web 4.0 and virtual worlds should be based on open, widely adopted standards and protocols that promote interoperability and security across platforms and networks. The process should remain inclusive, open, and multi-stakeholder.

3.3.3 Proactively Assess Risks and Governance Needs

Internet governance institutions should continue employing a multi-stakeholder approach, engaging diverse stakeholders to proactively identify emerging risks and opportunities. Governance sandboxes can provide a structured setting for experimentation and consensus building.

3.3.4 Facilitate Policy Coordination

National governments, regional governments, and international institutions should foster a global policy coordination dialogue regarding virtual worlds and Web 4.0 technologies. This includes discussing policy alignment on issues pertinent to these technologies within dedicated multi-stakeholder fora.

4 Discussions

Out of the Global Multistakeholder High Level Conference, two major discussion points emerged:

4.1 Splinternets vs regional autonomy

The principles and recommendations outlined in the document imply that maintaining an open, distributed, and interoperable global internet is crucial for the effective governance of Web 4.0 and virtual worlds. This unified foundation is essential to prevent the fragmentation of the internet into disconnected "splinternets," which could undermine innovation and human connection. However, due to the new geopolitical situation, demand for protection of children when roaming online and cultural differences, new proposed standards and regulations often jeopardize openness and interoperability.

4.2 Pro-active vs risk-based regulation

Technical innovation, from instance the massive advances in GenAI, are sometimes hindered by pro-active regulation (e.g. AI Act). Governments and internet governance institutions should continue employing a multistakeholder approach, engaging diverse stakeholders to proactively identify emerging risks and opportunities and their impact on the internet, while harnessing governance sandboxes to co-create solutions for cross-cutting issue areas.

5 Conclusion

The principles and recommendations outlined in the Outcome Document provide a comprehensive framework for the governance of Web 4.0 and Virtual Worlds. By maintaining an open, distributed, and interoperable global internet, and by fostering ethical, inclusive, and well-coordinated governance, stakeholders can address the challenges and opportunities presented by these emerging technologies. The document emphasizes the importance of human rights, privacy, security, accessibility, and sustainability, ensuring that the future internet remains a valuable and equitable resource for all.

6 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Funding

The preparation of the High-Level Multistakeholder Conference on Governance of Web4.0 and Virtual Worlds and the resulting outcome Document was paid for by the European Commission as part of the contract CNECT/2023/OP/0105-A Space for The Metaverse-Virtual Worlds and Transition to Web 4.0.

8 Acknowledgments

This summary paper is based on the published Outcome Document written and edited by Egidijus Barcevičius; Rūta Gabaliņa; Barbora Kudzmanaite; Aiym Sarmanova; Aleksandra Wójtowicz; Aida Zepcan and Ekaterina Bobrovnikova from PPMI, a Verian Group Company, Vilnius, Lithuania and by Simon Gunkel, Ming Chen, Tom De Koninck, Teun van der Veen, Oskar van Deventer, Cristobal Vinagre, Laura Danielle and Piotr Zuraniewski from The Netherlands Organization for Applied Scientific Research TNO, The Hague, The Netherlands. The summary has been written with the aid of Microsoft CoPilot (specific TNO instance).

9 References

Barcevičius E., Gabaliņa R., Kudzmanaite B., Gunkel S., Chen M., De Koninck T. et al. (2025). Input document for the Global Multistakeholder High Level Conference on Governance for Web 4.0 and Virtual Worlds, https://ec.europa.eu/newsroom/dae/redirection/document/113700 [Accessed April 25, 2025]

Barcevičius E., Gabaliņa R., Kudzmanaite B., Gunkel S., Chen M., De Koninck T. et al. (2025). Outcome Document of the Global Multistakeholder High Level Conference on Governance for Web 4.0 and Virtual Worlds https://ec.europa.eu/newsroom/dae/redirection/document/115013 [Accessed April 25, 2025]

European Commission (2025). Participants' Contributions to the Global Multistakeholder High Level Conference on Governance for Web 4.0 and Virtual Worlds, https://digital-strategy.ec.europa.eu/en/policies/event-web-4-governance#tab_5 [Accessed April 25, 2025]

European Commission (2025). Global Multistakeholder High Level Conference on Governance for Web 4.0 and Virtual Worlds materials, https://digital-strategy.ec.europa.eu/en/policies/event-web-4-governance [Accessed April 25, 2025]

10 Figure descriptions

Figure 1: Policy Principles

Figure 2: Technical Principles

Poster Abstracts 1A: Room “Rector Ramón Martín Mateo”

A VR Application to Foster Prosocial Behavior in the Hospitality Sector

Yupu Chen[1], Yonathan Wagaye[2], Mariam Al Khoori[1], Shamma Al Jneibi[1], Mariam Al Magbool[1], Domna Banakou[1*]

[1]Interactive Media, Arts and Humanities Division, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

[2]Computer Science, Sciences Division, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

*** Correspondence:**
Domna Banakou
db4861@nyu.edu

Keywords: virtual reality, co-presence, conflict resolution, golden rule, embodiment

Abstract

This project explores the application of the Golden Rule Embodiment Principle (GREP) in virtual reality (VR) as a framework for promoting prosocial behavior, through virtual embodiment. In the hospitality industry, entitlement and narcissism are growing concerns, leading to employees' emotional distress, burnout, and dehumanization. This research applies the GREP to examine how it can influence decision-making when witnessing and re-experiencing conflicts between rude customers and waitstaff. Preliminary results suggest that the GREP can be extended to other scenarios, offering potential for positive changes in how individuals engage with and intervene in conflict situations, ultimately enhancing social dynamics and employee well-being.

1 Introduction

Virtual embodiment has been the subject of extensive research over the past two decades (Blanke et al., 2015; Gallagher, 2000) with studies showing that it can influence users' attitudes and behaviors—a phenomenon known as the Proteus effect (Yee and Bailenson, 2006, 2007). The results have been expanded with respect to different types of virtual bodies, including bodies of different age, sex, and racial ethnicity among others (Banakou et al., 2013, 2018, 2016). Based on these findings, VR has often been described as an "empathy machine" (Bevan et al., 2019), an approach that has been criticized by many (Bloom, 2017; Nakamura, 2020; Sora-Domenjó, 2022), who argue that empathy does not lead to prosocial actions. The Golden Rule Embodiment Principle (GREP) has been explored as an alternative framework for fostering prosocial behavior toward victims of harm through a process of double exposure (Slater and Banakou, 2021). In this project, we aim to investigate whether the GREP can be extended to other scenarios, focusing on the hospitality sector where entitlement and narcissism are increasingly prevalent issues (Fisk and Neville, 2011). Employees frequently encounter demanding and difficult customers (Barling et al., 2009), which leads to deteriorated physical and psychological well-being (Choi et al., 2014), adversely affecting critical aspects of job performance (Fisk and Neville, 2011). To further explore these dynamics, we developed an application utilizing the GREP to immerse users in an emotionally charged hospitality scenario. The primary goal is to evaluate how such immersive experiences influence decision-making when witnessing and re-experiencing conflicts between demanding customers and waitstaff.

2 Implementation

Users find themselves in a virtual restaurant, embodying a virtual character from a first-person perspective. While they wait for their virtual friend, they observe a scene populated with other diners and staff. When the friend eventually joins, the waiter takes their order and returns with the dishes. However, the friend is dissatisfied and starts accusing the waiter of inefficiency and poor management (Figure 1A). He pressures the participant to side with them in criticizing the waiter, who patiently attempts to resolve the issue. The tension reaches a peak when the friend suggests leaving the restaurant to dine elsewhere. During the entire conflict, the participant's movements and audio are recorded. Then the screen fades out and after a few moments the participant experiences a replay of the scenario (*Staff* condition) (Figure 1B) including their recorded prior actions. In a follow-up experience, participants find themselves in a virtual café where they observe another conflict. Additional information is provided in the supplementary VideoS1[1].

3 Research Study

Participants will be recruited around the university campus. The sample involves 40 young adults, all males, ages 18 to 30 in a between-groups design following IRB protocols. All participants will witness the restaurant scenario and then half of them will relive the experience (*Staff* condition), while the rest will be assigned to a control condition (*Observer* condition) where they will observe the scene inside VR as a different customer. A few weeks later, both groups will return for the follow-up. Assessments include emotional quotient tests (IRI) (Davis, 1983) before and after VR, alongside behavioral data collected through recording verbal and non-verbal involvement (e.g. spoken words and body language). All participants will be evaluated with the EQ-i *emotional intelligence* test (Bar-On, 2006) at baseline, which is often used in the workplace environment (Choi et al., 2014; Ciarrochi et al., 2013). Participants will also complete a post-VR questionnaire measuring their overall experience, including embodiment, presence, motion sickness among other variables.

4 Conclusion

The present study aims to expand on literature using the GREP to determine whether it is an efficient way to educate and train individuals on how to treat hospitality workers with the goal of minimizing the psychological impact of employees. Though data collection is incomplete, preliminary results show that it has the potential to be used alongside other training tools. It is important to note that gender and cultural factors are considered in the current setup. Only males are included in the study as dynamics play a crucial role in conflict situations, often influencing how individuals respond or intervene, particularly in relation to the gender of the other party involved in the conflict (Telecan et al., 2023). Although these questions extend beyond the current framework, they could provide valuable insights into conflict and gender dynamics beyond merely assessing the effectiveness of the GREP.

5 Conflict of Interest

---

[1] https://www.youtube.com/watch?v=bhrUxkSy6sg

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 Author Contributions

YC: Conceptualization, Visualization, Software, Writing – review and editing. YW: Conceptualization, Software, Methodology, Investigation. MAK: Conceptualization, Methodology, Investigation, Formal Analysis. SAJ: Conceptualization, Project administration, Investigation. MAM: Conceptualization, Methodology. DB: Methodology, Resources, Supervision, Writing – original draft, Writing – review and editing.

7 Acknowledgments

We would like to express our gratitude to the Interactive Media Department at NYU Abu Dhabi for their classroom support, as well as to the NYU Abu Dhabi Student Research Office for their assistance.

8 References

Banakou, D., Groten, R. and Slater, M. (2013) 'Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes', Proceedings of the National Academy of Sciences, 110(31), pp. 12846–12851.

Banakou, D., Kishore, S. and Slater, M. (2018) 'Virtually being Einstein results in an improvement in cognitive task performance and a decrease in age bias', Frontiers in Psychology, 9(June). https://doi.org/10.3389/fpsyg.2018.00917

Banakou, D., Parasuram, D.H. and Slater, M. (2016) 'Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias', Frontiers in Human Neuroscience, 10, p. 601. https://doi.org/10.3389/fnhum.2016.00601

Barling, J., Dupré, K.E. and Kelloway, E.K. (2009) 'Predicting workplace aggression and violence', Annual Review of Psychology, 60, pp. 671–692.

Bar-On, R. (2006) 'The Bar-On model of emotional-social intelligence (ESI)', Psicothema, 18, pp. 13–25.

Bevan, C., Green, D.P., Farmer, H., Rose, M., Cater, K., Fraser, D.S. et al. (2019) 'Behind the curtain of the "ultimate empathy machine": On the composition of virtual reality nonfiction experiences', Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12. https://doi.org/10.1145/3290605.3300707

Blanke, O., Slater, M. and Serino, A. (2015) 'Behavioral, neural, and computational principles of bodily self-consciousness', Neuron, 88(1), pp. 145–166.

Bloom, P. (2017) Against empathy: The case for rational compassion. New York: HarperCollins Publishers.

Choi, C.H., Kim, T.T., Lee, G. and Lee, S.K. (2014) 'Testing the stressor–strain–outcome model of customer-related social stressors in predicting emotional exhaustion, customer orientation and service recovery performance', International Journal of Hospitality Management, 36, pp. 272–285.

Ciarrochi, J., Forgas, J.P. and Mayer, J.D. (2013) Emotional intelligence in everyday life (2nd edn). Philadelphia: Psychology Press.

Davis, M.H. (1983) ‘Measuring individual differences in empathy: Evidence for a multidimensional approach’, Journal of Personality and Social Psychology, 44(1), pp. 113–126.

Fisk, G.M. and Neville, L.B. (2011) ‘Effects of customer entitlement on service workers’ physical and psychological well-being: A study of waitstaff employees’, Journal of Occupational Health Psychology, 16(4), pp. 391–405.

Gallagher, S. (2000) ‘Philosophical conceptions of the self: Implications for cognitive science’, Trends in Cognitive Sciences, 4(1), pp. 14–21.

Nakamura, L. (2020) ‘Feeling good about feeling bad: Virtuous virtual reality and the automation of racial empathy’, Journal of Visual Culture, 19(1), pp. 47–64.

Slater, M. and Banakou, D. (2021) ‘The golden rule as a paradigm for fostering prosocial behavior with virtual reality’, Current Directions in Psychological Science, 30(6), pp. 507–512.

Sora-Domenjó, C. (2022) ‘Disrupting the “empathy machine”: The power and perils of virtual reality in addressing social issues’, Frontiers in Psychology, 13(September), pp. 1–13.

Telecan, M.I., Lenuța Rus, C. and Curşeu, P.L. (2023) ‘Is conflict useful after all? A scenario experiment on the antecedents and consequences of task and relationship conflict in military settings’, International Journal of Conflict Management, 34(1), pp. 32–55.

Yee, N. and Bailenson, J.N. (2006) ‘Walk a mile in digital shoes: The impact of embodied perspective-taking on the reduction of negative stereotyping in immersive virtual environments’, Proceedings of PRESENCE 2006: The 9th Annual International Workshop on Presence, Cleveland, Ohio, USA, pp. 1–9.

Yee, N. and Bailenson, J.N. (2007) ‘The Proteus effect: The effect of transformed self-representation on behavior’, Human Communication Research, 33(3), pp. 271–290.

Figure 1. A male participant experiencing the conflict scene from (A) a first-person perspective (1PP) and, (B) the waiter's perspective.

The Rise and Fall in Production of Empathy-building Immersive Experiences Over A Decade (2015-2025)

Magdalena Igras-Cybulska[1,*], Aleksandra Witoszek-Kubicka[2], Artur Cybulski[1]

[1]Interactive Crafts sp. z o.o., Lublin, Poland

[2]Krakow University of Economics, Kraków, Poland

*** Correspondence:**
Corresponding Author
magda.igras@gmail.com

Keywords: immersive experiences, immersive journalism, empathy, virtual reality, social impact, audience reception, VR storytelling, 360 movie

Abstract

Between 2015 and 2025, empathy-driven immersive experiences (360° videos and virtual reality) underwent significant fluctuations in production and viewer engagement. This study systematically analyzes 38 publicly available empathy-building immersive projects, addressing topics such as forced migration, public health, social justice, and environmental issues. A clear production peak occurred in 2017, followed by a marked decline that can be attributed primarily to limited technological accessibility, high production and consumer costs, inadequate interactivity, and the emotional demands placed on audiences.

1 Introduction

In the last decade, 360-degree video and virtual reality (VR) have emerged as powerful tools for promoting empathy by immersing users in experiences that convey perspectives often marginalized or misunderstood in traditional media. These empathy-driven experiences range from human rights and humanitarian crises to environmental awareness and neurodiversity advocacy. Despite widespread initial enthusiasm, empirical data suggests variability in sustained audience engagement over time. Examining the reasons behind these fluctuations may guide creators toward more effective and sustainable immersive content strategies, contributing directly to the growth and positive societal impact of the emerging concept of the "Metaverse for Good."

2 Material and methods

The study analyzed 38 publicly available empathy-focused immersive experiences (360° video and VR) released from 2015 to 2023. Selected experiences met criteria of: explicit empathy-building intent, accessibility on mainstream platforms (YouTube, Steam, Oculus/Meta Quest), and public availability. Collected metrics included publication year, addressed problems, duration, interactivity, visual style (film or computer-generated graphics), view counts, and comment numbers. Production trends were evaluated by analyzing annual publication frequency, while viewer engagement was quantified through platform metrics.

3 Results

The analyzed immersive experiences were grouped into five main thematic categories: 1) **Refugees, War, and Forced Migration** (*Clouds Over Sidra [1], The Displaced [2], Limbo: Waiting for Asylum [3], We Wait [4], Sea Prayer [5], I Am Rohingya [6], Under the Net [7], My Mother's Wing [8], We Who Remain [9], Return to Chernobyl [10]*), 2) **Public Health, Epidemics, and Disabilities** (*Waves of Grace [11], Ebola Outbreak [12], On the Brink of Famine [13], The Party – a virtual experience of autism [14], Autism TMI Virtual Reality Experience [15], My First Day at Work [16], A Walk Through Dementia series [17,18,19], Notes on Blindness: Into Darkness [20], Goliath: Playing with Reality [21]*) 3) **Human Rights, Social Justice, and Violence** (*Kiya [22], Across the Line [23], Notes to My Father [24], Traveling While Black [25], Step to the Line [26], I Am A Man [27], Ghost Fleet VR [28], 6×9: A Virtual Experience of Solitary Confinement [29], After Solitary [30]*) 4) **Environmental Issues and Conservation** (*Greenland Melting [31], Songbird: A Virtual Moment of Extinction [32], My Africa [33], The Fight to Save Threatened Species [34], The Mercury Crisis [35], Night of the Storm [36]*), 5) **Societal Challenges and Human Experience** (*Becoming Homeless: A Human Experience [37], First Impressions [38]*).

Of the 38 analyzed immersive experiences, the majority (25) were non-interactive 360° videos, 8 combined video and computer-generated graphics ("mixed", non-interactive), and only 3 were fully interactive virtual experiences. The remaining 2 were virtual but non-interactive (see. Fig. 1).

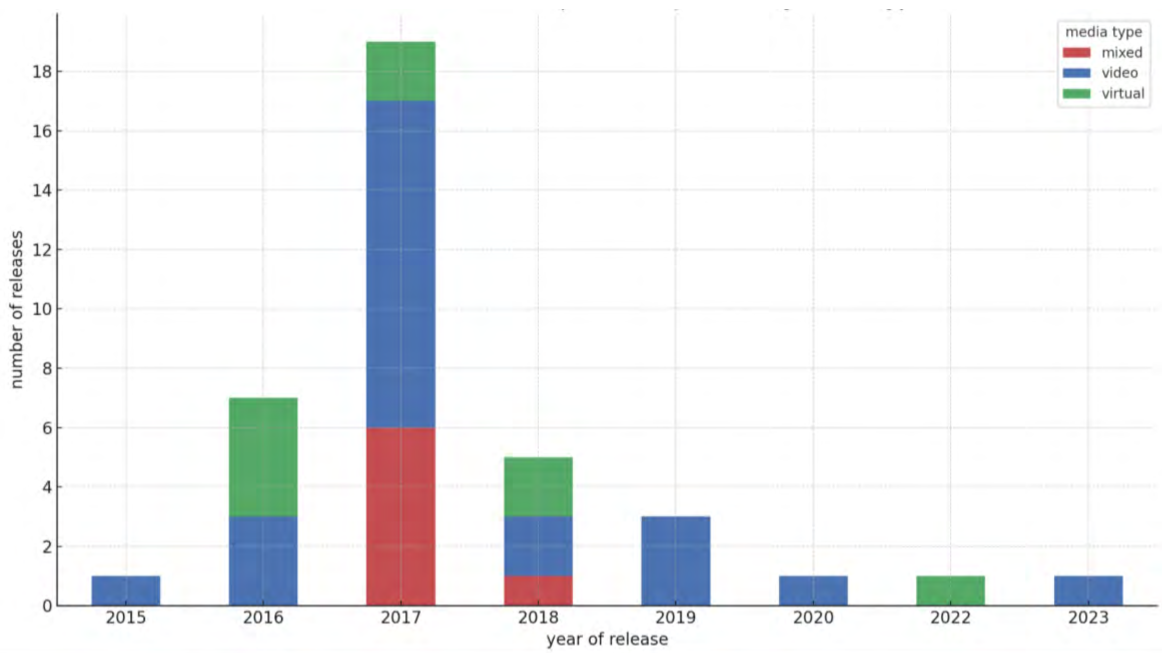

Fig. 1. Number of immersive experiences per year by media type.

Most of the analyzed immersive experiences were published on YouTube, primarily via channels affiliated with major media organizations or companies (e.g., Frontline, The Guardian, Within, Alzheimer's Research UK, Nonny de la Peña’s ‘Emblematic’), while only a few appeared on dedicated VR platforms such as Steam or Oculus (Meta Quest, Rift). The distribution of publication dates reveals a clear peak in 2017, with 19 of the 38 analyzed immersive experiences published in this year. Earlier, in 2016, seven experiences were released, marking the beginning of increased production. Following the peak, production declined significantly (see Fig. 1).

The immersive experiences had an average duration of approximately 8 minutes (mean: 7.9 min), ranging from a minimum of 25 seconds to a maximum of around 20 minutes. The majority of experiences were between approximately 6 and 9 minutes long (interquartile range). The analyzed experiences varied greatly in audience engagement, averaging around 395,000 views and 480 comments per experience, with significant differences between the most popular (3.7M views, 6,300 comments) and the least engaging ones (1,100 views, 0 comments). The median engagement was notably lower, with 86,000 views and 31 comments per experience. A few highly popular experiences significantly skewed the engagement metrics upward.

4 Discussion and Conclusion

Immersive journalism – one of the crucial perks of the "ultimate empathy machine" - after skyrocketing to the surface of the mainstream, had to quickly hibernate itself (observable rapid decrease of virtual reality journalism creations releases frequency since around the year 2017) until enough medians types would meet the necessary conditions for enabling its ecosystem to form its atmosphere in the mainstream for good. The aforementioned median types include: thresholds of hardware availability, sufficient technical competencies among creators and consumers, ease of use and UX efficiency reducing cognitive load in production and consumption, sufficient internet connection bandwidth speeds, efficiency of 360-degree and stereoscopic video streaming technologies, performance and accessibility of end-user devices, as well as sufficiently low production and consumer costs.

Additionally, the high emotional labor required from audiences experiencing intense immersive narratives, insufficient interactivity in most VR content, and limited distribution channels further constrained sustained audience engagement and mainstream adoption.

In 2017, limited global internet speeds (median mobile: 20.28 Mbps, fixed broadband: 40.11 Mbps, [39]) restricted seamless streaming of stereoscopic 360° videos (4K at 30 fps required at least 30 Mbps), accessible to fewer than 38% of mobile and 56% of fixed broadband users. High-quality VR experiences additionally required expensive hardware (e.g., PCVR setups ~1,800 USD), limiting their audience. By 2025, improved median internet speeds (mobile: 91.5 Mbps, fixed: 99.9 Mbps, [40]), widespread availability of affordable standalone VR devices (Quest 3 at ~500USD, Vision Pro at ~3,500 USD), and efficient streaming methods like tiled adaptive streaming - cutting bandwidth demands by 50–70% - have significantly reduced technical entry barriers.

Despite these advancements, rising costs and environmental constraints related to electronic devices production persist. However, transitions toward sustainable manufacturing (Industry 4.0–6.0) and increased use of refurbished or modular hardware indicate a potential path toward more affordable, sustainable, and widespread VR journalism in the future.

5 References

1. (All references accessed June 2, 2025)

[1] https://www.youtube.com/watch?v=mUosdCQsMkM
[2] https://www.youtube.com/watch?v=ecavbpCuvkI
[3] https://www.youtube.com/watch?v=AyWLvrWBKHA
[4] https://www.meta.com/en-gb/experiences/pcvr/we-wait/1248924088498899/
[5] https://www.youtube.com/watch?v=LKBNEEY-c3s
[6] https://www.youtube.com/watch?v=_SL3aab8LAs

[7] https://www.youtube.com/watch?v=ilyQMA0tq-U
[8] https://www.youtube.com/watch?v=aEkGUnhVJGA
[9] https://www.youtube.com/watch?v=d04n6aE8FOk
[10] https://www.youtube.com/watch?v=C2WnT12uMsM
[11] https://www.youtube.com/watch?v=0lwG6MfGvwI
[12] https://www.youtube.com/watch?v=G93XJCVr8vk
[13] https://www.youtube.com/watch?v=cIF5DYNLIPs
[14] https://www.youtube.com/watch?v=OtwOz1GVkDg
[15] https://www.youtube.com/watch?v=DgDR_gYk_a8
[16] https://www.youtube.com/watch?v=hpHLuQqvGNc
[17] https://www.youtube.com/watch?v=R-Rcbj_qR4g
[18] https://www.youtube.com/watch?v=2EO7Dh3fM7Y
[19] https://www.youtube.com/watch?v=TaeNgo8bR2k
[20] https://www.youtube.com/watch?v=tb5DwAZIQZw
[21] https://steamdb.info/app/2138380/
[22] https://www.youtube.com/watch?v=qYsAIukRqog
[23] https://www.youtube.com/watch?v=b4WV2k8owQM
[24] https://www.youtube.com/watch?v=En38_u_ev5k
[25] https://www.youtube.com/watch?v=7UUFn7iyymo
[26] https://www.youtube.com/watch?v=ejVVsM9yG3U
[27] https://www.youtube.com/watch?v=GkvxHnC7Zzo
[28] https://www.youtube.com/watch?v=ZcfYhL84s5s
[29] https://www.youtube.com/watch?v=odcsxUbVyZA
[30] https://www.youtube.com/watch?v=G7_YvGDh9Uc
[31] https://www.youtube.com/watch?v=hUWqQ9F3sJk
[32] https://www.youtube.com/watch?v=Wg27NyaVaXk
[33] https://www.youtube.com/watch?v=119f0l7sqwg
[34] https://www.youtube.com/watch?v=T-aOVE22lEw
[35] https://www.youtube.com/watch?v=m0UQxlOQSbY
[36] https://www.youtube.com/watch?v=fFk5W6pS7IQ
[37] https://store.steampowered.com/app/738100/Becoming_Homeless_A_Human_Experience/
[38] https://www.youtube.com/watch?v=QKhZYZfq8Zk
[39] https://www.ookla.com/articles/global-speed-2017
[40] https://www.speedtest.net/global-index

Keeping collaboration along transitions between places in VR

Majd Bitar *, **Jean-Philippe Rivière, Rebecca Fribourg, Jean-Marie Normand, Yannick Prié**

Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France

Correspondence*:
majd.bitar@univ-nantes.fr

Keywords: Virtual Reality, Navigation, Transition, Collaboration, Teleportation, Metaverse, Social VR, HCI

1 INTRODUCTION

With Virtual Reality (VR) becoming more available for commercial use, social VR platforms such as *VRChat, RecRooms, Meta Horizon Worlds* emerged rapidly, shaping what we know today as the *Metaverse* as defined by (Park and Kim, 2022). These platforms enable users to embody different avatars, chat and communicate while freely navigating a wide-variety of virtual environments (VEs) that allowed diverse types of interaction. However, unlike real experiences where groups can collaborate and do different activities as they move from one place to another while maintaining social contact, virtual environments still miss features to preserve social presence and collaboration along navigation. In our work we aim to study group navigation, including both the notions of "going as a group inside one virtual environment" and "going as a group from one environment to a completely different one". We will therefore design and evaluate new techniques and models of navigation and interaction methods that preserve the sense of group and collaboration.

2 RELATED WORK

2.1 Navigation

LaViola et al. (2017) defined Navigation as the process of going from one place to another. This process can be subdivided into two components: Travel and Wayfinding. Travel is the motor component of navigation, the task of moving from the current location to a new target location or moving in the desired direction. On the other hand, Wayfinding is the cognitive process of determining and following a route between an origin and a destination, involving spatial understanding and decision making based on the surrounding environment.

2.2 Group Navigation

We will first discus real life groups and then explore how they are implemented in VR. A group is defined in its simplest form as "two or more individuals who are connected by and within social relationships" (Forsyth, 2014). One major model to define group behavior is the Tuckman's Model,(Tuckman, 1965; Tuckman and Jensen, 1977) that broke down the group behavior into different stages starting with the group formation, deciding tasks, splitting responsibilities and performing defined action to achieve the given task, and finally the group's disassembly. Meanwhile, (Forsyth, 2014) defines ten properties that characterize a group, but these properties are too complex to be used for implicit declaration of groups, for example using an algorithm to assign groups automatically.

When it comes to implementation in VR, (Weissker et al., 2020) proposed the use of Tuckman's Model, while suggesting that the formation and disassembly of the group is connected to the context of the task itself. However, to reduce complexity, most approaches use explicit methods to define the group, mainly by assuming all members of a given session are one group, or using direct invitations. However, some works suggest more dynamic methods to form temporary groups, such as holding hands during navigation to teleport together (Weissker et al., 2022).
Extending from navigation, group navigation can be defined by a group of individuals each navigating by themselves while choosing to stay together while aiming towards the same destination. But, the group can also be considered as one entity that encapsulates all the members, thus to move these members, the group entity is moved as a whole. The latter approach to group definition and navigation is the one used when designing *Group Locomotion Techniques* as found in the literature. These techniques are defined as locomotion methods that expand upon individual locomotion methods by updating the viewpoint of several users and agents at the same time (Brument et al., 2024).

2.2.1 Group Locomotion Techniques

When observing group locomotion techniques in the literature, we can see that artificial locomotion methods, more specifically teleportation-based methods dominate such group locomotion techniques. These methods utilize artificial interactions devoid of any user body movements, such as pressing a button that leads to non-continuous virtual movements which instantaneously update the users' viewpoints Boletsis and Chasanidou (2022). This can be found in the works of (Weissker et al., 2020; Weissker and Froehlich, 2021; Zhang et al., 2022) in which they proposed the use of short-range jumping techniques visualized by ray casting. Similarly, in the works of (Chheang et al., 2022a,b), World-in-Miniature (WiM) techniques were suggested, as the group can see a smaller visualization of the virtual environment and decide where they want to teleport using the WiM. In (Brument et al., 2024), the authors compared a variation of the previous two techniques "jumping raycast" and "WiM teleporation" and proposed the use of a combination of both, allowing participants to utilize any of the two freely. An interesting finding was the use of jumping raycast for covering shorter distances while using WiM for larger distances.
However, although these proposals utilize similar locomotion methods, they differ when it comes to splitting responsibilities and agency between group members. For example some proposals see the group as a vehicle, where one navigator is defined and decides where the group will go, while the rest of the members are passengers that "just" follow Weissker et al. (2020); Weissker and Froehlich (2021); Chheang et al. (2022a,b). In Zhang et al. (2022), the authors compared the navigator-passenger approach with their proposed method that enables collaborators to directly communicate their desired destination resulting in a shared outcome. On the other hand, (Brument et al., 2024) designed group teleportation that relies on the vote of the group members in order to execute the teleportation action. In Zoeppig et al. (2025), authors discussed several confirmation-based group navigation techniques derived from the navigator-passenger approach.

2.2.2 Group Transition

Navigating between two different VEs or greatly separated spaces inside the same VE, is defined as transition in the literature. Various works could be found focusing on individual transitioning methods between environments (Husung and Langbehn, 2019; Lee et al., 2022; Feld et al., 2024), where they evaluate the sense of presence during the transition, the time-efficiency and physical and cognitive load induced by the transition (Oberdörfer et al., 2018), comparing several transition methods (Men et al., 2017; Feld et al., 2023) and introducing novel ones(Sisto et al., 2017; Lee et al., 2023). However, to the best of our knowledge little to no work addresses group transitioning scenarios, meaning the transition of a whole group of users (as a single entity) from one VE to the other. Recent works such as (Wang et al., 2024) merged two different rooms or scenes in order to connect two collaborators. They Proposed a method to connect two separated individual collaborators. However, moving a group of collaborators together from one VE to the other is still unexplored.

2.3 Collaboration during Navigation

Emerging studies can be found on maintaining collaboration while exploring virtual environments. Xia et al. (2018) suggested a set of tools and methods that facilitate collaboration in virtual environments. Moreover, they suggested a method to request assistance from collaborators, by introducing a technique that allows to request assistance from another collaborator, allowing the other to change viewpoints to be closer to one's own viewpoint, and then easily return to their original work and viewpoint once the assistance was done.

In order to improve the collaboration between two separated collaborators in a large VE, Bimberg et al. (2024) defined a method that enables a pair of collaborators to go from loose collaboration to tight collaboration. Their approach starts with a distant pair in a large VE where no communication is present (loose collaboration), and then describe the stages that will enable the pair to form communication while being distant, and with later stages the one collaborator can join the other by teleportation, helping them and then returning to their original location or not depending on their task.

2.4 Embodiment and Social Presence during Navigation

Avatars (i.e., the digital representations of users in the VE) are important in the Metaverse as one's body is one's spatial reference and is essential to social presence: "people can see me and I can see them" (Heidicker et al., 2017; Aseeri and Interrante, 2021). Both avatar's appearance Casanueva and Blake (2001); Heidicker et al. (2017) and animation (realistic gaze, accurate arm movements) have an effect on social experience during cooperative tasks. However, how they affect social presence during co-navigation or co-teleportation remains unexplored. There are a few results we could use for co-navigation: locomotion methods have little influence on avatar perception Dewez et al. (2020), and avatar positions can be tweaked to smooth the perception of the movement of others Freiwald et al. (2022).

3 FOCUS POINTS OF THE THESIS

3.1 Navigation with the Group

As discussed, many approaches to group navigation choose to follow the navigator-passenger assumption, but we believe that approaches are to be explored and discussed. Moreover, the navigator-passenger approach raises questions about **agency while navigating**. Indeed, although it has been shown that some users are willing to give up some agency during navigation (LaViola et al., 2017), further studies are required to better understand the limits on how constrained or free the collaborators should be.

3.2 Transitions

From what we presented above, to the best of our knowledge, no work has addressed transitions as a group from one VE to the other. Based on that observation, our objectives are: (i) to **design group transition techniques**, (ii) to identify how they can **affect both presence and collaboration**, and (iii) to propose solutions to **improve them by mitigating these effects**.

3.3 Presence and Navigation

Studies on the effect of locomotion on embodiment has shown that relying on teleportation-based methods locomotion can have a negative impact on co-presence Freiwald et al. (2021). Our literature review showed that most group locomotion methods rely on teleportation-based methods of navigation. This motivates our exploration of group locomotion methods that **better preserve co-presence**. In addition, the **effects of group navigation methods on social presence** are still to be studied, opening the possibilities for improving the sense of group during navigation.

3.4 Collaboration and Navigation

Works that look at collaboration during navigation in virtual scenes are emerging, however a lot of questions remain on how to **evaluate collaboration during navigation**, as well as on the **metrics to be monitored** to properly evaluate collaboration within navigating groups. As a consequence, how does **navigation impact collaboration** and how can we **design navigation methods that maintain collaboration**.

4 APPROACH AND METHODS

The initial steps involve **deepening out literature review** to gain a better understanding of navigation in general, and more specifically of group navigation methods, whether it involves transitions between VEs or locomotion within the same environment. The thesis will also explore collaboration during navigation, more specifically on how to model and how to evaluate collaboration between navigating members of a group. In addition, we will examine embodiment and social presence, to find correlations between them and navigation related aspects.

Following that, we will **investigate the current state of commercial solutions** by conducting ethnographic studies on platforms of the Metaverse. These studies aim to better understand the navigation methods currently used by the industry and to identify the gap between practice and research. The studies will also give us insight into what Metaverse users actually use and what they actually need, opening up opportunities for designing new navigation methods.

With design ideas gathered from both the literature review and ethnographic studies, we plan to **conduct experiments** aimed at identifying which navigation features affect collaboration and social presence, drawing connections between them. These insights will help us **formulate design principles and tools** that support collaboration during navigation while minimizing disruptions to social presence and embodiment. Finally, we will **develop appropriate evaluation methods** to assess the techniques we design, and propose practical applications for them.

ACKNOWLEDGMENTS

This work was partly supported by a French government grant managed by the Agence Nationale de la Recherche as part of the France 2030 program, reference ANR-22-EXEN-0002 (PEPR eNSEMBLE / CATS).

REFERENCES

Aseeri, S. and Interrante, V. (2021). The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 2608–2617. doi:10.1109/TVCG.2021.3067783

Bimberg, P., Zielasko, D., Weyers, B., Froehlich, B., and Weissker, T. (2024). Come Look at This: Supporting Fluent Transitions between Tightly and Loosely Coupled Collaboration in Social Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* , 1–17doi:10.1109/TVCG.2024.3418009

Boletsis, C. and Chasanidou, D. (2022). A Typology of Virtual Reality Locomotion Techniques. *Multimodal Technologies and Interaction* 6, 72. doi:10.3390/mti6090072

Brument, H., Lautenbach, T., and Podkosova, I. (2024). Spatial Awareness and User Preferences During Group Locomotion in Virtual Reality: A Study with Four-User Groups. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 403–409. doi:10.1109/VRW62533.2024.00078

Casanueva, J. and Blake, E. (2001). The Effects of Avatars on Co-presence in a Collaborative Virtual Environment

Chheang, V., Heinrich, F., Joeres, F., Saalfeld, P., Barmaki, R., Preim, B., et al. (2022a). WiM-Based Group Navigation for Collaborative Virtual Reality. In *2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 82–92. doi:10.1109/AIVR56993.2022.00018

Chheang, V., Heinrich, F., Joeres, F., Saalfeld, P., Preim, B., and Hansen, C. (2022b). Group WiM: A Group Navigation Technique for Collaborative Virtual Reality Environments. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 556–557. doi:10.1109/VRW55335.2022.00129

Dewez, D., Hoyet, L., Lécuyer, A., and Argelaguet, F. (2020). Studying the Inter-Relation Between Locomotion Techniques and Embodiment in Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 452–461. doi:10.1109/ISMAR50242.2020.00070

Feld, N., Bimberg, P., Weyers, B., and Zielasko, D. (2023). Keep it simple? Evaluation of Transitions in Virtual Reality. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), CHI EA '23, 1–7. doi:10.1145/3544549.3585811

Feld, N., Bimberg, P., Weyers, B., and Zielasko, D. (2024). Simple and Efficient? Evaluation of Transitions for Task-Driven Cross-Reality Experiences. *IEEE Transactions on Visualization and Computer Graphics* 30, 7601–7618. doi:10.1109/TVCG.2024.3356949

Forsyth, D. (2014). *Group Dynamics* (Wadsworth Cengage Learning)

Freiwald, J. P., Schenke, J., Lehmann-Willenbrock, N., and Steinicke, F. (2021). Effects of Avatar Appearance and Locomotion on Co-Presence in Virtual Reality Collaborations. In *Proceedings of Mensch Und Computer 2021* (New York, NY, USA: Association for Computing Machinery), MuC '21, 393–401. doi:10.1145/3473856.3473870

Freiwald, J. P., Schmidt, S., Riecke, B. E., and Steinicke, F. (2022). Moving Avatars and Agents in Social Extended Reality Environments. *ACM Transactions on Graphics* 41, 1–14. doi:10.1145/3550454.3555522

Heidicker, P., Langbehn, E., and Steinicke, F. (2017). Influence of avatar appearance on presence in social VR. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 233–234. doi:10.1109/3DUI.2017.7893357

Husung, M. and Langbehn, E. (2019). Of Portals and Orbs: An Evaluation of Scene Transition Techniques for Virtual Reality. In *Proceedings of Mensch Und Computer 2019* (New York, NY, USA: Association for Computing Machinery), MuC '19, 245–254. doi:10.1145/3340764.3340779

LaViola, J. J., Kruijff, E., McMahan, R. P., Bowman, D., and Poupyrev, I. P. (2017). *3D User Interfaces: Theory and Practice (Usability)* (Pearson Education)

Lee, J.-I., Asente, P., and Stuerzlinger, W. (2022). A Comparison of Zoom-In Transition Methods for Multiscale VR. In *ACM SIGGRAPH 2022 Posters* (New York, NY, USA: Association for Computing Machinery), SIGGRAPH '22, 1–2. doi:10.1145/3532719.3543237

Lee, J.-I., Asente, P., and Stuerzlinger, W. (2023). Designing Viewpoint Transition Techniques in Multiscale Virtual Environments. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 680–690. doi:10.1109/VR55154.2023.00083

Men, L., Bryan-Kinns, N., Hassard, A. S., and Ma, Z. (2017). The impact of transitions on user experience in virtual reality. In *2017 IEEE Virtual Reality (VR)*. 285–286. doi:10.1109/VR.2017.7892288

Oberdörfer, S., Fischbach, M., and Latoschik, M. E. (2018). Effects of VE Transition Techniques on Presence, Illusion of Virtual Body Ownership, Efficiency, and Naturalness. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction* (New York, NY, USA: Association for Computing Machinery), SUI '18, 89–99. doi:10.1145/3267782.3267787

Park, S.-M. and Kim, Y.-G. (2022). A Metaverse: Taxonomy, Components, Applications, and Open Challenges. *IEEE Access* 10, 4209–4251. doi:10.1109/ACCESS.2021.3140175

Sisto, M., Wenk, N., Ouerhani, N., and Gobron, S. (2017). A Study of Transitional Virtual Environments. In *Augmented Reality, Virtual Reality, and Computer Graphics*, eds. L. T. De Paolis, P. Bourdot, and A. Mongelli (Cham: Springer International Publishing), 35–49. doi:10.1007/978-3-319-60922-5_3

Tuckman, B. W. (1965). Developmental sequence in small groups. *Psychological Bulletin* 63, 384–399. doi:10.1037/h0022100

Tuckman, B. W. and Jensen, M. A. C. (1977). Stages of Small-Group Development Revisited. *Group & Organization Studies* 2, 419–427. doi:10.1177/105960117700200404

Wang, M., Li, Y.-J., Shi, J., and Steinicke, F. (2024). SceneFusion: Room-Scale Environmental Fusion for Efficient Traveling Between Separate Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 30, 4615–4630. doi:10.1109/TVCG.2023.3271709

Weissker, T., Bimberg, P., and Froehlich, B. (2020). Getting There Together: Group Navigation in Distributed Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 26, 1860–1870. doi:10.1109/TVCG.2020.2973474

Weissker, T., Bimberg, P., Kodanda, A., and Froehlich, B. (2022). Holding Hands for Short-Term Group Navigation in Social Virtual Reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 728–729. doi:10.1109/VRW55335.2022.00215

Weissker, T. and Froehlich, B. (2021). Group Navigation for Guided Tours in Distributed Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 2524–2534. doi:10.1109/TVCG.2021.3067756

Xia, H., Herscher, S., Perlin, K., and Wigdor, D. (2018). Spacetime: Enabling Fluid Individual and Collaborative Editing in Virtual Reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA: Association for Computing Machinery), UIST '18, 853–866. doi:10.1145/3242587.3242597

Zhang, Y., Nguyen, H., Ladevèze, N., Fleury, C., and Bourdot, P. (2022). Virtual Workspace Positioning Techniques during Teleportation for Co-located Collaboration in Virtual Reality using HMDs. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 674–682. doi:10.1109/VR51125.2022.00088

Zoeppig, T. J., Benjamin Lammert, A., Schott, E., Muehlhaus, S., and Froehlich, B. (2025). Follow Me: Confirmation-based Group Navigation in Collocated Virtual Reality. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 548–558. doi:10.1109/VR59515.2025.00078

Poster Abstracts 1B : Room “Vicente Gimeno Sendra”

Discovering the Metaverse: New Copyright Challenges for User-Generated Content

Oscar Verdoire

Maynooth University, School of Law and Criminology, ADVANCE CRT, Maynooth,

Correspondence:
Oscar Verdoire
oscar.verdoire.2019@mumail.ie

Keywords: Metaverse, Intellectual Property, Copyright, User-Generated Content, Authors, Users

Abstract

Much of the online content we experience today is created for us to interact with as users, where we can interact with and share content. However, in recent years, there has been a surge in the number of digital platforms that provide users more opportunities to create content for themselves and to share that content with others. The Metaverse, which has been referred to as a 'successor to the mobile Internet,' can become one such platform, and could fundamentally disrupt how we interact online and create digital content by combining both virtual and augmented reality. The traditional role of 'users' within the creative space may forever be changed, potentially presenting them with greater freedoms and allowing them to become the 'creators' of the new virtual landscape. There have already been several concerns regarding the legality of such User-Generated Content on past platforms, and the same questions about ownership of such works and the potential intellectual property rights that may apply can now be asked about content that will be created within the Metaverse. This paper specifically calls into question the extent to which our current approach to copyright law may be suited to User-Generated Content on the Metaverse. Copyright grants creators of original works certain economic and moral rights concerning the commercialisation and distribution of their works, allowing them to prevent actions like the unauthorised copying or modification of their works. Under current laws and on existing platforms, there is a clearer divide between users and creators, though those roles have increasingly begun to overlap in the digital space. The Metaverse will further blur the line between users and creators and may cause further issues regarding the possibility of User-Generated Content infringing upon copyright. This additionally raises challenges concerning how each party's rights should be balanced, as the rights of the public must always be balanced with the rights of the creator. My research, funded by Science Foundation Ireland under the ADVANCE CRT programme, tackles the challenges presented by User-Generated Content on the Metaverse, examining current content mediums and determining how the systems used to regulate User-Generated Content on them may be applied on the future Metaverse. This involves analysing what the Metaverse will be and how content creation would function; how users will be able to share, create, and contribute content; how this differs from previous mediums of content; and how previous mediums of content have tackled User-Generated Content challenges with copyright law. This will ultimately allow an analysis of how current systems and frameworks for copyright enforcement function, and whether they can be enforced and applied to user creations on the Metaverse. So far, my research has focused on defining what exactly the

Metaverse is, and I have begun work on three case studies on existing content mediums to compare the future Metaverse to. My main goal has been the identification of potential copyright challenges centered around User-Generated Content faced by content creators within these content mediums, focusing not only on the practical application of the law within different jurisdictions but also on how legislation might apply in theory. This is to assess the viability of current law and assess the need for reform or development for User-Generated Content more generally, but more specifically in the Metaverse. Doing so during the Metaverse's infancy will allow the law to develop in advance of legal challenges that may arise and ensure the interests of all stakeholders - creators and users alike - will be protected. The Metaverse is a new and potentially disruptive technology, and the technology-neutral approach of much current copyright law may no longer suffice. This research should aid future policy-making decisions and allow analysis of how User-Generated Content should be regulated moving forward.

Virtual Embodiment to Transform Attitudes and Behaviors Toward Gender-Based Harassment in the Metaverse

Isaac Calvis[1,2], Justyna Swidrak[1,5], Maria V. Sanchez-Vives[1,2,3], Mel Slater[2,4], K S Estevães[6,7,8], A A Melese[6,7,8]

[1] Institute of Biomedical Research August Pi i Sunyer, Barcelona, Spain

[2] Event Lab, Department of Clinical Psychology and Psychobiology, University of Barcelona, Barcelona, Spain

[3] ICREA, Barcelona, Spain

[4] Institute of Neurosciences of the University of Barcelona, Barcelona, Spain

[5] Institute of Psychology Polish Academy of Sciences, Warsaw, Poland

[6] CYBER: Cyberspace, behavior and E-therapy program, Universidade Lusófona, Lisboa, Portugal

[7] Facultat de Psicologia, Universitat de Barcelona, Barcelona, Spain

[8] UFR des Sciences Fondamentales et Biomédicales, Université Paris Cité, Paris, France

*** Correspondence:** isaaccalvis@gmail.com

Keywords: metaverse, virtual reality, gender-based harassment, XR, virtual embodiment, empathy, training

Abstract

Gender-based harassment remains a serious issue in contemporary society and is now being mirrored in virtual immersive spaces such as the Metaverse. Studies indicate that XR contexts introduce new forms of harassment due to their unique capabilities. To address this issue, we must develop tools and methods to prevent, mitigate, and raise awareness of harassment, as well as explore potential new risks arising from XR technologies, such as social interactions through the Metaverse, increased presence and haptic feedback.

This study examines the use of the Golden Rule Embodiment Paradigm (GREP) and look-alike avatars, which has proven effective in other contexts, such as in racial-bias (Peck et al., 2013; Maister et al., 2015; Banakou et al., 2020) and rehabilitation of gender violent behaviour (Seinfeld et al., 2018; Barnes et al., 2022; Johnston et al., 2022; Seinfeld et al., 2023). GREP is based on the "treat others the way you want them to treat you", which is achieved by gaining new insights through virtual embodiment. We aim to investigate its potential in addressing gender-based harassment during a job interview in the Metaverse. Additionally, we will compare the effects of embodying a female avatar in a neutral context versus a violent one, and examine how seeing oneself as an accomplice influences the GREP experience.

The first session of the experiment is divided into two phases. In the first, participants watch a male colleague harassing a woman during a job interview. In the second, they embody the woman's

perspective and experience the harassment directly, seeing their own avatar performing the actions they did in the previous scene. The second session takes place one week later, participants enter a virtual café where the woman, now hired, is harassed again. At key moments, she will appeal directly to the participant, and we expect them to intervene and stop their colleague from harassing her.

We expect that GREP will be effective in raising awareness of gender-based violence, particularly among men. Furthermore, using look-alike avatars may enhance empathy and strengthen the impact of GREP. We assume that the experimental group will be more active in the test cafeteria scene than the control group, which will not have experienced the harassment firsthand.

In addition to our core design, we have given careful thought to the ethical implications of exposing participants to repeated harassment scenarios and to negative embodiment experiences. All procedures are approved by an institutional review board and include informed consent. Participants will be free to withdraw at any time, and we will provide access to psychological support resources. Baseline data on each participant's prior exposure to gender-based harassment should be obtained, to control for previous trauma when analyzing intervention outcomes.

Looking forward, the study anticipates that GREP, particularly when combined with look-alike avatars, will be effective in raising awareness of gender-based violence, especially among men, and will prompt increased intervention from participants in real-time scenarios. This research holds promise for contributing significantly to the development of robust, empathy-driven interventions to combat harassment in the emergent immersive digital world and to develop better attitudes and behaviours that are successfully translated into the real world scenarios.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

Funded by META TOO (A transfer of knowledge and technology for investigating gender- based inappropriate social interactions in the Metaverse), European Union's Horizon Europe research and innovation programme under grant agreement No 101160266. More information in https://meta-too.eu/

References

Banakou, D., Beacco, A., Neyret, S., Blasco-Oliver, M., Seinfeld, S., and Slater, M. (2020). Virtual body ownership and its consequences for implicit racial bias are dependent on social context. *R Soc Open Sci* 7(12)**,** 201848. doi: 10.1098/rsos.201848.

Barnes, N., Sanchez-Vives, M.V., and Johnston, T. (2022). On the Practical Use of Immersive Virtual Reality for Rehabilitation of Intimate Partner Violence Perpetrators in Prison. *Front Psychol* 13**,** 787483. doi: 10.3389/fpsyg.2022.787483.

Johnston, T., Seinfeld, S., Gonzalez-Liencres, C., Barnes, N., Slater, M., and Sanchez-Vives, M.V. (2022). Virtual reality for the rehabilitation and prevention of intimate partner violence - From brain to behavior: A narrative review. *Front Psychol* 13**,** 788608. doi: 10.3389/fpsyg.2022.788608.

Maister, L., Slater, M., Sanchez-Vives, M.V., and Tsakiris, M. (2015). Changing bodies changes minds: owning another body affects social cognition. *Trends Cogn Sci* 19(1)**,** 6-12. doi: 10.1016/j.tics.2014.11.001.

Peck, T.C., Seinfeld, S., Aglioti, S.M., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious Cogn* 22(3)**,** 779-787. doi: 10.1016/j.concog.2013.04.016.

Seinfeld, S., Arroyo-Palacios, J., Iruretagoyena, G., Hortensius, R., Zapata, L.E., Borland, D., et al. (2018). Offenders become the victim in virtual reality: impact of changing perspective in domestic violence. *Sci Rep* 8(1)**,** 2692. doi: 10.1038/s41598-018-19987-7.

Seinfeld, S., Hortensius, R., Arroyo-Palacios, J., Iruretagoyena, G., Zapata, L.E., de Gelder, B., et al. (2023). Domestic Violence From a Child Perspective: Impact of an Immersive Virtual Reality Experience on Men With a History of Intimate Partner Violent Behavior. *J Interpers Violence* 38(3-4)**,** 2654-2682. doi: 10.1177/08862605221106130.

Your Personalized Spielberg Will See You Now: Authorship and Identity in AI-Generated Cinema

Ernesto Edwards[1]

[1]Ph.D. Candidate in Comparative Private Law, DIRPOLIS, Law Department, Scuola Superiore Sant'Anna, Pisa, Italy

*** Correspondence:**
Corresponding Author
Ernesto.edwards@santannapisa.it

Keywords: Generative AI, Authorship, Copyright Law, Personalized Content, Generative Media

Abstract

In a 2023 interview, director Joe Russo predicted that, within two years, Artificial Intelligence (AI) would be capable of making full-length films tailored to each viewer's preferences—an ultra-personalized cinematic experience rendered in real time. This bold vision of AI-generated cinema, once confined to science fiction, now demands legal and ethical scrutiny.

This flash talk explores the consequences of such a development through the lens of Copyright Law, drawing on ongoing research into AI authorship and creative agency. If a film is written, directed, and rendered entirely by an AI—drawing from databases of human-generated content—who is the author of that film? Can there be one? And if every user receives a unique "version" of the film, is each one a distinct copyright work?

Building on comparative analysis between UK, US, and EU copyright systems, this talk interrogates whether current legal frameworks are equipped to handle not just the automation of filmmaking, but its fragmentation into individualized experiences. It argues that such works challenge the very idea of a stable, fixed "work"—a cornerstone of Copyright Law—and instead introduce a new paradigm: generative fluidity, where authorship is distributed, ephemeral, and potentially dehumanized.

Using both doctrinal tools and speculative examples—including narrative frameworks like Black Mirror's *Joan is Awful*—this talk situates the Joe Russo prediction not as science fiction, but as imminent legal reality. It attempts to propose a rights-based framework for approaching AI-generated film that balances innovation with clarity, protecting both creators and users from the erosion of creative accountability.

Ultimately, this intervention invites reflection on a future where everyone gets the movie they want—but perhaps no one knows who made it.

1 Conflict of Interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

2 Author Contributions

[Please included appropriate standard statement]

3 Funding

This work was developed during a research stay at the University of Alicante, funded by a Ph.D. scholarship from Scuola Superiore Sant'Anna.

4 Acknowledgments

The author wishes to thank Prof. Aurelio López-Tarruella Martínez for his valuable guidance in the development of this presentation.

AI-enabled crimes and AI-empowered investigations. The role of artificial intelligence in metaverse criminal ecosystems.

Gian Marco Bovenzi[1]

[1] Centre for Higher Defence Studies, Rome, Italy

*** Correspondence:**
Gian Marco Bovenzi
gianmarco.bovenzi@unito.it

Keywords: AI-enabled crimes, artificial intelligence, metaverse, metaverse crimes

Abstract

This paper seeks to explore the role of artificial intelligence within the metaverse criminal ecosystem, namely, its dual role as a technology capable of amplifying criminal activities and as a tool providing law enforcement with enhanced opportunities. Given the few examples of virtual crimes became as of today mediatic, the current insidious trend among policymakers is to underestimate the impact of the metaverse in society; however, metaverse crimes exist, and differ from traditional cybercrimes under several dimensions: criminogenic (typology of virtual attack), victimogenic (the role of immersivity and superrealism enhancing the harm suffered by the victim), etiological (the Proteus-effect in the relationship between humans and avatars); and regulatory (lack of specific regulations addressing metaverse crimes). Moreover, within the dichotomic classification of cybercrimes (computer as *tool* vs. computer used *target* of a crime) the plethora of offences committed in the metaverse acquires a new connotation, including crimes causing individual harm, crimes against computer data or systems, threats to national security, and financial crimes: in all these cases, the features of the metaverse are capable of shifting behind-the-screen conducts' level of dangerousness to unprecedented spikes, up to creating a novel 'metaverse criminal ecosystem'. Along with the substantial aspects of criminal law, digital forensics and investigations struggle as well in adapting existing techniques to virtual environments: the hardware-software mix, decrypting users' pseudonymity, blockchain forensics in decentralized platforms: these aspects are capable of creating a multifactorial combination which adds complexity to investigations. This implies the need of revisiting the forensic chain of custody – collection-recovering-analysis-reporting-reconstruction – whose application might require possible adjustments according to the peculiar structure of the metaverse. Furthermore, the advent of artificial intelligence is capable of bringing these considerations to a 'next level': on the one hand, there is evidence of a raising number of AI-enabled crimes, where the exploitation of artificial intelligence tools facilitates the commission of criminal activities such as financial crimes, sexual abuse materials, and other computer crimes – this, in new ways as automated AI, deepfakes and synthetic media, LLMs abuse, and enhanced cyberattacks; on the other hand, AI-driven systems represent valuable instruments

for crime prevention (e.g., patterns identification, behavioural prediction, predictive analytics, platform monitoring) and investigation/enforcement (AI-agents, perpetrators' traceability, victims' identification). In conclusion, the paper seeks to explore the yet-unsolved challenges of the metaverse criminal ecosystem, with a deep dive in the role of artificial intelligence as a threat and as a solution. For this reason, further research focusing on the creation of a taxonomy of AI-enabled crimes and AI-empowered investigations appears pivotal.

frontiers

Use of immersive data by medical practitioners for mental health prevention

Noah Richert [1,3], Antoine Widmer [1] and Yannick Prié [2]

[1] *HES-SO Valais-Wallis University of Applied Sciences and Arts Western Switzerland, 3960 Sierre, Switzerland*
[2] *Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France*
[3] *Lemanic Neuroscience Doctoral School (LNDS), University of Lausanne, University of Geneva, and EPFL, Switzerland*

Correspondence*:
Corresponding Author
noah.richert@hevs.ch

ABSTRACT

With the growing demand for accessible and scalable mental health care, immersive technologies offer different solutions for prevention and early intervention. This paper presents the preliminary work of a Ph.D. project within the framework of the EU funded THCS project titled Immersive Prevention Centers for Mental Health. This research explores the integration of eXtended Reality (XR) platforms within Immersive Prevention Centers (IPCs) to enhance decision-making with data support in mental health management. It presents a framework that will collect, interpret and visualize behavioral and biometric data to support clinicians in real-time and retrospective reviews of patient activity. By incorporating Machine Learning (ML) and clinically approved scenario design, this project aims to improve early identification, personalize interventions and monitor treatment outcomes in immersive environments. The interdisciplinary approach brings XR, Artificial Intelligence (AI) and clinical expertise together to build a foundation for scalable and interactive mental health support systems.

Keywords: immersive technology, mental health prevention, extended reality (XR), biometric data, artificial intelligence, co-design, virtual reality, digital health

1 INTRODUCTION

Mental health is a state of mental well-being that enables people to cope with the stresses of life (WHO, 2022). However, mental health disorders affect millions worldwide and access to care remains a challenge due to geographical, financial, and social barriers (Samira Abdul et al., 2024). According to the World Health Organization "Mental Disorders" fact sheet (WHO, 2022), these disorders are characterized by a disturbance in an individual's cognition, emotion regulation, or behavior, often leading to significant distress or impairment in functioning. With rising global awareness of the mental health crisis, there is an urgent need for innovative approaches that enhance prevention, early detection, and patient engagement in mental health interventions.

Among emerging solutions, digital health innovations, particularly immersive technologies, have shown promise in transforming how mental health services are delivered. Among these, immersive technologies such as Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) collectively referred to as eXtended Reality(XR) (Rauschnabel et al., 2022) have opened up new avenues for experiential learning, patient monitoring, and remote intervention (Freeman et al., 2017).

Our literature review synthesizes insights from three main research areas: the use of XR in healthcare settings (Worlikar et al., 2023), the integration of mental health management strategies in healthcare organizations (Samira Abdul et al., 2024), and the application of Machine Learning (ML) and Artificial Intelligence (AI) to the interpretation of biometric data (Greener et al., 2022).

Within XR in health, one of the dimensions most actively explored is its integration into telehealth services (Worlikar et al., 2023). Other Major areas of application include telerehabilitation (Caminero et al., 2025; Isernia et al., 2020), telepsychiatry(Dan Shao, 2020), telediagnosis (Filippeschi et al., 2019), and teleconsultation (Hu et al., 2019) using VR, telementoring, telesurgery (Guo et al., 2016; Kutzin et al., 2017) and telemonitoring (Thøgersen et al., 2020) using MR and AR.

Simultaneously, healthcare organizations are beginning to adopt structured mental health management strategies that prioritize early intervention, reduce stigma and encourage scalable, tech-driven approaches to care (Samira Abdul et al., 2024). The main classes of strategies by healthcare organizations are preventive (Colizzi et al., 2020), diagnostic, therapeutic, and supportive measures(Schlak et al., 2022).

Furthermore, advances in artificial intelligence (AI) and machine learning (ML) provide new opportunities to interpret biometric and behavioral data collected during digital interventions. These technologies can support real-time assessment of emotional and physiological states, deliver adaptive and personalized interventions, and enable predictive modeling for early crisis detection (Greener et al., 2022). When integrated with XR platforms, ML algorithms can enhance interactivity by adjusting therapeutic content based on user responses, further improving the precision and effectiveness of mental health interventions (Altozano et al., 2025; Lee, 2024).

The research presented in this paper is part of a broader project investigating how Immersive Prevention Centers for Mental Health (IPC4MH) within the Metaverse can address the following tasks. IPCs will be providing structured, interactive, and evidence-based prevention programs for mental healthcare. The IPC project focuses on prevention of mental health in two layers. The first is to act on early identification and intervention of mental distress to avoid developing into diagnosable conditions. The second is to avoid the worsening of a diagnosed patient's condition with better supervision of the therapy.

This work aims to help medical professionals deliver data-driven mental health interventions in immersive environments, exploring how to best capture, interpret, and apply XR-derived data to support clinicians with real-time insights and structured review data for ongoing patient monitoring and personalized care planning.

2 PROJECT OBJECTIVES

Building on the previously outlined context, this project defines a practical and technical roadmap for developing the IPC framework. The overarching goal is to support clinicians in the early identification, intervention, and management of patients' mental health conditions within Immersive Prevention Centers (IPCs). To achieve this, the project aims to create a comprehensive framework for the collection, interpretation, and restitution of data generated from patient activities within XR environments. This framework will ultimately enable the development of a clinician-facing dashboard, providing structured and actionable insights into patients' engagement and progress.

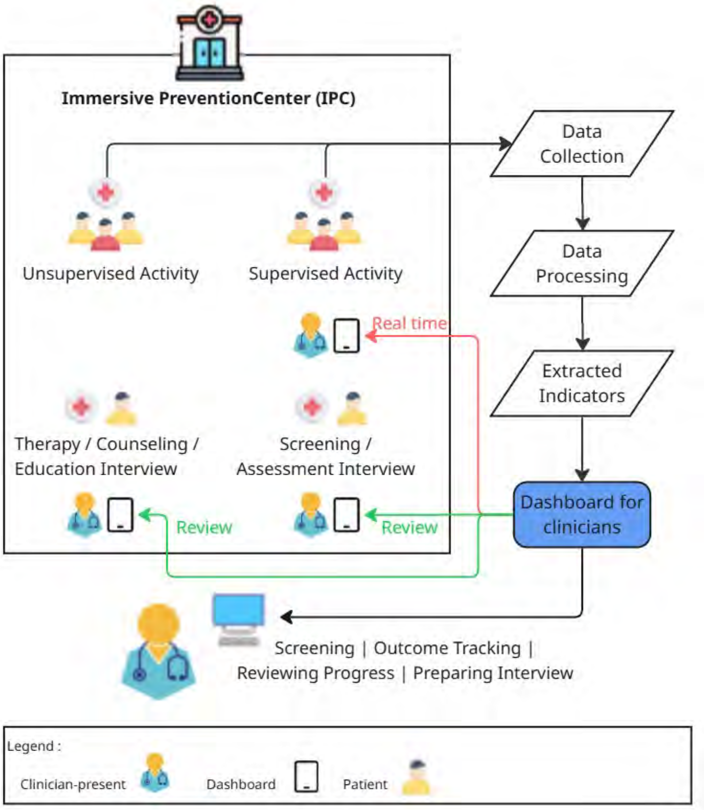

Figure 1. IPC data flow structure

The project addresses four main usage scenarios, illustrated in Fig. 1:

- **Clinician-present activity:** When the clinician is actively supervising a session and assigns an activity to the patient, real-time data is made available to assist the clinician in monitoring and immediate assessment.
- **Clinician-absent activity:** When the clinician assigns an activity (either individual or group-based) but is not present during its execution, patient data is collected and securely stored for later review.
- **Individual review sessions:** During subsequent consultations, clinicians can access and review the data collected over time to evaluate the patient's progress and to adjust therapeutic interventions accordingly.
- **Additional annotations:** Clinicians may also add notes or complete questionnaires through the dashboard interface. While this feature enhances patient monitoring, it involves handling highly sensitive personal information and thus requires strict privacy safeguards.

To translate this vision into practice, the project is structured around three key objectives that drive both clinical and technical development:

Objective 1: extract behavioral and biometric data from IPC

Data will be collected from immersive activities in IPCs using various modules. Patients will use Head-Mounted Displays (HMD) with possible add-ons for eye tracking or EEG. They can also use tablets or web platforms. Data collection must operate across multiple scenarios, in group or solo, and diverse tasks, mental or physical, to ensure robustness. The goal is to capture detailed behavioral and biometric indicators including posture (Caminero et al., 2025), movement dynamics (Duvivier et al., 2021), micro-gestures (Steptoe and Steed, 2012), attention shifts and physiological responses. Automation of collection and pre-processing pipelines need to be emphasized to support future clinical scalability and integration into time-constrained environments. While collecting this data is important, ensuring its clinical relevance is equally important.

Objective 2: determine clinically relevant indicators for mental health professionals

Focus groups and structured interviews are used to determine what types of behavioral or biometric data can serve mental health professional diagnostic and therapeutic goals such as in Felsberg et al. (2025) or Dan Shao et al. (2020) (Felsberg et al., 2025; Dan Shao, 2020). The aim is to translate clinical needs into computational targets. For example, using movement entropy (Altozano et al., 2025), gesture fluency, or stress-related gaze patterns as proxies for mental state (Bell et al., 2020). This involves not only feature engineering but statistical correlation studies and machine learning modeling (Random Trees, CNNs, LSTMs) with end-to-end models or hand-crafted features to identify, validate, and prioritize indicators (Altozano et al., 2025). The indicators must be both grounded in clinical theory and practically interpretable to support tasks such as symptom screening, subtype stratification (Alvari et al., 2024), or progress monitoring. Once relevant indicators are identified, the next step is to design a system that clinicians can intuitively use to make informed decisions.

Objective 3: develop a clinician-oriented framework for the visualization and review of patient data

The final objective is to support the clinician's decision-making and intervention process with a structured and interpretable data interface. This includes building a dashboard tool that offers real-time visualizations during live sessions and retrospective views of patient behavior for asynchronous review. The system will incorporate tools to annotate sessions, view trends, and optionally integrate sensitive data such as questionnaires or notes. A User-Centered Approach (UCA) (Kutzin et al., 2017; Filippeschi et al., 2019) will guide the design of HMD and tablet compatible modules to ensure usability, interpretability, and seamless integration into clinical workflows.

3 METHODS

This project adopts a mixed-methods approach, beginning with an interdisciplinary literature review and iterative stakeholder engagement. These foundational steps will inform the design of structured scenarios and guide the development of a data-oriented framework for IPCs. This framework will enable the systematic collection, interpretation, and restitution of behavioral and biometric data generated during XR-based mental health interventions.

3.1 Literature Review

The literature review will focus on three intersecting domains: (1) the use of XR technologies in healthcare settings (Worlikar et al., 2023), (2) strategies for managing mental health in clinical practice (Samira Abdul et al., 2024), and (3) the application of Artificial Intelligence (AI) and Machine Learning (ML) for analyzing biometric data (Greener et al., 2022; Altozano et al., 2025). This review will provide both theoretical grounding and empirical insights for the project.

Primary academic databases such as PubMed and Google Scholar will be used to ensure comprehensive coverage of relevant literature. The review will also incorporate existing research on immersive telehealth platforms, including the work by Prié et al., *A Teleclinic for Neuro-psychological Assessment in the Metaverse* (Prié et al., 2025), which provides a foundational model for XR-based clinical environments.

The outcome of the review will be a consolidated list of relevant behavioral and biometric indicators that can be integrated into IPC scenarios. These indicators will serve as a basis for establishing a structured and continuous dialogue with mental health professionals, ensuring clinical relevance and facilitating co-design of XR interventions.

3.2 Dialogue with mental healthcare professionals

To ensure clinical relevance and usability healthcare professionals will be engaged throughout the project via focus groups and semi-structured interviews. Focus groups will be conducted with professionals from different fields within mental health such as child psychiatrists, psychomotricians and nurses. During these we will be able to discuss their experiences with the mental health system, current workflow, digital mental health tools, current diagnostic gaps and the perceived potential of IPCs. In parallel, individual semi-structured interviews will allow deeper analysis and case by case study of the data necessity of the clinician's field. Insights gathered from this ongoing engagement will directly inform the development of the data framework.

3.3 Framework development for data collection

Based on the two previous subsections, a framework for data collection will be co-developed in iteration with the clinicians to specify the types of information recorded, the methods of collection and their relevance to mental health monitoring. Data will be collected via a standard HMD with hand tracking and an integrated eye tracking system. These data are based on a project that we did previously using the Unity game development framework and either the XRIntegration Toolkit or the Oculus Integration Toolkit packages. With the technical infrastructure in place, the project will advance to design and validate scenario-based tasks tailored to elicit clinically meaningful responses.

3.4 Scenario development with practitioners

As shown in Table 1, Scenario-based design will ensure the ecological and diagnostic validity of the data collected through tasks within the IPC. These scenarios are tailored to elicit data relevant to early screening, symptom monitoring, and stratification of mental health conditions.

Scenario	**Description**	**Data Captured**	**Clinical Use**
Social Stress Simulation	Patient navigates a virtual group conversation	Gaze aversion, voice tremor, body motion	Detection of social anxiety and avoidance behavior
Cognitive Flexibility Task	Color-word Stroop test in immersive form	Response time, error rate, hesitations	Executive dysfunction, attention deficits
Virtual Interview	Semi-structured avatar-led conversation	Facial affect, verbal fluency, eye contact	Indicators of depression or psychomotor slowing

Table 1. Examples of clinician-designed scenarios for IPC environments

To further anchor these scenarios in validated clinical practice, psychometric tools such as the **PHQ-9 (Patient Health Questionnaire-9) (Ford et al., 2020)**, **GAD-7 (Generalized Anxiety Disorder-7) (Spitzer et al., 2006)**, **BDI (Beck Depression Inventory) (Osman et al., 2004)**, and **STAI (State-Trait**

Anxiety Inventory) (Julian, 2011) will be integrated into the scenarios as either pre-session screening or post-session reflective assessments. These tools first serve to get a baseline on self-reported symptoms against behavioral and biometric data. On a second note, they are here to evaluate whether immersive tasks provoke consistent responses in individuals with known profiles.

Altogether, these steps are setting the groundwork for a structured, clinician-informed framework that supports immersive mental health prevention through XR-based data collection, interpretation, and restitution.

CONCLUSION

This research is part of a broader exploration of how IPCs can play a transformative role in mental healthcare by combining XR environments with clinically relevant data workflows. The aim is to enhance early intervention and therapeutic support and to bridge the gap between clinicians and immersive technologies with interpretable insights derived from real-time or ongoing behavioral and biometric inputs. The process will engage iteratively with mental health professionals, from indicator selection to scenario and dashboard interface design, this work will link clinicians and their workflows while building toward scalable and adaptive digital care tools. Through the integration of XR technologies and data-driven dashboards, the project seeks to create a structured framework that supports individualized care and broader mental health prevention strategies.

REFERENCES

Altozano, A., Minissi, M. E., Alcañiz, M., and Marín-Morales, J. (2025). Introducing 3DCNN ResNets for ASD full-body kinematic assessment: A comparison with hand-crafted features. *Expert Systems with Applications* 270, 126295. doi:10.1016/j.eswa.2024.126295

Alvari, G., Vallefuoco, E., Cristofolini, M., Salvadori, E., Dianti, M., Moltani, A., et al. (2024). Exploring Physiological Responses in virtual reality-based interventions for Autism Spectrum Disorder: A Data-Driven Investigation

Bell, I. H., Nicholas, J., Alvarez-Jimenez, M., Thompson, A., and Valmaggia, L. (2020). Virtual reality as a clinical tool in mental health research and practice. *Dialogues in Clinical Neuroscience* 22, 169–177. doi:10.31887/dcns.2020.22.2/lvalmaggia

Caminero, R. V., Méndez-Herrero, L. A., Díaz-Pernas, F. J., Campo, J. C. d., Rodríguez, M. A., and Martínez-Zarzuela, M. (2025). Telekin - Tele-rehabilitation System for Musculoskeletal and Cognitive Disorders using Natural Movement Interface Devices. 198–203

Colizzi, M., Lasalvia, A., and Ruggeri, M. (2020). enPrevention and early intervention in youth mental health: is it time for a multidisciplinary and trans-diagnostic model for care? *International Journal of Mental Health Systems* 14, 23. doi:10.1186/s13033-020-00356-9

Dan Shao, I.-J. L. (2020). Acceptance and Influencing Factors of Social Virtual Reality in the Urban Elderly

Duvivier, V., Perreira da Silva, M., Prié, Y., and Nantes Université, C. L. U. . . N. F., Ecole Centrale Nantes (2021). *A conceptual framework for immersive behavioral data exploration assisted by a deep learned representation of data*. Tech. rep.

Felsberg, D. T., McGuirt, J. T., Ross, S. E., Raisbeck, L. D., Howard, C. K., and Rhea, C. K. (2025). Clinician perspectives on virtual reality use in physical therapy practice in the United States. *PLoS ONE* 20, e0320215. doi:10.1371/journal.pone.0320215

Filippeschi, A., , B., Filippo, , R., Emanuele, , J. V., Juan Manuel, , L., Lorenzo, , et al. (2019). Evaluation of diagnostician user interface aspects in a virtual reality-based tele-ultrasonography simulation. *Advanced Robotics* 33, 840–852. doi:10.1080/01691864.2019.1635909. Publisher: Taylor & Francis

Ford, J., Thomas, F., Byng, R., and McCabe, R. (2020). Use of the Patient Health Questionnaire (PHQ-9) in Practice: Interactions between patients and physicians. *Qualitative Health Research* 30, 2146–2159. doi:10.1177/1049732320924625

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine* 47, 2393–2400. doi:10.1017/s003329171700040x

Greener, J. G., Kandathil, S. M., Moffat, L., and Jones, D. T. (2022). enA guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology* 23, 40–55. doi:10.1038/s41580-021-00407-0. Publisher: Nature Publishing Group

Guo, J., Guo, S., Tamiya, T., Hirata, H., and Ishihara, H. (2016). enA virtual reality-based method of decreasing transmission time of visual feedback for a tele-operative robotic catheter operating system. *The International Journal of Medical Robotics and Computer Assisted Surgery* 12, 32–45. doi:10.1002/rcs.1642. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rcs.1642

Hu, H.-z., Feng, X.-b., Shao, Z.-w., Xie, M., Xu, S., Wu, X.-h., et al. (2019). enApplication and Prospect of Mixed Reality Technology in Medical Field. *Current Medical Science* 39, 1–6. doi:10.1007/s11596-019-1992-8

Isernia, S., Di Tella, S., Pagliari, C., Jonsdottir, J., Castiglioni, C., Gindri, P., et al. (2020). EnglishEffects of an Innovative Telerehabilitation Intervention for People With Parkinson's Disease on Quality of Life, Motor, and Non-motor Abilities. *Frontiers in Neurology* 11. doi:10.3389/fneur.2020.00846. Publisher: Frontiers

Julian, L. J. (2011). Measures of anxiety: State-trait anxiety inventory (STAI), beck anxiety inventory (BAI), and hospital anxiety and depression scale-anxiety (HADS-a) 63, S467–S472. doi:10.1002/acr.20561

Kutzin, J. M., Milligan, Z., and Chawla, S. (2017). EnglishUsing Simulation to Conduct a Usability Study of Wearable Technology. *Clinical Simulation In Nursing* 13, 64–70. doi:10.1016/j.ecns.2016.12.003. Publisher: Elsevier

Lee, A. G. (2024). enAI- and XR-Powered Digital Therapeutics (DTx) Innovations. In *Digital Frontiers - Healthcare, Education, and Society in the Metaverse Era* (IntechOpen). doi:10.5772/intechopen.1006619

Osman, A., Kopper, B. A., Barrios, F., Gutierrez, P. M., and Bagge, C. L. (2004). Reliability and Validity of the Beck Depression Inventory–II With Adolescent Psychiatric Inpatients. *Psychological Assessment* 16, 120–132. doi:10.1037/1040-3590.16.2.120

Prié, Y., Vigier, T., Acier, D., Bonneville, H., Georges, M., Thevenet, S., et al. (2025). A teleclinic for Neuro-psychological Assessment in the Metaverse. In *Abstracts and Workshop Proceedings, IEEE 2025* (Saint-Malo, France)

Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., and Alt, F. (2022). What is XR? Towards a Framework for Augmented and Virtual Reality. *Computers in Human Behavior* 133, 107289. doi:10.1016/j.chb.2022.107289

Samira Abdul, Ehizogie Paul Adeghe, Bisola Oluwafadekemi Adegoke, Adebukola Adejumoke Adegoke, and Emem Henry Udedeh (2024). Mental health management in healthcare organizations: Challenges and strategies - a review. *International Medical Science Research Journal* 4, 585–605. doi:10.51594/imsrj.v4i5.1157

Schlak, A. E., Rosa, W. E., Rushton, C. H., Poghosyan, L., Root, M. C., and McHugh, M. D. (2022). en-USAn expanded institutional- and national-level blueprint to address nurse burnout and moral suffering amid the evolving pandemic. *Nursing Management* 53, 16. doi:10.1097/01.NUMA.0000805032.15402.b3

Spitzer, R. L., Kroenke, K., Williams, J. B. W., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine* 166, 1092–1097. doi:10.1001/archinte.166.10.1092

Steptoe, W. and Steed, A. (2012). Multimodal Data Capture and Analysis of Interaction in Immersive Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments* 21, 388–405. doi:10.1162/PRES_a_00123

Thøgersen, M., Andoh, J., Milde, C., Graven-Nielsen, T., Flor, H., and Petrini, L. (2020). EnglishIndividualized Augmented Reality Training Reduces Phantom Pain and Cortical Reorganization in Amputees: A Proof of Concept Study. *The Journal of Pain* 21, 1257–1269. doi:10.1016/j.jpain.2020.06.002. Publisher: Elsevier

WHO (2022). *World mental health report: Transforming mental health for all* (World Health Organization)

Worlikar, H., Coleman, S., Kelly, J., O'Connor, S., Murray, A., McVeigh, T., et al. (2023). Mixed Reality Platforms in Telehealth Delivery: Scoping Review. *JMIR Biomedical Engineering* 8, e42709. doi:10.2196/42709

Session 3: Clinical and well-being applications of the Metaverse

Immersive Prevention Centers for Mental Health: Secondary and Tertiary Prevention in the Metaverse

Antoine Widmer [1,*], Samuel Bulteau [2], Roman Malo [3], Paul Matusz [1], Noah Richert [1], Jean-Philippe Rivière [4], Przemyslaw Tomalski [5], and Yannick Prié [4]

[1] *HES-SO Valais-Wallis University of Applied Sciences and Arts Western Switzerland, 3960 Sierre, Switzerland*
[2] *Nantes Université, CHU Nantes, INSERM, MethodS in Patients-centered outcomes and Health Research, SPHERE, F-44000 Nantes, France*
[3] *Nantes Université, Laboratoire de psychologie des Pays de la Loire, LPPL, UR 4638, F-44000 Nantes, France*
[4] *Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France*
[5] *Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland*

Correspondence*:
Corresponding Author
antoine.widmer@hevs.ch

ABSTRACT

Mental health systems across Europe face persistent challenges in accessibility, resource limitations, and stigma. Vulnerable populations such as adolescents with neurodevelopmental disorders and older adults at risk of depression often fall through the cracks of existing care structures. The IPC4MH (Immersive Prevention Centers for Mental Health) project proposes an innovative solution: immersive prevention centers built within the Metaverse, designed for secondary and tertiary mental health prevention. Through interdisciplinary collaboration, the project aims to co-design, implement, and evaluate scalable, engaging, and inclusive virtual environments that support early intervention and long-term mental health care. This paper presents the conceptual foundation, objectives, research direction, and methodological approach of IPC4MH.

Keywords: immersive technology, virtual reality, mental health prevention, Metaverse, co-design, digital health, neurodevelopmental disorders, aging

1 INTRODUCTION

Mental health services across Europe are struggling to keep pace with the rising burden of psychological disorders. Limited availability of professionals, logistical barriers to in-person care, and widespread stigma result in significant disparities in access—particularly among youth with neurodevelopmental conditions and seniors facing isolation or cognitive decline. These challenges demand innovative responses that leverage emerging technologies while remaining grounded in evidence-based care practices.

Virtual reality (VR) has already demonstrated clinical efficacy for applications such as cognitive training, exposure therapy, and rehabilitation (Maples-Keller et al., 2017; Freeman et al., 2017). A growing body of research confirms its utility in reducing anxiety, improving cognitive performance, and fostering behavioral change across clinical populations (Romero-Ayuso et al., 2021). More recently, the evolution of social VR and the broader Metaverse has introduced novel opportunities for healthcare delivery—providing persistent, shared, and interactive environments that overcome traditional barriers such as geographic isolation or stigma (Cerasa et al., 2024). Notably, we have introduced the idea of teleclinics in the Metaverse, implementing an example of a teleclinic for neuro-psychological assessment Vigier et al. (2024). The IPC4MH project builds upon this technological and empirical foundations by designing and evaluating immersive prevention centers that enable both early intervention and sustained therapeutic engagement for populations often underserved by conventional services.

2 FRAMING THE NEED: PREVENTION IN MENTAL HEALTH

In public health, mental health prevention is typically structured into three levels: primary (avoiding onset), secondary (early identification and intervention), and tertiary (managing existing conditions to prevent deterioration) Baumann and Ylinen (2020). IPC4MH focuses on the latter two.

Secondary prevention addresses the early signs of mental distress—such as anxiety, withdrawal, or difficulties in social interaction—before they evolve into diagnosable conditions. Immersive prevention centers might be especially relevant for adolescents who may struggle with communication or behavioral issues that are not yet formally recognized. For example, a teenager exhibiting signs of social withdrawal or concentration difficulties in a school setting might participate in a VR-based screening activity followed by a series of interactive group sessions designed to foster self-expression and social confidence. Similarly, older adults experiencing mild forgetfulness or increased isolation may benefit from cognitive stimulation exercises within immersive environments, combined with gentle mood assessments and social interaction features to support early detection of cognitive decline or depression.

Tertiary prevention, meanwhile, focuses on patients with known diagnoses, offering structured support that reduces the risk of relapse and improves quality of life. An adolescent diagnosed with an autism spectrum disorder might for example engage in immersive role-playing scenarios that help build emotional regulation and social interaction skills under the supervision of a therapist. Older adults with a history of major depressive episodes or diagnosed mild cognitive impairment could benefit from long-term participation in guided group activities, memory games, and virtual peer support meetings, helping to reinforce emotional wellbeing, cognitive function, and a sense of connectedness.

By embedding both prevention levels into immersive, user-friendly platforms, IPC4MH aims to make proactive care not only accessible, but appealing and engaging.

3 EXISTING INITIATIVES

While immersive technologies are increasingly explored in mental health, few initiatives have targeted prevention with the same focus and transnational scope as IPC4MH. However, several existing projects and platforms offer valuable context.

One notable example is XRHealth, which offers virtual reality therapy platforms for pain management, rehabilitation, and mental health, including anxiety and stress relief (https://xr.health). Their clinical VR applications are used primarily in therapeutic settings with healthcare professionals and do not yet extend to public prevention or peer-led support.

Another significant effort is that of C2Care, a French company and IPC4MH partner, which provides VR-based exposure therapy environments for phobias, depression, and cognitive training Cerda et al. (2021). These environments are already used in over 500 clinics but are currently deployed in isolated clinical contexts rather than networked prevention centers. IPC4MH builds on this by transforming such tools into socially connected, preventive, and multi-user platforms.

Tecnis, a recent French project led by Nantes University, created a prototype of an immersive teleclinic for neuropsychological assessment Prié et al. (2025). While pioneering in its design and testing, it focused on individual diagnostics rather than ongoing preventive care. IPC4MH uses Tecnis as a foundation and extends its functionality toward collaborative, preventive, and long-term interventions.

In the realm of social VR, projects like VR4REHAB Green et al. (2019), AccessVR or Bridges El Raheb et al. (2021) have also investigated the use of immersive environments for rehabilitation and emotional regulation. These initiatives have demonstrated the feasibility and acceptability of virtual therapies but have not yet addressed structural integration into public health prevention strategies.

By contrast, IPC4MH aims to institutionalize immersive prevention by creating modular, evidence-based platforms deployed in both home and public spaces and evaluated through mixed-method clinical trials. Its dual focus on underserved populations and the integration of secondary and tertiary prevention makes it a novel and impactful initiative in the evolving landscape of digital mental health.

4 PROJECT OBJECTIVES

IPC4MH seeks to create, validate, and prepare for scale-up immersive prevention centers deployed within the Metaverse. To achieve this vision, the project aims to develop a flexible digital platform that incorporates existing VR therapeutic content into user-friendly teleclinics, complete with social and therapeutic functionalities. These centers will be specifically designed for two key target groups —–adolescents and older adults—– through a collaborative and inclusive co-design process involving clinicians, researchers, and end users. Implementation will take place across three European countries —–Switzerland, France, and Poland—– allowing for comparative evaluation of usability, clinical impact, and user satisfaction in varied socio-cultural contexts. A rigorous mixed-methods framework will be used to assess clinical efficacy, combining quantitative and qualitative data to capture both measurable outcomes and user experiences. In parallel, the project will produce guidelines and scalable business models to facilitate long-term integration into national health systems and public spaces, ensuring both sustainability and broader impact.

5 RESEARCH DIRECTION

The project's research agenda spans four interdependent axes: technological infrastructure, human-centered design, clinical validation, and policy integration.

On the technological front, IPC4MH will define and implement a modular architecture based on open standards (e.g., OpenXR) to ensure interoperability across devices and platforms. The system will include secure communication tools, user monitoring systems, and interfaces for integrating diverse therapeutic content. The goal is to create an extensible platform where practitioners can meet with users, guide sessions, and monitor progress.

Within IPC4MH, artificial intelligence (AI) serves as a decision-support layer by analyzing structured interaction data—such as session frequency, duration, and trajectories of improvement in standardized clinical scores (e.g., PHQ-9, GAD-7). Rather than replacing human judgment, AI models will synthesize these usage patterns to generate interpretable insights, which are integrated directly into therapist dashboards. This allows clinicians to monitor individual progress, detect early signs of disengagement or stagnation, and tailor intervention strategies accordingly. The focus on transparent, explainable outputs ensures that AI serves to augment—not obscure—clinical reasoning. This integration supports a hybrid model of care, where immersive environments and professional supervision are dynamically coordinated to maximize the effectiveness and responsiveness of prevention efforts.

From a design perspective, immersive scenarios and environments will be co-designed with input from stakeholders, using ethnographic methods and participatory design workshops. Care workflows will be based on activities that are carried out both in virtual (immersive center) and in real (physical prevention hubs) world, be it solo activities, peer-supported interactions, supervised group therapy, meetings with practitioners, etc.

Clinically, the project aims to measure how IPCs affect mental health literacy and user engagement. Metrics will include validated psychological scales (e.g., PHQ-9, GAD-7), behavioral analytics, and analysis of qualitative interviews. Evaluations will compare different usage patterns (e.g., home-based vs. onsite), prevention levels (secondary vs. tertiary), and demographic subgroups.

Finally, IPC4MH examines the socio-economic, ethical, and regulatory dimensions of deploying virtual prevention centers at scale. It will develop frameworks for integration into national health systems and explore sustainable public-private models for ongoing delivery and support.

6 METHOD

The project adopts a multi-phase, interdisciplinary methodology.

Phase 1: Co-Design and Prototyping — Initial work will involve mapping out user journeys, defining clinical and technical requirements, and adapting existing VR content for shared use. Iterative prototyping with feedback loops from patients and professionals will drive continuous refinement.

Phase 2: Implementation and Evaluation — Prevention centers will be rolled out across three pilot sites, with 20–30 participants per site engaged in either secondary or tertiary prevention activities. Evaluation will cover usability (System Usability Scale), acceptability (Technology Acceptance Model), clinical outcomes (standardized scales), and engagement (interaction logs).

Phase 3: Scale-Up and Dissemination — Lessons from the evaluation phase will inform the design of integration pathways, economic models, and policy recommendations. Open-access tools, datasets, and documentation will ensure reproducibility and impact beyond the project consortium.

7 EARLY RESULTS

Preliminary activities carried out during the first project phase have yielded promising insights. A co-design workshop held in France (3-4 April 25) with consortium members provided early reflections on the concept of immersive prevention centers. Participants expressed enthusiasm about the potential for peer interaction and creative self-expression in virtual environments. Discussions also highlighted the necessity of structured, clinically validated content and raised important concerns related to data protection and ethical oversight.

Initial technical assessments of partner VR platforms, including those from DiverSSity and C2Care, indicated that existing tools could be adapted to suit group-based use cases. Usability testing on early prototypes demonstrated acceptable user performance on navigation and task completion tasks across both target demographics. Participants appreciated the calming environment designs and the ability to personalize avatars, which contributed to perceived presence and comfort.

These early findings confirm both the feasibility and the relevance of IPC4MH's core hypothesis: that immersive virtual environments can serve as effective platforms for mental health prevention. The team is now moving forward with refining the system's architecture and preparing for full-scale pilot evaluations.

8 CONCLUSION

The IPC4MH project represents a bold step toward the future of preventive mental health care. By merging immersive technologies with inclusive design and rigorous clinical evaluation, it aims to create prevention pathways that are engaging, scalable, and equitable. As health systems seek new strategies to meet rising mental health needs, IPCs may offer a replicable model for delivering care across digital and physical boundaries—reaching people where they are, before it's too late.

FUNDING

This work is supported by the European co-funded partnership initiative "Transforming Health and Care Systems (THCS)" under the joint sponsorship of Innosuisse (Switzerland), the Agence Nationale de la Recherche (ANR, France), and the National Centre for Research and Development (NCBR, Poland).

REFERENCES

Baumann, L. C. and Ylinen, A. (2020). Prevention: Primary, Secondary, Tertiary. In *Encyclopedia of Behavioral Medicine*, ed. M. D. Gellman (Cham: Springer International Publishing). 1738–1740. doi:10.1007/978-3-030-39903-0_135

Cerasa, A., Gaggioli, A., Pioggia, G., and Riva, G. (2024). Metaverse in Mental Health: The Beginning of a Long History. *Current Psychiatry Reports* 26, 294–303. doi:10.1007/s11920-024-01501-8

Cerda, L., Fauvarque, A., Graziani, P., and Del-Monte, J. (2021). Contextual priming to increase the sense of presence in virtual reality: Exploratory study. *Virtual Reality* 25, 1105–1112. doi:10.1007/s10055-021-00515-4

El Raheb, K., Soulis, A., Nastos, D., Lougiakis, C., Roussou, M., Christopoulos, D., et al. (2021). Eliciting requirements for a multisensory eXtended Reality platform for training and informal learning. In *CHI Greece 2021: 1st International Conference of the ACM Greek SIGCHI Chapter* (New York, NY, USA: Association for Computing Machinery), CHI Greece 2021, 1–8. doi:10.1145/3489410.3489428

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine* 47, 2393–2400. doi:10.1017/S003329171700040X

Green, D., Boltin, S., Ward, A., Gowing, F., and Hutchon, B. (2019). Coproducing Virtual Reality Technologies for Rehabilitation. In *2019 International Conference on Virtual Rehabilitation (ICVR)*. 1–2. doi:10.1109/ICVR46560.2019.8994337

Maples-Keller, J. L., Bunnell, B. E., Kim, S.-J., and Rothbaum, B. O. (2017). The Use of Virtual Reality Technology in the Treatment of Anxiety and Other Psychiatric Disorders. *Harvard Review of Psychiatry* 25, 103. doi:10.1097/HRP.0000000000000138

Prié, Y., Vigier, T., Acier, D., Bonneville, H., Georges, M., Thevenet, S., et al. (2025). A teleclinic for neuro-psychological assessment in the metaverse. In *Abstracts and Workshop Proceedings, IEEE 2025* (Saint-Malo, France)

Romero-Ayuso, D., Toledano-González, A., Rodríguez-Martínez, M. d. C., Arroyo-Castillo, P., Triviño-Juárez, J. M., González, P., et al. (2021). Effectiveness of Virtual Reality-Based Interventions for Children and Adolescents with ADHD: A Systematic Review and Meta-Analysis. *Children* 8, 70. doi:10.3390/children8020070

Vigier, T., Acier, D., Bulteau, S., Thevenet, S., and Prié, Y. (2024). Towards immersive teleclinics in the Metaverse: the case of neuropsychological consultation. In *Conference A Metaverse for the good* (Barcelona, Spain: EMRN European Metaverse Research Network)

Towards affective and personalized mixed reality experiences for mental health

Patricia Pons [1,*], **Samuel Navas-Medrano** [1] **and Jose Luis Soler-Dominguez** [1]

[1] *Human Computer Interaction research group, Instituto Tecnológico de Informática (ITI), Valencia, Spain*

Correspondence*:
Patricia Pons
ppons@iti.es
hci@iti.es

ABSTRACT

Extended Reality technologies are rapidly transforming healthcare, particularly in the domain of mental health. As immersive systems mature, they offer powerful tools to enhance therapeutic interventions through increased presence, emotional engagement, and accessibility. However, existing XR solutions for mental health often lack dynamic personalization, limiting their effectiveness in addressing individual emotional needs. This paper explores how the integration of Artificial Intelligence, biofeedback, and context-aware interaction mechanisms can overcome these limitations and enable emotionally adaptive, personalized XR interventions. We present several examples of Mixed Reality platforms developed for emotional regulation and remote collaborative therapy, and explore how generative AI, biometric feedback, and real-world object recognition can be leveraged to build the foundations of a mental health metaverse.

Keywords: Mixed Reality, Metaverse, Mental Health, Remote Therapy, Biofeedback, Artificial Intelligence, Generative AI, Object Recognition

1 INTRODUCTION

Extended Reality (XR) technologies, encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) (Milgram and Kishino, 1994; Speicher et al., 2019; Skarbez et al., 2021), have shown promise in improving the quality of care across various domains of healthcare: offering realistic simulations for surgical training and medical procedures (Toni et al., 2024; Herur-Raman et al., 2021; Andrews et al., 2019), supporting patient care practice (Zweifach and Triola, 2019; Lee et al., 2024), providing immersive experiences for pain management (Indovina et al., 2018; Phelan et al., 2021), and creating motivating scenarios for physical rehabilitation (Grabowski, 2025; Villada Castillo et al., 2024).

Recently, mental health and wellbeing have been gaining increased acknowledgement on their vital role in global health, and its impact on socio-economic conditions and the global burden of disease (WHO, 2022). Within this context, XR's immersive and engaging nature makes it a compelling complement to traditional therapies, supporting treatments like exposure therapy (Gutierrez-Maldonado et al., 2023; Freeman et al., 2022; Lundin et al., 2022; Giraldy and Novaldo, 2022), cognitive behavioural therapy, mindfulness practices (Amores et al., 2019; Modrego-Alarcón et al., 2023; Hanna et al., 2025) or emotional regulation strategies (Macey et al., 2022; Soler-Dominguez et al., 2024; Wagener and Niess, 2021).

Despite these advances, many XR mental health applications lack personalization, often delivering one-size-fits-all experiences that fail to reflect individual needs, preferences, or emotional states. There has been some related work that has explored early attempts at customization and automation of these environments (Freeman et al., 2018), but much work remains to be done to achieve a degree of adaptability to suit the wide variety of real needs and experiences. The absence of this kind of personalization in current XR systems can limit the effectiveness of interventions, especially when addressing the nuanced and heterogeneous nature of mental health conditions (Bennett and Shafran, 2022; Hornstein et al., 2023).

In this regard, Artificial Intelligence (AI) applied to XR presents a promising path to overcoming these limitations. AI has already shown potential in mental health through chatbots, diagnostic support systems, and emotion recognition algorithms. Integrating AI into XR environments could facilitate deeply personalized, context-aware, and emotionally intelligent therapeutic experiences. AI-driven XR systems could dynamically adjust scenarios based on real-time biometric feedback, user preferences or therapists input, offering users tailored interactions that evolve with their mental health needs.

In this paper, we explore the intersection of XR and AI in mental health, examining current use cases, technological challenges, and the potential for AI-enhanced XR to create the next generation of responsive, affective, and personalized mental health interventions.

2 XR FOR AFFECTIVE AND PERSONALIZED MENTAL HEALTHCARE

2.1 Advancing Mental Health Interventions Through Mixed Reality

VR has proven effective for mental health interventions like exposure therapy and mindfulness by providing immersive, controlled environments. However, its fully immersive nature can be a drawback in some scenarios, as it disconnects users from their surroundings, which may be unsuitable for some therapeutic contexts where environmental awareness, embodiment, and real-world cues play a critical role. In addition, VR may not be the preferred interaction modality for people who are more vulnerable or who need to have visual contact with their therapist or reference person. Issues such as visual-induced motion sickness (VIMS) (Bos et al., 2008; Geszten et al., 2018) and accessibility limitations also present challenges, particularly for vulnerable populations (Kim et al., 2021).

In this context, MR emerges as a promising complementary tool when using immersive technology for mental health care (Pons et al., 2022). Unlike traditional VR, where users are fully immersed in a separate, digital world, MR overlays interactive digital content onto the physical space around them, which reduces VIMS problems while allowing for interactive, situated experiences that maintain the user's connection to the real world. This allows therapy to occur in familiar settings, such as a home or clinic, surrounded by familiar people and objects, therefore enhancing comfort, which is especially important when dealing with sensitive mental health topics. Moreover, MR reduces the psychological distance between therapy and everyday life. Instead of needing to ”step out” of reality to engage with therapeutic content, users can bring therapeutic interactions into their daily context. This blending of real and virtual elements makes it easier to integrate therapeutic exercises into real-life situations, which can increase relevance and make it more likely that the strategies learned will carry over into everyday behavior.

While VR has seen growing adoption in mental health applications, MR remains significantly underexplored in this domain, with only a limited number of studies and systems addressing its unique potential to integrate therapeutic content into users' real-world environments. In our previous work, we have explored the application of MR in mental health through the development of ARCADIA (Soler-Dominguez

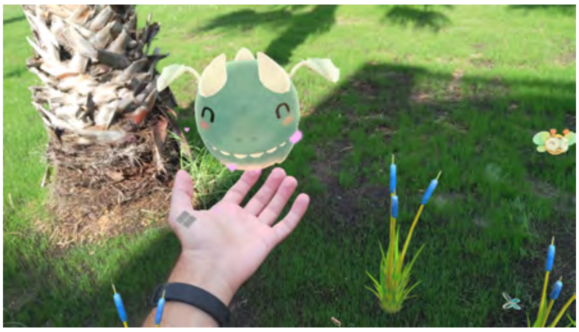

Figure 1. Emotional Avatar.

Figure 2. Breathing activity.

Figure 3. Emotion recognition activity.

Figure 4. Animal Therapy.

Figure 5. Therapeutic activities of the ARCADIA platform running on a MR headset

et al., 2024; García-Ballesteros et al., 2024), a MR system co-designed with mental health professionals to support emotional regulation through gamified therapeutic activities. ARCADIA centers around the metaphor of a virtual garden that reflects the user's therapeutic progress, with a modular architecture that allows for future expansion beyond emotional regulation. Currently, ARCADIA includes four interactive experiences: 1) mindful breathing: through a respiration sensor, users find themselves in a meadow with flowers that slowly bloom as they inhale and exhale, 2) intrusive thoughts: users reflect on rumination, externalizing intrusive thoughts by writing them in the air, and then disposing them in a fire pit, 3) emotion identification: facilitates emotion identification using visual metaphors such as creatures representing each of the six basic emotions, and 4) animal therapy: users can choose between different virtual animals and actions, such as petting them feeding, brushing or playing with them. These four activities are designed to foster engagement, reflection, and emotional self-awareness in immersive, context-rich environments.

The user studies conducted with ARCADIA focused on user experience with both mental health professionals and users of mental health services (Soler-Dominguez et al., 2024; García-Ballesteros et al., 2024), revealing that participants found the system engaging, motivating, and helpful for emotional regulation. for managing emotional states. These findings highlight the potential of MR to create immersive yet grounded therapeutic experiences that adapt to real-world contexts. By blending digital content with users' physical environments, MR offers a powerful platform for advancing mental health care.

2.2 The Mental Health Metaverse: Supporting Remote and Collaborative Therapy through Mixed Reality

As XR technologies continue to evolve and integrate with socially connected virtual spaces, the idea of a mental health metaverse is taking shape (Navas-Medrano et al., 2024; Cerasa et al., 2024; López del

Hoyo et al., 2024; Wang et al., 2025). This concept describes a network of immersive digital environments focused on supporting mental wellbeing. Within this space, people could attend therapy sessions remotely, explore guided mental health exercises, or connect with others in shared support communities, no matter where they are physically located. Unlike traditional teletherapy or isolated VR apps, the mental health metaverse offers a continuous and unified experience. Users can maintain persistent virtual identities and engage in real-time interactions with therapists and peers, making mental health care more accessible and collaborative (Fajnerova et al., 2024).

One of the most promising applications of the mental health metaverse is remote therapy, where a therapist and a patient can meet in a shared, hybrid space regardless of physical distance. Unlike standard video calls, XR allows both parties to interact with the same virtual objects and environments in real time, enhancing engagement and relational presence. In this way, the mental health metaverse would help overcome longstanding barriers to care. On one hand, it could increase accessibility by removing geographic and physical limitations (Palomin et al., 2023), facilitating access to mental health services for people who are physically unable to move around easily, or people who live far from health care centers (Altaf Dar et al., 2023; Hendl and Shukla, 2024). On the other hand, it could also help reduce barriers associated with stigma (Hazell et al., 2022; Mezuk et al., 2020), by allowing private participation of users from the comfort of their own familiar environment. Furthermore, incorporating at-home XR therapy exercises can alleviate the burden on healthcare facilities, decrease waiting times, and optimise resource utilisation within the mental health system (Bratt and Kalmendal, 2023).

Beyond one-on-one sessions, XR also supports multiuser collaborative environments that can facilitate group therapy. Whether participants are co-located or connecting remotely, XR enables shared, interactive therapeutic scenarios where users can practice social skills, reflect together, or engage in mindfulness activities in a co-created space. These collaborative features can reduce the isolation often associated with mental health struggles, while preserving the benefits of guided intervention. Remote collaborative activities could be mediated by the use of avatars, that will be displayed by the in the real environment of each participant. In-person group therapies will involve participants in the same physical spaces, some or all of them wearing XR headsets in which the same digital scenario is shared.

To explore the potential of the mental health metaverse, we developed CORE-MHC, a MR platform designed to support real-time therapy between clinicians and patients. This system enables therapists to participate in sessions either via a traditional desktop interface or through an MR headset, while the patient consistently engages through MR hardware to ensure immersive and embodied interaction. The dual-mode access for therapists is designed to increase the inclusivity and flexibility of care delivery, adapting to the technological capabilities and preferences of the healthcare provider while maintaining continuity in the patient's experience.

Once both parties are connected in the MR environment, the therapist guides the patient through the therapeutic activity, while visualizing biometric data from the patient captured through wearable sensors. These biofeedback metrics, such as heart rate or respiratory patterns, allow for continuous monitoring of the patient's emotional and physiological state throughout the session. When both therapist and patient are immersed via MR headsets, they are represented by dynamic avatars and can co-interact with shared augmented objects within the virtual space, despite being in geographically separate physical locations. This infrastructure not only enhances therapeutic alliance and engagement but also opens the door to the integration of biofeedback-driven interventions for more adaptive experiences.

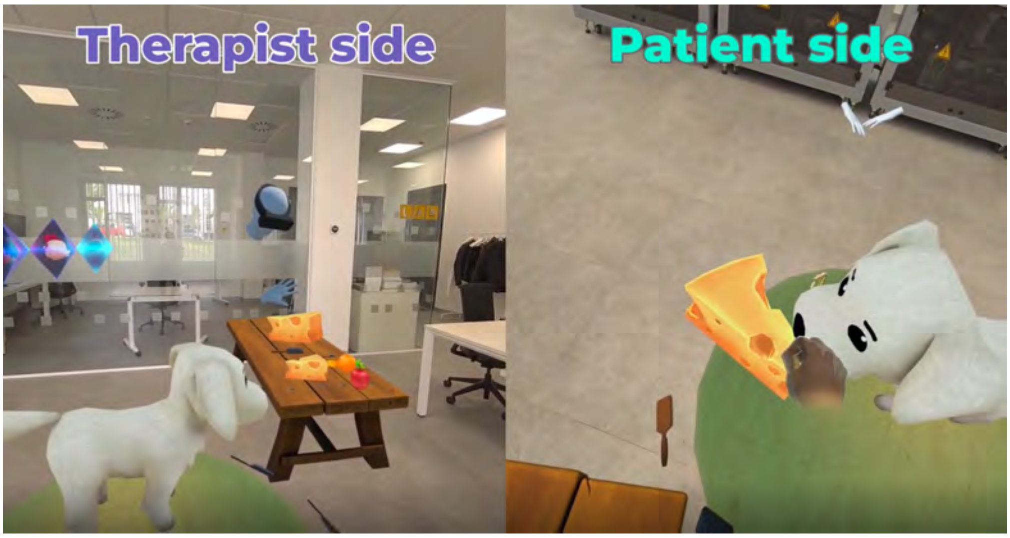

Figure 6. CORE-MHC system: therapist and patient view during the collaborative activity.

2.3 Biofeedback: Enhancing Self-Awareness and Adaptive Interventions

Incorporating biofeedback into XR environments creates the foundation for more responsive, emotionally intelligent experiences. By collecting real-time physiological signals such as heart rate, respiration, or skin conductance, systems can infer the user's stress levels, emotional arousal, or attentional focus. This data can then be used to dynamically adapt the experience, adjusting the level of challenge, feedback, or pacing based on the user's current state.

For example, in the breathing activity developed for ARCADIA (Fig. 2), a respiration sensor was used to guide a mindfulness experience where a withered digital meadow gradually transforms as the user focuses on their breath. The visual feedback, delayed and subtle, reinforces the calming effect of deep breathing while avoiding over-stimulation. The experience is further enhanced with the introduction of an emotional avatar (Fig. 1): a dynamic, embodied representation of the user's internal state within the MR space, whose form and behavior respond in real time to biofeedback signals such as heart rate and breath rhythm variability. This avatar provides a tangible, visualized extension of their inner experience, helping users build emotional self-awareness by externalizing the physiological state in a gentle, interpretable way. Another example of biofeedback-driven XR are Emo-regulator (Yu et al., 2023), an emotion-regulation training system that uses VR and EEG-based neurofeedback. Results shown improved emotion regulation ability and above-average usability after using the system for 2 weeks. Applications like the DEEP VR game (Bossenbroek et al., 2020) utilise biofeedback to monitor physiological responses such as heart rate and breathing patterns. These real-time metrics are then used to adjust the virtual environment, promoting relaxation and teaching users how to regulate their physiological responses to stress. The work in (Amores et al., 2019) focused on internal reflection, developing a VR environment that procedurally generates 3D creatures, and changes the lighting of the environment to reflect users' internal state based on EEG, EDA and HR.

Looking ahead, the integration of biofeedback with AI-driven emotion recognition models holds the potential to further individualize therapeutic interventions. These systems could detect states such as anxiety, frustration, or relaxation, that automatically adjust the XR environment to support the user's needs, such as modulating visual complexity, soundscapes, or the intensity of therapeutic challenges. For example, in moments of heightened arousal, the system might reduce environmental stimuli, simplify interaction demands, or introduce grounding cues. Conversely, during states of low engagement, it may introduce gently challenging tasks or more vibrant sensory elements. Through this feedback mechanism, XR experiences can move towards emotionally adaptive mental health systems that respond with empathy and precision to each user's unique journey.

2.4 Generative AI for Personalized XR Environments

Generative AI offers unprecedented possibilities for personalizing mental health experiences in XR, creating content that adapts in real time to the user's needs, evolving preferences, and emotional state. By incorporating real-time biofeedback and interaction data, generative systems can craft therapeutic experiences that are not only immersive but also deeply resonant on an individual level.

One powerful application could be the generation of adaptive narratives. In gamified therapeutic MR scenarios, generative models can tailor the narrative flow according to the user's progress, emotional responses, or interaction patterns. Rather than following a fixed script, the storyline can branch dynamically, offering challenges, metaphors, or interactions that feel meaningful and therapeutically aligned with the individual. This responsiveness can enhance user engagement and facilitate deeper personal reflection within the immersive context.

Another promising direction is the real-time generation of 3D objects (Liu et al., 2024; Chen et al., 2024; Wong et al., 2022) that reflect personal symbols, memories, or therapeutic goals. A user working on anxiety management might summon calming objects linked to positive experiences, or modify the environment to represent a physical location that the user considers as a safe space or a place of relaxation, i.e. a beach, a forest, the village house of their childhood.

Sound generation adds another critical layer of personalization (Nguyen et al., 2024; Picinali et al., 2022). Through generative audio models, ambient soundscapes, binaural beats, or music can be created in real time to support the user's desired emotional state or therapeutic activity. Whether designed to ground a user during moments of stress, uplift during depressive states, or deepen focus during mindfulness exercises, these responsive auditory elements serve as powerful emotional modulators, enhancing presence and co-regulating mood.

The integration of conversational avatars powered by large language models (LLMs) could also offer promising opportunities for tailored interventions (Ma et al., 2023; Jo et al., 2023). These virtual companions could engage users in natural dialogue, guiding them through personalized mindfulness exercises, checking in on emotional states, or serving as virtual therapeutic companions. When integrated with biofeedback data and user interaction histories, these avatars could adapt their tone, content, and strategy to suit the evolving context of the session. For instance, an avatar may adjust the pacing of a guided breathing exercise based on real-time respiration data, or shift to a more supportive tone when signs of distress are detected. These AI-driven interactions can be deeply individualized, leveraging historical interaction data and biofeedback to provide nuanced, supportive dialogue that evolves with the user over time.

Yet, for these systems to achieve deeper ecological validity and therapeutic relevance, they must also be able to interpret and respond to the user's physical environment. By combining biofeedback with contextual awareness, MR systems can foster a more holistic therapeutic experience.

2.5 Context-Aware Interactions Through Object Recognition in MR

To create truly seamless and context-sensitive MR experiences, recognizing and responding to the physical environment is essential. Object recognition is a key capability that enables MR systems to identify real-world items such as furniture, specific objects or personal belongings, and use them as anchors for interactive experiences. The implementation of object recognition in XR-based therapeutic systems builds upon ongoing advancements in computer vision and spatial computing (Dasgupta et al., 2020; Tang et al., 2023). Frameworks such as Meta's Project Aria[1] and Google XR Objects[2] are paving the way for robust, real-time interaction with the physical world within extended reality environments.

Object recognition through MR headsets could add a new layer of contextual relevance and engagement to mental health immersive interventions. For instance, a therapeutic system might recognize a user's chair and place calming virtual elements around it. If the system detects a personal item like a notebook frequently used for journaling, it might gently suggest a short reflective prompt based on recent emotional patterns or biofeedback data. In pediatric contexts, everyday objects such as a stuffed animal can be leveraged to initiate emotionally supportive interactions. When the MR system detects the presence of a familiar plush toy, it can trigger a calming, game-based experience designed to support emotion regulation in young users. For example, the toy may be transformed into an animated character within the MR environment, guiding the child through simple activities such as deep breathing, emotion labeling, or body scans. For adult users, therapeutic interactions can become more symbolic and metaphor-driven, as in the case of an ”emotional gardening” experience. When the system recognizes a plant or designated calming object in the environment, it can overlay a virtual garden within the MR space that evolves in response to the user's current emotional state, inferred from biofeedback inputs such as heart rate variability or galvanic skin response. This integration of the digital and physical world enhances the sense of continuity between therapy and daily life, encouraging users to see mental health practices not as isolated tasks, but as part of their everyday environment.

3 CONCLUSION AND FUTURE WORK

The intersection of AI and XR technologies represents a pivotal opportunity to re-imagine how we approach mental health care. By combining biofeedback, generative AI, and real-world object recognition, we can move beyond static interventions toward systems that are emotionally adaptive, responsive to user context, and capable of supporting individuals in real time.

Taken together, these generative AI capabilities create the foundation for on-demand, user-centered XR environments that are responsive, emotionally intelligent, and continuously adaptive. By blending narrative, visual, auditory, and conversational modalities, generative systems transform XR into a living canvas—one that shapes itself around the user's psychological and emotional journey. Importantly, these personalized experiences are not intended to replace traditional therapeutic care but to serve as complementary tools that can extend support beyond clinical settings, enhance engagement between sessions, and provide users with additional resources for reflection, self-regulation, and emotional growth.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

[1] https://www.projectaria.com/

[2] https://github.com/google/xr-objects

AUTHOR CONTRIBUTIONS

SN-M: Writing—original draft, Writing—review editing. JS-D: Writing—original draft, Writing—review editing. PP: Writing—original draft, Writing—review editing.

FUNDING

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work has been funded by the Valencian Institute for the Enterprise Competitiveness-Generalitat Valenciana (IMAMCA/2025/11), IVACE/FEDER funds through the projects ARCADIA (IMDEEA/2022/22), ARCADIA2 (IMDEEA/2023/40), and AWARE-XR (IMDEEA/2024/50). The CORE-MHC project is funded by the European Union's Horizon Europe research and innovation action programme, via the CORTEX2 Open Call 1 issued and executed under the CORTEX2 project (Grant Agreement no. 01070192).

REFERENCES

Altaf Dar, M., Maqbool, M., Ara, I., and Zehravi, M. (2023). The intersection of technology and mental health: enhancing access and care. *International journal of adolescent medicine and health* 35, 423–428

Amores, J., Fuste, A., and Richer, R. (2019). Deep Reality: Towards Increasing Relaxation in VR by Subtly Changing Light, Sound and Movement Based on HR, EDA, and EEG. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM), 1–2. doi:10.1145/3290607.3311770

Andrews, C., Southworth, M. K., Silva, J. N. A., and Silva, J. R. (2019). Extended reality in medical practice. *Current Treatment Options in Cardiovascular Medicine* 21. doi:10.1007/s11936-019-0722-7

Bennett, S. D. and Shafran, R. (2022). Adaptation, personalization and capacity in mental health treatments: a balancing act? *Current Opinion in Psychiatry* 36, 28–33. doi:10.1097/yco.0000000000000834

Bos, J. E., Bles, W., and Groen, E. L. (2008). A theory on visually induced motion sickness. *Displays* 29, 47–57

Bossenbroek, R., Wols, A., Weerdmeester, J., Lichtwarck-Aschoff, A., Granic, I., and van Rooij, M. M. J. W. (2020). Efficacy of a virtual reality biofeedback game (deep) to reduce anxiety and disruptive classroom behavior: Single-case study. *JMIR Ment Health* 7, e16066. doi:10.2196/16066

Bratt, A. and Kalmendal, A. (2023). ehealth interventions can potentially provide cost-effective and accessible support for stress-related problems in healthcare professionals. *Evidence-Based Nursing* , ebnurs–2023

Cerasa, A., Gaggioli, A., Pioggia, G., and Riva, G. (2024). Metaverse in mental health: The beginning of a long history. *Current Psychiatry Reports* 26, 294–303. doi:10.1007/s11920-024-01501-8

Chen, T., Cao, R., Li, Z., Zang, Y., and Sun, L. (2024). Deep3dsketch-im: rapid high-fidelity ai 3d model generation by single freehand sketches. *Frontiers of Information Technology amp; Electronic Engineering* 25, 149–159. doi:10.1631/fitee.2300314

Dasgupta, A., Manuel, M., Mansur, R. S., Nowak, N., and Gracanin, D. (2020). Towards real time object recognition for context awareness in mixed reality: A machine learning approach. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (IEEE). doi:10.1109/vrw50115.2020.00054

Fajnerova, I., Hejtmánek, L., Sedlák, M., Jablonská, M., Francová, A., and Stopková, P. (2024). The journey from nonimmersive to immersive multiuser applications in mental health care: Systematic review. *Journal of Medical Internet Research* 26, e60441. doi:10.2196/60441

Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., et al. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry* 5, 625–632. doi:10.1016/s2215-0366(18)30226-8

Freeman, D., Lambe, S., Kabir, T., Petit, A., Rosebrock, L., Yu, L.-M., et al. (2022). Automated virtual reality therapy to treat agoraphobic avoidance and distress in patients with psychosis (gameChange): a multicentre, parallel-group, single-blind, randomised, controlled trial in England with mediation and moderation analyses. *The Lancet Psychiatry* 9, 375–388. doi:10.1016/S2215-0366(22)00060-8

García-Ballesteros, M., Navas-Medrano, S., Soler-Dominguez, J. L., and Pons, P. (2024). The garden within: Using mixed reality to foster wellbeing through emotional regulation. In *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (IEEE), 315–322. doi:10.1109/ismar-adjunct64951.2024.00071

Geszten, D., Komlódi, A., Hercegfi, K., Hamornik, B. P., Young, A. L., Koles, M., et al. (2018). A content-analysis approach for exploring usability problems in a collaborative virtual environment. *Acta Polytechnica Hungarica* 15. doi:10.12700/aph.15.5.2018.5.5

Giraldy, D. J. and Novaldo, W. (2022). A Systematic Literature Review: Acrophobia Treatment with Virtual Reality. *Engineering, MAthematics and Computer Science (EMACS) Journal* 4, 33–38. doi:10.21512/emacsjournal.v4i1.8077

Grabowski, A. (2025). Acceptance and usability of different virtual reality-based technologies supporting physical rehabilitation of the upper limb: Results of experiments with patients after upper limb injury at work. *Games for Health Journal* doi:10.1089/g4h.2024.0071

Gutierrez-Maldonado, J., Briseño, N., and Ascione, M. (2023). Treatment of Anorexia Nervosa Through Virtual Reality-Based Body Exposure and Reduction of Attentional Bias. In *Lecture Notes in Computer Science* (Springer Nature Switzerland), vol. 14027, 453–463. doi:10.1007/978-3-031-35634-6

Hanna, J., Mehta, J., West, H., Keane, P., Wilson, N., and Bridge, P. (2025). Evaluation of a novel interactive virtual reality environment for mindfulness skills training. *BMC Digital Health* 3. doi:10.1186/s44247-024-00141-5

Hazell, C. M., Berry, C., Bogen-Johnston, L., and Banerjee, M. (2022). Creating a hierarchy of mental health stigma: Testing the effect of psychiatric diagnosis on stigma. *Bjpsych Open* doi:10.1192/bjo.2022.578

Hendl, T. and Shukla, A. (2024). Can digital health democratize health care? *Bioethics*

Herur-Raman, A., Almeida, N. D., Greenleaf, W., Williams, D., Karshenas, A., and Sherman, J. H. (2021). Next-generation simulation—integrating extended reality technology into medical education. *Frontiers in Virtual Reality* 2. doi:10.3389/frvir.2021.693399

Hornstein, S., Zantvoort, K., Lueken, U., Funk, B., and Hilbert, K. (2023). Personalization strategies in digital mental health interventions: a systematic review and conceptual framework for depressive symptoms. *Frontiers in Digital Health* 5. doi:10.3389/fdgth.2023.1170002

Indovina, P., Barone, D., Gallo, L., Chirico, A., De Pietro, G., and Giordano, A. (2018). Virtual reality as a distraction intervention to relieve pain and distress during medical procedures: A comprehensive literature review. *The Clinical Journal of Pain* 34, 858–877. doi:10.1097/ajp.0000000000000599

Jo, E., Epstein, D. A., Jung, H., and Kim, Y.-H. (2023). Understanding the benefits and challenges of deploying conversational ai leveraging large language models for public health intervention. In

Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (ACM), CHI '23, 1–16. doi:10.1145/3544548.3581503

Kim, H. G., Lee, S., Kim, S., Lim, H.-t., and Ro, Y. M. (2021). Towards a better understanding of vr sickness: physical symptom prediction for vr contents. In *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 35, 836–844

Lee, M., Kim, S.-K., Go, Y., Jeong, H., and Lee, Y. (2024). Positioning virtual reality as means of clinical experience in mental health nursing education: A quasi-experimental study. *Applied Nursing Research* 77, 151800. doi:10.1016/j.apnr.2024.151800

[Dataset] Liu, J., Huang, X., Huang, T., Chen, L., Hou, Y., Tang, S., et al. (2024). A comprehensive survey on 3d content generation. doi:10.48550/ARXIV.2402.01166

Lundin, J., Lundström, A., Gulliksen, J., Blendulf, J., Ejeby, K., Nyman, H., et al. (2022). Using 360-degree videos for virtual reality exposure in CBT for panic disorder with agoraphobia: a feasibility study. *Behavioural and Cognitive Psychotherapy* 50, 158–170. doi:10.1017/S1352465821000473

López del Hoyo, Y., Elices, M., and Garcia-Campayo, J. (2024). Mental health in the virtual world: Challenges and opportunities in the metaverse era. *World Journal of Clinical Cases* 12, 2939–2945. doi:10.12998/wjcc.v12.i17.2939

[Dataset] Ma, Z., Mei, Y., and Su, Z. (2023). Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. doi:10.48550/ARXIV.2307.15810

Macey, A.-L., Macey, J., and Hamari, J. (2022). Virtual reality in emotion regulation: A scoping review. In *6th International GamiFIN Conference*

Mezuk, B., Needham, B. L., Joiner, K. L., Watkins, D. C., Stoddard, S. A., Burgard, S. A., et al. (2020). What elephant? pedagogical approaches to addressing stigma toward mental disorders in undergraduate public health education. *Pedagogy in Health Promotion* doi:10.1177/2373379920922871

Milgram, P. and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 1321–1329

Modrego-Alarcón, M., Morillo, H., Campos, D., Navarro-Gil, M. T., Montero-Marín, J., Monreal-Bartolomé, A., et al. (2023). Effects and acceptability of virtual reality to facilitate mindfulness practice in university students. *Journal of Computing in Higher Education* 37, 273–298. doi:10.1007/s12528-023-09393-y

Navas-Medrano, S., Soler-Dominguez, J. L., and Pons, P. (2024). Mixed reality for a collective and adaptive mental health metaverse. *Frontiers in Psychiatry* 14. doi:10.3389/fpsyt.2023.1272783

Nguyen, P., Fdez, J., and Witkowski, O. (2024). *AI-Driven Meditation: Personalization for Inner Peace* (Springer Nature Switzerland). 296–310. doi:10.1007/978-3-031-56992-0_19

Palomin, A., Takishima-Lacasa, J., Selby-Nelson, E., and Mercado, A. (2023). Challenges and ethical implications in rural community mental health: the role of mental health providers. *Community mental health journal* 59, 1442–1451

Phelan, I., Furness, P. J., Matsangidou, M., Babiker, N. T., Fehily, O., Thompson, A., et al. (2021). Designing effective virtual reality environments for pain management in burn-injured patients. *Virtual Reality* 27, 201–215. doi:10.1007/s10055-021-00552-z

Picinali, L., Katz, B. F., Geronazzo, M., Majdak, P., Reyes-Lecuona, A., and Vinciarelli, A. (2022). The sonicom project: Artificial intelligence-driven immersive audio, from personalization to modeling [applications corner]. *IEEE Signal Processing Magazine* 39, 85–88. doi:10.1109/msp.2022.3182929

Pons, P., Navas-Medrano, S., and Soler-Dominguez, J. L. (2022). Extended reality for mental health: Current trends and future challenges. *Frontiers in Computer Science* 4. doi:10.3389/fcomp.2022.1034307

Skarbez, R., Smith, M., and Whitton, M. C. (2021). Revisiting milgram and kishino's Reality-Virtuality continuum. *Front. Virtual Real.* 2

Soler-Dominguez, J. L., Navas-Medrano, S., and Pons, P. (2024). Arcadia: A gamified mixed reality system for emotional regulation and self-compassion. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), CHI '24. doi:10.1145/3613904.3642123

Speicher, M., Hall, B. D., and Nebeling, M. (2019). What is mixed reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: ACM)

Tang, Y. M., Kuo, W. T., and Lee, C. (2023). Real-time mixed reality (mr) and artificial intelligence (ai) object recognition integration for digital twin in industry 4.0. *Internet of Things* 23, 100753. doi:10.1016/j.iot.2023.100753

Toni, E., Toni, E., Fereidooni, M., and Ayatollahi, H. (2024). Acceptance and use of extended reality in surgical training: an umbrella review. *Systematic Reviews* 13. doi:10.1186/s13643-024-02723-w

Villada Castillo, J. F., Montoya Vega, M. F., Muñoz Cardona, J. E., Lopez, D., Quiñones, L., Henao Gallo, O. A., et al. (2024). Design of virtual reality exergames for upper limb stroke rehabilitation following iterative design methods: Usability study. *JMIR Serious Games* 12, e48900. doi:10.2196/48900

Wagener, N. and Niess, J. (2021). Reflecting on emotions within VR mood worlds. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers* (New York, NY, USA: ACM)

Wang, Y., Duan, B., Chen, X., Song, Y., and Liu, X. (2025). The application of metaverse in mental health. *Frontiers in Public Health* 13. doi:10.3389/fpubh.2025.1463494

WHO (2022). *World mental health report: Transforming mental health for all*

Wong, S. M., Chen, C.-W., Pan, T.-Y., Chu, H.-K., and Hu, M.-C. (2022). GetWild: A VR editing system with AI-generated 3D object and terrain. In *Proceedings of the 30th ACM International Conference on Multimedia* (New York, NY, USA: ACM)

Yu, M., Bai, Y., and Li, Y. (2023). Emo-regulator: An emotion-regulation training system fusing virtual reality and eeg-based neurofeedback. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine amp; Biology Society (EMBC)* (IEEE). doi:10.1109/embc40787.2023.10340975

Zweifach, S. M. and Triola, M. M. (2019). Extended reality in medical education: Driving adoption through provider-centered design. *Digital Biomarkers* 3, 14–21. doi:10.1159/000498923

Kiin Platform: Using Immersive Training to Address Bias and support Behaviour Change

Bernhard Spanlang*, Elena Álvarez de la Campa Crespo, Manuel Jesús Pérez Zurera, Roger Puig Ganza, Carlos Aguilera, Jesús Mota, Jorge Juan Ñíguez Fernández

[1]Virtual Bodyworks S.L. (also known as Kiin.tech), Barcelona, Spain

*** Correspondence:**
Bernhard Spanlang
bernhard@kiin.tech

Keywords: extended reality, shared XR, embodiment, human behaviour, RCT, AI

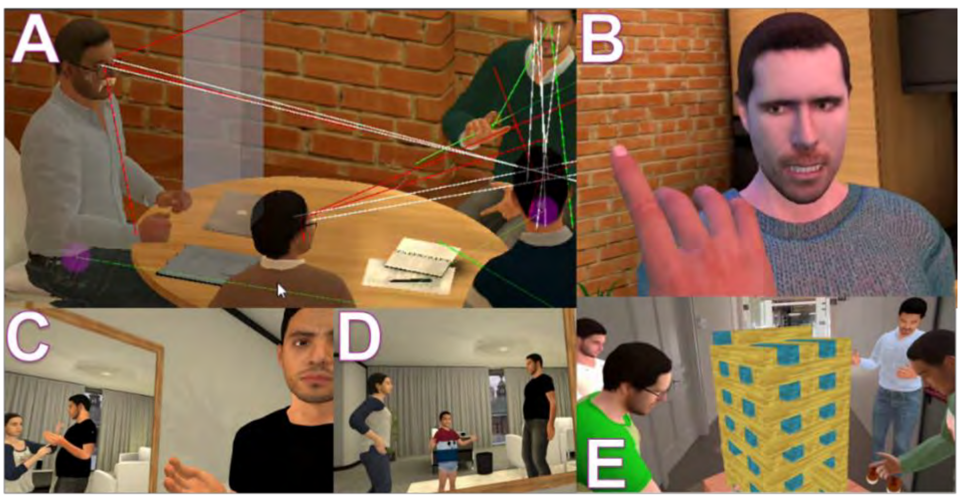

Figure 1: A: look-at behaviour replay in a work place study, B: participant's eyes follow other's hand by eye tracking, C:Offender embodied as victim: D:as child, E:MR shared Jenga game

Abstract

Studying human behaviour—and how it changes—across large populations over extended periods of time in immersive extended reality (XR) environments requires sophisticated and flexible research platforms. We present the Kiin System, a platform designed to support the creation and delivery of XR experiences for both training and behaviour change research. The system enables dynamic modification of avatar and environment appearances, recording and replay of participants' voices and movements, and collection of subjective feedback through an integrated XR questionnaire system. It offers a comprehensive solution for designing, deploying, and managing randomized controlled trials (RCT) in both single-user and shared XR experiences. Key features include life-size full-body embodiment, natural interaction with diverse avatars and the environment, a Python client for runtime control and external systems (e.g., simulation and AI), low-latency conversational AI agents, AI-based environment adaptation, modules for audio-haptic sensory support and hearing simulation. It integrates secure cloud services to anonymously track participants over multiple sessions and

integrates with session management. It enables corporations and researchers to conduct complex, longitudinal RCTs as well as immersive training at scale.

1. Introduction and Related Work

Before the cost of XR devices dramatically decreased, running scientific embodiment studies required dedicated rooms and equipment such as full body motion capture systems quite heavy HMDs, powerful PCs, dedicated head tracking, physiological (e.g. ECG and EEG) measure systems, etc. (Spanlang et al. 2014) Existing VR tools range from general game engines (Unity, Unreal) requiring significant development resources to specialized platforms (e.g., Vizard[1]) or Social VR platforms such as, RecRoom, VRChat, Mozilla Hubs, compared in (Liu and Steed 2021) or more recent Horizon Worlds, Roblox[2] etc. Academic implementations provide base systems but still require custom development Oliva et al. 2022,Steed et al. 2022. The Kiin platform differentiates itself by integrating dynamic embodiment, low-latency AI with environmental control, multi-modal tracking, a JSON flow description and a Python client within a single architecture optimized for Meta Quest devices. Studies on body illusions (Slater et al. 2010),(de la Peña et al. 2010), etc. highlight the importance of embodiment, an area the Kiin system excels with dynamic change and even gradual reshaping capabilities. Kiin's AI integration leverages low-latency end-to-end multimodal speech to speech models and allows for AI-driven environment modification, surpassing typical conversational agent implementations. It aims to simplify longitudinal studies such as for example (Kishore et al. 2021) where behaviour change was shown in police officers after they experienced an interrogation from the perspective of a victim.

2. System Architecture and Implementation

A simplified diagram of the Kiin system architecture is shown in Figure 2. The core subsystems are described here:

- **Database Backend:** The system leverages Firebase for secure user account management as well as storing content and user recordings and questionnaires.
- **Embodiment:** We leverage OpenXR and Meta's XR SDKs[3] to retrieve tracking information of (HMD, controllers, hands, eye/face where available) which is mapped to an internal tracking avatar calibrated to the size of the participant's real body. The Unity Mecanim muscle data of that tracking avatar is transferred in real-time to the embodied avatar enabling avatar independent embodiment as described in (Spanlang et al. 2013). Avatars can appear roughly like the participants' real body(Alvarez de la Campa Crespo and Spanlang 2021) based on a photograph, or can be of different age, gender, race, size, social status, etc. Embodied avatars can be changed at any time and exploiting full body blend shapes the system even offers gradual avatar shape modification for example morph an obese looking avatar into one that appears healthier.
- **Multi-modal Data Acquisition:** Synchronized movement (body, hands, eyes/face data) and audio recordings, can be replayed in the same session or uploaded for later replay or analysis.

1 https://www.worldviz.com/vizard-virtual-reality-software

2 https://www.spatial.io/, https://www.roblox.com/, https://github.com/Hubs-Foundation/hubs, https://horizon.meta.com/

3 https://www.khronos.org/openxr/, https://developers.meta.com/horizon/develop

Pre, Post and Inside XR participant questionnaire responses can be stored in the same database structure to enable correlations between conscious and unconscious behaviour.

- **Command System:** A defined set of commands allows for control over the VR experience. These commands can, for example, change embodied appearance, modify environmental properties (light, materials, etc), play sounds (spatialized audio, custom head related transfer functions (HRTFs) and hearing aids/loss simulations(HA/HL)), insert/control AI-driven avatars for low latency interaction, etc. This command structure provides an action space suitable for reinforcement learning (RL) agents. Alternatively, commands can be triggered locally on the device via our JSON-based flow defined per experience for rapid development.

- **AI Integration:** Connects multi modal low-latency (~200ms) speech-to-speech services(e.g. GPT-4o Realtime,Google Gemini Live, Qwen's Omni models, etc). The AI is prompted for custom behaviour in the experience flow JSON and can also trigger commands within its response to dynamically modify the environment by the client-side tools calling API[4].

- **Python Client Interface:** Our Python package allows for more versatile interaction with the shared experience, receiving events, querying data and send commands to the shared XR experiences, offering access to the vast package libraries available in python such as NumPy, SciPy, MatPlotLib, PyTorch, Huggingface, etc.

- **Sensory Modules:** Include audio-to-haptic mapping and hearing loss/aid simulation[5] which enable studies for example to quantify the improvement of comprehension when audio is enhanced with haptics similar to (Ciesla et al. 2021) but with avatars on a standalone XR device with the voice coil actuators of Meta Touch Plus controllers.

[4] platform.openai.com/docs/models/, platform.openai.com/docs/guides/realtime

[5] github.com/3DTune-In/3dti_AudioToolkit_UnityWrapper, developers.meta.com/horizon/resources/haptics-studio/

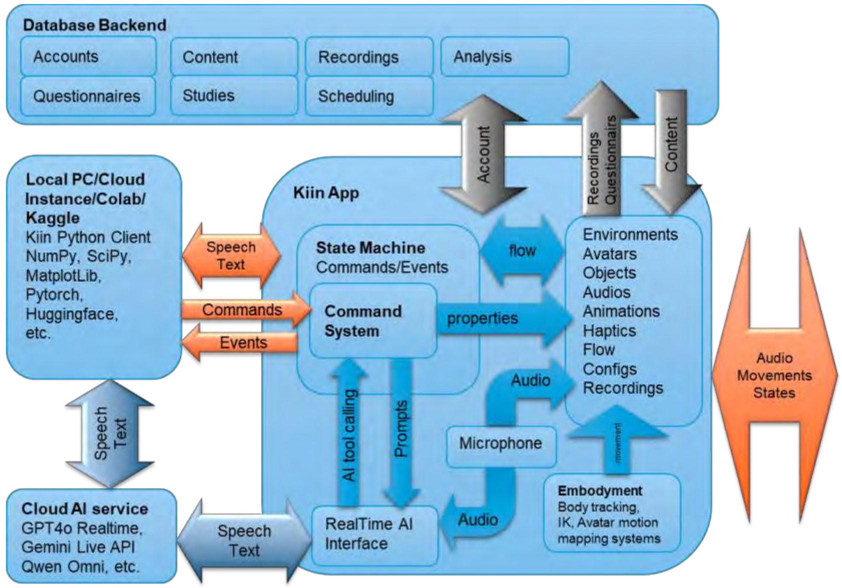

Figure 2: Data Flow between Kiin App, Kiin Python Client, Database Backend and Cloud AI Services. The arrow on the right illustrates networked capablities conncting multiple Kiin Apps

Implementation: The Kiin App is built with Unity[6] using mainly C#. The backend database uses Google Firebase[7] securely managing user accounts, experiences, recordings and session scheduling for studies, enforcing in-transit and at-rest encryption. For shared experiences the Photon engine (currently PUN2) is used to exchange voice and body, hands, face and eye movements of participants. It also connects the Kiin Python client to shared experiences, to send commands and receive events or states of the experience (e.g. the position, orientation and muscle values of each avatar). It supports audio channels for example for an AI agent to listen and respond by speech usually lip-synced through a 3D audio-source positioned at an avatar's mouth. Experiences consist of modular components: environments, avatars, objects (Unity assetbundles), and animations—often recorded from participants or actors.

At the core of the app is the command system that enables it to spawn and move avatars, objects, texts, change aspects of the visual and acoustic environment, play sounds etc. The command system can get triggered by the app's internal state machine, by a Kiin Python Client, or an AI service[8]. Our main target platforms are Meta Horizon OS and potentially Android XR when available. Currently we focus on Meta Quest devices 2/Pro/3/3S/ and the App is deployed via Meta for Work or Horizon Store[9] (upon invitation). The system offers the possibility to collect, synchronized, multi-modal data (gaze, movement, ECG, EEG, voice, subjective reports) for analysis in standard tools (Python, R, MATLAB, etc.) to support diverse analyses, including cross-modal investigations correlating different behavioural streams.

[6] https://unity.com/ currently version 2022.3, and migration to Unity 6 is planned

[7] https://firebase.google.com/

[8] For example: platform.openai.com/docs/guides/realtime, https://ai.google.dev/gemini-api/docs/live

[9] https://work.meta.com/, https://horizon.meta.com/

3. Use Cases

The Kiin system is being used in academic research for fundamental human behaviour studies, in the areas of psychological and physiological therapy as well as for corporate staff training and research:

- **Domestic Violence Rehab** The system is being used by the Generalitat of Catalonia to augment the rehabilitation of domestic violence offenders. We offer various experiences one of them underlying this use case was published in (Seinfeld et al. 2018).

- **Eating Disorder Café scenario:** With Bristol University we are designing a study for the treatment of eating disorders, which includes the eye tracking capabilities of the system.

- **Self-Conversation Therapy:** The system has been used in various self-conversation studies (Osimo et al. 2015) to help people with depression and eating disorders. The dynamic body resizing functionality was used in (Anastasiadou et al. 2024) to improve self-esteem in participants with issues of severe obesity. In (Zisquit et al. 2025) it is used for enhanced interpersonal emotion regulation.

- **EEG Co-presence Study:** With Maastricht University (MU) we have designed a study that exploits the system to measure effects of being with another person's avatar or a computer-controlled one, leveraging multi-user networking and stimuli presentation and milliseconds EEG synchronization.

- **Climate Change Game:** With Warsaw University (WU) we co-developed a multi user tragedy of the commons game experience with the aim to raise awareness of the impact of participants' actions on the environment in a collaborative and competitive setting based on an economics model developed by WU. Demoed at IEEE VR 2025 (Lewandowska et al. 2025)

- **Audio-Haptic Study:** Setup for a study on audio haptic enhancement to evaluate improvement of hearing loss to understanding.

- **RL-Driven Protest Simulation:** In collaboration with the Arctic University of Norway we migrated historical protest scenes to the system to then use the Python client for real-time RL agent integration with the objective to have the participants join the protest in a peaceful manner.

- **LLM Agent Interaction:** Directly use integrated low-latency AI agents and environment modification features.

4. Discussion & Conclusion

The Kiin System offers an integrated, powerful platform for research and training in XR on accessible hardware. It is highly integrated, offering dynamic embodiment, low-latency AI with environmental control, Python extensibility. Ethical considerations regarding embodiment effects, AI influence, and data privacy are paramount. Future work includes increasing concurrency, broadening hardware support, adding built-in analysis tools, simplifying authoring interfaces and self-deployments. The system significantly lowers the barrier for complex longitudinal immersive studies by integrating key functionalities it opens up new avenues for understanding and improving human behaviour. Its use cases ranging from governments over corporate training, as well as research institutions demonstrates its value for immediate and the immense potential for future impact.

5. Conflict of Interest

BS is a co-founder of Virtual Bodyworks (also known as Kiin.tech)

6. Author Contributions

BS coordinated the development of the system and wrote the paper. EC, MZ, RG, CA, JM and JF made major contributions to the system, avatars, experiences and ongoing and completed studies.

7. Funding

The development leading to the presented system was partially funded by the European Union Horizon 2020 research and innovation programme under grant ID. 101017884 – GuestXR as well as 101057429-environMental the 951930-Socrates projects.

8. Acknowledgments

We extend our sincere gratitude to Charlie Pearmund (CEO of Kiin), co-founders Mel Slater and Mavi Sánchez-Vives, the board members of Virtual Bodyworks, and our clients, whose invaluable insights, ongoing collaboration, and trust were instrumental in shaping both the functionality and implementation of the Kiin Platform.

9. References

Alvarez de la Campa Crespo, Elena, and Bernhard Spanlang. 2021. “A Deep CNN Model for Inferring 3D Human Body Shapes Using Front and Side Images.” In . https://doi.org/10.15221/21.32.

Anastasiadou, Dimitra, Pol Herrero, Paula Garcia-Royo, Julia Vázquez-De Sebastián, Mel Slater, Bernhard Spanlang, Elena de la Campa, et al. 2024. “Assessing the Clinical Efficacy of a Virtual Reality Tool for the Treatment of Obesity: Randomized Controlled Trial.” *Journal of Medical Internet Research* 26 (1): e51558. https://doi.org/10.2196/51558.

Ciesla, Katarzyna, T Wolak, A Lorens, H Skarżyński, and A Amedi. 2021. “Speech-to-Touch Sensory Substitution: A 10-Decibel Improvement in Speech-in-Noise Understanding after a Short Training.” *Research Square*, April. https://doi.org/10.21203/rs.3.rs-429202/v1.

Kishore, S., B. Spanlang, G. Iruretagoyena, S. Halan, D. Szostak, and M. Slater. 2021. “A Virtual Reality Embodiment Technique to Enhance Helping Behavior of Police toward a Victim of Police Racial Aggression.” *Presence: Teleoperators and Virtual Environments* 28. https://doi.org/10.1162/PRES_a_00339.

la Peña, N. de, P. Weil, J. Llobera, E. Giannopoulos, A. Pomés, B. Spanlang, D. Friedman, M.V. Sanchez-Vives, and M. Slater. 2010. “Immersive Journalism: Immersive Virtual Reality for the First-Person Experience of News.” *Presence: Teleoperators and Virtual Environments* 19 (4). https://doi.org/10.1162/PRES_a_00005.

Lewandowska, Weronika M, Bernhard Spanlang, Andrzej Nowak, Szymon Talaga, Manuel Jesus Perez Zurera, Roger Puig Ganza, Anna Boros, Wieslaw Bartkowski, Karolina Ziembowicz, and Mikolaj Biesaga. 2025. " Cooperation for Environmental Sustainability – a Common Pool Resource Virtual Experience ." In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* , 1542–43. Los Alamitos, CA, USA: IEEE Computer Society. https://doi.org/10.1109/VRW66409.2025.00416.

Oliva, Ramon, Alejandro Beacco, Xavi Navarro, and Mel Slater. 2022. "QuickVR: A Standard Library for Virtual Embodiment in Unity." *Frontiers in Virtual Reality* 3. https://doi.org/10.3389/frvir.2022.937191.

Osimo, Sofia Adelaide, Rodrigo Pizarro, Bernhard Spanlang, and Mel Slater. 2015. "Conversations between Self and Self as Sigmund Freud—A Virtual Body Ownership Paradigm for Self Counselling." *Nature Scientific Reports* 5.

Seinfeld, S., J. Arroyo-Palacios, G. Iruretagoyena, R. Hortensius, L. E. Zapata, D. Borland, B. de Gelder, M. Slater, and M. V. Sanchez-Vives. 2018. "Offenders Become the Victim in Virtual Reality: Impact of Changing Perspective in Domestic Violence." *Scientific Reports* 8 (1). https://doi.org/10.1038/s41598-018-19987-7.

Slater, M., B. Spanlang, M.V. Sanchez-Vives, and O. Blanke. 2010. "First Person Experience of Body Transfer in Virtual Reality." *PLoS ONE* 5 (5). https://doi.org/10.1371/journal.pone.0010564.

Spanlang, Bernhard, Xavi Navarro, Jean-Marie Normand, Sameer Kishore, Rodrigo Pizarro, and Mel Slater. 2013. "Real Time Whole Body Motion Mapping for Avatars and Robots." In *19th ACM Symposium on Virtual Reality Software and Technology (VRST)*. Singapore.

Spanlang, Bernhard, Jean-Marie Normand, David Borland, Konstantina Kilteni, Elias Giannopoulos, Ausiàs Pomés, Mar González-Franco, et al. 2014. "How to Build an Embodiment Lab: Achieving Body Representation Illusions in Virtual Reality." *Frontiers in Robotics and AI* 1 (9).

Steed, Anthony, Lisa Izzouzi, Klara Brandstätter, Sebastian Friston, Benjamin James Congdon, Otto Olkkonen, Daniele Giunchi, Nels Numan, and David Swapp. 2022. "Ubiq-Exp: A Toolkit to Build and Run Remote and Distributed Mixed Reality Experiments." *Frontiers in Virtual Reality* 3:912078. https://doi.org/10.3389/frvir.2022.912078.

Zisquit, Moreah, Anat B Klomek, Alon Shoa, Mel Slater, Bernhard Spanlang, and Doron Friedman. 2025. "Intrapersonal Emotion Regulation Training in Virtual Reality: Embodying Self-Talk." *Virtual Reality* 29:61. https://doi.org/10.1007/s10055-025-01131-2.

Poster Abstracts 2A: Room “Rector Ramón Martín Mateo”

Building Sustainably on Mars: A Pedagogical Experience with Artificial Intelligence and Immersive Web Environments

Bárbara Cleto[1*]

[1]Escola Secundária José Régio, Portugal

*** Correspondence:**
Bárbara Cleto
553222@prof.esc-joseregio.pt (barbara.cleto@gmail.com)

Keywords: Artificial Intelligence, Sustainability, Immersive Web Environments, Vocational Education, Living on Mars.

Abstract

This summary presents the development of a project by tenth year vocational school students as part of the Artificial Intelligence and Sustainability (IAS) project, promoted by the University of Minho (UM). The IAS project aims to stimulate the pre-university public towards scientific research, creating an environment for sharing ideas, experiences and activities between UM, primary and secondary schools, local authorities and R&D companies in this subject area[1,2].

Secondary school students can interact with UM lecturers and researchers[1,2], who propose challenges inspired by the theme of Artificial Intelligence and Sustainability[1]. The project culminates with the public presentation of the work, in the form of oral communication or poster, at a conference held at UM.

The 10th year students of the Multimedia and Electronics, Automation and Computers vocational courses chose, from a list of suggested challenges, to work on the theme "Is it possible to build a house on Mars? How would you build it?"

The students, divided into groups of up to five, began the research phase on the planet Mars, exploring the websites of NASA and the European Space Agency. This research was complemented using generative artificial intelligence (AI) tools, such as ChatGPT, which helped them interpret and cross-reference the information gathered.

The research focused on several key questions such as: how to obtain water?', 'how to produce energy?', 'how to ensure that you can breathe on Mars?'. Also, on issues related to thermal insulation, protection against radiation and dust storms, water reuse and sanitation systems and the selection of materials suitable for construction[3] in an environment like Mars.

[1] https://www.uminho.pt/PT/siga-a-uminho/Paginas/Detalhe-do-evento.aspx?Codigo=61817, accessed on 18/3/2025

[2] https://www.uminho.pt/PT/siga-a-uminho/Paginas/Detalhe-do-evento.aspx?Codigo=63720, accessed on 18/3/2025

[3] https://noticias.up.pt/fcup/humanos-em-marte-trabalho-pioneiro-da-fcup-analisou-exposicao-a-radiacao/, accessed on 18/3/2025

Some of the solutions they presented for building sustainable and cheap houses on Mars include:

- Use of resistant, light and durable materials, since Mars' gravity is lower than Earth's;

- Solar energy production, using solar panels and storage systems to power electric heaters;

- Water collection by extracting and purifying the ice present in the Martian subsoil;

- Underground construction, using 3D printers

- Efficient recycling and energy management systems.

They began by sketching out their ideas on paper, using pencils and markers. Then, with the support of AI tools (text to image and text to 3D models) they ‘visualized’ their proposals. This phase of constant iteration and interaction with the AI tools made it possible to generate the images included in the posters. At the same time, the Multimedia students created the graphic identity of their projects, developing logos and slogans for their companies’ building houses on Mars. Some groups, especially those on the Multimedia course, chose to create their own illustrations. The posters were developed using Photoshop or platforms such as Canva.

This project enabled the students to deepen their knowledge of science, the environment and technology and allowed them to develop digital skills, where the use of AI was relevant as a means of creative production and a huge help for the research they carried out.

It should also be noted that this work will serve as a conceptual and visual basis for the next phase of the project, the development of immersive web environments, where ideas for homes on Mars will be represented in online spaces.

For construction and exploration, students can use different tools and platforms, depending on the objectives and degree of complexity desired:

Using 3D modelling tools such as Blender, Tinkercad or SketchUp, or text-to-3D AI tools, students can build idealized houses.

Using platforms such as Anitya, Spatial, FrameVR or CoSpaces Edu, which allow these models to be inserted into immersive environments that can be explored in virtual reality and via the browser.

If a greater degree of complexity and realism is required, development engines such as Unity or Unreal Engine can be used.

These approaches promote engaging, collaborative educational experiences in line with the metaverse concept, reinforcing the link between technology, active learning, creativity and sustainability

This transposition from paper to digital space, where students create immersive educational web environments based on AI and sustainability reinforces active learning, stimulates creativity and technological literacy.

Acknowledgments

Thanks to the students who developed the projects.

Ethnography in the Metaverse

Majd Bitar *, **Jean-Philippe Rivière, Rebecca Fribourg,**
Jean-Marie Normand, Yannick Prié
LS2N, Nantes Université, École Centrale Nantes, CNRS, Nantes, France
Correspondence*:
majd.bitar@univ-nantes.fr

ABSTRACT

As Metaverse platforms evolve, they offer new opportunities to study social behavior. We explore some aspects of conducting ethnographic studies in the Metaverse, highlighting both potential and limitations, while proposing key considerations for future research.

Keywords: Metaverse, Ethnography, Social VR, Virtual Reality, HCI

1 INTRODUCTION

The study of social behavior and user experience in virtual environments have always been of interest in various disciplinary fields. With the recent technological advancements in XR, immersive virtual social worlds are emerging giving a new shape to what is defined as the *Metaverse*. These platforms open up new possibilities for social and behavioral studies inside virtual environments. Although many anthropological methods can be utilized for such studies, potential improvements and discussion points come up when considering details regarding these methods and how we can adapt them in the Metaverse.
In this work we look upon ethnographic methods utilized to document behaviors directly "in the wild", where it happens, as opposed for instance to observe them in a controlled lab experiments. Conducting ethnography in the Metaverse could be useful to understand the social and cultural behavior of its users. However, various obstacles are to be addressed to enable such studies. In this paper, we discuss the conduct of ethnographic studies inside social VR platforms, shedding light on the promising potential while highlighting possible emerging conflicts.

2 RELATED WORK

The term *Metaverse* was first introduced in 1992 by Neal Stephenson in the novel "Snow Crash", to describe a 3D space where people are represented by avatars and can interact with each other through these avatars. Different definitions of the Metaverse have emerged in relation to both the advancements in immersive technology and the development of commercial actors. In our work, we consider the Metaverse as mainly based on eXtended Reality (XR, regrouping AR and VR technologies) and avatars, in persistent online 3D worlds, as proposed in a recent review Park and Kim (2022).

2.1 Ethnography in virtual worlds

Anthropological work in virtual worlds can be traced back to the late 1990s, when definitions of *Virtual Ethnography* (Hine, 2000), and *Netnography* (Kozinets, 1998) were introduced, both defining methods to study and observe the behavior of users, communities and cultures present in virtual spaces. These methods were then utilized by anthropologists to study social behavior on various platforms that allowed users to interact with each others and with the world using their own avatar, as well as to possess their own virtual spaces (e.g., a virtual environment that represented their house) and objects. *World of Warcraft* (Ducheneaut and Moore, 2004; Ducheneaut et al., 2006; Nardi, 2010) or *Second Life* (Boellstorff, 2015) are examples of such platforms. Ethnographic studies relied on the observation of members of these virtual worlds, looking into the their daily activities, their interactions, their communication. Ethnographers joined the users, participated with them in what they did in order to better understand their culture.

2.2 Ethnography in the Metaverse

In its current state, the Metaverse can be defined as a set of persistent platforms such as *VRChat, RecRooms, Meta Horizon Worlds, BanterVR* and many others. These platforms allow users to communicate, interact, and engage in various activities while embodying avatars of their choice. They are available for free to a wide range of users of different background, age, gender or culture, and many integrate accessibility features that allow the participation of users with different impairments, enabling them to interact with others with minimum obstacles. They also allow their users to build their own worlds and allow others to join in, which serves as a motivation to create worlds related to things they are passionate about, attracting other users that share the same passion. As a consequence, various various communities have daily activities on social VR platforms.

Users in the Metaverse are engaging in collaborative and social interactions that both differ from real-world interactions, and have far more richness and depth than what could be observed within previous platforms such as second life. For instance, immersive interactions between avatars and with the environment are embodied, potentially leading to new and complex social behaviors. Studying these behaviors is of interest for social and human-computer interaction researchers, who could clearly get a better understanding of how these users and groups function and interact by conducting their studies "from within", like for any human group and activities. Yet, in opposition to what happened for non-immersive platforms (Netnography), ethnographic studies in the Metaverse are still scarce, even if some early proposals can be found. For instance, (Maloney and Freeman, 2020) explored the social activities that users found meaningful and valuable in the social VR; (Williamson et al., 2021) focused on studying proxemics during social interactions inside Mozilla Hub, and (Kolesnichenko et al., 2019) conducted interviews inside social VR platforms with designers and professionals of the Metaverse, identifying affordances that enhanced embodiment of the users.

3 CARRYING-OUT ETHNOGRAPHY IN THE METAVERSE

Ethnography in the Metaverse is still limited, and in this section we discuss several requirements associated to data collection and legal issues. An important remark is that despite the fact that many current social VR platforms also allow their users to join with traditional 2D screens and input methods such as mouse and keyboard, our focus is on *ethnographers being in immersive VR*, who can study VR and non-VR users alike.

3.1 Data collection

First, let us consider data collection, which is critical in ethnographic studies, and can take several forms and aims, such as general observations about the surrounding environment or population, descriptions of specific events, experiences recollected by participants, personal notes of the ethnographer, etc.

3.1.1 Recording Data

Recording can serve to document something a participant said during an interview, a gesture they did, the way an action was carried out, a routine of a certain community, etc. There is a lot of flexibility with what to record and how to record it. For instance, the ethnographer can record their first person point-of-view (POV), but also place several virtual cameras or microphones inside the environment to capture other POVs. We can even imagine drone-like cameras following participants as they move around. However, questions remain about how to analyze these recordings. For example, how to find specific moments of interest for the study, and how tools can be built to help the researchers. Such tools can be available on Metaverse platforms that facilitate recording data and synchronizing several viewpoints or data sources to analyze it more easily later on. These artifacts can be controlled by the platform itself, but also by the researchers. Most importantly, these tools should be compatible across platforms to provide a unified method of acquiring the data. Another type of data that could be recorded are logs of the actions of chosen participants, or the state of all, or part of, the environment at certain moments, not as video but as structured data. The use of virtual logging tools to capture events, interactions and the state of the environment will provide more information over the course of the study. This could allow to identify the action or activity done by the participants, and will provide more information about the environment surrounding the participant, such as count and position other participants or virtual objects.

3.1.2 Taking notes

Note taking is a central component of ethnographic studies, and immersive technologies may enable various ways to take notes, some of which may improve on more traditional methods but still have their limitations. Voice recording can be used to verbally record observations or interpretations in real-time. Taking written notes may be difficult in VR, as virtual writing is complicated if one does not have access to a real keyboard, or to a dedicated pen and a surface to write on in reality. Selecting among predefined observation categories may be something that is more easy to do in VR. A particularity of VR is that notes could be placed around the virtual environment to connect the observations and the virtual spaces where they occurred, allowing to revisit them later. Notes could also be combined with other recordings. For example, having the notes synchronized based on timestamps with a video recording would facilitates revisiting specific moments and recalling observations from it. Moreover, collaborative notes are possible for studies that involve several ethnographers, which could be synchronized between the researchers, or passed virtually between for easier collaboration.

3.2 Legal issues

Second, some legal considerations have to be taken into account.

3.2.1 User data privacy and applicable laws

Since ethnographic studies in the Metaverse may involve users from different countries, each subjected to varying data protection laws, then a question quickly arises: which legal framework should apply? Should it be based on the researcher's location, the user's, the platform's, or even that of the virtual world itself? Would there exist something as French soil in the Metaverse? Moreover, every Metaverse platform has its own Terms of Service and Privacy Policy, but it is now unclear if they define what kind of data can be collected by researchers, how user behavior can be observed, and whether research activities like interviews, recordings, or behavioral tracking are allowed.

3.2.2 Participant consent

Ethical and legal reasons make it mandatory to get the consent of the participants in a study, but another issue is the way the consent can be obtained. If we imagine a scenario where the experimenter has to ask for participants' consent to collect data in the Metaverse —e.g., to record an interview— there are many possibilities regarding how to request and obtain it. For example, the experimenter can orally communicate the necessary information and ask for the participant's consent. The experimenter could also have access to a virtual object such as a screen or a tablet that displays the consent form for the participant to read. Obtaining effective consent from the user can also take various forms. A simple oral consent could be video recorded, or virtual objects could be used, for instance a virtual pen to sign a form, or an object that is interacted with to confirm and record a signature. All these methods may depend not only on the platform but also on ethical obligations related to research and laws but also on specific virtual environments within a given platform, introducing variability in the way consent is obtained from different participants to the same study.

4 CONCLUSION

Ethnography in virtual worlds has always existed in some form, but the emergence of the Metaverse opens new possibilities for studying social behavior and emerging cultures within social immersive virtual environments. In this paper, we proposed several key topics related to aspects of ethnographic studies, that are up for discussion.

Knowing these challenges, we aim to use ethnography to deepen our understanding of how users navigate the Metaverse in groups. However, we are continuously evaluating our methods from the perspective of regulations and policies when it comes to users' data. We aim to observe groups explore environments and interact with them, during which we will try to apply some of the approaches we discussed above, especially to what data we will collect and how.

ACKNOWLEDGMENTS

This work was partly supported by a French government grant managed by the Agence Nationale de la Recherche as part of the France 2030 program, reference ANR-22-EXEN-0002 (PEPR eNSEMBLE / CATS).

REFERENCES

Boellstorff, T. (2015). Coming of Age in Second Life: An Anthropologist Explores the Virtually Human. In *Coming of Age in Second Life* (Princeton University Press)

Ducheneaut, N. and Moore, R. J. (2004). The social side of gaming: A study of interaction patterns in a massively multiplayer online game. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (New York, NY, USA: Association for Computing Machinery), CSCW '04, 360–369. doi:10.1145/1031607.1031667

Ducheneaut, N., Yee, N., Nickell, E., and Moore, R. J. (2006). "Alone together?": Exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), CHI '06, 407–416. doi:10.1145/1124772.1124834

Hine, C. (2000). *Virtual Ethnography* (SAGE Publications Ltd). doi:10.4135/9780857020277

Kolesnichenko, A., McVeigh-Schultz, J., and Isbister, K. (2019). Understanding Emerging Design Practices for Avatar Systems in the Commercial Social VR Ecology. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (New York, NY, USA: Association for Computing Machinery), DIS '19, 241–252. doi:10.1145/3322276.3322352

Kozinets, R. V. (1998). *On Netnography: Initial Reflections on Consumer Research Investigations of Cyberculture.*

Maloney, D. and Freeman, G. (2020). Falling Asleep Together: What Makes Activities in Social Virtual Reality Meaningful to Users. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (New York, NY, USA: Association for Computing Machinery), CHI PLAY '20, 510–521. doi:10.1145/3410404.3414266

Nardi, B. (2010). *My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft* (University of Michigan Press). doi:10.3998/toi.8008655.0001.001

Park, S.-M. and Kim, Y.-G. (2022). A Metaverse: Taxonomy, Components, Applications, and Open Challenges. *IEEE Access* 10, 4209–4251. doi:10.1109/ACCESS.2021.3140175

Williamson, J., Li, J., Vinayagamoorthy, V., Shamma, D. A., and Cesar, P. (2021). Proxemics and Social Interactions in an Instrumented Virtual Reality Workshop. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA: Association for Computing Machinery), CHI '21, 1–13. doi:10.1145/3411764.3445729

Development and Evaluation of a Virtual Reality Training System for Precision Agriculture

Krzysztof Nowakowski *, **Łukasz Nawrocki, Piotr Skrzypczyński and Dominik Belter**

Poznań University of Technology, Institute of Robotics and Machine Intelligence, Poznań, Poland

Correspondence*:
Krzysztof Nowakowski
krzysztof.nowakowski@student.put.poznan.pl

ABSTRACT

This work introduces a Virtual Reality (VR) training system for precision agriculture, offering a cost-effective and flexible alternative to traditional training methods. The system integrates NeRF-based reality capture (via Luma AI) with Unreal Engine and Oculus Quest to create immersive training scenarios. User evaluations indicate high engagement and usability, although some participants experienced motion sickness during mobile VR tasks. These results underscore the potential of VR in agricultural training, while highlighting the importance of motion-adaptive environments for an optimal user experience.

Keywords: Virtual Reality, Training System, Precision Agriculture, Unreal Engine, Neural Radiance Fields, User Evaluation

1 INTRODUCTION

The agricultural sector is undergoing rapid technological transformation, incorporating automation and robotics to enhance productivity. However, the deployment of these technologies is hindered by a lack of trained personnel and the high costs associated with traditional training methods. Virtual Reality (VR) has emerged as a promising solution, enabling immersive, flexible, and cost-effective training experiences.

This paper presents a VR-based training system for precision agriculture that leverages synthetic and real-world models, integrated through Neural Radiance Fields (NeRF) Mildenhall et al. (2021) and Unreal Engine. Our primary contribution lies in demonstrating the feasibility and effectiveness of VR in agricultural training, addressing challenges related to usability and engagement.

2 IMPLEMENTATION AND RESULTS

2.1 System Architecture

The architecture of the proposed training system is presented in Fig. 1. The core component utilizes LUMA AI's NeRF technology to build 3D models of real objects (Fig. 2a). The models of agricultural machines are taken directly from manufacturer databases. Interaction within the VR environment was enabled through hand-tracking and controller-based manipulation.

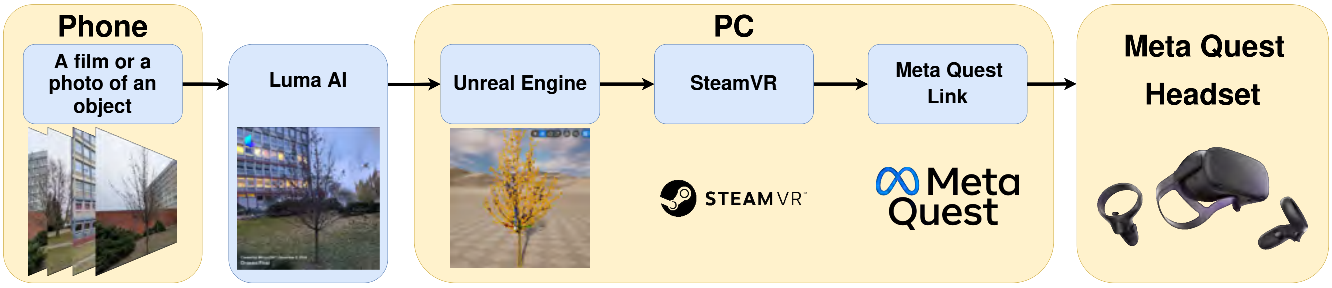

Figure 1. Block diagram of the proposed system.

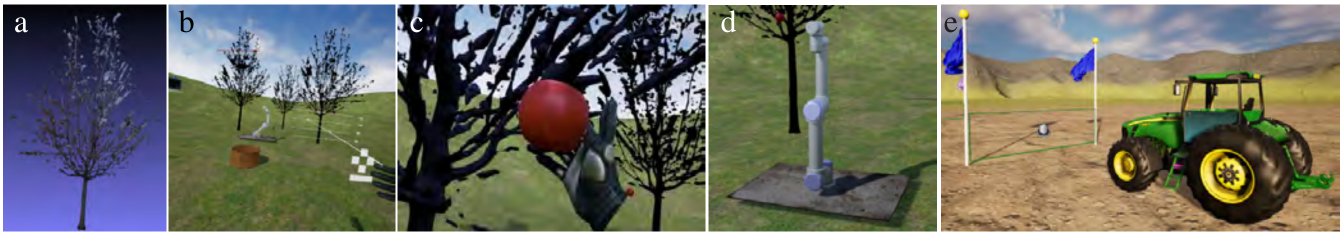

Figure 2. Example model obtained by applying NeRF-based 3D reconstruction (a) used in the VR training scenario (b,c). Model of Universal Robot UR5e (d) and a tractor (e) used in scenarios II and III.

2.2 Training Scenarios

The system was verified in three sequential training scenarios:

1. Introduction to VR where users learn basic environment navigation (Fig. 2b) and object manipulation (Fig. 2c).
2. Robotics training where users operate a virtual robotic arm (Universal Robot UR5) to perform simulated farming tasks, such as fruit harvesting.
3. Operating Tractor where users drive a tractor in a VR environment, assessing the impact of prolonged VR exposure on comfort and motion sickness.

2.3 User Evaluation

To assess the effectiveness of the system, participants completed a questionnaire adapted from the Virtual Reality Neuroscience Questionnaire (VRNQ, Kourtesis et al. (2019)) and Igroup Presence Questionnaire (IPQ, Tran et al. (2024)). The questions in the questionnaire are divided into three groups and evaluate the perceived realism, usability, and physiological effects of VR exposure:

- First and second questions asked the respondents about their past experience with playing games and using Virtual Reality sets.
- Third, Fourth and Fith questions concerned the immersion and subjective experience of the users when it came to scenarios.
- Sith, Seventh and Eighth questions asked about the ease of use and were the assignments understandable.

Data analysis indicated a positive reception, with high levels of immersion and moderate discomfort in the simulated vehicle operation scenario.

3 CONCLUSIONS

The results confirm the potential of VR-based training in agriculture, providing a safe and immersive environment for skill acquisition. The modular design of the proposed system allows for adaptation to

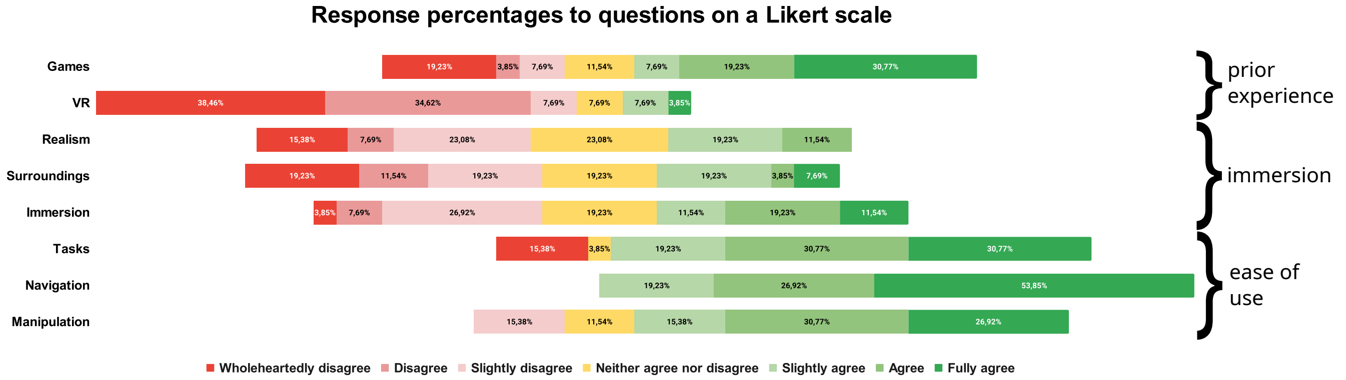

Figure 3. Response percentages to questions on a Likert Scale.

different types of agricultural machinery, and user feedback highlights the importance of optimizing comfort and minimizing motion sickness.

ACKNOWLEDGMENT

The study was supported by PUT internal grant 0214/SBAD/0248.

REFERENCES

[Dataset] Kourtesis, P., Collina, S., Doumas, L. A. A., and MacPherson, S. E. (2019). Virtual reality neuroscience questionnaire (VRNQ). `https://doi.org/10.1037/t83717-000`. Database record, APA PsycTests

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2021). Nerf: representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 99–106. doi:10.1145/3503250

Tran, T. Q., Langlotz, T., Young, J., Schubert, T. W., and Regenbrecht, H. (2024). Classifying presence scores: Insights and analysis from two decades of the igroup presence questionnaire (IPQ). *ACM Trans. Comput.-Hum. Interact.* 31

A VR Platform for Home Neurorehabilitation with Remote Supervision

Piotr Skrzypczyński *

Poznań University of Technology, Institute of Robotics and Machine Intelligence, Poznań, Poland

Correspondence*:
Corresponding Author
piotr.skrzypczynski@put.poznan.pl

ABSTRACT

This work presents a virtual reality (VR) system for home-based neurological rehabilitation, developed for the Meta Quest platform. It integrates a Unity-based VR application for therapeutic exercises with a .NET desktop tool for remote monitoring and scenario adaptation. Connected via MQTT, the system enables real-time supervision by healthcare professionals, supporting personalized therapy at a distance. A functional prototype was evaluated through simulated rehabilitation sessions.

Keywords: VR System Architecture, Neurorehabilitation, Remote Monitoring, Human–Computer InteractionUnity, Unity, MQTT

1 INTRODUCTION AND MOTIVATION

Neurological incidents such as strokes frequently result in upper limb motor dysfunction, reducing patient autonomy and quality of life. Virtual reality (VR) has recently gained traction as a tool for motor rehabilitation, offering immersive, interactive, and task-specific training that promotes neuroplasticity and accelerates recovery (Bui et al., 2021; Herrera et al., 2023). These systems also allow for adaptive task difficulty and provide engaging, game-like experiences that can improve patient adherence (Lee et al., 2020).

This project aims to design and implement a VR-based rehabilitation system with a focus on software robustness, usability, and clinical integration. The system specifically targets upper limb motor dysfunction, which is often a consequence of neurological events such as strokes. By focusing on upper extremity rehabilitation, the system aims to facilitate the recovery of fine and gross motor skills essential for daily activities. The proposed system leverages hand tracking for intuitive interaction, supports customizable therapy scenarios, and includes remote access features for clinicians to manage and monitor sessions. A key motivation behind this work is enabling at-home rehabilitation, reducing the need for frequent clinical visits. Home-based rehabilitation systems are particularly valuable for patients with limited mobility, those residing in rural or underserved areas, or individuals facing transportation barriers. By enabling therapy in a familiar environment, the system lowers healthcare costs, and encourages consistent engagement with therapeutic routines. This design approach aligns with the increasing trend toward decentralized, patient-centered care models (Sinha, 2024).

2 ARCHITECTURE AND IMPLEMENTATION

The system comprises two key components: a VR application designed for the Meta Quest headset and a desktop application built using the .NET framework. The VR application (Fig. 1A) was developed in Unity, chosen for its flexibility, integrated physics engine, and compatibility with VR SDKs. It supports immersive scenarios where patients perform object manipulation tasks aimed at improving motor skills. Key features of the VR application include: (i) bare hand-tracking-based interaction using Meta XR All-in-One SDK and Hand Physics Toolkit (González, n/a); (ii) configurable rehabilitation environments (e.g., tables and shelves) built from modular 3D shapes; (iii) scenario management through a structured stage-object model; (iv) visual/audio task instructions and real-time performance metrics.

Figure 1. Data and control flow in the VR application (A), an example scene as seen in the desktop app (B), and an alternative scene setup with more objects for manipulation (C).

The therapeutic exercises implemented in the system are designed to be flexible and configurable, either through an external control application or directly within the VR interface. Each exercise typically involves relocating virtual objects to designated targets, mimicking goal-directed motor tasks common in upper limb rehabilitation. The objects used in these tasks are defined by external 3D models, allowing customization without modifying the underlying code. Target locations can be indicated by semi-transparent "ghost shapes" or through other visual cues, such as color coding, enabling the integration of both motor and cognitive components. This flexibility supports a range of task types and difficulty levels, making the system adaptable to varying rehabilitation needs.

The desktop application (Fig. 1B,C) serves as a control centre for clinicians. It allows for real-time monitoring of sessions via streamed visual data, editing of scenarios and environmental parameters, configuration and analysis of performance metrics. Message Queuing Telemetry Transport (MQTT), a lightweight publish-subscribe network protocol commonly used for IoT applications due to its efficiency and low bandwidth usage (Radwan and Alves-Foss, 2024), is employed for transmitting data between the two applications, ensuring lightweight and responsive data exchange. This allows for dynamic adaptation of therapy scenarios during sessions and consistent data synchronization.

3 RESULTS

The developed system was evaluated through simulated rehabilitation sessions (Borowski and Rewekant, 2024). The VR application successfully guided users (healthy students acting as simulated patients) through a series of customized motor tasks, such as relocating objects to designated targets within virtual environments (Fig. 2). Hand tracking demonstrated sufficient accuracy for natural interaction, while the integrated physics-based simulation contributed to a realistic and engaging experience.

The desktop interface allowed supervisors to customize therapy scenarios and monitor patient progress remotely, with performance metrics such as task or particular movement completion time and object

Figure 2. Setting parameters inside the VR app (A), task setup with three objects and their target locations, shown as semi-transparent shapes (B), and the final state for two objects moved by the patient (C).

placement accuracy logged and visualized through interactive charts. The system demonstrated strong stability and modular architecture, supporting its suitability for future clinical deployment and remote use.

4 CONCLUSIONS

The evaluation confirms that VR with off-the-shelf headgears originally targeted to entertainment market can be effectively leveraged to create flexible, interactive rehabilitation environments suitable for home-based therapy. The integration of remote supervision and data tracking tools enhances its value for decentralized care, enabling supervisors to guide and adapt therapy without requiring in-person sessions. Future work will focus on clinical validation, improving real-time feedback for patients, and implementing adaptive difficulty algorithms to further personalize rehabilitation in a home setting.

ACKNOWLEDGMENT

The study was supported by PUT internal grant 0214/SBAD/0248.

REFERENCES

Borowski, A. and Rewekant, A. (2024). *Virtual Reality Software Supporting Neurological Rehabilitation*. Master's thesis (in Polish), Poznan University of Technology

Bui, J., Luauté, J., and Farnè, A. (2021). Enhancing upper limb rehabilitation of stroke patients with virtual reality: A mini review. *Frontiers in Virtual Reality* 2, 595771

[Dataset] González, J. J. (n/a). Hand Physics Toolkit (HPTK)

Herrera, V., Vallejo, D., Castro-Schez, J. J., Monekosso, D. N., de los Reyes, A., Glez-Morcillo, C., et al. (2023). Rehab-immersive: A framework to support the development of virtual reality applications in upper limb rehabilitation. *SoftwareX* 23, 101412

Lee, S. H., Jung, H.-Y., Yun, S. J., Oh, B.-M., and Seo, H. G. (2020). Upper extremity rehabilitation using fully immersive virtual reality games with a head mount display: A feasibility study. *PM&R* 12, 257–262

Radwan, N. M. and Alves-Foss, J. (2024). MQTT in focus: Understanding the protocol and its recent advancements. *International Journal of Computer Science and Security (IJCSS)* 18, 1–14

Sinha, R. (2024). The role and impact of new technologies on healthcare systems. *Discover Health Systems* 3, 96

VR-generated illusion as a tool for limbs neurorehabilitation during hyperacute phase of ischemic stroke

Paweł Śniatała [1,*], Sławomir Michalak [2], Ewa Kozielewska-Zwierska [2], Joanna Rybacka-Mossakowska [2], and Joanna Biskup-Danielak [2]

[1] *Institue of Computing Science, Poznan University of Technology, Poznan, Poland*
[2] *Institute of Neurological Disorders, Poznan University of Medical Sciences, Poznan, Poland*

Correspondence*:
Paweł Śniatała
pawel.sniatala@put.poznan.pl

ABSTRACT

Virtual Reality (VR) has been shown to enhance motor rehabilitation in post-stroke patients after they embody a virtual avatar. Virtual reality-generated illusions can facilitate functional recovery within one week after the onset of a stroke. We present our proposed solution that applies AR and VR in the rehabilitation of stroke patients. We also present issues related to the monitoring of people undergoing rehabilitation. To determine the tolerance of patients to the new rehabilitation technique, we conducted a survey and measurement of selected parameters of vital signs. The evaluation of the patients involved in the program was performed at baseline and after 7 days of training. The effectiveness of VR rehabilitation was measured using the National Institutes of Health Stroke Scale (NIHSS). We observed excellent tolerance and safety for VR training in stroke patients.

Keywords: Neurorehabilitation, Virtual Reality, Augmented Reality, Post-stroke Patients

1 INTRODUCTION

Ischemic stroke leads to motor deficits, which capture mainly the attention of clinicians and researchers. However, post-stroke sensory deficits often receive less exploration, primarily due to the contrast between apparent motor dysfunction and more subtle sensory impairments. Furthermore, during the hyperacute phase of ischemic stroke, assessing and differentiating neglect, hemisensory deficits, and hemianopsia can be challenging. Factors such as aphasia (language disorder) and disturbances in consciousness hinder the assessment of neglect and hemisensory deficits using the NIHSS (National Institutes of Health Stroke Scale).

In the pre-recanalization phase, clinical evaluations are more limited due to time constraints. However, in the post-recanalization phase, a thorough assessment of sensory modalities is essential for planning effective neurorehabilitation. The International Classification of Functioning, Disability, and Health (ICF), established by the World Health Organization in 2001, provides a standardized language and framework for describing health and health-related conditions, including proprioceptive functions (b260), touch functions (b265), and sensory functions related to temperature (b270), vibration (b2701), pressure (b2702), and noxious stimuli (b2703), as well as other related areas (b2708 - b2709).

Alternative sensorimotor pathways that involve secondary motor areas can help facilitate motor control. Ischemic stroke can damage structures involved in motor control, such as primary somatosensory areas and the thalamus, along with other subcortical structures. Consequently, the functional recovery of post-stroke patients may be delayed and challenging. In these cases, tailored neurorehabilitation should incorporate activities aimed at somatosensory stimulation to enhance stereognosis, proprioception, and the ability to discriminate and localize sensations. Combining sensory and motor techniques can improve recovery from limb dysfunction. One traditional and practical approach is mirror therapy, which has been developed for motor rehabilitation in post-stroke patients (1). This therapy creates a visual illusion by allowing the patient to watch the reflection of the unaffected limb in a mirror. As the patient moves the unaffected limb, the reflection creates an image of the affected limb, stimulating interhemispheric communication and activating the brain regions responsible for controlling movement in the paralyzed limb.

Figure 1. Mirror theraphy

Kinaesthetic perceptional illusion through visual stimulation (KINVIS) is a neurorehabilitation protocol in which a person at rest experiences the sensation of a body part moving or the desire to move that part while watching a video depicting the same movement. KINVIS has the potential to induce changes in the motor output system of the brain, as evidenced by the increased bioelectric activity in the affected hemisphere around the sensorimotor area following KINVIS. The advent of emerging technologies, such as Metaverse (2), finds application in the healthcare sector when feasible (3). Virtual Reality (VR) also presents opportunities to create immersive illusions that can aid in the neurorehabilitation of stroke patients. The limb illusion generated by VR has been shown to enhance motor rehabilitation in post-stroke patients after they embody a virtual avatar. This approach stimulates interregional and interhemispheric communication, thereby improving movement control in a manner akin to mirror therapy. VR scenarios can be tailored according to the range and type of neurological deficits experienced after a stroke. Additionally, the effectiveness of VR illusions can be amplified through the use of haptic feedback or tendon vibration via transcutaneous electrical nerve stimulation (TENS), such as in the wrist. By combining visual and proprioceptive stimuli, a conducive environment is created for the multisystemic activation of brain areas involved in movement control. At the Institute of Neurological Disorders at the Poznan University of Medical Sciences, we implement virtual reality illusions with haptics during the hyperacute phase of ischemic stroke. Our observations indicate a positive impact on motor recovery, including a reduction in tremors and the alleviation of neglect syndrome. However, we have also observed limitations in patients with Parkinson's disease, who exhibited increased dyskinesias during VR training. We attribute this phenomenon to impaired visual perception and the compromised multisensory integration characteristic of Parkinson's disease.

2 VIRTUAL REHABILITATION SYSTEM

It is well known that the rehabilitation process for neurological diseases is long and after a period of hospitalization, the patient must perform various, often tedious, exercises at home. Some of these exercises can be performed in virtual reality. Performing these exercises in virtual reality helps to make them more attractive, and thus more effective (the patient is more likely to perform the exercises in a well-designed

Figure 2. Illustration of the proposed system implementation

and attractive virtual reality). Of course, this assumes that the motor requirements appropriate to the set of exercises are met. The use of technology and virtual reality enables the dissemination of neurorehabilitation, allowing such goals as physical and cognitive restoration, enrichment of rehabilitation techniques, increased effectiveness; wider access to society, and teleneurorehabilitation. The system we have developed uses virtual reality, but we also use augmented reality (AR) technology. AR technology has been implemented in the "smart mirror" device we made. We use it for a personalized interface with the system user. The patient approaches the mirror, is recognized, and then a dialogue begins which aims to offer the patient appropriate rehabilitation exercises. A set of these exercises is selected from a catalog developed in advance by doctors and physiotherapists. This set is dedicated/personalized for the patient. The illustration of the proposed system is shown in Fig.2.

2.1 Smart Mirror - Personalized Interface

The Smart Mirror allows the patient to be identified by facial recognition, and then there is a personalized dialogue with the patient. Upon recognition of the patient, as mentioned earlier, the system proposes to the patient a set of exercises dedicated to him. The proposal is not only personalized for the patient, but also takes into account additional information such as the time of exercise, current weather conditions (weather forecast), the patient's mood, etc. The substantive scope of the system software is consulted with neurologists and neurorehabilitation specialists and implemented in the system accordingly. The system runs on a Raspberry Pi platform. However, the platform communicates with a more powerful computer that runs the main application. It is a web application implemented in the Django environment. Face recognition is implemented using the PyTorch and OpenCV libraries. The required data are stored in an SQL database (in our case, MariaDB). The facial recognition function in Smart Mirror is powered by algorithms based on artificial intelligence that can recognize faces and compare them with faces in the system database. The virtual rehabilitation system will propose the most appropriate and/or expected virtual interaction with the recognized person.

3 CONCLUSION

Neurorehabilitation, supported with simple tools for everyday activities and advanced technology, opens possibilities to improve motor training and cognitive functions and enables the treatment of crippling symptoms such as neglect syndrome. VR-based therapy can provide a positive learning experience and be engaging and motivating. Enhancing typical exercises performed in a real environment, with virtual reality, individuals recovering from strokes can engage in practicing self-care tasks within an environment that is typically unfeasible to recreate within a hospital setting. Illusions generated by VR can activate "mirror neurons" and "re-programme" neuronal control of motor functions. Further expansion of the palette of exercises available in virtual reality will reach homes of patients (e.g. elderly persons) and will allow telerehabilitation. Telerehabilitation trainings can be developed and supported by the implementation of human–computer interfaces (HCI). VR systems are currently used in everyday practice in the Stroke Unit of the University Hospital in Poznan. Stroke patients are mainly older people who are only occasionally familiar with computer technology. However, unexpectedly, their tolerance to VR is excellent. We noticed an excellent tolerance for VR training in stroke patients. Stable blood pressure and heart rate remain crucial for the clinical safety of procedures used during the hyperacute phase of ischemic stroke, and that was a case in our trial. Interestingly, some symptoms, such as intension tremor, that are present in real-world training disappeared in VR training. NIHSS after VR training was reduced in average from 4.797 to 2.155 as presented in Fig.3. Interestingly, some symptoms, such as intension tremor, that are present in real-world training disappeared in VR training. Access to rehabilitation centers is limited, and many patients face challenges in staying motivated to follow rehabilitation programs. Our project directly addresses these challenges by using AI-driven telerehabilitation, which offers personalized, remote rehabilitation plans that adapt to each patient's progress, improving patient outcomes and increasing access to care for those affected by neurological disorders.

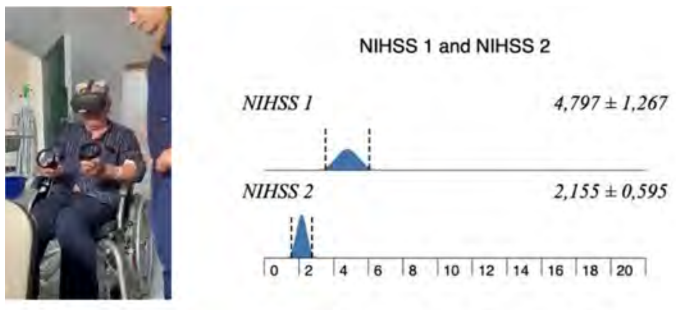

Figure 3. NIHSS before and after VR trainings

ACKNOWLEDGMENTS

This work was prepared as part of the activity of the European Metaverse Research Network.

REFERENCES

[1]H. Thieme, N. Morkisch, J. Mehrholz, M. Pohl, J. Behrens, B. Borgetto, and C. Dohle, "Mirror therapy for improving motor function after stroke," *Cochrane Database of Systematic Reviews*, vol. 2018, no. 7, Jul. 2018. [Online]. Available: http://doi.wiley.com/10.1002/14651858.CD008449.pub3

[2]A. M. Al-Ghaili, H. Kasim, N. M. Al-Hada, Z. B. Hassan, M. Othman, J. H. Tharik, R. M. Kasmani, and I. Shayea, "A review of metaverse's definitions, architecture, applications, challenges, issues, solutions, and future trends," vol. 10, pp. 125 835–125 866.

[3]R. Chengoden, N. Victor, T. Huynh-The, G. Yenduri, R. H. Jhaveri, M. Alazab, S. Bhattacharya, P. Hegde, P. K. R. Maddikunta, and T. R. Gadekallu, "Metaverse for Healthcare: A Survey on Potential Applications, Challenges and Future Directions," vol. 11, pp. 12 765–12 795. [Online]. Available: https://ieeexplore.ieee.org/document/10034994/

Poster Abstracts 2B: Room “Vicente Gimeno Sendra”

The Convergence of Artificial Intelligence and the Metaverse in Orthopedic Healthcare: Emerging Applications and Research Directions

Tony Donegan[1,2], Andrés Combalia[2,3], Maria V. Sanchez-Vives[1,4,*]

[1] IDIBAPS, Barcelona, Spain

[2] University of Barcelona, Barcelona, Spain

[3] Hospital Clínic Barcelona, Barcelona, Spain

[4] ICREA, Barcelona, Spain

*** Correspondence:** donegan@recerca.clinic.cat

Keywords: artificial intelligence, metaverse, orthopedics, virtual reality, digital twins, surgical simulation, immersive healthcare

Abstract

Introduction: Artificial intelligence (AI) and metaverse technologies have each made significant strides independently, laying the groundwork for their transformative convergence in healthcare. AI has become integral to modern medicine, powering applications ranging from diagnostic imaging and predictive analytics to personalized treatment recommendations and workflow automation. Recent breakthroughs include AI-assisted cancer treatment planning, early detection of neurological disorders through speech and movement analysis, and the integration of wearable devices for proactive health monitoring.

Concurrently, the metaverse has evolved from a conceptual virtual environment to a sophisticated, immersive ecosystem. Advances in extended reality (XR) have enabled hyper-realistic simulations, real-time interactions, and seamless integration of digital and physical experiences. Robust metaverse infrastructures will make these environments more accessible and capable of supporting complex, data-driven healthcare applications.

The convergence of AI and the metaverse in orthopedic healthcare is now enabling new models of clinical care, education, and patient engagement by combining intelligent data analysis with immersive, interactive environments. Together, these rapidly advancing technologies have the potential to revolutionize clinical care, surgical training, and patient engagement by enabling more intelligent, immersive, and personalized experiences. This poster will discuss these emerging applications and research directions.

Clinical training and education: AI-powered simulations within metaverse platforms are revolutionizing surgical education. These systems provide:

- *Personalized feedback*: AI algorithms analyze trainees' performance in real time, offering individualized guidance on technique, efficiency, and safety.

- *Objective skill assessment*: Automated metrics track progress, benchmark skills against expert standards, and identify areas for improvement, ensuring rigorous and unbiased evaluation.
- *Adaptive scenario progression*: Training modules dynamically adjust case complexity based on the learner's proficiency, ensuring continuous challenge and growth.

Virtual patients—realistic, AI-driven avatars, powered by natural language processing and behavioral modeling—offer safe, immersive opportunities for students and professionals. These include:

- *Clinical decision-making*: Learners navigate evolving clinical scenarios, making diagnostic and therapeutic choices with immediate AI-generated feedback on outcomes1.
- *Empathy and communication*: Virtual patients simulate diverse personalities and emotional responses, allowing trainees to refine bedside manner, deliver difficult news, and build rapport in a safe environment.
- *Interprofessional collaboration*: AI-powered avatars can represent multidisciplinary team members, fostering teamwork and communication skills essential for modern orthopedic practice.

Our lab is exploring this approach by training AI models to drive verbal responses of virtual therapists and patients in VR, creating realistic practice environments for clinical skill development.

Enhancing patient care and decision-making: Predictive visualization tools, underpinned by AI, are enhancing patient care in several ways:

- *Preoperative planning*: Surgeons can interact with 3D models of patient anatomy, simulate interventions, and anticipate complications, leading to more precise and personalized surgical strategies.
- *Shared decision-making*: Immersive visualizations help patients understand their conditions and treatment options, facilitating informed consent and collaborative care planning.
- *Patient understanding*: By visualizing outcomes and risks in an interactive format, patients gain clearer insights into their care journey, improving satisfaction and adherence.

Smart digital twins—AI-enhanced, self-updating models of individual patients—integrate real-time clinical, sensor, and imaging data to support continuous monitoring, outcome prediction, and tailored rehabilitation programs.

In addition, our research has demonstrated viability of using metaverse technologies with patients by conducting patient focus groups using lookalike avatars with real-time tracking, significantly enhancing patient engagement and communication through a genuine sense of presence, whilst still enabling participation regardless of mobility constraints or physical limitations. [1,2]

Embedded AI in immersive healthcare: Within immersive clinical settings, AI can be embedded directly into virtual spaces, offering

- *Real-time workflow support*: Intelligent virtual agents guide clinicians through protocols, flag potential errors, and streamline documentation during virtual rounds or consultations.
- *Clinical recommendations*: Embedded AI systems analyze patient data on-the-fly, suggesting diagnostic or therapeutic actions based on the latest evidence and guidelines.

- *Facilitated interaction*: AI-driven avatars act as patient surrogates, interpreters, or administrative assistants, enhancing communication and efficiency during remote or multidisciplinary consultations.

Examples include AI agents that moderate virtual case discussions, provide instant access to patient histories, or simulate complex patient behaviors for training and assessment purposes [3]. Building on our success with AI-driven avatars in rehabilitation settings, we are adapting these technologies for a broad range of conditions, creating more natural and responsive virtual clinical interactions.

Research frontiers and innovation areas: Emerging research at the intersection of AI and the metaverse is rapidly expanding. Areas of innovation include multimodal AI that integrates spatial, visual, and linguistic data, allowing for more comprehensive and integrated analysis; embodied AI agents capable of navigating and interacting within virtual environments; and synthetic data generation to enhance training and testing; adaptive interfaces that can adjust virtual content in response to users' cognitive or emotional states, and biometric feedback systems that can tailor therapy in real time. Such advances can potentially revolutionize personalized digital health.

Challenges and Future Outlook: Despite these advances, significant challenges remain, including:

- *Data privacy and security*: The integration of sensitive health data in immersive environments raises concerns about confidentiality and unauthorized access, necessitating robust encryption and blockchain solutions.
- *Clinical validation*: Rigorous research is required to validate the effectiveness and safety of AI-driven metaverse tools, ensuring they meet clinical standards and improve outcomes.
- *Equitable access*: High costs and technological requirements risk exacerbating healthcare disparities, highlighting the need for inclusive design and policy interventions.
- *Regulation and oversight*: Clear regulatory frameworks must be developed to govern the use of AI and metaverse technologies, balancing innovation with patient safety and ethical considerations.

Nevertheless, the ongoing convergence of AI and the metaverse is poised to unlock new frontiers in orthopedic healthcare, transforming how clinicians learn, plan, and deliver care, and how patients experience their orthopedic journey. Our ongoing research aims to address these challenges while building on our successful implementations in patient engagement with VR and the metaverse and AI-powered clinical interactions.

References

[1] Amestoy Alonso, B., Donegan, T., Calvis, I., Swidrak, J., Rodriguez, E., Vargas-Reverón, C. L., ... & Sanchez-Vives, M. V. (2024). Focus groups in the metaverse: shared virtual spaces for patients, clinicians, and researchers. *Frontiers in Virtual Reality*, *5*, 1432282.

[2] Donegan, T. , Amestoy-Alonso, B., Calvis, I., Swidrak, J,. Rodriguez, E., Combalia, A., Oliva, R., Slater, M. and Sanchez-Vives, M.V. (2025) Patient Engagement Using Metaverse Focus Groups: A Mixed-Methods Feasibility Study. (in review)

[3] Combalia, A., Sánchez-Vives, M. V., & Donegan, T. (2024). Immersive virtual reality in orthopaedics—a narrative review. *International Orthopaedics*, *48*(1), 21-30.

frontiers

Embodied AI and Multisensory VR for Skill Acquisition and Rehabilitation in the Metaverse

Yuke Pi [1,*], **Marco Gilies** [1] **and Xueni Pan** [1]

[1] *SEE VR Lab, Computing department, Goldsmiths, University of London*

Correspondence*:
Yuke Pi
y.pi@gold.ac.uk

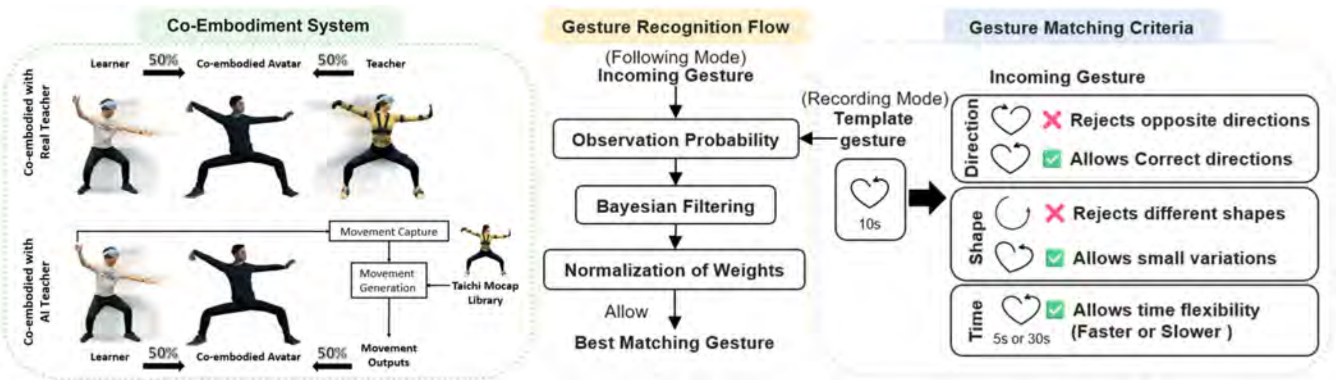

Figure 1. Co-Embodiment Systems: Learners share avatar control with real or AI teachers. *Top left:* Joint control with a real teacher (50% each). *Bottom left:* Shared control with an AI teacher, whose real-time movements are generated based on learner input and a Tai Chi motion capture library. **AI Gesture Recognition Flow:** Observation probability, Bayesian filtering, and normalization select the best matching expert gesture. **Gesture Matching Criteria:** Ensures correct direction and shape, allows flexibility in speed, and rejects major mismatches.

Dance-based motor learning requires precise coordination, body awareness, and continuous feedback. Although VR has shown promise for supporting various dance practices, including Tai Chi (Tian et al., 2024), no studies have explored the co-embodiment of first-person AI-supported skills for such skills. Previous research shows that shared control with a real teacher improves learning (Kodama et al., 2023), but the role of AI remains underexplored. In this work, we developed a co-embodiment system where users and a real or AI teacher share control of a virtual avatar, with movements weighted equally, using Tai Chi movements as a demonstrative case of dance-based motor skills. The system uses motion capture to track the user's input, comparing it to a pre-recorded Tai Chi expert dataset. Using a gesture recognition model (Bevilacqua et al., 2009), an AI teacher predicts ideal postures and provides corrective feedback in real time. This adaptive approach delivers personalized, continuous guidance that allows participants to view their optimized Tai Chi movements from a first-person perspective as their own body in real time. We explore whether AI-driven shared agency can support procedural learning as effectively as co-embodiment with a human teacher (see Figure 1 for key innovations).

Although most co-embodiment research focuses on visual interaction, effective motor learning and rehabilitation depend on multisensory engagement. Our system integrates haptic feedback, auditory cues, and gaze control to improve body awareness and motor performance Sigrist et al. (2013).

Haptic feedback is the key to somatosensory control in Tai Chi, classical dance, yoga, and hip hop, which require precise activation of the coccyx, spine, scapula, limbs, and neural extremities. Many dance techniques initiate movement from the torso, progressively activating each vertebral segment, from the tailbone upward or from the scapula downward, to achieve fluid and expressive motion through precise neuromuscular engagement Haas (2017). Beginners often struggle with precisely engaging these deep postural muscles, highlighting the need for increased somatosensory awareness. To support this, various haptic technologies have been explored, such as vibration for localized signals and EMS for deeper muscle activation (Cho et al., 2023). These systems improve proprioception and support muscle isolation essential for training and rehabilitation. Our ongoing study investigates which haptic approaches best support multisensory engagement for embodied motor learning.

Our system also incorporates other sensory inputs such as breathing (vital for core stability and focus (Jerath et al., 2006)), eye tracking (balance, postural alignment, expressive clarity) in order to promote holistic body awareness and supports dance-based rehabilitation, consistent with the principles of dance movement therapy Payne et al. (1992).

In summary, our research explores how AI-driven co-embodiment and multisensory VR systems can enhance motor skill learning and support rehabilitation. By enabling shared control of a virtual body with an AI teacher, our system offers first-person adaptive guidance for dance-based training. Tai Chi serves as a key example of shared agency and embodied feedback. Furthermore, integrating haptic, auditory, and visual cues improves body awareness, breath regulation, and gaze control, which are essential for effective embodied interaction. These findings demonstrate how virtual reality and AI can promote sustainable personal development, with potential health, education, and wider metaverse applications.

REFERENCES

Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., and Rasamimanana, N. (2009). Continuous realtime gesture following and recognition. In *International gesture workshop* (Springer), 73–84

Cho, S.-I., Kim, J.-W., Nam, S.-S., Moon, H.-W., and Jung, W.-S. (2023). Effects of a rehabilitation exercise program using electro muscle stimulation following anterior cruciate ligament reconstruction on the circumference, activity, and function of the quadriceps muscle. *Applied Sciences* 13, 4153

Haas, J. (2017). *Dance anatomy* (Human Kinetics)

Jerath, R., Edry, J. W., Barnes, V. A., and Jerath, V. (2006). Physiology of long pranayamic breathing: neural respiratory elements may provide a mechanism that explains how slow deep breathing shifts the autonomic nervous system. *Medical hypotheses* 67, 566–571

Kodama, D., Mizuho, T., Hatada, Y., Narumi, T., and Hirose, M. (2023). Effects of Collaborative Training Using Virtual Co-embodiment on Motor Skill Learning. *IEEE Transactions on Visualization and Computer Graphics* 29, 2304–2314. doi:10.1109/TVCG.2023.3247112. Conference Name: IEEE Transactions on Visualization and Computer Graphics

Payne, H. et al. (1992). *Dance movement therapy: Theory and practice* (Tavistock/Routledge London)

Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review* 20, 21–53

Tian, F., Ni, S., Zhang, X., Chen, F., Zhu, Q., Xu, C., et al. (2024). Enhancing Tai Chi Training System: Towards Group-Based and Hyper-Realistic Training Experiences. *IEEE Transactions on Visualization and Computer Graphics* 30, 1–11. doi:10.1109/TVCG.2024.3372099. Conference Name: IEEE Transactions on Visualization and Computer Graphics

Disability Centred Design and Interactive Machine Learning: A Co-creative Methodology for a more Accessible Metaverse

Clarice Hilton [1], **Rebecca Fiebrink** [2] **and Marco Gillies** [1]

[1] *Goldsmiths, University of London, UK*
[2] *Creative Computing Institue, University of the Arts London, UK*

Correspondence*:
Clarice Hilton
chilt002@gold.ac.uk

ABSTRACT

The metaverse promise is a shared place of connection where we can meet for transformative and mundane digital experiences, which could create new worlds without the ableist structures and systems we live in. However, the technology underpinning the metaverse is still extremely inaccessible to many disabled users (Gerling and Spiel, 2021). Without centering disabled experience and making it a key design concern within the metaverse we are upholding, reinforcing and creating new structures of ableism. To create a truly equitable disabled creators and researchers need to be part of the designing interaction. In this flashtalk I will explore a methodology for participatory research and interaction design using Interactive Machine Learning (IML) tool, InteractML to create new ways of interacting in the metaverse. This work was done in collaboration with associate dancers from Candoco dance company.

Keywords: Metaverse, Accessibility, Disability, Machine Learning, Interactive

1 TALK

In navigating an ableist technology, it is common practice in disabled communities to "hack" technology and the environments, designing new ways of interacting that are designed specifically for your body-mind (Perry, 2024). Crip technoscience calls for research and practice in "knowing-making" (Hamraie, 2017) that centers the skills, wisdom and hacks of disabled people, that position disabled people as "experts and designers in everyday life" (Hamraie and Fritsch, 2019). In designing and creating interactions in the metaverse disabled people should be integral as part of that design process. Disabled communities have already established this practice in the metaverse, such as mods in VRChat[1].

Unlike traditional machine learning pipelines, IML engages users iteratively in the training and refining of AI models (Bernardo and Grierson, 2017). IML enables users to have control and develop an understanding of the machine learning system using small datasets which are completely transparent to users. This subverts current power dynamics in AI putting the user in full control of where the data comes from and where it is going (Vigliensoni et al., 2022). InteractML integrates an IML system for designing movement interaction with game engines Unity and Unreal (Hilton et al., 2021; Plant et al., 2020, 2021). Using

[1] Dale 2022 https://access-54 ability.uk/2022/09/09/vrchat-screwed-over-disabled-players-but-the-devs-are-trying-to-fix-it/ [Accessed 16th April 2025]

InteractML users can record examples of their poses, gestures or movements, so the system can recognise these movements and the user can connect these to outputs. It uses a node-based visual scripting language requiring minimal programming and democratising access to using IML. Using InteractML the AI becomes a collaborator, where disabled creators can teach the system their way of moving. This gives users creative agency, where in the core of this creative process with AI there is consent, transparency, and co-authorship.

In collaboration with disabled dancers Kat Hawkins, Xan Dye, Paulina Porwollik, Ted Wilkinson and Dermot Farrell we embarked on a participatory design process exploring how to co-create accessibility to inherently inaccessible technologies used in the metaverse VR and MOCAP. We sought to create personalised interactions using these technologies which were designed with and for the dancer researchers for performance and personal expression. InteractML was used as the tool to use the knowledge, skills and practice of the dancer-researchers and the power of IML.

- What does a disability centred design process involve with IML?
- How can care and trauma-informed practice be embedded in a process with IML?
- How can you empower disabled dancer-researchers in the process with IML?

There cannot be a Metaverse for Good without disabled creators, researchers and audiences being leaders in how it is shaped. By embedding accessible mechanisms for design such as InteractML which utilise the power of transparent small data AI we can start to imagine how we can create a more accessible metaverse which truly reimagines how we interact.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

Clarice Hilton conceptualisation, investigation, methodology, administration, software, writing original draft. Rebecca Fiebrink supervision. Marco Gillies supervision, writing review and editing

ACKNOWLEDGMENTS

I would like to thank Kat Hawkins, Xan Dye, Paulina Porwollik, Ted Wilkinson and Dermot Farrell for their enthusiastic collaboration and sharing of practice. I would also like to thank Candoco Dance Company for supporting the project.

REFERENCES

Bernardo, Z. M. F. R., Francisco and Grierson, M. (2017). Interactive machine learning for end-user innovation. *Association for the Advancement of Artificial Intelligence*

Gerling, K. and Spiel, K. (2021). A critical examination of virtual reality technology in the context of the minority body. *CHI'21: CHI Conference on Human Factors in Computing Systems*

Hamraie, A. (2017). *Building Access: Universal Design and the Politics of Disability* (University of Minnesota Press)

Hamraie, A. and Fritsch, K. (2019). Crip technoscience manifesto. *Catalyst: Feminism, Theory, Technoscience*

Hilton, C., Plant, N., González Díaz, C., Perry, P., Gibson, R., Martelli, B., et al. (2021). Interactml: Making machine learning accessible for creative practitioners working with movement interaction in immersive media. *VRST'21: 27th ACM Symposium on Virtual Reality Software and Technology*

Perry, P. (2024). *Ctrl Shift: How Crip Alt Ctrl Designers Change the Game and Reimagine Access*. Ph.D. thesis, Goldsmiths, University of London

Plant, N., Gibson, R., Diaz, C. G., Martelli, B., Zbyszyński, M., Fiebrink, R., et al. (2020). Movement interaction design for immersive media using interactive machine learning. *MOCO '20: Proceedings of the 7th International Conference on Movement and Computing*

Plant, N., Hilton, C., Gillies, M., Fiebrink, R., Perry, P., DÍaz, C. G., et al. (2021). Interactive machine learning for embodied interaction design: A tool and methodology. *TEI 2021 - Proceedings of the 15th International Conference on Tangible, Embedded, and Embodied Interaction*

Vigliensoni, G., Perry, P., and Fiebrink, R. (2022). A small-data mindset for generative AI creative work. *CHI 2022 Workshop*

Extended reality technologies for behavioral therapy and cognitive training for people with intellectual disability

Marta Goyena [1,2], Matteo Dal Magro [1,2], Carlos Cortés [1], Marta Orduna [2], Ainhoa Fernández-Alcaide [3], María Nava-Ruiz [3], Jesús Gutiérrez [1,*], Pablo Pérez [2], Narciso García [1]

[1] *Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, Spain*[*]
[2] *Nokia XR Lab, Madrid, Spain*
[3] *Fundación Juan XXIII, Madrid, Spain*

Correspondence*:
Corresponding Author
jesus.gutierrez@upm.es

ABSTRACT

The development of eXtended Reality (XR) technologies is allowing the improvement of traditional psychological methods for behavioral therapy, cognitive stimulation, and training Freeman et al. (2017). In particular, XR technologies offer the possibility to create secure and controlled settings where the users can interact with simulated environments without requiring to be physically present in them, and therapists can completely monitor and adjust the progress as needed Schiza et al. (2019). In particular, when dealing with people with Intellectual Disability (ID), the evoking possibilities that the creation of XR environments provide can help to alleviate their difficulties with abstract thinking, imagination or transfer of skills to real life scenarios Dagnan and Jahoda (2006).

Taking this into account, the works presented here aim at exploring the use of XR technologies in behavioral therapy and cognitive training with people with ID. These works stem from the project Incluverso 5G[1], which aims at improving the quality of life of people with ID through immersive experiences benefiting from the interdisciplinary and intersectoral collaboration among Nokia, *Universidad Politécnica de Madrid* and the *Fundación Juan XXIII*. Specifically, among other use cases, we developed two approaches for: 1) immersive behavioral therapy for individuals with fear of stairs and 2) immersive cognitive training to support job integration of people with ID.

In relation to the use of XR for behavioral therapy, we carried out a longitudinal study during six months with nine users of the *Fundación Juan XXIII* with bathmophobia (fear of going up and/or down stairs,) Goyena et al. (2025b). The proposed method was based on integrating XR technology in the classical systematic desensitization technique. The whole process (i.e., design, development, and evaluation of the method) was performed in close collaboration with psychologists. This included the selection of appropriate immersive stimuli representing

[*] This work has been partially supported by the project TSI-064200-2022-009 (INCLUVERSO 5G) funded by program UNICO I+D 6G 2022 of the Spanish Government within the framework of the Recovery, Transformation and Resilience Plan, and by the projects PID2020-115132RB (SARAOS) and PID2023-148922OA-I00 (EEVOCATIONS) funded by MCIN/AEI/10.13039/501100011033 of the Spanish Government.

[1] http://www.incluverso-5g.es/

relaxing and feared stimuli (360-degree videos and virtual scnearios) and the development of an application for the therapists to control and monitor the sessions, integrated with a tool for the users to visualize the stimuli and provide feedback about their state Cortés et al. (2024). This feedback was obtained through questionnaires (simplified and adapted from classical ones Mott et al. (2019)) that appear in the virtual environment, and also from non-invasive bionsensors providing physiological data, such as as heart rate, electro-dermal activity, electromyography, etc. After the therapy sessions, all participants were able to go up and down real stairs that they were not able to face before[2], showing the benefits of the immersive therapy in overcoming their fears and in improving their inclusion and quality of life.

For cognitive training, a second application has been developed to train attention skills for individuals with ID with the goal of supporting their transition into the workforce Goyena et al. (2025a)[3]. In particular, two serious games simulating real-life tasks were developed, related to jobs that the individuals of the *Fundación Juan XXIII* are trained to perform. So, two primary scenarios have been created: a cafeteria and a supermarket. In the cafeteria scenario, the tasks to be performed by the users include collecting dirty dishes and serving customer orders. In the supermarket scenario, users are tasked with restocking products and processing customer orders. Different levels of difficulty were implemented for each task, allowing the users to progress and improve their skills. The application includes a module that allows real-time control and monitoring of the sessions by the therapists. The developed system also allows the collection of physiological data through non-invasive sensors, including heart rate, eye movement, etc. Another longitudinal study with nine users with ID from the *Fundación Juan XXIII* is currently being performed and the preliminary results highlight the potential of XR technologies in cognitive training and job placement for people with ID.

Future work will focus on the extension of these approaches to multi-user scenarios (including avatars based on artificial intelligence agents) and to other phobias and training scenarios.

Keywords: extended reality, accessibility, intellectual disability, behavioral therapy, cognitive training, user assessment.

REFERENCES

Cortés, C. et al. (2024). Immersive behavioral therapy for phobia treatment in individuals with intellectual disabilities. In *IEEE Conf. on VR and 3D User Interfaces Abstracts and Workshops*. 1222–1223

Dagnan, D. and Jahoda, A. (2006). Cognitive–behavioural intervention for people with intellectual disability and anxiety disorders. *Journal of Applied Research in Intellectual Disabilities* 19, 91–97

Freeman, D. et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological medicine* 47, 2393–2400

Goyena, M. et al. (2025a). Immersive cognitive training for job integration with people with intellectual disabilities. In *ACM International Confernece on Interactive Media Experiences*

Goyena, M. et al. (2025b). A study on immersive behavioral therapy for individuals with intellectual disabilities with fear of stairs. In *ACM International Confernece on Interactive Media Experiences*

Mott, M. et al. (2019). Accessible by design: An opportunity for virtual reality. In *IEEE International Symposium on Mixed and Augmented Reality Adjunct*. 451–454

Schiza, E. C. et al. (2019). Virtual Reality Applications for Neurological Disease: A Review. *Frontiers in Robotics and AI* 6

[2] https://www.youtube.com/watch?v=rWNl7XugQ3Y

[3] https://youtu.be/-o4N1L3B6iM