Self-Disclosure in Metaverse-based Mental Healthcare: The Role of Avatars

Runjie Xie
University of Bamberg
runjie.xie@uni-bamberg.de

Siao Fang Chen
Netlight Consulting
siaofang.chen@gmail.com

Sandra Birnstiel
Friedrich-Alexander-Universität Erlangen-Nürnberg
sandra.birnstiel@fau.de

Benedikt Morschheuser
University of Bamberg
Benedikt.Morschheuser@uni-bamberg.de

Abstract

Given the rise of mental health issues and healthcare system limitations, the metaverse emerges as a promising platform for care delivery. Patients and physicians can meet virtually, overcoming constraints of traditional online communication by using avatars. While patient self-reporting is crucial for diagnosis and treatment, little is known about how avatars, core elements of the metaverse, influence communication behavior in such situations. This study explores how avatar identification impacts self-disclosure intention in VR-based mental healthcare, focusing on the roles of self-presence and perceived anonymity. In a lab experiment conducted in Germany using VRChat, 62 participants engaged in simulated one-on-one mental healthcare sessions. Results show that both similarity and wishful identification are significantly associated with self-presence and perceived anonymity. However, only self-presence has a significant positive relationship with self-disclosure intention. These findings advance research on metaverse and telehealth, highlighting the importance of user-avatar dynamics in designing future metaverse mental healthcare interventions.

Keywords: Avatar identification, virtual worlds, Virtual Reality (VR), metaverse therapy, privacy

1. Introduction

In recent years, mental health issues have increased drastically, particularly among young people (Morini et al., 2025; Orr et al., 2023). This widespread decline in psychological wellbeing is driven by global stressors such as economic and political instability, alongside the lingering impact of the Covid-19 pandemic (Zahedi et al., 2022). Traditionally, mental healthcare support has been provided through clinical settings and support groups (Morini et al., 2025). However, like the broader healthcare sector, the mental health industry is fragile and overstretched. Workforce shortages, high costs, and limited resources constrain access to care, especially in in-person settings (Orr et al., 2023). As a result, current healthcare systems in many countries are ill-equipped to meet the growing demand for mental health services.

Simultaneously, many individuals in need of support avoid seeking help due to practical barriers, shame, or the stigma associated with mental illness (McLellan et al., 2022). Concerns over medication side effects have also raised interest in non-pharmacological alternatives to complement or replace conventional treatments (Orr et al., 2023). Generally, the pandemic has underscored the urgent need for the healthcare sector to accelerate digital transformation and establish effective remote mental healthcare solutions (Riches et al., 2024; Zahedi et al., 2022). In response, the mental health industry is steadily exploring new technologies.

One promising avenue for patient-physician interaction is the metaverse. It is an immersive, three-dimensional virtual space where users connect and interact with each other in social, economic, and entertainment activities within extended reality contexts (Kim et al., 2025; Riches et al., 2024). A central element of the metaverse is the avatar – a virtual representation of users. Avatars serve as bridges between the physical and digital worlds, enabling real-world actions to shape virtual experiences and vice versa. Many existing platforms, such as Meta Horizon, Second Life, and VRChat, offer extensive avatar customization options, allowing users to modify appearance and attributes. These affordances enable individuals to transcend offline limitations, creating a digital self that expresses their authentic identity (Hooi & Cho, 2014). While some opt for avatars that resemble their offline identity, others adopt entirely different forms, such as human-like animals or fantasy creatures, to explore alternative identities. As a result, avatars are often perceived not merely as tools or puppets, but as extensions of the self, with which users strongly identify (Freeman & Maloney, 2020). This makes identification with avatars a key link between virtual appearance and social behavior in virtual spaces (Takano & Taka, 2022).

The metaverse is emerging as a promising platform for immersive and interactive therapeutic interventions,

URI: https://hdl.handle.net/10125/111750
978-0-9981331-9-5

welcomed by both patients and physicians (Digulet al., 2021). Research shows that VR-based therapies can effectively address a wide range of mental health conditions, such as phobias, anxiety, depression, trauma, posttraumatic stress disorders, autism, substance abuse, and addiction (Baghaei et al., 2021; Dellazizzo et al., 2020; Orr et al., 2023; Riches et al., 2024). Moreover, studies have demonstrated the healing effect of experiencing VR forest environment, boosting mood, vitality, and restoration (Mattila et al., 2020).

In medical settings, the quality of interpersonal interaction is vital for system acceptance. Successful mental health interventions rely heavily on patient self-disclosure, the act to share personal information. Traditional telehealth technologies lack the naturalness that patients experience in face-to-face office visits (Zahedi et al., 2022). Physicians expressed concerns about limited interactivity and conversational immediacy, primarily due to the absence of nonverbal cues (Atanasova et al., 2018; Mora et al., 2008). Yet, these remote technologies can provide a supportive environment, as patients often prefer anonymity when discussing sensitive medical issues (Morini et al., 2025). Unlike traditional telehealth applications, metaverse combines immersive qualities of face-to-face communication with the protective anonymity of online interaction (Baccon et al., 2019; Sykownik et al., 2022). Through avatars, the metaverse recreates non-verbal cues and presence while offering users a protective layer of identity. This unique combination promotes greater openness and honesty, encouraging patients to disclose more authentically (Dilgul et al., 2021).

While avatars are central in the metaverse, their psychological impact, particularly on self-disclosure behavior, remains underexplored. One study has examined the effect of avatar similarity on self-disclosure in gaming contexts (Hooi & Cho, 2014), but it neglected the concept of wishful identification. Moreover, the direct transferability of findings from deeply engaging gaming environments to serious applications is limited (Birk et al., 2016). In addition, most avatar identification studies have focused on non-immersive settings (e.g., Messinger et al., 2019; Takano & Taka, 2022), overlooking the heightened embodiment afforded by VR. So far, little is known about how avatar identification shapes self-disclosure in VR therapeutic settings. Understanding this relationship is crucial for designing metaverse healthcare applications that foster patient openness, thereby facilitating effective diagnosis and treatment. This leads to our research question: ***How does avatar identification impact users' self-disclosure intention in metaverse mental healthcare?***

To address it, we conducted a controlled lab experiment with 62 participants, simulating one-on-one patient-physician mental healthcare consultations in VR. Building on prior research that highlights the metaverse's potential to combine immersion of face-to-face communication with the anonymity of online interaction, we examined self-presence and perceived anonymity as predictors of self-disclosure intention.

2. Theoretical Background

2.1. Avatar Identification

Avatar identification refers to a temporary shift in self-perception, where users mentally associate with their avatars and internalize its valued traits (Van Looy et al., 2012). This process can lead to cognitive changes as users momentarily lose real self-awareness and adopt their avatars' identity (Kim et al., 2025). Such identification is reinforced through direct avatar control and a first-person view, enabling users to merge their sense of self with the avatar (Klimmt et al., 2009).

Avatar identification can be divided into similarity and wishful identification (Van Looy et al., 2012), both rooted in self-concept theory. This theory explains individuals' drive to present themselves as their actual or ideal self to satisfy needs for self-verification or self-enhancement (Kim et al., 2025; Messinger et al., 2019). ***Similarity identification*** is the degree to which users perceive their avatars as resembling themselves (Hooi & Cho, 2014). This includes not only physical traits, but also personality and life experiences. A strong visual resemblance can further enhance the perceived psychological alignment (Hooi & Cho, 2014). ***Wishful identification*** refers to the extent to which users project their ideal self onto the avatar, perceiving it as possessing desirable traits (Van Looy et al., 2012). Driven by the inherent desire to enhance one's self-concept and ego, this identification form helps compensate for perceived discrepancies between the real and virtual self, enabling a temporary detachment from reality and everyday challenges (Kim et al., 2025). Similarity and wishful identification are not mutually exclusive. In practice, users often design avatars that combine real and idealized traits, trying to satisfy both self-verification and self-enhancement needs at the same time (Messinger et al., 2019).

2.2. Self-Presence

Presence is a core characteristic of the metaverse, referring to the illusion of being physically and psychologically present in the mediated virtual space which is perceived as real and unaffected by its artificial nature (Behm-Morawitz, 2013). According to Lee (2004), presence can be categorized into three dimensions: physical, social, and self-presence.

Physical presence describes the sense of immersion in a virtual environment, including realistic interaction and a perceived detachment from the physical world. ***Social presence*** refers to the feeling of being with other users in a virtual space and experiencing interactions as natural and real. ***Self-presence*** is the perception of one's virtual representation as an extension of the self into the virtual world. In this psychological state, the para-authentic virtual self is experienced as the actual self (Lee, 2004), making sensory inputs from the virtual space feel real and vivid (Hooi & Choi, 2014). Thus, self-presence is crucial for effective interaction and communication through avatars (Kim et al., 2025).

2.3. Anonymity

Anonymity is the extent to which a communicator perceives the message source as unknown and undefined (Scott, 1998). It can take various forms, such as ***visual anonymity***, where the physical appearance of the sender is not identifiable, and ***discursive anonymity***, where no personal cues are embedded in the content itself. While visual anonymity can be a unique feature of metaverse healthcare, discursive anonymity is not feasible, as effective treatment relies on the disclosure of sensitive personally identifiable information.

Anonymity is a key affordance of the ***online disinhibition effect***, which describes people's tendency to behave more openly and intensely online than in face-to-face interactions (Clark-Gordon et al., 2019). This is largely attributed to a heightened sense of comfort and reduced accountability. While much of the research has focused on its negative consequences, such as toxicity, aggression and antisocial behavior (Joinson, 2001; Suler et al., 2004), anonymity can also foster positive social outcomes, such as increased self-disclosure (Clark-Gordon et al., 2019). In social VR, anonymity strongly drives self-disclosure (Maloney et al., 2020). Moreover, anonymity predicts satisfaction and use intention in metaverse therapies (Turan Akdag & Wahl, 2024).

2.4. Self-Disclosure

Self-disclosure refers to the act of revealing personal information that was previously unknown to others, thereby making it shared knowledge (Derlega & Chaikin, 1977). It is vital in social communication and is closely linked to trust, liking, and the development and maintenance of interpersonal relationships (Baccon et al., 2019). In mental health contexts, self-disclosure is crucial for emotional release, stress reduction, and various forms of support seeking, such as advice on managing symptoms, emotional empathy, understanding, and validation of experiences and feelings (Atanasova et al., 2018; Morini et al., 2025).

Individuals often feel less vulnerable when sharing personal experiences in anonymous, mediated spaces (Morini et al., 2025). For example, users often create throwaway accounts in online social networks like Reddit to seek support on sensitive mental health subjects (Andalibi et al., 2016). Such anonymous online communities enable more extensive and intimate self-disclosure, particularly on stigmatized mental health topics (McLellan et al., 2022; Morini et al., 2025), such as mental health challenges after sexual abuse (Andalibi et al., 2016) and self-harm (Williams et al., 2018).

However, these platforms also present drawbacks. A lack of professional oversight often leads to poor information quality and the spread of misinformation, raising concerns about the reliability of advice received (Atanasova et al., 2018). Moreover, negative or dismissive responses can discourage participation and even worsen mental health issues (Morini et al., 2025). The presence of trolling and bullying on such platforms further poses serious risks. Data protection might also be a concern (Atanasova et al., 2018).

As in other communication forms, metaverse users naturally disclose a wide range of personal information and self-concepts, like emotions, experiences, beliefs, lifestyle, sexuality, dreams, and fantasies (Maloney et al., 2020). Despite skepticism and privacy concerns, many users perceive virtual spaces as conducive for forming authentic connections (Sykownik et al., 2022). Prior research also indicates that individuals are more willing to disclose sensitive topics to avatars than to human interviewer due to a reduced fear of judgement (Pickard et al., 2016). A key distinction of the metaverse from previous forms of computer-mediated communication is that users interact within immersive virtual worlds from a first-person perspective through controllable, embodied avatars (Sykownik et al., 2022). Avatars are powerful communication tools, as they enable more naturalistic self-disclosure by conveying rich non-verbal cues, such as facial expressions, gestures, posture, co-presence, and proxemics (Maloney et al., 2020). These elements simulate intimate offline interactions, such as nodding, hugging, or handshakes.

Beyond transmitting verbal and non-verbal signals, avatars themselves act as a central medium of self-disclosure. They can be visually customized to reflect either the users' physical appearance or an idealized version of themselves (Sykownik et al., 2022). Lookalike avatars may convey height, appearance, race, and gender, while idealized ones can reveal users' desires and feelings (Xie et al., 2024).

Research shows that avatars are not only a communication channel for self-disclosure but can also shape it in subtle and unconscious ways (Sykownik et al., 2022). The ***Proteus Effect*** describes how users' perception and behavior in virtual environments can be

influenced by their avatar's traits, often aligning with expectations or stereotypes linked to them (Yee et al., 2009). For example, an avatar's appearance can affect how users behave, interact, and communicate (Oberdörfer et al. 2024; Yee et al., 2009). Users tend to be more extroverted and confident in self-disclosure when appearing as attractive avatars (Yee et al., 2009).

As self-disclosure is central to mental healthcare, by studying the influence of metaverse's core element – avatars – on it, we aim to advance understanding of user behavior in these environments and inform the design of future metaverse mental healthcare interventions.

3. Hypotheses and Research Model

For users to experience self-presence, strong identification with their avatars is essential (Kim et al., 2025). Avatars resembling users can enhance self-presence. This is likely due to a reduced discrepancy between the actual and virtual self, a key predictor of self-presence (Jin, 2012). Feeling similar to one's avatar is vital for experiencing a genuine sense of existence in the virtual environment (Kim et al., 2025). Research on metaverse users indicates that avatars with physical resemblance to users and similarity in beliefs, values, and attitudes heighten self-awareness, resulting in increased feelings of self-presence (Hooi & Choi, 2014; Kim et al., 2025). Thus, we propose:

H1. *Similarity identification is positively associated with self-presence.*

When users admire or aspire to their avatars' characteristics, the sense of self-presence can also be enhanced. This occurs as wishful identification aligns with users' ideal self-concepts and can temporarily satisfy their intrinsic needs, thereby intensifying their engagement and immersion in the virtual environment (Fraser et al., 2023; Kim et al., 2025). Moreover, this gap closing between the actual and ideal self via avatars may alleviate real-life negative feelings like depression or low self-esteem (Bessière et al., 2007), fostering better self-presence. Therefore, we propose:

H2. *Wishful identification is positively associated with self-presence.*

Perceived anonymity is the belief that one's identity is unknown to others. When an avatar closely resembles the user, it can make the user feel identifiable and less anonymous due to the reduced psychological distance between the avatar and the self (Hooi & Cho, 2014). Physically similar avatars are associated with greater alignment in beliefs, values, and attitudes with their real selves, fostering a sense of homophily and reducing perceived anonymity (Hooi & Cho, 2014). In contrast, dissimilar avatars can increase perceived anonymity. Prior research in the non-immersive context of Second Life has shown that greater avatar-user similarity increases perceived identifiability in virtual teams (Midha & Nandedkar, 2012). Thus, we propose:

H3. *Similarity identification is negatively associated with perceived anonymity.*

Wishful identification occurs when users view their avatars as a better or idealized version of themselves. These avatars compensate for perceived deficiencies or unmet needs in real life, offering an escape from reality and everyday stressors (Birk et al., 2016; Fraser et al., 2023). Adopting this alternate identity facilitates psychological detachment from the real self, which can heighten the perception of anonymity. Thus, we derive:

H4. *Wishful identification is positively associated with perceived anonymity.*

Self-presence can induce flow, a mental state characterized by deep absorption and intrinsic motivation (Jin, 2012). Research on social networks indicates that flow experiences can enhance self-disclosure by boosting enjoyment, trust, and engagement, while reducing uncertainty (Kwak et al., 2014). This can encourage more open communication and a greater willingness to share personal information. Moreover, the immersive and deeply absorbed state of self-presence and flow leads to reduced self-consciousness and a stronger sense of control (Wang et al., 2017). This mental state can cause individuals to unconsciously ignore and downplay risks and negative consequences, thereby lowering inhibitions around disclosing sensitive information (Wang et al., 2017). Moreover, self-presence may enhance perceived security, further encouraging self-disclosure (Hooi & Cho, 2014). Therefore, we propose:

H5. *Self-presence is positively associated with self-disclosure intention.*

Previous studies indicate that individuals prefer anonymity for reporting mental health issues (Andalibi et al., 2016; Morini et al., 2025) and evidence a positive link between anonymity and self-disclosure (Joinson, 2001). Driven by the benign online disinhibition effect, anonymity helps individuals to reduce anxiety, enabling more open self-expression (Clark-Gordon et al., 2019; Suler, 2004). Furthermore, anonymity protects privacy by shielding individuals from unwanted publicity or discrimination (Hooi & Cho, 2014), enabling patients to express themselves without fear of judgment. This promotes the sharing of intimate information and encourages self-disclosure (Baccon et al., 2019), which is especially crucial in mental healthcare, where discussing stigmatized topics can be especially challenging (McLellan et al., 2022). Hence, we propose:

H6. *Perceived anonymity is positively associated with self-disclosure intention.*

The complete research model is illustrated in Figure 1.

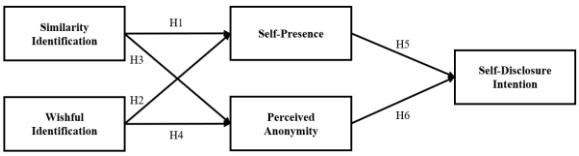

Figure 1. Proposed Research Model

4. Research Methodology

4.1. Experimental Design

To test our research model, we used a between-subjects experiment with two conditions (lookalike vs. idealized avatar) to induce different levels of avatar identification. To simulate a realistic mental healthcare setting, we conducted a wellbeing workshop featuring relaxation exercises and a talk on participant wellbeing. This approach was chosen because meditation and breathing exercises are established interventions to improve mental health, allowing users to immerse themselves in calming virtual environments and practice self-soothing, mindfulness, and behavioral activation (Orr et al., 2023; Riches et al., 2024). The experiment was conducted in a controlled lab setting using Meta Quest 3 headsets for both participants and the instructor, allowing real-time mirroring of physical movements in avatars and thereby enhancing the sense of merging with avatars. We selected VRChat as the metaverse platform due to its popularity and extensive avatar customization options (Maloney et al., 2020; Sykownik et al., 2021). VRChat offers a vast library of community-built avatars and is compatible with third-party avatar editing tools like Ready Player Me for fine-grained customization.

4.2. Procedure

Before the experiment, participants were informed about the study's purpose, procedure, and the data collection, and provided signed consent. They were then randomly assigned to one of two conditions and tasked to create a custom avatar on a laptop within six minutes to ensure consistency. Allowing users to design their avatars is a well-established method for enhancing the psychological bond between user and avatar, thereby reinforcing avatar identification (Birk et al., 2016; Takano & Taka, 2022). In the similarity condition, participants were instructed to create a lookalike avatar using Ready Player Me, customizing body shape, facial features, hairstyle, and clothing. In the wishful condition, participants were instructed to either enhance their lookalike avatar or select a predesigned fictional or fantastical avatar from the VRChat library.

Participants then received an introduction to the VR headset and entered a private instance of the world "VR Yoga" for the wellbeing workshop, ensuring only the participant and instructor were present. The virtual yoga studio featured mirrors, allowing participants to view their avatars and movements. To reduce distraction, participants had three minutes to explore the studio before the approximately 15-minute workshop began.

The second author consistently acted as the instructor, using the same avatar and maintaining a uniform communication style. Prior to the workshops, she familiarized herself with the topic through research to gain sufficient knowledge. The workshop structure was developed under expert guidance and accounted for physical movement limitations imposed by VR. It included five simple relaxation exercises, focusing on breathing, upper-body stretching, and mindfulness. Following an instructional approach, the instructor verbally explained and demonstrated each exercise.

After the exercises, a talk on participant wellbeing was held in the studio, simulating a realistic patient-physician exchange. To ensure consistency, we adapted guiding questions from Herbert et al. (2021) to cover topics like sociodemographics, body metrics, health habits, life satisfaction, and mental health. Responses were voluntary and not evaluated due to privacy and ethical considerations. After the session, participants completed a questionnaire and received a debriefing.

The study procedure was approved by the university's ethical board, and data collection was authorized by the university's Data Protection Officer.

4.3. Measurements

Table 1. Measurements

Code	Items
SI1	My avatar is like me in many ways.
SI2	My avatar resembles me.
SI3	I identify with my avatar.
SI4	My avatar is an extension of myself.
SI5	My avatar is similar to me.
SI6	I resemble my avatar.
WI1	If I could become like my avatar, I would.
WI2	I would like to be more like my avatar.
WI3	My avatar is an example to me.
WI4	My avatar is a better me.
WI5	My avatar has characteristics that I would like to have.
SP1	How much did you feel like your avatar was an extension of yourself?
SP2	How much did you feel like your avatar was just a character and was not really you?
SP3	During the virtual workshop, how much did you feel like you were actually in the world as your avatar was in the world?
SP4	During the virtual workshop, how much did you feel like you are experiencing what your avatar is experiencing?
PA1	I feel that my avatar makes me indistinguishable from others in the virtual world.
PA2	I feel that I am not identifiable from others in the virtual world through my avatar.
PA3	I feel that I can blend in with other avatars unnoticed through my avatar.
SDI1	I will disclose my information when health professionals in virtual worlds ask me to provide personal health information.
SDI2	I am likely to provide more personal health information in virtual worlds when I feel it is necessary for my health.
SDI3	If my information disclosure helps to maintain health, I am likely to provide my information to healthcare professionals in virtual worlds.

As depicted in Table 1, all measurement items in this study were derived from prior studies and have been modified to align with the specific context of our investigation. The items for similarity identification (SI) and wishful identification (WI) were adapted from Van Looy et al. (2012), self-presence (SP) from Behm-Morawitz (2013), perceived anonymity (PA) from Hooi & Choi (2014), and self-disclosure intention (SDI) from Yuchao et al. (2021). Each item was measured on a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). As control variables, we included single-item measures for gender, age, education, VR experience, prior privacy violations, and perceived motion sickness.

4.4. Data Collection

Table 2. Profiles of Respondents

Characteristics	N	Characteristics	N
Gender		**Familiarity with VR**	
Male	29	Not familiar at all	10
Female	33	A little familiar	21
Age		Moderately familiar	18
18-20	4	Fairly familiar	9
21-25	15	Extremely familiar	4
26-30	26	**Past privacy violation experience**	
31-35	6	Very rarely	29
36-45	4	Rarely	22
46-55	3	Sometimes	6
56-60	4	Often	5
Education		Very often	0
High school or equivalent	12	**Perceived motion sickness**	
Completed vocational training	4	Not at all	18
Bachelor's degree	16	A little	18
Master's degree or equivalent	27	Moderately	13
Doctoral degree	2	Strong	9
Not specified	1	Very strong	4

Healthy participants over 18 were recruited via university mailing lists, social media, and on-site at the university and city center. In total, 62 participants (31 per condition) took part in our study and completed the questionnaire in German or English between June and August 2024. Table 2 depicts participant demographics.

5. Data Analysis and Results

We evaluated our research model using partial least squares structural equation modeling (PLS-SEM) with SmartPLS 4. PLS-SEM was chosen for its ability to analyze complex relationships between observed and latent variables, its robustness with non-normally distributed data, and its suitability for smaller sample sizes (Hair et al., 2021; Henseler et al., 2015).

To ensure sufficient sample size, we applied the ten times rule of thumb, which requires the minimum sample size to be ten times the maximum number of inner or outer model paths leading to any latent variable (Hair et al., 2021). Following this, the model fulfils the minimum requirement of 60. Moreover, we employed the more precise gamma-exponential method in WarpPLS (Kock & Hadaya, 2018; Kock, 2024), using standard IS research parameters (power = 0.8, alpha = 0.05). This approach yielded a comparable minimum sample size of 61, which was also met.

5.1. Measurement Model

We evaluated the reliability, convergent validity, and discriminant validity of our measurement items. For reliability (see Table 3), all item loadings exceeded the recommended threshold of 0.7 (Hair et al., 2021), except for the item SP2 measuring self-presence. This item was retained due to its origin from a validated scale. All composite reliability (CR) and average variance extracted (AVE) values met their recommended criteria of 0.7 and 0.5 (Fornell & Larcker, 1981).

Discriminant validity was confirmed through three methods. First, cross-loadings confirmed that items loaded higher on their corresponding constructs than other constructs (Hair et al., 2021). Second, the Fornell-Larcker criterion (see Table 4) showed that the square root of each construct's AVE on the diagonal was notably greater than its inter-construct correlations. Third, all heterotrait-monotrait ratio of correlations (HTMT) values ranged from 0.124 to 0.668, well below the recommended threshold of 0.85 (Henseler et al., 2015). Jointly, these results indicate good reliability, convergent, and discriminant validity (Hair et al., 2021).

Table 3. Psychometric Properties of Items

	CR	AVE	Items	SI	WI	SP	PA	SDI
SI	0.908	0.623	SI1	**0.790**	0.452	0.378	-0.156	0.021
			SI2	**0.791**	0.175	0.236	-0.445	-0.034
			SI3	**0.726**	0.559	0.609	-0.251	0.274
			SI4	**0.756**	0.564	0.539	-0.068	0.005
			SI5	**0.846**	0.294	0.365	-0.296	0.011
			SI6	**0.818**	0.127	0.262	-0.463	0.066
WI	0.913	0.679	WI1	0.440	**0.878**	0.453	0.157	0.095
			WI2	0.460	**0.866**	0.427	0.026	0.089
			WI3	0.333	**0.782**	0.314	0.012	-0.046
			WI4	0.402	**0.820**	0.420	0.040	-0.083
			WI5	0.288	**0.768**	0.293	0.018	0.079
SP	0.836	0.563	SP1	0.520	0.679	**0.861**	-0.093	0.248
			SP2	0.426	0.270	**0.665**	-0.163	0.083
			SP3	0.317	0.170	**0.712**	-0.224	0.431
			SP4	0.267	0.109	**0.749**	-0.148	0.350
PA	0.745	0.595	PI1	-0.177	0.126	-0.099	**0.716**	-0.171
			PI2	-0.355	-0.005	-0.195	**0.822**	-0.081
SDI	0.838	0.633	SDI1	0.235	0.027	0.280	-0.204	**0.821**
			SDI2	-0.091	0.058	0.300	0.006	**0.734**
			SDI3	0.058	0.002	0.295	-0.164	**0.830**

Table 4. Correlation Matrix

	Mean	SD	VIF	SI	WI	SP	PA	SDI
SI	3.20	0.926	1.294	**0.789**				
WI	2.49	1.011	1.294	0.477	**0.824**			
SP	3.13	0.799	1.040	0.527	0.474	**0.750**		
PA	3.21	0.832	1.040	-0.355	0.069	-0.196	**0.771**	
SDI	3.64	0.826	-	0.091	0.035	0.365	-0.157	**0.796**
SD=Standard Deviation; Bold diagonal elements are square roots of AVE								

We then assessed multicollinearity by calculating the variance inflation factors (VIF). All values, ranging from 1.040 to 1.294, were below the threshold of 3.3 (Kock, 2015), indicating that multicollinearity is less of

an issue. Moreover, common method bias was examined using Harman's single-factor test (Podsakoff et al., 2003). The first factor accounted for 26.8% of the variance, which is below the 50% threshold, suggesting common method bias is less of a concern in this study.

5.2. Structural Model

A bootstrapping procedure with 5000 resamples was performed to determine the significance of path coefficients (β) and the coefficients of determination (R^2) for dependent variables. Figure 2 illustrates the results, supporting four of five hypotheses. Similarity identification was positively associated with self-presence (β = 0.390, p < 0.001) and negatively associated with perceived anonymity (β = -0.503, p < 0.01). Wishful identification was positively correlated with both self-presence (β = 0.289, p < 0.01) and perceived anonymity (β = 0.309, p < 0.05). Self-presence was positively associated with self-disclosure intention (β = 0.397, p < 0.01), while no significant correlation was identified between perceived anonymity and self-disclosure intention (β = -0.081, p > 0.05).

We also examined the effect of control variables. Past privacy violation experience was negatively associated with the self-disclosure intention (β = -0.320, p < 0.05), while perceived motion sickness was positively associated with the self-disclosure intention (β = 0.241, p < 0.05). Other control variables showed no significant effect on self-disclosure intention.

The model's explanatory power (R^2) was 34.3% for self-presence, 20.0% for perceived anonymity, and 28.7% for self-disclosure intention. All Q^2 values for dependent variables were above zero, indicating the model's predictive relevance (Hair et al., 2021).

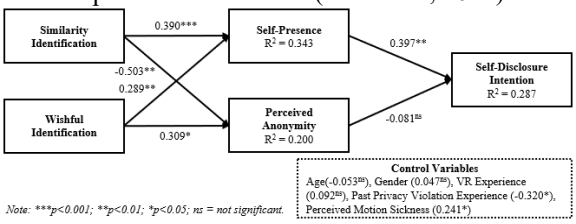

Figure 2. Results of the Research Model

6. Discussion

This study examined how avatar identification influences self-disclosure intention in metaverse mental healthcare. In a between-subjects experiment conducted in VR using VRChat to simulate realistic consultations, we found that both similarity and wishful identification positively impact self-presence, which in turn positively affects self-disclosure. While both identification types also influenced perceived anonymity, perceived anonymity did not predict self-disclosure intention.

The significant effects of both avatar identification types on self-presence align with recent metaverse marketing research (Kim et al., 2025), suggesting that reducing gaps between the actual and virtual self and between the actual and ideal self enhances the avatar's role as an extension of the self. In line with previous studies on Second Life (Hooi & Cho, 2014; Midha & Nandedkar, 2012), similarity identification significantly reduced perceived anonymity, indicating that psychological closeness to avatars increases senses of identifiability. In contrast, the relationship between wishful identification and perceived anonymity has not been explored in other contexts. The significant positive effect may reflect that the use of idealized avatars can escape reality (Fraser et al., 2023) and foster a sense of alienation from one's actual self. Mirroring prior virtual world research (Hooi & Cho, 2014), self-presence predicts self-disclosure intention, highlighting the influence of immersive, absorptive mental states on willingness to share sensitive information.

The non-significant effect of perceived anonymity on self-disclosure intention in our study contrasts with findings from non-immersive virtual worlds in leisure contexts (Hooi & Cho, 2014), online mental health communities (e.g., McLellan et al., 2022; Morini et al., 2025), and metaverse group therapies (Digul et al., 2021). Empirical findings on anonymity and self-disclosure are often inconsistent, suggesting that the relationship is context-dependent, influenced by specific definitions, measurement methods, and interaction settings (Clark-Gordon et al., 2019). Unlike prior studies that focused on peer interactions, our research examined a less explored one-on-one patient-physician interaction. In such direct encounters, the absence of peer observers may lessen the relevance of perceived anonymity, as concerns about self-image are reduced. Additionally, this specific context often involves a baseline of trust in the physician, which may further encourage self-disclosure. While trust in medical professionals is not unconditional even in face-to-face interactions (Atanasova et al., 2018), individuals tend to disclose sensitive personal information more openly online, especially towards avatars (Pickard et al., 2016). Nevertheless, the individual decision to self-disclose personal information in stigmatized mental health contexts, whether anonymously or identifiably, may involve a more complex trade-off. While anonymity can protect against criticism or shame and lower inhibition, identifiability can help individuals to be perceived as more credible and trustworthy by others, impacting the support that might be received (Andalibi et al., 2016).

Analysis of control variables revealed that prior experience of privacy violations and motion sickness significantly influenced self-disclosure intention. While the former aligns with existing research, the latter raises

ethical concerns, as physical discomfort may heighten the sense of acute vulnerability, prompting them to disclose more to seek help or expedite the session.

6.1. Theoretical Contributions

Our study offers several valuable contributions to the literature on metaverse mental healthcare, avatar identification, and technology-mediated self-disclosure. Specifically, we address a critical gap by integrating and extending prior work on avatar identification and self-disclosure into the context of psychologically sensitive interactions within metaverse-based mental healthcare.

First, this study contributes to understanding how avatars influence behavioral intentions, shedding light on the avatar's impact on the actual self in VR-based therapy. Prior literature on avatar identification has largely focused on leisure contexts, which differ in motivation and engagement from serious applications (Birk et al., 2016) such as mental healthcare. Moreover, much of this work was conducted in non-immersive desktop and mobile virtual worlds, such as Second Life and Zepeto (e.g., Hooi & Cho, 2014; Kim et al., 2025). Yet, our findings indicate that many effects observed in non-immersive leisure contexts also apply to VR-based serious applications. This suggests that the research and design of emerging serious metaverse applications can draw on the extensive knowledge established in entertainment-focused virtual environments.

Second, prior research on avatar identification was mainly survey-based, relying on participants' memory and imagination of past behavior (e.g., Kim et al., 2025; Midha & Nandedkar, 2012). Therefore, by conducting a controlled lab experiment within an immersive VR setting that simulated realistic mental healthcare interactions and capturing immediate post-experience responses, this study delivers more accurate evidence.

Third, this study extends understanding of self-concept in the metaverse by distinguishing between two different avatar identification types. While prior research has mainly focused on similarity identification (e.g., Hooi & Choi, 2014; Midha & Nandedkar, 2012), our findings highlight the often-overlooked role of wishful identification. This is especially relevant, as users often blend real and ideal self-aspects in avatars to satisfy both self-verification and self-enhancement needs (Messinger et al., 2019), and younger users tend to favor idealized avatars (Kim et al., 2025).

Fourth, this study contributes to the understanding of key predictors of online self-disclosure. It highlights the significant role of self-presence, aligning with prior findings from desktop-based virtual worlds (Hooi & Cho, 2014). Moreover, this research advances the contextual understanding of how anonymity affects online self-disclosure (Clark-Gordon et al., 2019; Joinson, 2001). While anonymity is widely recognized as a key facilitator of self-disclosure in various digital settings, our results indicate that its influence is less salient in private, one-on-one patient-physician metaverse mental healthcare settings. Previous vignette research (Turan Akdag & Wahl, 2024) identified perceived anonymity, manipulated via avatar similarity, as a driver of satisfaction and subsequent use intention of metaverse therapy. However, in trusted therapeutic interactions, perceived anonymity may not influence self-disclosure and therefore may not support the effectiveness of metaverse mental health consultations.

6.2. Practical Implications

Our study offers several practical insights for stakeholders involved in metaverse mental healthcare.

First, application designers should be cautious in relying on anonymity as a universal driver of self-disclosure in metaverse mental healthcare, as its effect may be more context-dependent. While prior research in traditional digital contexts emphasized anonymity's benefits, our findings from a confidential one-on-one patient-physician setting suggest its influence may be limited. Instead, fostering a stronger sense of self-presence appears more critical, which can be shaped by both similarity and wishful avatar identification. Therefore, designers should integrate flexible avatar customization tools that allow users to create avatars they genuinely identify with. They should also enhance self-presence by incorporating additional factors such as realistic sensory feedback, real-time full-body tracking, or highly interactive, visually appealing environments that closely mirror offline settings (Kim et al., 2025).

Second, clinicians should rethink metaverse mental healthcare procedures, rather than simply replicating offline methods. They must recognize the subconscious psychological effects of avatar uses and guide patients to become aware of them. Patients should have enough time to create avatars they are comfortable using, as strong avatar identification enhances self-presence, which in turn promotes self-disclosure. This is vital, as accurate diagnosis and effective treatment heavily rely on patient self-disclosure. In addition, physicians should carefully select physical exercises for virtual sessions, as patients may experience motion sickness.

Third, while we studied self-disclosure as a desirable therapeutic behavior, its enhancement through avatar identification is a double-edged sword with potential privacy and ethical risks. Strong avatar identification may be exploited to manipulate users into oversharing sensitive personal information, increasing their vulnerability to social engineering (Xie et al., 2024). Therefore, policymakers should mandate serious metaverse applications such as in healthcare contexts to

clearly inform users about the psychological effects of avatar choices, supporting informed decisions and protecting against manipulative design practices.

6.3. Limitations and Future Research

Like any research, our study presents limitations and offers avenues for future research. First, although the required sample size was met, the generalizability of our findings may be limited by a relatively skewed composition. Despite diverse recruitment channels, most participants were young university graduates. Future research should aim for more diverse samples, including patients with actual mental health conditions.

Second, the lab experiment's physical nature made full anonymity between participants and experiment staff impossible, as in-person contact was unavoidable. This likely created a basic level of trust towards the staff involved in the experiment. Yet, our findings remain valuable, as digital healthcare is expected to accompany, not replace, in-person interactions (Atanasova et al., 2018; Mora et al., 2008). Future research should extend this study into a field experiment, where both patient and physician identities remain unknown to each other.

Third, platform-specific traits of VRChat may limit the findings' generalizability. For instance, comic-like avatars from Ready Player Me tend to be preferred in stressful, serious situations, as they can mask negative traits, boost confidence (Laskowitz & Huffstadt, 2023), and influence self-disclosure through the Proteus Effect (Yee et al., 2009). Their playful nature may reduce anxiety around sharing sensitive information (Van FC et al., 2025). Future studies should investigate the effect of different platforms and avatar styles on self-disclosure.

Fourth, our study focused solely on self-presence and anonymity to explain self-disclosure intention. Given the moderate explained variance, future research should explore additional factors, such as other forms of presence, and variations in the physician's avatar, and perceived empathy in patient-physician interactions.

Fifth, our study did not investigate peer interactions between patients. As group therapy is common and may become popular in the metaverse, future research should explore these settings, since the presence of peer avatars could influence self-disclosure in different ways.

Acknowledgements
The underlying research project PRIME is funded by the German Federal Ministry of Research, Technology and Space (BMFTR) under funding code 16KIS1894K.

8. References

Andalibi, N., Haimson, O. L., De Choudhury, M. & Forte, A. (2016). Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. CHI Conf. Hum. Factor Comput. Syst.

Atanasova, S., Kamin, T., & Petric, G. (2018). The benefits and challenges of online professional-patient interaction: Comparing views between users and health professional moderators in an online health community. Computers in Human Behavior, 83, 106-118.

Baccon, L. A., Chiarovano, E., & MacDougall, H. G. (2019). Virtual Reality for Teletherapy: Avatars May Combine the Benefits of Face-to-Face Communication with the Anonymity of Online Text-Based Communication. Cyberpsychol. Behav. Soc. Netw., 22(2), 158-165.

Baghaei, N., Chitale, V., Hlasnik, A., Stemmet, L., Liang, H.-N., & Porter, R. (2021). Virtual Reality for Supporting the Treatment of Depression and Anxiety: Scoping Review. JMIR Mental Health, 8(9), e29681.

Behm-Morawitz, E. (2013). Mirrored selves: The influence of self-presence in a virtual world on health, appearance, and well-being. Comput. Hum. Behav., 29(1), 119-128.

Bessière, K., Seay, A, F., & Kiesler, S. (2007). The ideal elf: Identity exploration in World of Warcraft. CyberPsychology & Behavior, 10(4), 530-535.

Birk, M. V., Atkins, C., Bowey, J. T., & Mandryk, R. L. (2016). Fostering Intrinsic Motivation through Avatar Identification in Digital Games. ACM CHI, 2982-2995.

Clark-Gordon, C. V., Bowman, N., Goodboy, A., & Wright, A. (2019). Anonymity and Online Self-Disclosure: A Meta-Analysis. Communication Reports, 32(2), 98-111.

Dellazizzo, L., Potvin, S., Luigi, M., & Dumais, A. (2020). Evidence on Virtual Reality-Based Therapies for Psychiatric Disorders: Meta-Review of Meta-Analyses. Journal of Medical Internet Research, 22(8), e20889.

Derlega, V. J., & Chaikin, A. L. (1977). Privacy and Self-Disclosure in Social Relationships. J. Soc. Issues, 33(3).

Dilgul, M., Hickling, L. M., Antonie, D., Priebe, S., & Bird, V. J. (2021). Virtual Reality Group Therapy for the Treatment of Depression: A Qualitative Study on Stakeholder Perspectives. Frontiers in Virtual Reality, 1.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. J. Mark. Res., 18(1), 39-50.

Fraser, R., Slattery, J., & Yakovenko, I. (2023). Escaping through video games: Using our avatar to find meaning in life. Comput. Hum. Behav., 144, 107756, 1-12.

Freeman, G., & Maloney, D. (2020). Body, Avatar, and Me: The Presentation and Perception of Self in Social Virtual Reality. ACM Hum. Comput. Interact., 4(CSCW3), 239.

Hair, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2021). A primer on partial least squares structural equation modeling (PLS-SEM) (3rd ed.). Sage Publications.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115-135.

Herbert, C., Marschin, V., Erb, B., Meißner, D., Aufheimer, M., & Bösch, C. (2021). Are You Willing to Self-Disclosure for Science? Effects of Privacy Awareness and Trust in Privacy on Self-Disclosure of Personal and Health Data in Online Scientific Studies – An Experimental Study. Frontiers in Big Data, 4, 763196.

Hooi, R., & Cho, H. (2014). Avatar-driven self-disclosure: The virtual me is the actual me. Comput. Hum. Behav.

Jin, S.-A. A. (2012). The virtual malleable self and the virtual identity discrepancy model: Investigative frameworks for virtual possible selves and others in avatar-based identity construction and social interaction. Computers in Human Behavior, 28(6), 2160-2168.

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. Eur. J. Soc. Psychol., 3(2), 177-192.

Kim, M., Oh, H. J., Choi, J. H., & Jung, Y. (2025). Decoding millennials and generation Z consumers' brand behaviors in the Metaverse: The relationships among avatar identification, self-presence, and psychological dynamics. Journal of Consumer Behaviour, 24(1), 44-57.

Klimmt, C., Hefner, D., & Vorderer, P. (2009). The Video Game Experience as "True" Identification: A Theory of Enjoyable Alterations of Players' Self-Perception. Communication Theory, 19(4), 351-373.

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. Int. J. e-Collab., 11(4).

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. Inf. Syst. J., 28(1).

Kock, N. (2024). WarpPLS User Manual: Version 8.0. Laredo.

Kwak, K., Choi, S., & Lee, B. (2014). SNS flow, SNS self-disclosure and post hoc interpersonal relations change: Focused on Korean Facebook user. Computers in Human Behavior, 31, 294-304.

Laskowitz, L., & Huffstadt, K. (2023). Between Reality and Fiction: Self-Representation as an Avatar and Its Effects on Self-Presence. IJCAI, 193-203.

Lee, K. (2004).Presence, Explicated. Commun. Theory, 14(1).

Maloney, D., Zamanifard, S., & Freeman, G. (2020). Anonymity vs. Familiarity: Self-Disclosure and Privacy in Social Virtual Reality. ACM VRST, 25, 1-9.

Mattila, O., Korhonen, A., Pöyry, E., Hauru, K., Holopainen, J., & Parvinen, P. (2020). Restoration in a virtual reality forest environment. Comput. Hum. Behav., 107, 106295.

McLellan, A., Schmidt-Waselenchuk, K., Duerksen, K., & Woodin, E. (2022). Talking back to mental health stigma: An exploration of YouTube comments on anti-stigma videos. Computers in Human Behavior 131, 107214.

Messinger, P. R., Ge, X., Smirnov, K., Stroulia, E., & Lyons, K. (2019). Reflections of the extended self: Visual self-representation in avatar-mediated environments. Journal of Business Research, 100, 531-546.

Midha, V., & Nandedkar, A. (2012). Impact of similarity between avatar and their users on their perceived identifiability: Evidence from virtual teams in Second Life platform. Comput. Hum. Behav., 28(3), 929-932.

Mora, L., Nevid, J., & Chaplin, W. (2008). Psychologist treatment recommendations for Internet-based therapeutic interventions. Comput. Hum. Behav., 24(6).

Morini, V., Sansoni, M., Rossetti, G., Pedreschi, D., & Castillo, C. (2025). Participant behavior and community response in online mental health communities: Insights from Reddit. Comput. Hum. Behav., 165, 108544.

Oberdörfer, S., Birnstiel, S., & Latoschik, M. (2024). Proteus effect or bodily affordance? The influence of virtual high-heels on gait behavior. Virtual Reality, 28(2), 81.

Orr, E., Arbel, T., Levy, M., Sela, Y., Weissberger, O., Liran, O., & Lewis, J. (2023). Virtual reality in the management of stress and anxiety disorders: A retrospective analysis of 61 people treated in the metaverse. Heliyon, 9, e17870.

Pickard, M. D., Roster, C. A., & Chen, Y. (2016). Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? Comput. Hum. Behav., 65, 23-30.

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. J. Appl. Psychol., 88(5), 879-903.

Riches, S., Williams, G., & Saidel, S. (2024). The Metaverse as a Therapeutic Space: Opportunities for Public Health Benefits and Clinician-Industry Partnerships. Int. J. Hum. Comput. Interact., 40(21), 6825-6828.

Scott, C. R. (1998). To reveal or not to reveal: A theoretical model of anonymous communication. Commun. Theory.

Suler, J. (2004). The Online Disinhibition Effect. CyberPsychology & Behavior, 7(3), 321-326.

Sykownik, P., Maloney, D., Freeman, G., & Masuch, M. (2022). Something Personal from the Metaverse: Goals, Topics, and Contextual Factors of Self-Disclosure in Commercial Social VR. ACM CHI, 632, 1-17.

Takano, M., & Taka, F. (2022). Fancy avatar identification and behaviors in the virtual world: Preceding avatar customization and succeeding communication. Computers in Human Behavior Reports 6, 100176, 1-12.

Turan Akdag, M., & Wahl, N. (2024). Behind the Veil: Exploring Anonymity's Effect in Emerging Metaverse for Mental Health Therapy. HICSS.

Van FC, L., L., Suci, A., Defit, S., Wang, H.-C., & Purwanegara, M. S. (2025). Dare to blow the whistle? The role of cute avatars in a university's whistleblowing system. Computers in Human Behavior, 166, 108567.

Van Looy, J., Courtois, C., De Vocht, M., & De Marez, L. (2012). Player Identification in Online Games: Validation of a Scale for Measuring Identification in MMOGs. Media Psychology, 15(2), 197-221.

Wang, L., Yan, J., Lin, J., & Cui, W. (2017). Let the users tell the truth: Self-disclosure intention and self-disclosure honesty in mobile social networking. International Journal of Information Management, 37(1), 1428-1440.

Williams, A., Nielsen, E., & Coulson, N. (2018). "They aren't all like that": Perceptions of clinical services, as told by self-harm online communities. J. Health Psychol., 25.

Xie, R., Kirchner-Krath, J., & Morschheuser, B. (2024). Towards an Ethical Metaverse: A Systematic Literature Review on Privacy Challenges. Proceedings of the European Conference on Information Systems (ECIS), 6.

Yee, N., Bailenson, J. N., & Ducheneaut, N. (2009). The Proteus Effect: Implications of Transformed Digital Self-Representation on Online and Offline Behavior. Communication Research, 36(2), 285-312.

Yuchao, W., Ying, Z., & Liao, Z. (2021). Health Privacy Information Self-Disclosure in Online Health Community. Frontiers in Public Health, 8, 602792.

Zahedi, F. M., Zhao, H., Sanvanson, P., Walia, N., Jain, H., & Shaker, R. (2022). My Real Avatar has a Doctor Appointment in the Wepital: A System for Persistent, Efficient and Ubiquitous Medical Care. Inf. Manag., 59.