

The Impact of Human-Likeness and Self-Disclosure on Message Acceptance in Virtual AI Influencers

Cherie Sew
The University of Melbourne
Melbourne, Australia
csew@student.unimelb.edu.au

Saumya Pareek
The University of Melbourne
Melbourne, Australia
saumya.pareek@student.unimelb.edu.au

Jarod Govers
The University of Melbourne
Melbourne, Australia
jarod.govers@unimelb.edu.au

Sarah Schömbbs
The University of Melbourne
Melbourne, Australia
sschombs@student.unimelb.edu.au

Ryan M. Kelly
RMIT University
Melbourne, Australia
ryan.kelly@rmit.edu.au

Jorge Goncalves
The University of Melbourne
Melbourne, Australia
jorge.goncalves@unimelb.edu.au

Abstract

Virtual AI-generated Influencers (VAIs) are increasingly being used by corporations and public agencies, raising questions about how their visual design and communication strategies impact end-users' propensity to accept the messages they deliver. We examined the impact of human-likeness (how closely a VAI resembles a human) and self-disclosure (whether the message contains personal information) on message acceptance, alongside dispositional factors like empathy and anthropomorphising tendencies. In a mixed-methods experiment, participants (N=120) watched short-form videos featuring VAIs of varying human-likeness (High/Moderate-High/Moderate-Low/Low) and self-disclosure (present/absent). We observed the strongest message acceptance from the VAIs with the lowest human-likeness, and message rejection for VAIs with moderate to low human-likeness. Additionally, participants' message acceptance was influenced by their empathy tendencies. Our qualitative analysis revealed further insights into participants' perceptions of the human-likeness of VAIs, their discomfort with self-disclosure, and their tendency to anthropomorphise VAIs. These findings provide important implications for the design of VAIs.

CCS Concepts

• Human-centered computing → Empirical studies in HCI.

Keywords

Virtual Influencers, Message Acceptance, Human-likeness, Self-disclosure, Empathy, Artificial Intelligence, Social Media

ACM Reference Format:

Cherie Sew, Saumya Pareek, Jarod Govers, Sarah Schömbbs, Ryan M. Kelly, and Jorge Goncalves. 2025. The Impact of Human-Likeness and Self-Disclosure on Message Acceptance in Virtual AI Influencers. In *Designing Interactive Systems Conference (DIS '25)*, July 5–9, 2025, Funchal, Portugal. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3715336.3735756>

1 Introduction

Virtual influencers have recently emerged as a new trend in the digital landscape, providing unique advantages over traditional online influencers [7]. Virtual influencers are computer-generated characters with personas that are carefully curated by an individual or a management company. People typically encounter these influencers on short-form social media platforms such as Instagram and TikTok. Notably, younger individuals are displaying an interest in these influencers, with those between the ages of 18–44 reportedly following at least one virtual influencer on social media [62]. The advent of generative artificial intelligence (GenAI) has led to the creation of virtual AI influencers (VAIs)—a type of virtual influencer whose voice, words, and likeness are entirely generated by AI. Recent advancements have facilitated the creation of VAIs by simplifying and automating the generation of graphical images, characters, and related content, lowering the barrier to entry for those interested in using VAIs to influence the views of others.

The growing popularity and reach of virtual influencers has attracted the attention of organisations beyond corporate or retail sectors. For instance, during the height of the COVID-19 pandemic, the World Health Organization (WHO) collaborated with a virtual influencer to spread awareness related to the pandemic and disseminate critical public health information [16]. Furthermore, the Singapore Tourism Board recently advertised a role for a collaborator to develop a VAI focused on creating travel content for their country [56]. These examples underscore the increasing willingness of public agencies to leverage virtual influencers as a means to reach tech-savvy and/or younger audiences.

However, despite the potential for virtual influencers to achieve higher levels of engagement compared to human influencers with similar follower counts [2], their credibility remains a challenge [10, 43]. As such, researchers propose using emotional narratives to increase the perceived credibility of virtual influencers [33]. For human influencers, sharing intimate content that contains appropriate self-disclosure, the act of revealing personal information, thoughts, or feelings [36, 37], can improve their credibility and online engagement with followers [32]. Similarly, prior research on virtual agents has found that empathising with users by responding with self-disclosure can be an effective way of increasing perceived intimacy and enjoyment [31] as well as accepting the messages being delivered by the virtual agents [64]. Likewise, virtual influencers participate in self-disclosure through online content as a form of

storytelling in order to build a connection with their followers [48]. However, there is limited research on the effects of self-disclosure on message acceptance for VAIIs.

Moreover, virtual influencers are not confined to a single, human-like appearance. Depending on the influencer's theme and persona, their visual representation can range from a highly realistic human form to an animated object with humanoid features. However, prior work has often grouped *all* virtual influencers into a single category of human-likeness, with the purpose of comparing their effectiveness to that of traditional human influencers [40, 74]. However, this approach overlooks the fact that variations in the degree of visual human-likeness can exist even within the category of virtual influencers designed to mimic humans. The spectrum ranges from virtual characters that closely resemble real humans to those that take the form of 2D, animated, cartoon-like characters.

In this study, we aim to investigate the factors that influence the acceptance or rejection of messages delivered by VAIIs. Specifically, we examine how different levels of visual human-likeness, and the inclusion of self-disclosure in the message, affect message acceptance. Furthermore, we explore the role of dispositional factors, such as empathy and the tendency to anthropomorphise technology, in shaping individuals' perceptions of VAIIs. We pose the following research questions:

RQ1: How do different levels of human-likeness in VAIIs affect individuals' acceptance of their messages?

RQ2: How does the presence or absence of self-disclosure in messages from VAIIs impact individuals' acceptance of those messages?

RQ3: How do dispositional factors, such as empathy and the tendency to anthropomorphise, influence the relationship between human-likeness or self-disclosure and message acceptance?

With these research questions guiding our work, we conducted an online survey-based experiment with 120 participants. We generated the stimuli used in our study with real Generative AI tools including Midjourney and ChatGPT-4. For the purpose of this paper, we refer to the stimuli as VAIIs for clarity and consistency. Each participant viewed four short-form videos, each featuring a VAII delivering a message. To explore the effects of human-likeness, we manipulated the appearance of each VAII character, ensuring a range of human-likeness (High, Moderate-High, Moderate-Low, Low), while maintaining overall consistency in presentation. Additionally, we generated two video versions for each message topic: one with self-disclosure and one without it. Before participants were exposed to our stimuli, we measured their dispositional cognitive empathy, affective empathy, empathic response to fantasy characters, and tendency to anthropomorphise. After displaying each stimuli, we measured participants' message acceptance.

Our results indicate that human-likeness plays a significant role in message acceptance, with the strongest message acceptance observed for the VAII with the lowest human-likeness, appearing as a 2-D animated character. Interestingly, participants were less likely to accept the message from the moderate-low human-like VAII, which was presented as a 3-D cartoon character. Further, participants' dispositional factors, specifically cognitive and affective empathy, influenced their message acceptance, with these effects varying based on the level of human-likeness presented.

This study makes the following contributions for usage of VAIIs which can also be broadly applied to virtual influencers. First, we discuss how our findings on varying degrees of human-likeness for different kinds of VAIIs can be utilised by commercial entities and governmental or public agencies. Further, we reflect on the impact of the Uncanny Valley effect on High human-like VAIIs, and how advances in VAII design or video animation can lead to higher message acceptance. Third, we highlight the need for caution when using self-disclosure as a strategy to enhance message acceptance for VAIIs due to the disbelief that VAIIs can understand human experiences. Fourth, we demonstrate the positive effects of dispositional empathy on message acceptance, and discuss balancing visual human-likeness and eliciting empathetic responses for better message acceptance. We conclude by discussing the implications of VAII usage.

2 Related Work

2.1 Virtual Human and AI Influencers

Virtual humans are computer-generated avatars designed to exhibit human-like behaviours and characteristics. The concept of virtual humans dates back to the 1950s, with early examples including humanoid figures that were used to simulate various scenarios for the aeronautical industry [17]. Over the decades, virtual humans have found diverse applications across different industries, such as digital twins of patients for medical purposes, animated characters in media, and virtual assistants for the elderly.

More recently, virtual humans have made their way onto social media platforms, expanding their influence beyond traditional animations in advertisements. The first virtual human with beginnings on a social platform was Ami Yamato - a virtual vlogger on YouTube presented as a 3D animation [49]. The launch of Lil' Miquela Sousa on Instagram in 2016 further cemented the potential of virtual humans as influencers, as she amassed a significant following (2.6 million on Instagram) and collaborated with various brands and celebrities. Partnerships between corporations and virtual influencers are growing due to the flexibility, exclusivity, brand safety, and perceived innovation that these virtual influencers provide [12]. Consequently, numerous virtual influencers have been created in an attempt to capture a similar success [12]. Moreover, while some agencies creating these virtual influencers brand them as 'AI influencers', their 'storyline' or content remains reliant on the direction of human teams.

However, there are several virtual influencers with AI-generated content who are actual AI influencers. While these influencers still require human programming and moderation, their reliance on human intervention is much more limited compared to virtual influencers like Lil' Miquela or Aitana [34, 54]. For instance, Kuki AI is a VAII chatbot that allows users to chat with her 'live' video, wherein her 3-D human avatar answers questions on video as well as through the chatbox [26]. Similarly, Neuro-sama is a 2-D virtual streamer with AI-powered chat capabilities and the ability to play games like Minecraft or Osu [71]. Although the four aforementioned influencers currently fall under the broad category of virtual influencers, Kuki AI and Neuro-sama distinguish themselves with their limited reliance on humans for content generation. Further, as Generative AI (GenAI) technologies have streamlined the process of

creating and generating virtual influencers, as well as their content, making them more accessible, this is likely to lead to a rise in VAIIs similar to the partially autonomous Kuki AI and Neuro-sama.

Existing literature primarily examines popular virtual influencers without differentiating between those driven by human input and those powered by AI. Despite this, their findings remain important as they enable us to understand the implications of using virtual influencers. For example, Lu et al. [35] found that audiences separate the real human from the character they portray when engaging with VTubers—a type of virtual influencer where a human controls a Japanese animation-inspired avatar. The study also revealed that audiences have different expectations for the character and the human actor behind it. The reason for this distinction could be rooted in escapism, as followers—especially those with higher empathy—may seek emotional diversion from everyday life through their engagement with virtual influencers [40]. Furthermore, these virtual influencers are fundamentally technology posing as social actors, with the purpose of eliciting social responses from their audience [18, 52].

In addition, previous work has found that humans are particularly sensitive towards the “humanness” and uncanniness of virtual influencers [46]. Interestingly, they also found that people are less influenced by perceived trust when deciding to follow virtual influencers. Instead, emotional reactions, or how these influencers make them feel, has a more significant impact on motivating people to follow them. The Uncanny Valley is a phenomenon where people feel discomfort from entities that approach, but do not fully obtain, human-like appearance [42]. However, Choudhry et al. [10] posit that the absence of feature abnormalities—flaws in the design or animation—could mitigate the Uncanny Valley effect as they found that individuals continue to engage with virtual influencers even when they were perceived as *‘too real’*.

2.2 Message Acceptance

Prior work evaluated the effectiveness of messages across numerous factors, including source credibility [24, 47, 51, 70], psychological strategies on how to convey a message [23], the persuasiveness of the content and its structure [8, 23, 58], and the attitude towards the messages themselves [55]. Attitude towards messages is often measured as message acceptance or message rejection. Moreover, there are situational or dispositional factors that may impact individuals’ perception of messages [1, 23]. Situational factors refer to external circumstances that can influence perceptions, such as self-disclosure [45]. Conversely, dispositional factors are those inherent to an individual, such as how their personality traits, characteristics, or tendencies may influence their trust or persuasion from online messages [11, 21, 23, 29, 63].

2.2.1 Situational Factors Impacting Message Acceptance. One potential situational factor that can impact message acceptance is self-disclosure, defined as the act of sharing personal information or experiences with others [37]. Self-disclosure can vary in valence, intention, as well as frequency [36]. Previous literature found that messages from unknown actors are perceived as more authentic when they include self-disclosure [45], leading to higher message acceptance. In addition, self-disclosure by AI chatbots has been found to enhance the positive effects of providing emotional support to

conversation partners [38], as well as increase perceived intimacy and enjoyment [31]. Tsumura and Yamada’s [64] study on embodied AI chatbots further supports the importance of self-disclosure by anthropomorphic agents for user acceptance, regardless of the agent’s appearance. Further, the same study hypothesised that self-disclosure with high relevance to the specific context in which the virtual agent operates can lead to better user acceptance, compared to self-disclosure that is less contextually relevant.

Additionally, self-disclosure can significantly impact parasocial interaction with a virtual influencer, regardless of whether their message is emotionally positive or negative [33]. However, the same study also found that parasocial interaction is correlated with the disclosure that the virtual influencer is not a real human, particularly when the valence of the message is also taken into account (i.e., parasocial interaction reduces when the message is negative and participants are aware that the virtual influencer is not human). This phenomenon is exemplified by the real-world scenario in which Lil’ Miquela faced backlash in 2019 after sharing a ‘personal’ experience of being sexually assaulted, which was met with criticism due to her virtual nature [59].

Importantly, the aforementioned study by Lim and Lee [33] examined a virtual influencer characterised by a high degree of human-likeness, presenting content through still images and written text on Instagram. In contrast, Lu et al. [35] explored virtual influencer streamers who, rather than striving for human-likeness, adopt Japanese anime-inspired personas. This approach allowed for greater freedom in expressing negative sentiments. Additionally, previous literature has found that perceived human-likeness has a positive influence on the formation of the social presence, defined as ‘the sense of being with another’ by Biocca and Harms [4], which can directly influence trust [30, 44]. Thus, we extend this previous work by exploring how the degree of human-likeness in VAIIs affects the acceptance of their messages.

2.2.2 Dispositional Factors Impacting Message Acceptance. A dispositional factor that may impact an individual’s message acceptance is their empathy. The concept of empathy is a well-researched topic, with the term first introduced over a century ago by Titchener [68]. According to a meta-analysis on empathy by Cuff et al. [13], despite inconsistencies in the academic definitions of empathy, it can be divided into two types: affective empathy (i.e., emotional contagion) and cognitive empathy (i.e., understanding the emotions of a stimulus). Previous work has found that affective empathy can facilitate perceived interactions with on-screen performers [14]. Further, individuals with high affective and cognitive empathy have been found to be more willing to follow virtual influencers [40]. However, this study did not differentiate between empathy toward human influencers and virtual influencers, who are arguably fictional characters.

Another potential dispositional factor is an individual’s tendency to anthropomorphise. The degree to which individuals attribute human characteristics to technology-based objects is a well-researched area, especially for robots and virtual agents. Within the space of virtual influencer research, Yang et al. [73] found that people reported stronger trust in virtual influencers that were highly anthropomorphic compared to cartoon-like virtual influencers due to the perception that they can control their actions. However, Yan et al.

[72] suggest that there is a limit on the level of anthropomorphism a virtual influencer should embody as social presence decreases when the virtual influencers become extremely anthropomorphic. In addition, as research on the impact of virtual influencers is still in its early stages, there are limited studies examining how individuals' tendency to anthropomorphise impacts their perception of virtual influencers and the content they deliver.

Thus, in our research, we examine the impact of varying degrees of human-likeness and self-disclosure on the acceptance of messages delivered by VAIIs specifically. Further, we investigate how dispositional empathy and tendency to anthropomorphise technology influence individuals' perceptions of messages from VAIIs.

3 Method

We conducted a mixed-methods study using a counterbalanced design. Our study aimed to investigate the impact of human-likeness and self-disclosure on message acceptance from VAIIs. To achieve this, we created four virtual AI influencers (VAIIs) that differed in their degree of human-likeness (see Figure 1). Participants watched the VAIIs deliver spoken messages in four short-form videos, where two videos used self-disclosure and two did not. All of our stimuli were created using AI tools, mirroring the manner in which VAIIs would be created in practice.

3.1 Experimental Stimuli

3.1.1 VAI Character. To create the VAIIs for our study, we began by using the Generative AI tool Midjourney¹ to create four static digital characters with varying degrees of human-likeness. According to IZEA's Influencer Equality report [28], the majority of participants in influencer marketing identify as female. Given these findings, and considering that the majority of TikTok creators are reportedly aged 18-24 [27], we determined that our VAIIs should be female, and appear to be in their 20s. This demographic and visual alignment ensures that our VAIIs reflect the characteristics of influencers that users are most likely to come across on social media platforms.

To ensure consistency across the four VAIIs, we controlled different aspects of their appearance and presentation, keeping everything except for our independent variable, the degree of human-likeness, identical in the prompts given to Midjourney. The consistent characteristics included ethnicity (east asian), facial appearance (natural, no makeup), hair colour and style (neat up-do, dark brown to black hair), as well as the VAIIs' positioning (head and body facing camera), and type of portrait (half-length). For the human-likeness, we adjusted the prompt wording for each VAI by specifying the following descriptors, listed in decreasing order of human-likeness: 'female human', 'female avatar', '3D cartoon character', and '2D animation character'. Additionally, the realism of the VAIIs was specified, with the 'female human' being 'hyper-realistic' while the other three digital renders as 'slightly realistic' to prevent Midjourney from producing renders that were overly imaginative. While each VAI generated is visually human-like, they represent different degrees of human-likeness within the spectrum of looking human.

Some of these differences are depicted visually in the VAI's skin texture and the facial portions. For example, the High VAI resembles a human and has an uneven skin-tone with natural lighting and facial asymmetry, while the other three VAIIs have glossier, almost plastic skin and symmetrical faces. Further, for the Moderate-High VAI, the facial proportions are more grounded in human anatomy whereas the Moderate-Low VAI has exaggerated features that have been stylised similar to Pixar animated characters. Finally, the Low VAI is highly stylised, also resembling a cartoon character with exaggerated features but with simplified 2D textures. The prompts used for generating each VAI can be found in Appendix A.

3.1.2 Self-disclosure Messages. We used ChatGPT-4² to produce the message content for our study. We prompted the language model to generate popular topics that online influencers might create content on, in order to share personal experiences or opinions to connect with their audience. From this list of topics, we selected 'Personal Challenges and Growth' as our focus, and further prompted ChatGPT to craft messages as if it were an online influencer sharing a video on this topic, including personal experiences. We generated four variations of messages within this topic: Overcoming Fear of Failure, Embracing Change, Finding Balance, and Nurturing Self-Compassion. These message contents contained several instances of first-person references such as 'I've learned (...)' or 'I felt (...)'. Further, we used ChatGPT to create another version of each message *without* self-disclosure or personal anecdotes—giving us a total of eight AI-generated messages. The message content can be found in Appendix B.

3.1.3 VAI Audio and Video. After creating the four VAI characters and the message scripts, we used ElevenLabs³ to produce the text-to-speech human voice audio with the message content. Due to the within-subjects nature of our study, we opted to select different voice profiles for each VAI to avoid potential biases that may arise if participants recognised the same voice across all stimuli. As the participants of our study were based in the U.S., we selected voice profiles labelled as having American accents, all with a clear and deliberate narration tone.

Finally, we combined the VAIIs and the audio narration using HeyGen⁴, an AI video generator platform. HeyGen works by receiving the static images and animating the VAIIs while trying to match the mouth movement to the uploaded audio. This resulted in a total of eight videos for the main experiment (4 VAIIs x 2 levels of self-disclosure).

3.2 Manipulation and Confounding Checks

Prior to the main experiment, we conducted manipulation checks to ensure our VAIIs and the messages reflected the prescribed degrees of perceived human-likeness and self-disclosure. Further, we conducted confounding checks to ensure that messages with and without self-disclosure did not inadvertently affect the perceived appropriateness and valence of the content. These confounding checks are necessary to ensure that perceived appropriateness and message valence did not have significant differences, as that can

¹<https://www.midjourney.com/home>

²<https://chatgpt.com/>

³<https://elevenlabs.io/>

⁴<https://www.heygen.com/>

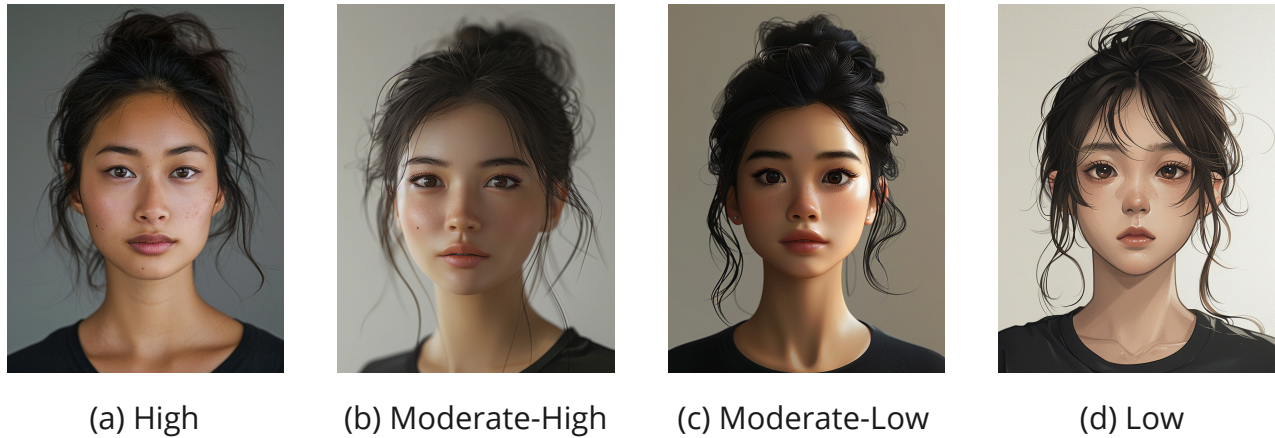


Figure 1: VAIs created for the purpose of this study. From left to right: High human-like (H), Moderate-High human-like (MH), Moderate-Low human-like (ML), Low human-like (L).

interfere with the interpretation of the manipulation. This isolates the aforementioned confounding factors from the influences of self-disclosure on the messages [57].

Following Nah's [45] method of validating self-disclosure messages, we confirmed the presence of self-disclosure in our messages by measuring their perceived intimacy. Intimacy [45] of the message was measured using two items, each with a 7-point Likert scale (1 = *not at all*, 7 = *very*), with the expectation that messages with self-disclosure should be perceived as significantly more intimate than messages that did not contain self-disclosure. Appropriateness [3] was measured using four items, each with a 7-point semantic differential scale, while valence [45] was measured with a single item on a 7-point semantic differential scale. According to previous literature, non-verbal cues from the messenger can influence perceptions of the message itself [60]. Thus, we measured the messages' intimacy, appropriateness, and valence separate from the videos in order to prevent confounds.

Table 1: Manipulation and confounding check results from Welch's Two Sample t-tests for Message Intimacy, Appropriateness, and Valence.

	Self-disclosure video		No-disclosure video	
	M	SD	M	SD
Manipulation check				
Message intimacy***	4.60	1.38	3.55	1.49
Confounding check				
Message appropriateness	5.63	1.04	5.81	1.07
Message valence	6.15	0.71	6.09	0.78

We used Prolific⁵ to recruit a total of 48 independent raters and compensated them based on the hourly rate recommended by the platform (\$ 15.66 USD) at the time of data collection. Half of the

raters reviewed the messages with self-disclosure, and the other half reviewed the four messages without self-disclosure. After reading each message, the raters were required to complete three surveys measuring perceived intimacy of the message, appropriateness of the message, and valence of the message, respectively.

Table 1 displays the descriptive statistics from our manipulation and confounding check results. The manipulation for the self-disclosure messages was verified as our raters evaluated the messages with self-disclosure as significantly more intimate (Welch's $t(183.08) = 4.98$, $p < 0.001$, $d = 0.730$) than those without self-disclosure.

Additionally, no significant difference was observed in message appropriateness, with both conditions being rated as appropriate (Welch's $t(181.71) = -1.12$, $p = 0.266$, $d = -0.165$). Similarly, the valence of messages with and without self-disclosure was rated to be positive, with no statistically significant difference between the two (Welch's $t(179.93) = 0.61$, $p = 0.546$, $d = 0.089$). These results together validated the intimacy of messages with self-disclosure, and confirmed the intended lack of change in appropriateness and valence in messages with absent/present self-disclosure.

Next, we recruited another group of 41 raters to evaluate the human-likeness of the generated VAI videos compared to our intended categorisation. Participants were split into two groups: one group viewed videos of our VAIs with the self-disclosure messages, while the other group viewed videos without. Each participant was asked to rank the videos from most human-like to least human-like. The videos were randomised to mitigate order effects.

We measured the level of agreement by comparing the raters' ranking with our designated human-likeness classification. Figure 2 shows the frequency with which each VAI was assigned to the intended level of human-likeness by the raters. Based on the results, we concluded that the manipulations of self-disclosure as well as human-likeness were successful, allowing us to confidently use the stimuli in our experiment.

⁵<https://www.prolific.com/>

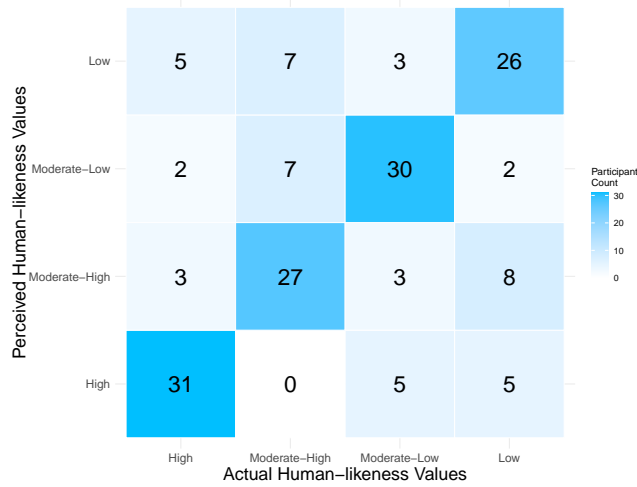


Figure 2: Degree of human-likeness confusion matrix from our manipulation check

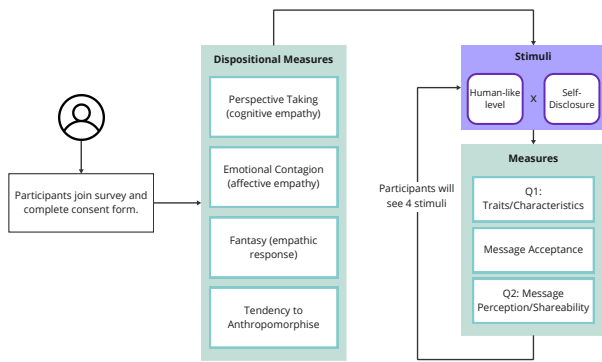


Figure 3: Experimental procedure of our study

3.3 Participants

We recruited a total of 120 U.S.-based participants for our main study through the Prolific platform with a mean age of 39.7 (SD = 11.1). We filtered for participants who have an approval rating of $\geq 99\%$, and at least 500 Prolific submissions. The median time taken to complete the study was 22 minutes, and participants were compensated with \$ 15.66 USD per hour. The study was approved by the Human Ethics Committee of our university.

3.4 Procedure

Figure 3 shows the experimental procedure we adopted for this study. The survey began with a plain language statement that informed participants what the study was about. This was followed by measuring participants' dispositional empathy and anthropomorphising tendencies. To ensure a robust measurement of empathy, we drew inspiration from previous studies that separately measured participants' cognitive empathy, affective empathy, and relevant

empathic responses [64]. We measured cognitive empathy and empathic response by adapting the subscales of perspective taking and fantasy respectively, from the Interpersonal Reactivity Index (IRI) [15]. Both subscales contained 7 items measured on a 5-point Likert scale ('Strongly Disagree' to 'Strongly Agree'). Affective empathy was assessed using the emotional contagion subscale that was adapted from Stiff et al. [61]. The tendency to anthropomorphise was measured with a 5-item, 11-point numerical rating scale (0 = *not at all*; 10 = *very much*), adapted from the technology device subclass of the bigger Individual Differences in Anthropomorphism Questionnaire (IDAQ) [67].

Subsequently, we briefed the participants that they would be watching four short-form videos. In our instructions, we asked participants to imagine that they were scrolling through social media and had come across these 'lifestyle influencers' sharing life advice relevant to the participants' current experience of facing personal challenges. Next, participants saw each of the four videos, which were counter-balanced to include all possible sequences of our stimuli to mitigate order effects. To prevent participants from skipping the videos, we included a 30 second delay timer for the display of the 'Next' button. After each video, participants answered an open-ended question, Q1: 'What traits or characteristics do you find most appealing or off-putting about [VAII Character]?'.

Following this, we measured the dependent variable—**message acceptance**—with a 3-item, 5-point likert scale adapted from Shen [55]. Participants were asked to indicate their agreement for the following items: 'I support what the message was trying to accomplish', 'I agree with the position advocated in the message', and 'I am favourable toward the main point of the message'.

Further, participants were asked another open-ended question, Q2: 'How do you feel about the message shared by [VAII Character]? Would you share it with a friend? Please elaborate.' to understand their thoughts on the message delivered by the VAII. Finally, participants were thanked with a short message and redirected back to Prolific.

3.5 Analysis

3.5.1 Quantitative Analysis. As the data collected is ordinal, we utilised a Cumulative Link Mixed Model (CLMM) to investigate the relationships between message acceptance and each of the following factors: human-likeness (RQ1), presence/absence of self-disclosure (RQ2) and dispositional factors (RQ3). Further, we used Estimated Marginal Means (EMM) to analyse the effects of our predictors on message acceptance to account for population dynamics compared to raw means. Our model was also tested for multicollinearity by computing the Variance Inflation Factors (VIF) across all predictors, with all values below 5 which indicated an absence of multicollinearity [25].

3.5.2 Qualitative Analysis. Two independent coders analysed the qualitative data using an inductive approach. The analysis aimed to understand how the VAIIs' human-likeness and self-disclosure may have affected message acceptance and sought to identify additional issues that may explain participants' appraisal of the VAII. The two coders began by familiarising themselves with the data independently to ensure a deep understanding of its contents. Following an inductive approach, they identified a set of high-level themes based

on the data, which served as a framework for clustering the codes. Afterwards, they met to discuss and refine these themes, addressing any discrepancies to ensure consistency in the coding process. The qualitative responses were grouped into the following themes: the realism of a VAI, participants' resonance with the messages, and their perceived human-like attributes of a VAI.

4 Results

We addressed our three research questions through one CLMM to predict participants' message acceptance by VAIs using the following variables:

- Human-likeness: High (H), Moderate-High (MH), Moderate-Low (ML), Low (L)
- Self-disclosure: Present, Absent
- Cognitive Empathy: Perspective Taking. Averaged from 7-item Likert scale.
- Affective Empathy: Emotional Contagion. Averaged from 7-item Likert scale.
- Fantasy: Empathic response towards fictional characters. Averaged from 7-item Likert scale.
- Tendency to Anthropomorphise: Averaged from 5-item rating scale.

Next, we describe our main findings in detail with regards to each of the research questions. The full model results can be found in Appendix C.

4.1 Quantitative Findings

4.1.1 RQ1: Influence of Human-likeness on Message Acceptance. Our results indicate that human-likeness is a main effect for participants' message acceptance. Figure 4 displays the statistically significant estimated contrasts between the different human-likeness levels from our model. We report the coefficients from our logit link CLMM models which we present as standardised effect sizes using Cohen's d . Additionally, we present the magnitude of change via estimated marginal mean values, as this approach allows us to understand the magnitude of the effect of human-likeness on message acceptance.

For the low human-like VAI, we observe a significant positive effect (Cohen's $d = 0.330$, $SE = 0.115$, $p < 0.01$). This implies that participants had a higher probability of accepting the message when they encountered the low human-like VAI (EMM = 4.66, $SE = 0.361$), as shown in Figure 4.

Conversely, we observed a negative effect on message acceptance for the moderate-low human-like VAI. This effect was statistically significant (Cohen's $d = -0.336$, $SE = 0.111$, $p < 0.01$). Thus, participants rejected the message when the messenger's human-likeness was moderately-low (EMM = 3.99, $SE = 0.350$), as depicted in Figure 4.

4.1.2 RQ2: Impact of Self-disclosure on Message Acceptance. We did not observe a statistically significant impact of self-disclosure on the acceptance of the messages ($\beta = 0.296$, $SE = 0.309$, $p = 0.337$).

4.1.3 RQ3: Impact of Dispositional Factors on Message Acceptance. We observed that participants' cognitive empathy impacted their message acceptance as visualised in Figure 5. Cognitive empathy,

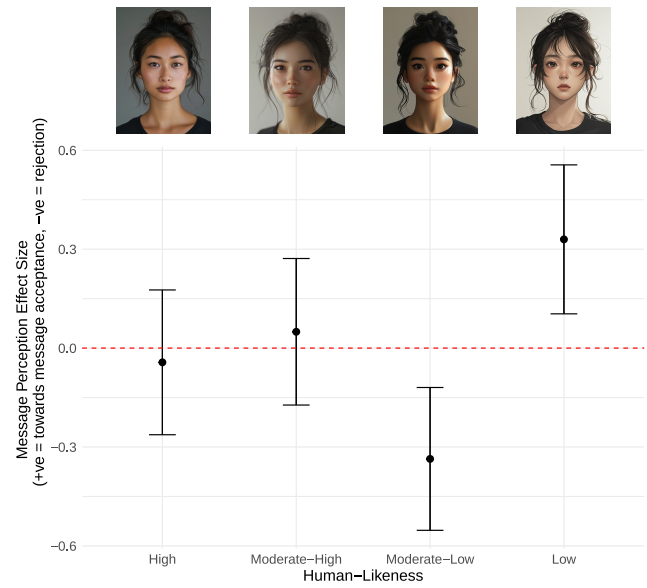


Figure 4: Comparison of effect size for different levels of human-likeness

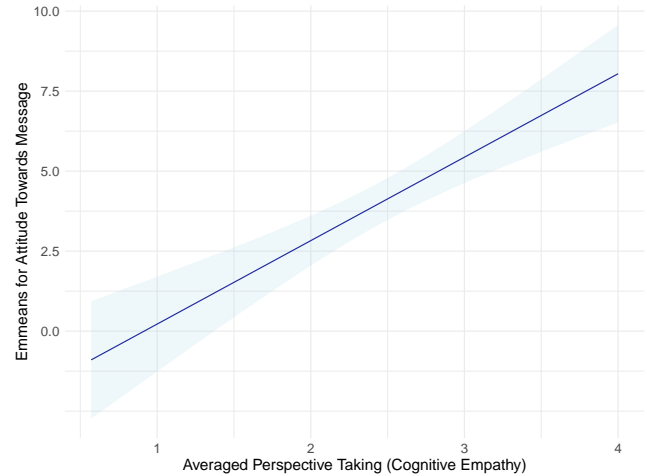


Figure 5: Estimated Marginal Means of Message Acceptance across participants' reported tendency for Perspective Taking

which was measured as a dispositional factor of participants' reported tendency of perspective taking, was statistically significant for participants' decision for message acceptance ($\beta = 2.608$, $SE = 0.460$, $p < 0.001$).

Additionally, Figure 6 highlights the trends of participants' predicted message acceptance when both participants' affective empathy and the human-likeness of the VAIs were taken into account. Affective empathy was measured as participants' tendency of emotional contagion. We measured the moderating effect of human-likeness on the tendency of emotional contagion as affective empathy can be caused by the visual exposure to the emotional

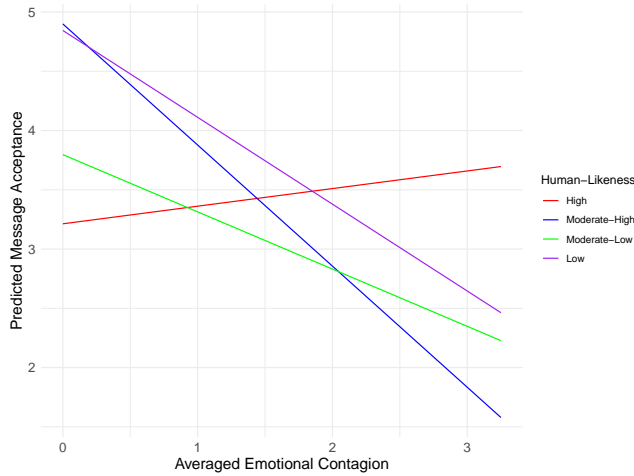


Figure 6: Predicted message acceptance across participants' reported tendency for emotional contagion and human-likeness

expressions of others [20]. We observe a significant negative effect of affective empathy for the moderate-high VAI on message acceptance in Figure 6.

Subsequently, we conducted a pairwise contrast to examine the differences between the levels of human-likeness—with affective empathy—on message acceptance, as shown in Table 2. We observed that participants with higher reported affective empathy in the form of emotional contagion were more likely to accept the message from the high human-like VAI compared to the moderate-high human-like VAI (EMM = 1.170, SE = 0.282, $p < 0.001$). Similarly, participants with higher reported emotional contagion were more likely to accept the message from the high human-like VAI compared to the low human-like VAI (EMM = 0.881, SE = 0.283, $p = 0.01$). The empathic response of fantasy and tendency to anthropomorphise were not significant factors in message acceptance.

Table 2: Contrasts of model estimates

Contrast	Estimate	SE	p-value
H - MH***	1.170	0.282	< 0.001
H - ML	0.632	0.270	0.090
H - L**	0.881	0.283	0.010
MH - ML	-0.539	0.275	0.202
MH - L	-0.289	0.283	0.737
ML - L	0.250	0.272	0.795

4.2 Qualitative Findings

Our qualitative analysis revealed that participants had differing perceptions about each of the VAIs. Participants were reluctant to accept messages from the High human-likeness VAIs, which they perceived as being 'too realistic' and therefore masquerading as a person. Messages shared by the Low human-like VAI were seen

as more acceptable because this VAI was clearly distinguishable from a human. Regarding self-disclosure, we found that participants viewed the VAIs as unable to have human experiences, potentially explaining why self-disclosure did not have an effect on message acceptance. Lastly, we found evidence that the participants tended to anthropomorphise the VAIs, attributing empathetic personality traits and evidencing a tendency towards dispositional perspective taking.

4.2.1 Realism of the VAIs. Our analysis suggests that the realism of the VAIs was perceived negatively by some participants, potentially explaining the lack of message acceptance among the VAIs with higher human-likeness. For example, after viewing the Moderate-High VAI's video, P92 said that the character "doesn't look real, but is trying to be. It makes me uncomfortable." (P92, Moderate-High, No-SD). In fact, the increasing realism of the VAIs created a sense of discomfort: "This one was more realistic, and I found that to be even more unsettling than the previous ones." (P9, High, No-SD).

Participants' discomfort over the realism of the VAIs could explain their message acceptance for the Low human-like VAI. For example, P60 commented that "because you can tell she was not a person, it wasn't as bad as previous virtual humans." (P60, Low, SD). In fact, some found the artistic style of the Low human-likeness VAI to be more appealing: "even though you can tell it was an AI, I liked the style of it better" (P87, Low, No-SD) and "I like the art style" (P12, Low, SD). Participants also found the VAI to be relatable when mirrored with their personal experience: "I liked her messy appearance. Reminded me what I look like when facing difficulties." (P116, Low, SD). The perceived age of the VAI was also noted as something that made them relatable: "She looks like an anime character. Her voice matches her looks. She looks like someone close to my age so it is easier to relate to her." (P81, Low, SD).

We also identified the possible reason participants reject messages from the Moderate-Low human-like VAI. Participants perceived the Low human-like VAI as attempting to mimic a human: "[VAI] is a cartoon character who pretends to be a person and attempts to give serious life advice. I find it hard to find something to like about her." (P65, Moderate-Low, No-SD). Further, the style of the VAI mattered as participants disliked the 3-D animated design of the Moderate-Low human-like VAI: "Now [VAI] was the worst digital rendering with too large of a head. Her entire face looked like a Pixar animation, the least realistic besides the fixated eyes and the out-of-sync mouth to speech." (P30, Moderate-Low, No-SD).

Furthermore, the artificial nature of the VAI disrupted participants' sense of realism as well. Although some participants perceived the messages positively, they were unsettled by the VAIs regardless of whether they were the most human-like: "I found the message appealing but the AI image unsettling." (P18, High, No-SD) or the least human-like: "The message is good, but the character is AI. It's off-putting." (P26, Low, No-SD). Some participants mentioned the uncanny valley effect as a possible explanation for their feelings: "The uncanny valley effect made it difficult for me to concentrate on the message." (P3, Moderate-High, SD).

4.2.2 Resonance with the Message. Participants' responses suggest that the presence of self-disclosure was not taken seriously because of the belief that the VAIs are unable to have human experiences. For example, P55 said that "all I could think of was that she was

making up things she had experienced." (P55, High, SD). Likewise, participants struggled with the idea that a VAI could have a human experience: *"I think the message is fine and support it, but coming from an AI it means nothing, because the hardships it describes it can never understand at its current level"* (P66, Moderate-Low, SD).

Nevertheless, participants did view the content of the messages as agreeable and logical, even if they were reluctant to accept them from some of the VAIs. For instance, P41 wrote: *"I agree with the message and can relate to it. I would also share this with a friend if they needed to hear it."* (P41, Moderate-High, No-SD). P7 found some of the messages relatable: *"The message resonate with me and yes, I would share it with a friend."* (P7, High, SD). Some participants acknowledged that while the messenger was in this case AI-generated, the message should still be shared for its perceived positive value: *"Honestly yeah I would share it with a friend. Although it's AI generated, the message is good. She's clearly trying to spread positively and that's a good thing, especially coming from AI."* (P4, Moderate-High, No-SD).

However, some participants are less inclined to share the video as they state their preference for a human messenger: *"The message should be delivered by a human if it involves emotion. I would not share with a friend."* (P14, Moderate-High, SD). Some stated they would share the message separately from the VAI messenger: *"I wouldn't share it because it's an AI but I might tell my friends what [the VAI character] said but put it in my own words."* (P29, Low, No-SD).

4.2.3 Perception of Human-like Attributes in the VAIs. Although we did not observe statistical significance in our quantitative results for participants' tendency to anthropomorphise, there were nevertheless traces of anthropomorphisation in participants' comments about the VAIs. Participants often referred to the characters using personal pronouns as well as assigning a theory of mind to the VAIs through perceiving them as having their own beliefs: *"I like that [the VAI character] believes in showing love towards oneself and not being so hard on yourself."* (P67, Moderate-High, No-SD).

Besides the physical attributes of the VAIs, we found indications of participants perceiving human attributes in the VAIs. Some participants responded to the messages positively, attributing their reactions to the VAIs' perceived authenticity: *"I would share it but only with a close friend. she delivers her message with simplicity and authenticity, therefore it's easy to comprehend and to agree with."* (P118, Moderate-High, SD), with a participant pointing out that this could induce feelings of relatability: *"[...] She sounded interested and authentic in her expression of change and it would definitely be something that would resonate with people."* (P112, High, SD).

Moreover, the VAIs were perceived as confident: *"I found [VAI]'s confidence and straight-forwardness appealing [...]"* (P108, Moderate-High, SD); intelligent: *"[VAI] understood what needs to be done in our fast paced world."* (P7, Moderate-Low, SD); and honest: *"I liked how [VAI] was extremely honest and open about her experiences regarding fear of failure."* (P93, Low, SD). Further, participants attributed personality traits to the VAIs which aligned with the messages, like empathy: *"She seems energetic, empathetic and more in touch with society."* (P35, High, No-SD) as well as compassion: *"She came across as compassionate and caring I found her to be believable."* (P17, Moderate-High, No-SD), which complements our

results in which participants' dispositional perspective taking was statistically significant.

On the other hand, some participants perceived negative traits from the VAIs, like not being genuine: *"She doesn't seem very genuine, not likeable."* (P70, Moderate-Low, No-SD) or having a patronising attitude: *"The condescending tone is off-putting."* (P42, High, SD).

5 Discussion

5.1 Impact of Human-Likeness (RQ1) and Self-Disclosure (RQ2) on Message Acceptance

We found that human-likeness has a significant impact on message acceptance (RQ1). Our results indicate that the Low human-like VAI held the strongest message acceptance, which corroborates existing claims that less human-like virtual influencers are more advantageous for social media users [43]. Beyond finding the 2D art style more appealing, our participants were more willing to accept the message from the VAI when they perceived the VAI as not attempting to visually mimic human appearance. Interestingly, within the lower spectrum of human-likeness, we observed an adverse reaction to message acceptance for the Moderate-Low VAI where participants perceived it as attempting to mimic a human. We found that participants were more sensitive to the 3D animation style of the VAI, which supports previous literature that found a negative relationship between audiences' emotional bond and perceived immersion with realistic 3D characters as opposed to less realistic characters [6].

In addition, our participants also experienced the Uncanny Valley effect as they directly referred to the phenomenon as well as perceiving some of the more human-like VAIs as 'creepy' or 'unsettling'. Current technologies are limited as AI-driven animation has not arrived at the point of being fully realistic or indistinguishably human. Hence, our results support the findings of existing research that the Uncanny Valley effect experienced by participants is further exacerbated by slight design and movement abnormalities—otherwise known as feature abnormalities—in the VAIs [10, 53]. As Generative AI technologies continue to advance, social media users are more likely to accept messages from High human-like VAIs, particularly if abnormalities with AI video animation of human-like VAIs are eliminated.

We did not find a statistically significant effect of self-disclosure on the acceptance of VAIs' messages (RQ2). Previous research has shown that self-disclosure by virtual agents is an effective method for inducing relatability and acceptance from people [64]. However, our qualitative analysis revealed that participants referenced VAIs as being unable to have had a lived human experience. The human-like appearances and feature abnormalities of the VAIs could have been a factor for participants' disbelief that VAIs can comprehend human experiences or hardships. Notably, participants still found the message content to be agreeable, and they would share the messages, provided it was coming from a human messenger. Hence, when High human-like VAIs become indistinguishable from human influencers, self-disclosure has the potential to have a significant positive effect on message acceptance.

5.2 Influence of Dispositional Factors (RQ3) on Message Acceptance for VAIs

We found that perspective taking—cognitive empathy—had a significant positive effect on message acceptance. Even though participants held strong beliefs that VAIs have the inability to undergo human experiences, many still managed to relate strongly to the VAI and the message they delivered. Participants with higher cognitive empathy perceived positive values from the messages and supported the distribution, as well as perceived the VAIs as possessing beliefs or values on the message they deliver. This supports previous work that found people with high cognitive empathy have a greater willingness to follow virtual influencers [40].

However, participants with higher reported affective empathy—measured as emotional contagion—had a higher probability of accepting messages from the High human-like VAI compared to the Moderate-High human-like VAI and Low human-like VAI. This finding aligns with existing literature suggesting that individuals with high affective empathy are generally more receptive to emotionally expressive agents [66]. Moreover, although we did not observe any statistical significance for the tendency to anthropomorphise predictor, our qualitative analysis indicated that participants strongly anthropomorphise the VAIs. Our participants consistently referred to them with feminine pronouns as opposed to the neutral form of ‘it’ for an inanimate object. One possible explanation for the lack of significance is that this measure, which has been used to assess the tendency of individuals to anthropomorphise an array of technological agents, may not adequately capture the nuances of individuals’ tendency to anthropomorphise VAIs. This highlights the need for the development of a more comprehensive measure specific to virtual characters in online communities. While existing literature suggested that self-disclosure in a parasocial interaction will increase message acceptance [45], our participants displayed signs of parasocial interaction with the VAIs regardless of self-disclosure in the message delivered by the VAIs. Participants attributed human personalities to the VAIs, which was occasionally based on the message content delivered by the VAIs. The act of attributing characteristics to the VAIs led some participants to further personify them as a mentor or friend.

5.3 Implications for the Use of VAIs

VAIs offer a new avenue to provide marketable digital mascots designed to propagate messages in a scalable and customisable manner. The application of virtual influencers itself is becoming increasingly frequent, with prior work indicating that anthropomorphic chatbots on company websites can increase a brand’s perceived likeability, trustworthiness, and likelihood to purchase a product [9, 63]. Our findings contribute to industry research by identifying that the uncanny effect of digital VAIs can undermine the acceptance of the messages they deliver, thus industries and promoters should seek *more highly* animated VAIs to help propagate their messages. Likewise, digital marketing should rely on evoking feelings of *empathy* and balanced visual human-likeness within their VAIs, as this further enhances message acceptance relevant for advertising. Likewise, VAIs remain most effective at propagating a convincing message when they self-recognise that they *are not capable of human experiences*. Our findings reiterate

those of Govers et al. [23], who identified that social media chatbots with perceived objectivity and ethereal non-human presence enhanced message acceptance.

Beyond commercial, our findings relate to the utility of anthropomorphic visual agents to act as audio-visual ‘message mascots’ to convey information—akin to mascots such as ‘smokey bear’ in the U.S. to link animals and fire safety [5], or ‘Sofia’, a digital ambassador to convey messages from the United Nations Development Programme [41]. In addition, public agencies and governments have been experimenting with the use of virtual influencers as a different method of engaging with their audience. It is particularly compelling for these stakeholders to use virtual influencers for information propagation as they can attract the attention from “anti-intellectuals” and elicit positive reactions towards messages they would otherwise disregard from experts [19]. Besides seeking collaboration with established virtual influencers, as demonstrated by the World Health Organization (WHO) [16], some organisations are also exploring the creation of their own VAIs. For instance, the Singapore Tourism Board (STB) is seeking to develop their own VAI for producing trustworthy destination content. Likewise, World Wildlife Fund (WWF) Spain experienced success in their collaboration with Meta to develop their own virtual influencer—Habichuela—to raise awareness on consumption behaviours as social media users reported changes in behaviour based on the content published by Habichuela [39].

Importantly, social media platforms should develop codes of practice and conduct which attribute human-responsibility to their virtual personas rather than targeting the technology itself. For instance, the underlying AI technology companies, such as Midjourney, GPT-4, and ElevenLabs, used in our study should not require editorial liability (i.e., the attribution of responsibility to any derivatives of their content, such as a VAI that promotes harmful messages, as occurred with Microsoft’s Tay experiment [69], and other adversarial bots that can mimic social content [21, 22]). Instead, VAIs should be accountable to their specific creator—given the variety of benevolent and malevolent use cases for audio-visual AI technology. Currently, few countries or platforms have policies for VAIs—an issue also seen by the lack of deepfake regulation/law [50]. Given that participants saw their VAI as “someone [they] would hang out with”, and “like a friend”, it is vital for research, industries, and governments to ascribe human-accountability to the creators of autonomous VAIs to prevent abuse.

5.4 Limitations and Future Work

While the VAIs were designed based on traditional influencer statistics, it is plausible that homophily may inform message acceptance. Future work should examine human-like VAIs with different demographics. Furthermore, our study deliberately used VAIs in order to compare the impact of situational and dispositional factors on message acceptance for VAIs. Although essential to the purpose of our study as our intention was to examine message acceptance across VAIs of different human-likeness, this may have limited the generalisability of our findings to message acceptance for human-like virtual influencers. Future work can explore message acceptance for VAIs contrasted with a human influencer baseline to examine the effectiveness of VAIs compared to a ‘traditional’ influencer.

Additionally, we also acknowledge the limitation of validating the reflected self-disclosure of our stimuli using an indirect method of measuring perceived intimacy as the manipulation check. Future research related to self-disclosure of VAIIs should consider an annotation approach to validate absent or present self-disclosure as a more direct method of measuring whether the manipulated self-disclosure is behaving as intended [65]. Finally, virtual influencers can come in the form of anthropomorphic animals or objects, which could be advantageous for suppressing the Uncanny Valley effect [10]. We encourage future research to investigate the differences in message acceptance between Low human-like VAIIs and non-human VAIIs for better message propagation.

6 Conclusion

In this study, we examined how human-likeness, self-disclosure, and individual differences influence the acceptance of messages from VAIIs. Our findings highlight the impact of different human-likeness in VAIIs on message acceptance. Further, we found that message acceptance from VAIIs is influenced by the Uncanny Valley effect. In addition, we note that contrary to previous work on virtual agents' self-disclosure for better acceptance of the agents [64], self-disclosure did not significantly impact message acceptance as participants believed that VAIIs are unable to understand human experiences. Our findings also highlight the positive moderating effects of users' dispositional cognitive empathy and affective empathy on message acceptance, with high cognitive empathy individuals more likely to accept messages from VAIIs, and high affective empathy individuals being more likely to accept messages from High human-like VAIIs over Moderate-High human-like or Low human-like VAIIs. We envision that VAIIs will play an important role in various future applications, including entertainment, marketing, and government or public agency communications. Our findings are highly valuable in these areas for effective message dissemination and engagement.

References

- [1] Irwin Altman and Dalmas A. Taylor. 1973. *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston, Oxford, England. viii, 212–viii, 212 pages.
- [2] Nick Baklanov. 2019. The Top Instagram Virtual Influencers in 2019 | Hype - Journal | HypeAuditor Blog. <https://hypeauditor.com/blog/the-top-instagram-virtual-influencers-in-2019/>
- [3] Natalya N Bazarova. 2012. Public intimacy: Disclosure interpretation and social judgments on Facebook. *Journal of Communication* 62, 5 (2012), 815–832.
- [4] Frank Biocca and Chad Harms. 2002. Defining and measuring social presence: Contribution to the networked minds theory and measure. *Proceedings of PRESENCE 2002* (2002), 7–36.
- [5] Lyndsie Bourgon. 2019. *A Brief History of Smokey Bear, the Forest Service's Legendary Mascot*. Smithsonian Magazine. <https://www.smithsonianmag.com/history/brief-history-smokey-bear-180972549>
- [6] Willie Bouwer and Francois Human. 2017. The impact of the uncanny valley effect on the perception of animated three-dimensional humanlike characters. *The Computer Games Journal* 6 (2017), 185–203.
- [7] Alison Bringé. 2024. Council Post: The Rise Of Virtual Influencers And What It Means For Brands. (Aug 2024). <https://www.forbes.com/councils/forbescommunicationscouncil/2022/10/18/the-rise-of-virtual-influencers-and-what-it-means-for-brands/>
- [8] Yu-Ting Chang, Hueiju Yu, and Hsi-Peng Lu. 2015. Persuasive messages, popularity cohesion, and message diffusion in social media marketing. *Journal of Business Research* 68, 4 (2015), 777–782.
- [9] Jengchung Victor Chen, Huyen Thi Le, and Sinh Thi Thu Tran. 2021. Understanding automated conversational agent as a decision aid: matching agent's conversation with customer's shopping task. *Internet Research* 31, 4 (2021), 1376–1404.
- [10] Abhinav Choudhry, Jinda Han, Xiaoyu Xu, and Yun Huang. 2022. "I Felt a Little Crazy Following a 'Doll'" Investigating Real Influence of Virtual Influencers on Their Followers. *Proceedings of the ACM on human-computer interaction* 6, GROUP (2022), 1–28.
- [11] Shauna Concannon and Marcus Tomalin. 2023. Measuring perceived empathy in dialogue systems. *AI & SOCIETY* (2023). <https://doi.org/10.1007/s00146-023-01715-z>
- [12] Mauro Conti, Jenil Gathani, and Pier Paolo Tricomi. 2022. Virtual Influencers in Online Social Media. *IEEE Communications Magazine* 60, 8 (2022), 86–91. <https://doi.org/10.1109/MCOM.001.2100786>
- [13] Benjamin MP Cuff, Sarah J Brown, Laura Taylor, and Douglas J Howat. 2016. Empathy: A review of the concept. *Emotion review* 8, 2 (2016), 144–153.
- [14] R Glenn Cummins and Boni Cui. 2014. Reconceptualizing address in television programming: The effect of address and affective empathy on viewer experience of parasocial interaction. *Journal of Communication* 64, 4 (2014), 723–742.
- [15] Mark H Davis. 1980. Interpersonal reactivity index. (1980).
- [16] Lindsay Dodgson. 2020. The World Health Organization has recruited a CGI influencer to get young people interested in safe practices around the coronavirus. <https://www.businessinsider.com/who-using-cgi-influencer-to-spread-safe-coronavirus-practices-2020-4>
- [17] William A Fetter. 1982. A progression of human figures simulated by computer graphics. *IEEE Computer Graphics and Applications* 2, 09 (1982), 9–13.
- [18] Andrew Gambino, Jesse Fox, and Rabindra A Ratan. 2020. Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication* 1 (2020), 71–85.
- [19] Maximilian HEE Gerrath, Hossein Olya, Zahra Shah, and Huaiyu Li. 2024. Virtual influencers and pro-environmental causes: the roles of message warmth and trust in experts. *Journal of Business Research* 175 (2024), 114520.
- [20] Amit Goldenberg and James J Gross. 2020. Digital emotion contagion. *Trends in cognitive sciences* 24, 4 (2020), 316–328.
- [21] Jarod Govers, Philip Feldman, Aaron Dant, and Panos Patros. 2023. Down the Rabbit Hole: Detecting Online Extremism, Radicalisation, and Politicised Hate Speech. *ACM Comput. Surv.* 55, 14s, Article 319 (jul 2023), 35 pages. <https://doi.org/10.1145/3583067>
- [22] Jarod Govers, Philip Feldman, Aaron Dant, and Panos Patros. 2023. Prompt-GAN—Customisable Hate Speech and Extremist Datasets via Radicalised Neural Language Models. In *Proceedings of the 2023 9th International Conference on Computing and Artificial Intelligence (Tianjin, China) (ICCAI '23)*. Association for Computing Machinery, New York, NY, USA, 515–522. <https://doi.org/10.1145/3594315.3594366>
- [23] Jarod Govers, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2024. AI-Driven Mediation Strategies for Audience Depolarisation in Online Debates. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3613904.3642322>
- [24] Bradley S Greenberg and Gerald R Miller. 1966. The effects of low-credible sources on message acceptance. *Communications Monographs* 33, 2 (1966), 127–136.
- [25] J. Hair, W. Black, B. Babin, and R. Anderson. 2010. *Multivariate data analysis* (7th ed.). Pearson, Upper Saddle River, NJ.
- [26] Astrid Hiort. 2023. Understanding The Role of AI and Virtual Influencers Today. <https://www.virtualhumans.org/article/understanding-the-role-of-ai-and-virtual-influencers-today>
- [27] Josh Howarth. 2022. TikTok User Age, Gender, & Demographics (2022). <https://explodingtopics.com/blog/tiktok-demographics#tiktok-users-age>
- [28] IZEA. 2023. The State of Influencer Equality, 2023 Report. <https://izea.com/resources/insights/2023-state-of-influencer-equality/>
- [29] Andreas Janson. 2023. How to leverage anthropomorphism for chatbot service interfaces: The interplay of communication style and personification. *Computers in Human Behavior* 149 (2023), 107954. <https://doi.org/10.1016/j.chb.2023.107954>
- [30] S Venus Jin and Seounmi Youn. 2023. Social presence and imagery processing as predictors of chatbot continuance intention in human-AI-interaction. *International Journal of Human-Computer Interaction* 39, 9 (2023), 1874–1886.
- [31] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I hear you, I feel you": encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.
- [32] Fernanda Polli Leite, Nicolas Pontes, and Paulo de Paula Baptista. 2022. Oops, I've Overshared! When Social Media Influencers' Self-Disclosure Damage Perceptions of Source Credibility. *Computers in Human Behavior* 133 (2022), 107274. <https://doi.org/10.1016/j.chb.2022.107274>
- [33] Rachel Esther Lim and So Young Lee. 2023. "You are a virtual influencer!": Understanding the impact of origin disclosure and emotional narratives on parasocial relationships and virtual influencer credibility. *Computers in Human Behavior* 148 (2023), 107897.
- [34] Laura Llach. 2024. Meet Spain's first AI model who earns up to €10,000 a month. <https://www.euronews.com/next/2024/05/04/meet-the-first-spanish-ai-model-earning-up-to-10000-per-month>

- [35] Zhicong Lu, Chenxinran Shen, Jiannan Li, Hong Shen, and Daniel Wigdor. 2021. More kawaii than a real-person live streamer: understanding how the otaku community engages with and perceives virtual YouTubers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [36] Mufan Luo and Jeffrey T Hancock. 2020. Self-disclosure and social media: motivations, mechanisms and psychological well-being. *Current opinion in psychology* 31 (2020), 110–115.
- [37] Mohan Masaviru. 2016. Self-disclosure: Theories and model review. *Journal of Culture, Society and Development* 18, 1 (2016), 43–44.
- [38] Jingbo Meng and Yue Dai. 2021. Emotional support from AI chatbots: Should a supportive partner self-disclose or not? *Journal of Computer-Mediated Communication* 26, 4 (2021), 207–222.
- [39] Meta. 2023. Meta for government and charities, case studies: World Wildlife Fund Spain. <https://www.facebook.com/government-nonprofits/case-studies/2023-07-24-WWF-Spain>
- [40] Agata Mirowska and Jbid Arsenyan. 2023. Sweet escape: The role of empathy in social media engagement with human versus virtual influencers. *International Journal of Human-Computer Studies* 174 (Jun 2023), 103008. <https://doi.org/10.1016/j.ijhcs.2023.103008>
- [41] Cedric Monteiro, Mahtab Haider, and Jeanne Lim. 2017. *UNDP in Asia and the Pacific Appoints World's First Non-Human Innovation Champion*. United Nations Development Programme. <https://web.archive.org/web/20180709173848/http://www.asia-pacific.undp.org/content/rbap/en/home/presscenter/pressreleases/2017/11/22/rbfsingapore.html>
- [42] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>
- [43] Evangelos Moustakas, Nishtha Lamba, Dina Mahmoud, and C Ranganathan. 2020. Blurring lines between fiction and reality: Perspectives of experts on marketing effectiveness of virtual influencers. In *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*. IEEE, 1–6.
- [44] J. Munnukka, D. Maity, H. Reinikainen, and V. Luoma-aho. 2019. Thanks for watching". The effectiveness of YouTube vlog endorsements. *Computers In Human Behavior* 93 (2019), 226–234. <https://doi.org/10.1016/j.chb.2018.12.014>
- [45] Hye Soo Nah. 2022. The appeal of "real" in parasocial interaction: The effect of self-disclosure on message acceptance via perceived authenticity and liking. *Computers in Human Behavior* 134 (2022), 107330.
- [46] Anika Nissen, Colin Conrad, and Aaron Newman. 2023. Are You Human? Investigating the Perceptions and Evaluations of Virtual Versus Human Instagram Influencers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [47] Chanthika Pornpitakpan. 2004. The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of applied social psychology* 34, 2 (2004), 243–281.
- [48] Makena Rasmussen. 2021. Why Storytelling Can Make or Break a Virtual Influencer. <https://www.virtualhumans.org/article/whystorytelling-can-make-or-break-a-virtual-influencer>
- [49] Makena Rasmussen. 2022. Who Was the First Virtual Influencer? <https://www.virtualhumans.org/article/who-was-the-first-virtual-influencer#:~:text=Despite%20being%20fictional%2C%20Max%20Headroom,influencers%20for%20the%20first%20time>
- [50] Andrew Ray. 2021. *UNIVERSITY OF NEW SOUTH WALES LAW JOURNAL* 44, 3 (2021), 983–1013. <https://search.informit.org/doi/10.3316/informit.20211005054413>
- [51] Cherie Sew and Jorge Goncalves. 2023. Effects of External Factors on User Behaviour towards Influencer Generated Online Misinformation. In *Proceedings of the 35th Australian Computer-Human Interaction Conference*. 86–98.
- [52] Cherie Sew, Sarah Schömbis, Ryan M. Kelly, and Jorge Goncalves. 2025. User Willingness to Engage with AI-Generated Influencer Content. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*.
- [53] Jun'ichiro Seyama and Ruth S Nagayama. 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence* 16, 4 (2007), 337–351.
- [54] Jasmine Sheena. 2023. Brands are still figuring out virtual influencers. <https://www.marketingbrew.com/stories/2023/09/12/brands-are-still-figuring-out-virtual-influencers>
- [55] Lijiang Shen. 2010. Mitigating psychological reactance: The role of message-induced empathy in persuasion. *Human Communication Research* 36, 3 (2010), 397–422.
- [56] Singapore Tourism Board. 2024. Virtual Influencer Open Call for Collaboration. <http://web.archive.org/web/20240910041223/https://www.stb.gov.sg/content/stb/en/trade-events-and-resources/virtual-influencer-open-call-for-collaboration.html>
- [57] Gary Smith. [n. d.]. Chapter 5 - Sampling. In *Essential Statistics, Regression, and Econometrics*, Gary Smith (Ed.). Academic Press, 141–162. <https://doi.org/10.1016/B978-0-12-382221-5.00005-2>
- [58] Stephen M Smith and Richard E Petty. 1996. Message framing and persuasion: A message processing analysis. *Personality and Social Psychology Bulletin* 22, 3 (1996), 257–268.
- [59] Sandra Song. 2019. Lil Miquela Criticized For "Sexual Assault" Vlog. <https://www.papermag.com/lil-miquela-sexual-assault-vlog#rebellitem11>
- [60] Michiel Spapé, Ville Harjunen, Imtiaz Ahmed, Giulio Jacucci, and Niklas Ravaja. 2019. The semiotics of the message and the messenger: How nonverbal communication affects fairness perception. *Cognitive, Affective, & Behavioral Neuroscience* 19, 5 (Jul 2019), 1259–1272. <https://doi.org/10.3758/s13415-019-00738-8>
- [61] James B Stiff, James Price Dillard, Lilnabeth Somera, Hyun Kim, and Carra Sleight. 1988. Empathy, communication, and prosocial behavior. *Communications Monographs* 55, 2 (1988), 198–213.
- [62] The Influencer Marketing Factory. 2024. Virtual Influencers Survey + INFOGRAPHIC. <https://theinfluencermarketingfactory.com/virtual-influencers-survey-infographic/>
- [63] Wan-Hsiu Sunny Tsai, Yu Liu, and Ching-Hua Chuan. 2021. How chatbots' social presence communication enhances consumer engagement: the mediating role of parasocial interaction and dialogue. *Journal of Research in Interactive Marketing* 15, 3 (2021), 460–482. <https://doi.org/10.1108/JRIM-12-2019-0200>
- [64] Takahiro Tsumura and Seiji Yamada. 2023. Influence of agent's self-disclosure on human empathy. *PLOS ONE* 18, 5 (May 2023), e0283955–e0283955. <https://doi.org/10.1371/journal.pone.0283955>
- [65] Mina Valizadeh, Pardis Ranjbar-Noiey, Cornelia Caragea, and Natalie Parde. 2021. Identifying medical self-disclosure in online communities. (2021).
- [66] Matias Volante, Sabarish V. Babu, Himanshu Chaturvedi, Nathan Newsome, Elham Ebrahimi, Tania Roy, Shaundra B. Daily, and Tracy Fasolino. 2016. Effects of Virtual Human Appearance Fidelity on Emotion Contagion in Affective Inter-Personal Simulations. *IEEE Transactions on Visualization and Computer Graphics* 22, 4 (2016), 1326–1335. <https://doi.org/10.1109/TVCG.2016.2518158>
- [67] Adam Waytz, John Cacioppo, and Nicholas Epley. 2010. Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science* 5, 3 (2010), 219–232.
- [68] Lauren Wispé. 1986. The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of personality and social psychology* 50, 2 (1986), 314.
- [69] M. J. Wolf, K. Miller, and F. S. Grodzinsky. 2017. Why we should have seen that coming: comments on Microsoft's tay "experiment," and wider implications. *SIGCAS Comput. Soc.* 47, 3 (sep 2017), 54–64. <https://doi.org/10.1145/3144592.3144598>
- [70] Paul CS Wu and Yun-Chen Wang. 2011. The influences of electronic word-of-mouth message appeal and message source credibility on brand attitude. *Asia Pacific Journal of Marketing and Logistics* 23, 4 (2011), 448–472.
- [71] Chloe Xiang. 2023. This Virtual Twitch Streamer is Controlled Entirely By AI. <https://www.vice.com/en/article/this-virtual-twitch-streamer-is-controlled-entirely-by-ai/>
- [72] Ji Yan, Senmao Xia, Amanda Jiang, and Zhibin Lin. 2024. The effect of different types of virtual influencers on consumers' emotional attachment. *Journal of Business Research* 177 (2024), 114646.
- [73] Jeongwon Yang, Ployyin Chuentarawong, Heejae Lee, and T Makana Chock. 2023. Anthropomorphism in CSR endorsement: A comparative study on humanlike vs. cartoonlike virtual influencers' climate change messaging. *Journal of Promotion Management* 29, 5 (2023), 705–734.
- [74] Jeongwon Yang, Ployyin Chuentarawong, Heejae Lee, Yu Tian, and T Makana Chock. 2023. Human versus virtual influencer: The effect of humanness and interactivity on persuasive CSR messaging. *Journal of Interactive Advertising* 23, 3 (2023), 275–292.

A Avatars

We generated the avatars of our stimuli using Midjourney with the following prompts:

High human-like

an east asian female human in her 20s, with light skin, natural, no make up, russet brown colored eyes, neat updo hairstyle, dark brown to black hair, wearing plain black t-shirt, image in close-up portrait, shallow depth of field to emphasize subject, subject head and body directly facing camera, gray blurred background, hyper realistic photography, half body, color correction –ar 9:16

Moderate-High human-like

an east asian female avatar in her 20s, with light skin, natural, no make up, russet brown colored eyes, neat updo hairstyle, dark brown to black hair, wearing

plain black t-shirt, image in half-length portrait, shallow depth of field to emphasize subject, subject head and body directly facing camera, uncanny valley, slightly realistic, half body, color correction -ar 9:16

Moderate-Low human-like

an east asian female cartoon character in her 20s, with light skin, natural, no make up, russet brown colored eyes, neat updo hairstyle, dark brown to black hair, wearing plain black t-shirt, image in half-length portrait, shallow depth of field to emphasize subject, subject head and body directly facing camera, slightly realistic, half body, color correction -ar 9:16

Low human-like

an east asian female human in her 20s, with light skin, natural, no make up, russet brown colored eyes, neat updo hairstyle, dark brown to black hair, wearing plain black t-shirt, image in half-length portrait, shallow depth of field to emphasize subject, subject head and body directly facing camera, slightly realistic, half body, color correction -ar 9:16

B Message Content

We generated the message content of our stimuli using ChatGPT (GPT-4) with the following prompt:

Imagine you are an influencer, give me four different variations you might say related to personal challenges and growth in a 30 second video in an attempt to relate to your followers. You should have two versions with each variation: one with a self-disclosure message in which you are sharing something very personal about yourself, like a personal experience or personal opinion, and a second version that's the same message but stripped of anything personal.

The generated messages used for the study are as follows:

Overcoming Fear: Present Self-Disclosure

Hey! Today, I want to open up about something that's been a big part of my life: my fear of failure. I used to be terrified of making mistakes or not meeting expectations. It held me back in so many ways, and I know I'm not alone in this. I've learned that failure isn't the end of the world; it's actually a stepping stone to success. Each time I've stumbled, I've gained invaluable lessons that have helped me grow stronger and more resilient. So, I want to encourage you to embrace your fears and see them as opportunities for growth. Remember, it's not about avoiding failure, it's about how we bounce back and keep moving forward.

Overcoming Fear: Absent Self-Disclosure

Hey! Today, I want to talk about something that many of us face: the fear of failure. It's a common feeling, and it can hold us back in so many ways. But here's the thing - failure isn't the end of the world; it's actually a stepping stone to success. Each time we stumble, we gain invaluable lessons that can help us grow stronger and more resilient. So, I want

to encourage you to embrace your fears and see them as opportunities for growth. Remember, it's not about avoiding failure, it's about how we bounce back and keep moving forward.

Embracing Change: Present Self-Disclosure

Hello. Let's talk about change. Change can be scary, right? I've had my fair share of changes, from moving to a new city to switching careers. Each time, I felt a mix of excitement and anxiety. But I've come to realize that change is a natural part of life and growth. I want to share with you that it's okay to feel uncertain about change. What's important is to stay open-minded and adaptable. Change can lead to new opportunities and experiences that we might never have imagined. So, let's embrace change together and see where it takes us!

Embracing Change: Absent Self-Disclosure

Hello! Let's talk about change. It's a natural part of life and growth, but it can also be scary. It's normal to feel a mix of excitement and anxiety when facing change, whether it's moving to a new city, switching careers, or any other major life event. What's important is to stay open-minded and adaptable. Change can lead to new opportunities and experiences that we might never have imagined. So, let's embrace change and see where it takes us. Remember, it's about staying flexible and making the most of the opportunities that come our way.

Finding Balance: Present Self-Disclosure

Hey everyone! Today, I want to talk about something we all struggle with: finding balance. In our fast-paced world, it's easy to feel like we're constantly juggling responsibilities, and it can be overwhelming. I've been there, trying to balance work, personal life, and self-care. It's been a journey, but I've learned that it's okay to take a step back and prioritize what truly matters. Remember, it's not about doing everything perfectly, but about finding harmony in our lives. Let's strive for balance together and support each other along the way.

Finding Balance: Absent Self-Disclosure

Hey everyone! Today, I want to talk about finding balance. In our fast-paced world, it's easy to feel like we're constantly juggling responsibilities, and it can be overwhelming. Finding balance is a challenge for many of us, but it's crucial for our well-being. It's not about doing everything perfectly, but about finding harmony in our lives. Taking a step back to prioritize what truly matters can make all the difference. Let's strive for balance together and support each other along the way.

Nurturing Self-Compassion: Present Self-Disclosure

Hey everyone! Today, I want to get personal and talk about self-compassion. I've always been my own toughest critic, constantly pushing myself to do more and be better. But I've learned that being kind to myself is just as important as striving for success. I've realized that self-compassion is not about being easy on myself,

but about treating myself with the same kindness and understanding that I would offer a friend. It's been a game-changer for me, and I want to encourage you to embrace self-compassion too. Remember, you are enough, just as you are.

Nurturing Self-Compassion: Absent Self-Disclosure

Hi! Let's talk about something important: self-compassion. In a world where we're often our own toughest critics, it's crucial to remember to be kind to ourselves. Self-compassion is not about being easy on ourselves, but about treating ourselves with kindness and understanding. Embracing self-compassion can be a game-changer. It's about acknowledging that we are enough, just as we are. So, let's practice being kind to ourselves and remember that it's okay to be gentle with our own hearts.

C Model Results

Table 3: Model estimates with standard errors and p-values for human-likeness, self-disclosure, and interaction terms.

Predictor	Estimate	Std. Error	p-value
Human-likeness (MH)	1.857	0.664	0.005
Human-likeness (ML)	0.999	0.649	0.124
Human-likeness (L)	2.882	0.678	<0.001
Self-disclosure (Absent)	0.296	0.309	0.338
Perspective Taking	2.608	0.460	<0.001
Tendency to Anthropomorphise	-0.036	0.176	0.840
Human-likeness (H):Fantasy	-0.005	0.455	0.992
Human-likeness (MH):Fantasy	-0.087	0.456	0.849
Human-likeness (ML):Fantasy	-0.205	0.453	0.650
Human-likeness (L):Fantasy	-0.609	0.459	0.185
Self-disclosure (Present):Emotional Contagion	0.202	0.484	0.677
Self-disclosure (Absent):Emotional Contagion	0.096	0.493	0.847
Human-likeness (MH):Emotional Contagion	-1.170	0.282	<0.001
Human-likeness (ML):Emotional Contagion	-0.632	0.270	0.019
Human-likeness (L):Emotional Contagion	-0.881	0.283	0.002