How does a virtual human earn your trust? Guidelines to improve willingness to self-disclose to intelligent virtual agents

Christopher You
christopheryou@ufl.edu
University of Florida

Rashi Ghosh
rashighosh@ufl.edu
University of Florida

Andrew Maxim
amaxim@ufl.edu
University of Florida

Jacob Stuart
jacobstuart@ufl.edu
University of Florida

Eric Cooks
ecooks@ufl.edu
University of Florida

Benjamin Lok
lok@cise.ufl.edu
University of Florida

ABSTRACT

Virtual humans demonstrate the ability to act as non-judgmental conversational partners, eliciting greater self-disclosure. However, it is unclear what virtual human and conversational characteristics are important when self-disclosing. To address this gap, we conducted a set of qualitative, semi-structured interviews (n = 17) among computer science students to investigate participant mental models of willingness to disclose to virtual humans and characteristics of virtual humans that affect their self-disclosure. Our findings indicate that participants' mental models of virtual humans are largely inconsistent with current literature. This inconsistency appears to eliciting hesitancy and discomfort with virtual humans. Furthermore, trust and listening were identified as two primary characteristics of a virtual human interaction that are valuable towards willingness to disclose. Additionally, these characteristics were also valued in different ways for virtual humans in comparison to real humans. From the interviews, we identify and provide guidelines of designing virtual human interactions and conversations to elicit greater willingness to disclose.

CCS CONCEPTS

• **Human-centered computing** → *HCI theory, concepts and models*; Virtual reality.

KEYWORDS

virtual humans, self-disclosure, virtual agents, willingness, trust, listening

ACM Reference Format:
Christopher You, Rashi Ghosh, Andrew Maxim, Jacob Stuart, Eric Cooks, and Benjamin Lok. 2022. How does a virtual human earn your trust? Guidelines to improve willingness to self-disclose to intelligent virtual agents. In *ACM International Conference on Intelligent Virtual Agents (IVA '22), September 6–9, 2022, Faro, Portugal.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3514197.3549686

IVA '22, September 6–9, 2022, Faro, Portugal

ACM ISBN 978-1-4503-9248-8/22/09…$15.00
https://doi.org/10.1145/3514197.3549686

1 INTRODUCTION

Self-disclosure is the sharing of any information about oneself to another person [9]. Self-disclosure is beneficial in numerous settings such as healthcare, therapy, well-being, or even with friends and family. However, eliciting self-disclosure can be challenging due to factors such as fear of self-disclosure [11] and the desire to maintain a good impression [30]. To support self-disclosure, previous work has utilized virtual humans to serve as the partner in a conversation, rather than a real human. Virtual humans are computer-generated characters that attempt to act, look, and talk like real humans [12]. Interacting with virtual humans has lead to greater self-disclosure and willingness to self-disclose [10, 18].

While self-disclosure and willingness to disclose has been shown to increase with virtual humans, it is unclear what attributes are valuable to increase people's willingness to self-disclose. Some attributes that have been investigated to impact one's willingness to self-disclose to virtual humans include listening [13], humor [15], empathy [26], and mimicry [26]. These attributes are often based on previous psychology literature in self-disclosure. By utilizing these attributes, we make the assumption that one's willingness to disclose is affected in the same way for conversations with virtual humans as it is with real humans. In reality, one's willingness to disclose may change due to factors that are only present with virtual humans. Thus, we conducted semi-structured interviews among computer science students to determine relevant attributes that impact their willingness to disclose to a virtual human. To ensure that participants had the same frame of reference for self-disclosure, we only discussed self-disclosure in the context of situations that are beneficial to the participant, such as in health and well-being. In conducting our interviews, we aim to answer the following RQs:

- **RQ1.** What is the current mental model of virtual humans and perceptions of their usage for self-disclosure?
- **RQ2.** What are the attributes that impact one's willingness to disclose to virtual humans?
- **RQ3.** What are current barriers of self-disclosure to virtual humans?

Our work contributes to the field of intelligent virtual agents by uncovering current mental models of virtual humans and what attributes are valuable to increase self-disclosure among computer science students. We propose guidelines to address current participant-identified barriers to utilize virtual humans for self-disclosure and improve willingness to disclose to virtual humans.

Through our interview process, we found that current mental models of virtual humans are largely inconsistent with actual virtual human capabilities (e.g., virtual humans were simplified to personal voice assistants). This lack of understanding contributes to potential hesitancy in interacting with virtual humans. Furthermore, of attributes that impact willingness to disclose, participants largely agreed that trust and listening were the attributes that needed to be present to encourage their willingness to disclose to a virtual human. Interestingly, trust and listening were discussed differently when a virtual human was the conversational partner as opposed to a real human. Whereas with a human trust meant confidentiality, relatable experience and expertise, and honesty, with a virtual human, trust more closely aligned with data security, accurate information and expertise of the organization, and naturalness.

2 RELATED WORKS

As our goal is to leverage one's willingness to disclose to promote healthy behavior, willingness to disclose, in our context, only refers to situations in which one's willingness to disclose is ultimately beneficial to them. Thus, willingness to disclose in beneficial scenarios may include interactions with one's doctor about their own health or conversations with a friend regarding personal matters to receive comfort and advice. While willingness to disclose can increase with virtual humans [18], conversations with virtual characters or other entities often differ in nature from conversations with humans [7]. We discuss willingness to disclose in the context of humans and virtual humans in the following section.

2.1 Willingness to Disclose to Humans

The willingness to disclose to humans in dyadic (two-person) conversations can be positively impacted by either member in the dyad. The amount of disclosure is described as a product of both the people within a conversation and the interaction between them [2]. For instance, when both doctor and patient felt greater perceptions of rapport, greater disclosure of the patient was produced [14]. Rapport can be increased by both members of the dyad through verbal and non-verbal cues. This is evidenced by Miller et al. [23] who found that greater disclosure can occur when a low discloser is paired with a partner who is a high discloser, as more attentive listeners can draw out further responses from the responder. In terms of non-verbal cues, mirrored posture, body orientation, and directed gaze from one member of a dyadic conversation can elicit greater disclosure from the other member and aid in rapport building [23]. As a result, in dyadic conversations, a persons' willingness to disclose can be impacted by the characteristics of the other member in the dyad. Besides rapport and verbal and non-verbal cues, Pope & Siegman [27] note that the outward personality of a member of the dyad can elicit greater disclosure. In their study, the overall "warmth" of the interviewer (e.g., smiling, nodding, and speech), allowed for participants to disclose more information in comparison to a "cold" interviewer. Other aspects exuded by a member of a dyad such as trust [36], a sense of humor [17], and familiarity [20] have also demonstrated the ability to increase a member's willingness to disclose. Additionally, anonymity and a judgment-free space can positively impact one's willingness to disclose [25] [31]. Thus, one's willingness to disclose is often impacted by numerous factors.

2.2 Willingness to Disclose to Virtual Humans

Conversations with a virtual human can elicit greater self-disclosure than conversations with humans [18]. It was observed that when participants believed they were interacting with a human, participants reported higher fear of self-disclosure, more impression management, and were also observed to disclose more by a third-party observer. Despite the fact that one may feel anonymous when interacting with a computer [35], computers (and in particular, virtual humans) are often treated as humans still. The Computers Are Social Actors paradigm demonstrated that social responses typical for human conversation is commonplace and easily instigated when conversing with a computer [24]. As a result, virtual humans can elicit social responses appropriate for human conversation, even if they lack photo-realism or extremely high fidelity. A virtual human achieves "the best of both worlds" according to Lucas et al. [18], as virtual humans can be perceived as having characteristics that improve willingness to disclose human-human conversation (e.g., personality, verbal and non-verbal cues, humor, etc.) while achieving a natural sense of anonymity and non-judgment.

The desired dimensions of conversation in a human-virtual human conversation are not necessarily the same as in a human-human conversation. While it is true that computers can be perceived as human in conversation, other work has found that there are discrepancies. For instance, Clark et al. investigated attributes of conversation with humans and found that traits such as trustworthiness and a sense of humor were valuable [7]. However, when conversing with a chatbot, participants' perceptions of these traits were changed; instead, chatbots needed to provide functional trustworthiness (privacy), and a sense of humor was seen as a novelty feature rather than a necessary feature. This was largely due to the fact that conversations with chatbots (and computers as a whole) are seen as transactional – not conversational [8][28]. Thus, it is important to frame the conversation with a virtual human. In our work, we aim to discover what attributes in conversation with people allow them to feel more willing to disclose in contexts that are beneficial to them (e.g., healthcare). We also aim to discover how these attributes are perceived in the contexts of virtual humans and the nature of virtual human conversations overall.

3 METHOD

3.1 Participants

We recruited 17 participants (M = 8; F = 9) who were students at the University of Florida. Recruited participants were compensated with credit in one of their courses within the University of Florida. Following qualitative interviewing procedures, participants were recruited until saturation had occurred. Participants had a mean age of 24.3 years (SD = 5.06) and self-described as follows: 52.9% Asian, 23.5% Black or African, 11.8% White or Caucasian, and 11.8% Hispanic, Latino, or Latina. All participants spoke English fluently. Additionally, all participants were students recruited from a computer science class within the computer science department at the University of Florida (by major): 58.8% Computer Science, 29.4% Human-Centered Computing, 6.88% Data Science, and 6.88% Curriculum and Instruction (in Education). 41.2% of participants were undergraduate, and 58.8% were graduate students. Thus, all participants were students in computer science or related fields.

All participants had experience with automated voice interfaces, with a large majority 58.8% (10/17) using them daily or several times a week. The (41.2%) remainder of the population used automated voice interfaces several times a month or year. However, 41.2% participants stated they never interact with a virtual human, and only 17.6% stated interaction with a virtual human several times a month. No one stated that they had interactions with a virtual human more than several times a month. Of those who had interacted with a virtual human, the interaction was usually due to a former research study the participant had taken a part in or as a web assistant. Only 17.6% participants stated the usage of a virtual human for conversation or self-disclosure.

3.2 Procedure and Data Analysis

Interviews were conducted in order to understand the attributes of conversations and people that impact one's willingness to disclose. Furthermore, interviews gauged participants' mental models of a virtual human and attributes that may impact their willingness to disclose in virtual human contexts. Each interview lasted approximately 40 minutes, with a maximum of 50 minutes per interview. Interviews were semi-structured in order to allow for follow-up and adjustment of questions based on participant responses. Each interview covered a total of five topics: (1) Mental models of self-disclosure, (2) Valued attributes that impact participants' willingness to disclose to other people, (3) Mental models of virtual humans, (4) Valued attributes that impact participants' willingness to disclose to virtual humans, and (5) Barriers to self-disclosure to virtual humans. Following a similar approach to [7], attributes described in (2) were written down by the interviewer and re-discussed during topic (4). Immediately after (3), participants were given a standardized definition of a virtual human to ensure all participants had the same frame of reference for a virtual human. For our contexts, a virtual human was defined as a "computer-generated virtual character that looks and behaves like a human, capable of robust conversation through semi-scripted conversations." Additionally, they were shown an example of what a virtual human and interaction may look like. Participants were thoroughly informed that this definition was just *one* example of a virtual human for our purposes, and ultimately they can vary in appearance, conversational nature, and purpose. Additionally, participants were then all given the same context for self-disclosing to a virtual human: the self-disclosure to a virtual human was conducted such that a trained professional (e.g., doctor) could later review their responses and best serve their underlying needs.

Audio from each interview was recorded and transcribed to be analyzed through the Inductive Thematic Analysis [5]. Notes during the interview were also recorded for analysis. Thematic analysis was performed by a total of three researchers (Anon). While each of the three researchers had a background in HCI and qualitative data analysis, an initial set of data was used to train researchers. Afterward, the remaining thematic coding was performed independently by two (Anon) of the three researchers. Researchers began by listening to the audio recordings and reading the transcriptions. The process was robust, generating as many codes as possible, as described by Braun et al. [5]. The interview guide and questions helped shape the initial themes discovered from the codes. Using virtual sticky notes, an initial theme map was generated and piled with the codes. Themes were reviewed and refined appropriately by relating themes to one another and drawing relationships when possible (Anon). Additionally, quotes that best represented the themes being drawn were extracted and saved for later analysis, following a process like that of Clark et al. [7] and Cowan et al. [8].

4 RESULTS

4.1 Mental Models of Self-Disclosure

We gauged participants' mental models on willingness to disclose, as well as to whom and when they value disclosure. There were two main purposes for self-disclosure: transactional and social. This is similar to work in [6, 7]. Transactional purposes for self-disclosure include situations where the conversational partner is there to assist, such as with a doctor or therapist. Social purposes for self-disclosure include situations where the participant wanted to disclose with no real objective, such as when ranting or venting. However, regardless of the purpose, the level of self-disclosure in a conversation was primarily based on three factors: (1) context, (2) relationship, and (3) urgency.

Context. Through the interviews, it became apparent that the topic of conversation and why that topic was being discussed were important for greater willingness to disclose. In social conversations, participants stated that it may not be necessary to disclose large amounts of personal information, but when discussing pertinent health-related topics, participants indicated that they generally would have little issue giving as much information as needed.

"If I'm talking to a therapist, I'm going to be 100 percent honest because I know the therapist is very professional, so I will always tell other details and the truth to a therapist in order to get as much as help from the therapist." [P007]

Relationship. As our scenarios for disclosure were primarily dyadic conversations, participants described that the relationship between them and the other member of the dyad influenced their level of disclosure. Participants agreed that when the other member of the dyad was well-trusted and familiar, their willingness to self-disclose increased, echoing previous work such as with Wheeless et al. [36].

"Years of knowing these people of being with these people through the good parts and the bad parts of the certain moments of life tell me enough so that I know that anything else I could tell them or I could need to disclose to get some opinion or feedback from there wouldn't make an issue for me." [P008]

While participants stated that disclosure to a trusted person was beneficial to their willingness to disclose, some participants also stated that disclosure to less familiar people was not out of the question. Ultimately though, participants described that the disclosure increases the stronger the relationship.

"If the friend is very close to you, then both of you have formed a sense of trust between the two of you. Then you're obviously willing to share a lot more than you would be willing to share with an acquaintance." [P001]

Urgency. The urgency of the topic discussed was described as important for the quantity of disclosure. Particularly when disclosing to someone who can assist in the participants' situation, the

urgency of the situation takes precedence over the discomfort that comes from self-disclosure.

"It depends on like if, the urgency of the situation... If it's like a matter of life and death or something I really need help with, the urgency, then I would say probably that [would make me more willing to disclose]." [P016]

4.2 Mental Models of Virtual Humans

To our knowledge, there is little work determining current perceptions of a virtual human and reactions towards conversations with a virtual human. Thus, prior to defining what a virtual human was, participants were asked what they perceived virtual humans to be, how virtual humans could be used, and what their reactions toward disclosing to a virtual human for their own benefit were.

The most prevalent mental model of a virtual human was simply an FAQ chatbot or intelligent personal assistant (e.g., Alexa, Siri), often without mention of any sort of graphical representation nor intelligent conversation. Virtual humans were perceived to be equipped primarily for simple tasks such as FAQ, phone services, automated messages, or scheduling appointments. Virtual humans were viewed as personal assistants that could replace humans in low-risk, non-complex scenarios. Virtual humans were sometimes described as artificial intelligence, but the dialogue capabilities were perceived as generally unintelligent, as participants still expected simple, purely pre-generated answers like with an FAQ chatbot.

"A virtual human might be used as a digital assistant . . . , where you want to maybe imitate human behavior but you can't get actual humans" [P004]

These mental models of a virtual human capture some correct elements but are generally not aligned with the current ideas and purposes of virtual humans as seen in literature. In literature, virtual humans have demonstrated usage for structured conversations such as with self-disclosure [18], but they have also been utilized for complex dialogue such as social conversations [37], therapy [29], and health conversations [10]. Virtual humans were not pictured as capable of sensitive or intelligent conversation for self-disclosure. The vast majority of participants described discomfort in self-disclosing to a virtual human, and participants almost always preferred a human conversational partner. Some of these discomforts stemmed from unfamiliarity. However, participants described that a larger discomfort stemmed from the belief that virtual humans were simply incapable of supporting self-disclosure. Thus, the discovered mental models of virtual humans appear to be inconsistent with currently known virtual human capabilities, leading to greater reluctance.

[The virtual human] cannot convince me it has the advanced technology... when we reuse [these] systems, they're really bad. It cannot even compare [to a real therapist]. It cannot give you useful advice. [P007]

After determining initial sentiments in self-disclosing to virtual humans, the remainder of the interview was framed around conversation with a virtual human as a way to elicit self-disclosure from the participant such that a trusted professional could later review the information to better support the participant. When participants were given this context, nearly all described greater willingness to disclose and far less discomfort in self-disclosing.

I would not really be mindful about ... other issues if that specific virtual human was hosted by an institution I trust. [I would be] completely willing [to disclose] once I have trust in the context. [P008]

4.3 Self-Disclosure Attributes for Humans and Virtual Humans

We gauged participants on their willingness to self-disclose to humans and virtual humans. This section describes how the mental models of conversation and valued attributes overlap and differ when self-disclosing to a virtual human as opposed to a human.

4.3.1 Conversation Nature.

Human-Human Conversations. Self-disclosure was seen as a *dialogue* between the two members in a dyadic conversation. While participants were open to disclosing to those close to them or trusted authoritative figures (e.g., doctor), the willingness to disclose was described to increase when participants felt that the other person in the conversation was either also disclosing or assisting.

"I guess the number one thing would be that the other person is also sharing some information, be it like personal or relevant to the topic." [P004]

Self-disclosure conversations with real people were seen to be organic and natural, with each member contributing, being able to elaborate and infer, and providing personal anecdotes or advice.

"You feel more comfortable when you're around people you can relate to. It just kind of [feels] natural." [P007]

"If you're talking with a doctor, he can not only pick up on things that you're saying, but ... while [you] didn't say it, it looks like [you're] experiencing XYZ." [P009]

Human-Virtual Human Conversations. Participants described that the self-disclosure to a virtual human as opposed to a human varied in conversation type. For instance, the purpose of disclosure to a virtual human was described as largely transactional, as seen in similar literature [6, 7]. Furthermore, even though the conversations with a virtual human are dyadic, participants described that the conversation to a virtual human felt more similar to a *monologue* rather than a dialogue as there was a lack of mixed initiative. Participants perceived talking to virtual humans as either primarily virtual human-lead (e.g., a digital educator that shares large amounts of information) or primarily human-lead (e.g., to rant or vent) but rarely an equally-lead conversation.

"As for me personally, maybe to vent, like I wouldn't wanna drag my friends over and have them deal with all my problems. So having a virtual human that's a little isolated, separated from the whole situation for me to just y'know vent or rant about something and they would be able to respond to that. It would be nice." [P015]

Conversations, where one needed to self-disclose to a virtual human, were seen to be inorganic and structured, with a lack of room for open interpretation. Self-disclosure to a virtual human was perceived to be appropriate when the task was straightforward (e.g., in lieu of a form or FAQ).

Virtual humans make lots of sense ... getting information for specific, very structured circumstances or situations. If it was a health

incidence for instance, like the first ... interview you get from a [nurse] before you meet the doctor ... that first draft of questions ... your family heath situation, or how are you feeling right now, what hurts, when did it [start], why – those kinds of things ..." [P008]

Ultimately, the risk involved with a conversation would elevate the need to speak with a real person. For human conversations, participants were focused on urgency rather than convenience. If a situation demanded immediate disclosure (e.g., in life-threatening situations), participants reasonably did not feel willing to disclose to a virtual human, especially if a human conversational partner were available. Instead, with virtual humans, participants were more focused on convenience rather than urgency. If the situation allowed for convenient disclosure to a virtual human, participants were more willing to disclose.

I would prefer to talk to the virtual human especially on like a day-to-day thing ... where you can't always [get] access to a therapist. In emergency situations ... I would bypass a virtual human and probably demand to speak to a human life. [P016]

4.3.2 Attributes.

Trust and listening were identified as the primary valued attributes that would improve people's willingness to disclose when conversing with a human or a virtual human. However, trust and listening were discussed differently for virtual humans than with humans. Additionally, trust was described in three different ways by participants: (1) trust in terms of security, (2) trust in terms of credibility, and (3) trust in terms of sincerity. We discuss the attributes of trust and listening and their differences in the context of virtual humans in the following section.

Trust in terms of Security

Human-Human Self-Disclosure. Participants felt that it was valuable to have a conversational partner who could provide security with the disclosed information. When there was knowledge that the information would be kept *confidential* between the participant and their conversational partner, participants described they would feel more willing to disclose. This echoes findings from previous literature such as Anestis et al. [3], who found that that confidentiality increased self-disclosure from military personnel.

"With medical professionals ... I can have trust upon them ... So [they] will not share my experience with others. It'll just be the one-to-one talk ... But with friends, it can be a thing they could share my experience, pass on the things with my name attached, maybe some of the things I don't even wanna share." [P003]

Human-Virtual Human Self-Disclosure. When the same idea of security was posed for virtual humans, participants described that security with a virtual human was valuable in the form of *data security*, rather than confidentiality. Some participants noted that confidentiality did not make much sense in the context of virtual humans as it was simply expected.

"Like [the virtual human] can't go talk to anybody – it doesn't have friends to tell like, oh my gosh let me tell you what this [person] talked about in therapy today!" [P017]

Data security was a more reasonable conceptualization of confidentiality. Knowledge that one's information would be stored securely and kept from non-authorized figures was identified as a key attribute towards improving willingness to disclose to virtual humans. Furthermore, participants wanted transparency on data storage.

"I think as long as like, the developers are transparent with how or where my data is stored, or if it is even stored ... As long as they're like, transparent about what's done with the data I think I'd be fine." [P002]

Trust in terms of Credibility

Human-Human Conversations. Trust was also described in the context of credibility. As participants described that self-disclosure was often for transactional purposes, they expressed the need for advice or aid based on the disclosed information. Because of this, credibility of the conversational partner was identified as a valuable attribute to improve the willingness to disclose. Participants described when their conversational partner was credible, it was easier to trust and therefore self-disclose. Credibility was identified into two components by participants: *relatability* and *expertise*. First, a conversational partner was deemed as credible by participants when the the conversational partner could provide relatable or personal advice through similar experiences or learning.

"If I tell [a doctor] something's wrong, I know I'm not going to just get an answer I could've got off Web MD or something ... He gives me the impression that he's going through his user experience and through ... his knowledge base and seeing what matches what I'm talking about." [P016]

Second, credibility was dependent on the expertise of the conversational partner. For instance, when speaking with a doctor, many participants felt that their willingness to disclose would be greater as they have the assurance that they're speaking to a certified professional.

"Obviously, if it's ... someone I feel isn't really qualified to diagnose my issues, then I might be less likely to disclose every bit of information. But if I'm maybe with maybe like a doctor or the nurse or the PA – people of much higher, I guess, expertise – then I'm definitely more open to telling them every little thing." [P013]

Human-Virtual Human Conversations. When disclosing to virtual humans, credibility was discussed under the context of *accuracy* and the *quality of the organization behind the virtual human*. First, rather than being able to give relatable advice, it was more important to be accurate with the information that it provided.

"It should have the ability that we can trust upon that virtual person; we could have faith that it would give us a thing which is fruitful to us; it would help us." [P003]

The second aspect of credibility described focused on the quality of the organization rather than the virtual human itself. Participants felt that virtual humans did not necessarily have a reputation that could be assessed like that of a real human. Participants described the need to know who was behind the virtual human and what they would do with the information, tying back to data security concerns. Ultimately, participants felt that the virtual human needed to be situated within a reputable organization to achieve credibility.

"I would disclose with a virtual human [from a doctor's office] just because like I know where it's going ... I know it would help me, so I'd

be willing to because I know at the end of the day, it's like going to a doctor and they're going to look over it." [P011]

Trust in terms of Sincerity

Human-Human Conversations. Besides security and credibility, trust was also described in terms of sincerity or honesty. Participants described that self-disclosure to a *honest* conversational partner was valuable, especially when that partner could give advice or share experiences. Sincerity was also described as valuable when a conversational partner was *non-judgmental*, being receptive to the participant's self-disclosure and vulnerability.

"I like knowing that people aren't as judgemental. I have a group of friends that are like, way less judgemental than another group of my friends. So they know more about me than the other group does, only because I feel more comfortable talking to them about those kind of things." [P010]

Other factors were encompassed within sincerity such as being able to provide a safe space, being friendly, and being amicable.

"It's a feeling of this is a safe space and you can say the things you need to say. And if there's not enough trust than you just say less." [P016]

Human-Virtual Human Conversations. When self-disclosing to a virtual human, participants described that an honest and a non-judging virtual human were also valuable to their willingness to disclose. However, participants also felt that those characteristics were naturally present. This was echoed by some participants saying one would need to go out of their way to design a virtual human that was judgmental and insincere.

"I think you'd have to purposefully make the virtual human judging in order for it to be judging; I think the default would be not judging, because they're not going to react negatively to anything your saying." [P005]

Participants described that sincerity in a virtual human was not demonstrated purely through its words like with humans but through its ability to demonstrate *naturalness* and *genuine expression*. Comfort in a virtual human conversation was described to be partially dependent on the naturalness of the virtual human's behaviors such as animations, facial expressions, responses, and appearance. These behaviors were considered sincere in that they were realistic behaviors that one would expect in human-human conversations.

I would say, genuine facial expressions and sincerity like that [are important], because if that virtual human is not sharing those facial expressions, like the sensitive type of movement in eyebrows or whatever, ... it's like I'm talking to a robot. [P012]

Listening

Human-Human Conversations. As self-disclosure can be challenging for some participants, active listening was highlighted as a valuable attribute towards increasing their willingness to disclose, echoing work seen in. [7, 23, 34]. Listening was framed as valuable towards willingness to disclose because it *demonstrated engagement*. Participants described that being engaged meant that each member was deeply involved in the conversation.

"I would say the most important is the fact that they seem like they are actively listening ... Like I'll say something, and they'll disregard it, that's one of my biggest pet peeves ... as long as you're actively paying attention and ... building on top of the conversation ..." [P015]

	Self-Disclosure Comparison	
Attribute for Self-Disclosure	**Humans**	**Virtual Humans**
Conversation Nature	Dialogue Urgency over Convenience	Monologue Convenience over Urgency
Trust - Security	Confidentiality	Data Security
Trust - Credibility	Quality of Information Expertise of Human	Accuracy of Information Quality of Organization
Trust - Sincerity	Honest and Non-Judgmental	Natural and Genuine
Listening	Demonstrate Engagement	Demonstrate Plausibility

Table 1: A brief summary of each attribute of self-disclosure from Section 4.3 when conversing with a human as opposed to a virtual human. This figure illustrates the difference in discussion of each attribute for humans and virtual humans.

Human-Virtual Human Conversations. For virtual humans, participants felt that listening to demonstrate engagement was valuable as well. This overlapped with their ideas of sincerity, saying that a virtual human should be genuine and expressive to demonstrate listening. However, listening in terms of a virtual human went a step deeper as there was a layer of technology that interfered with the plausibility of the interaction. For instance, participants felt that if a conversation felt scripted, it seemed as if the virtual human wasn't listening. Therefore, the virtual human needed to *demonstrate plausibility*. If the virtual human did not acknowledge the direct thoughts shared by the participant, it felt as if the system was not truly listening but instead replying with pre-generated responses, generating a lack of plausibility. In turn, the resulting willingness to disclose would be negatively impacted.

"I would say active listening and actually getting feedback from them while I talk ... I think I would be able to like, disclose more ... So they're repeating what I said back to me and making sure I said that, like nodding their head, body language ... a head nod, just to make sure they understand what I'm saying." [P011]

4.4 Barriers

We identified barriers that participants felt existed that prevented or reduced their willingness to disclose to virtual humans. The primary factors described by participants included security barriers and technology barriers.

For security barriers, the primary concerns were who would see the data, how it would be used, and how it would be securely stored. Some participants felt that mitigating such security barriers would be impossible for themselves as there was such a large distrust for technology as is. However, other participants stated that aspects of both the virtual human and the platform that the virtual human was hosted on could mitigate the security barriers.

For technology barriers, participants described discomfort with current virtual human technologies due to the lack of realism and uncanny valley effects, causing a resulting disbelief that the virtual human is helpful in the first place. A large majority of the population either described the uncanny valley effect or literally stated the term "uncanny valley" as a major barrier for them from using virtual humans in self-disclosure conversations. Related to that was realism, both in terms of graphical appearance and conversational technology (e.g., NLP, NLU). These technology barriers lead to broken plausibility. Participants felt that the virtual human was fake, not intelligent, and could not understand them nor help, similar to sentiments shared in Section 4.2.

5 DISCUSSION

Based on the results, our interviews and resulting thematic analysis addressed our research questions.

- **RQ1.** *What is the current mental model of virtual humans and perceptions of their usage for self-disclosure?* We uncovered that people's current mental models of virtual humans are inconsistent with current literature, potentially impacting their willingness to self-disclose to virtual humans.
- **RQ2.** *What are the attributes that impact one's willingness to disclose to virtual humans?* Trust in terms of security (data security), credibility (accuracy and organizational credibility), and sincerity (genuineness and expressiveness) as well as listening (demonstrating plausibility) were identified as valued attributes self-disclosure to virtual humans.
- **RQ3.** *What are current barriers of self-disclosure to virtual humans?* Current barriers to self-disclosure with virtual humans were described as security-based (data security) and technology-based (quality of virtual human interactions).

We identify additional **guidelines** based on these findings for designing virtual humans to increase users' willingness to disclose.

Establish the Virtual Human with a Reputable Organization. Establishing the virtual human with a reputable organization can improve user's willingness to disclose as it can improve trust. Trust in terms of credibility and security were two attributes that participants felt would improve their willingness to disclose. Zalake et al. suggested virtual human can improve in credibility through association with reputable brands (e.g., healthcare organizations) [38]. Trust-building principles are also suggested for this guideline [19]. This association can be drawn through the virtual human or even the platform that the virtual human is hosted on, as participants described that peripheral aspects (e.g., site professionalism) can play a role in their willingness to even use the system.

Be (Honestly) Transparent About Privacy and Data Usage. Without upfront knowledge that participants' data was secure, participants described that they would be less willing to disclose to a virtual human. Participants nearly all noted that data security was an integral aspect of their willingness to disclose to technology, echoing similar findings [1, 22]. While it may be helpful to have the virtual human state its intentions with the participants' data, participants also preferred to see cues for security in modern technology such as explicit and succinct terms and conditions and transparency on who would see the stored data. On top of that, common visual cues in UX were described by participants to help participants feel secure about their data. This included visual aids such as digital locks, opt-in options, and visually aesthetic websites or applications that are hosting the virtual human. Ultimately, security was noted to be not guaranteed, but these guidelines gave participants some assurance that security was attempting to be maintained.

Make Virtual Humans Responsive to the Individual. To increase willingness to disclose, virtual humans need to be able to respond directly to the information provided by a user when possible. Participants described a need to have conversations that are not just scripted or generic but rather specific to the user in robust ways. The value in robust conversational systems has been demonstrated in assisting in mental health scenarios [10] and building rapport [13]. Additionally, the virtual human should be responsive in its listening by being expressive in its responses. The reactions in terms of movements, animations, and facial expressions should be genuine and strive to be true to a real conversation [15, 16].

Bridge the Mental Model Gap. Researchers should aim to bridge the gap in mental model between the average user's understanding and the current understanding of a virtual human and its capabilities. Currently, virtual humans are largely perceived as incapable of meaningful and intelligent self-disclosure conversation – even amongst a population familiar with AI and voice assistants. However, developments in virtual humans [10, 29, 37] have demonstrated that virtual humans are capable of aiding in self-disclosure. Similar to Clark et al. [7], then, a collective goal should be to reframe the concept of a virtual human to help relieve discomforts and disbeliefs in virtual human capabilities. Participants described that many of their discomforts may be simply relieved over time as virtual human technologies become more advanced and familiar. The technology acceptance model best supports this as it describes that technologies will be adopted based on perceived usefulness and perceived ease-of-use (amongst other external factors) [21]. As virtual humans continue to grow in capability, the capabilities of virtual humans should be reframed for users for greater acceptance.

When considering the findings of the work, it is important to consider the theoretical nature of the interviewing process. From a surface level, findings based on controlled observations and think-aloud tasks during an experiment have shown to differ from findings procured via interview (e.g., in terms of quantity of discovery [4] or correlation between interviews and observations [33]). Self-described behavior from interviews can also differ from a participants' actual behavior in practice. Clark et al. point this out as well by describing that participant views may be affected by a lack of familiarity with conversational agents or may differ if interaction with a conversational agent was presented to participants [7]. However, incorporating participant-derived guidelines has shown to be beneficial even with a theoretical nature. Zalake et al. utilized a series of focus groups to identify design characteristics for virtual human health interventions [38]. These characteristics, when applied, positively influenced participants' intent to engage with and relieve discomforts with the virtual human. Vilaro et al. performed a similar interviewing procedure and determined that interviewed characteristics can be utilized to tailor virtual humans to enhance engagement [32]. Thus, while the results from interviews to determine attributes for self-disclosure may differ in practice, it still stands that there is benefit in utilizing participant-lead feedback to foster virtual human engagement.

6 CONCLUSION

Eliciting self-disclosure from people proves to be a continuing challenge. With virtual humans, we can potentially increase willingness to disclose. Thus, we aimed to determine what attributes are valuable for self-disclosure to virtual humans. Our work determined that trust and listening are valuable attributes for a virtual human to portray; however, the way in which trust and listening are portrayed differ for a virtual human than with a human. Additionally, our findings show that the perceptions of virtual humans are largely limited to personal assistants. These limited mental models of virtual humans can be harmful to interactions as participants perceive

the virtual human to be incapable of supporting their needs, hurting plausibility and acceptance. As our population was composed of students in computer science, one limitation is that their perceptions on self-disclosure and virtual humans may be inconsistent with a more general population (e.g., concerns with NLP and NLU). This would also explain some of the keen interest in popular topics in AI such as "uncanny valley" and data security. Furthermore, our population was quite familiar with automated voice interfaces and assistants, but had little experience with virtual humans themselves. Because of this, some participants' answers were theoretical in nature rather than grounded in previous virtual human experience, as discussed above. While their concerns and thoughts with virtual humans may have been biased by other personal assistant technologies (e.g., Alexa, Siri), attributes were only identified after a more pertinent mental model of a virtual human was established.

The purpose of this study was to determine the perceptions of attributes and mental models from a typical tech-familiar user. Future work should aim to work with other populations and determine any differences in valued attributes and mental models. Furthermore, the identified attributes should be evaluated in follow-up work to determine the real effect that they have when present in a self-disclosure intervention with a virtual human.

REFERENCES

[1] Tawfiq Alashoor, Sehee Han, and Rhoda C Joseph. 2017. Familiarity with big data, privacy concerns, and self-disclosure accuracy in social networking websites: An APCO model. *Communications of the Association for Information Systems* 41, 1 (2017), 4.

[2] Irwin Altman and Dalmas A Taylor. 1973. *Social penetration: The development of interpersonal relationships.* Holt, Rinehart & Winston.

[3] Michael D Anestis and Bradley A Green. 2015. The impact of varying levels of confidentiality on disclosure of suicidal thoughts in a sample of United States National Guard personnel. *Journal of Clinical Psychology* 71, 10 (2015), 1023–1030.

[4] Ester Baauw and Panos Markopoulous. 2004. A comparison of think-aloud and post-task interview for usability testing with children. In *Proceedings of the 2004 conference on Interaction design and children: building a community*. 115–116.

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[6] Christine Cheepen. 1988. *The predictability of informal conversation.* Continuum.

[7] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. 2019. What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[8] Benjamin R Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. " What can i help you with?" infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.

[9] Paul C Cozby. 1973. Self-disclosure: a literature review. *Psychological bulletin* 79, 2 (1973), 73.

[10] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, et al. 2014. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. 1061–1068.

[11] Barry Alan Farber. 2006. *Self-disclosure in psychotherapy.* Guilford Press.

[12] Maia Garau, Mel Slater, David-Paul Pertaub, and Sharif Razzaque. 2005. The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators & Virtual Environments* 14, 1 (2005), 104–116.

[13] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *International workshop on intelligent virtual agents*. Springer, 125–138.

[14] Judith A Hall, Jinni A Harrigan, and Robert Rosenthal. 1995. Nonverbal behavior in clinician—patient interaction. *Applied and preventive psychology* 4, 1 (1995), 21–37.

[15] Sin-Hwa Kang and Jonathan Gratch. 2010. The effect of avatar realism of virtual humans on self-disclosure in anonymous social interactions. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. 3781–3786.

[16] Sin-Hwa Kang and Jonathan Gratch. 2010. Virtual humans elicit socially anxious interactants' verbal self-disclosure. *Computer Animation and Virtual Worlds* 21, 3-4 (2010), 473–482.

[17] Sin-Hwa Kang, David M Krum, Peter Khooshabeh, Thai Phan, Chien-Yen Chang, Ori Amir, and Rebecca Lin. 2017. Social influence of humor in virtual human counselor's self-disclosure. *Computer Animation and Virtual Worlds* 28, 3-4 (2017), e1763.

[18] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100.

[19] Wenhong Luo and Mohammad Najdawi. 2004. Trust-building measures: a review of consumer health portals. *Commun. ACM* 47, 1 (2004), 108–113.

[20] Divine Maloney, Samaneh Zamanifard, and Guo Freeman. 2020. Anonymity vs. familiarity: Self-disclosure and privacy in social virtual reality. In *26th ACM Symposium on Virtual Reality Software and Technology*. 1–9.

[21] Nikola Marangunić and Andrina Granić. 2015. Technology acceptance model: a literature review from 1986 to 2013. *Universal access in the information society* 14, 1 (2015), 81–95.

[22] Grzegorz Mazurek and Karolina Małagocka. 2019. What if you ask and they say yes? Consumers' willingness to disclose personal data is stronger than you think. *Business Horizons* 62, 6 (2019), 751–759.

[23] Lynn C Miller, John H Berg, and Richard L Archer. 1983. Openers: Individuals who elicit intimate self-disclosure. *Journal of personality and social psychology* 44, 6 (1983), 1234.

[24] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 72–78.

[25] Matthew D Pickard and Catherine A Roster. 2020. Using computer automated systems to conduct personal interviews: Does the mere presence of a human face inhibit disclosure? *Computers in Human Behavior* 105 (2020), 106197.

[26] Matthew D Pickard, Catherine A Roster, and Yixing Chen. 2016. Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior* 65 (2016), 23–30.

[27] Benjamin Pope and Aron W Siegman. 1968. Interviewer warmth in relation to interviewee verbal behavior. *Journal of Consulting and Clinical Psychology* 32, 5p1 (1968), 588.

[28] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[29] Albert Rizzo, Belinda Lange, John G Buckwalter, Eric Forbell, Julia Kim, Kenji Sagae, Josh Williams, JoAnn Difede, Barbara O Rothbaum, Greg Reger, et al. 2011. SimCoach: an intelligent virtual human system for providing healthcare information and support. (2011).

[30] Barry R Schlenker. 1980. *Impression management.* Monterey, CA: Brooks/Cole.

[31] Lee Taber and Steve Whittaker. 2020. " On Finsta, I can say'Hail Satan'": Being Authentic but Disagreeable on Instagram. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.

[32] Melissa J Vilaro, Danyell S Wilson-Howard, Mohan S Zalake, Fatemeh Tavassoli, Benjamin C Lok, François P Modave, Thomas J George, Folakemi Odedina, Peter J Carek, and Janice L Krieger. 2021. Key changes to improve social presence of a virtual health assistant promoting colorectal cancer screening informed by a technology acceptance model. *BMC Medical Informatics and Decision Making* 21, 1 (2021), 1–9.

[33] Eva Ejlersen Wæhrens, Henning Bliddal, Bente Danneskiold-Samsøe, Hans Lund, and Anne G Fisher. 2012. Differences between questionnaire-and interview-based measures of activities of daily living (ADL) ability and their association with observed ADL ability in women with rheumatoid arthritis, knee osteoarthritis, and fibromyalgia. *Scandinavian journal of rheumatology* 41, 2 (2012), 95–102.

[34] Harry Weger Jr, Gina Castle Bell, Elizabeth M Minei, and Melissa C Robinson. 2014. The relative effectiveness of active listening in initial interactions. *International Journal of Listening* 28, 1 (2014), 13–31.

[35] Suzanne Weisband and Sara Kiesler. 1996. Self disclosure on computer forms: Meta-analysis and implications. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3–10.

[36] Lawrence R Wheeless and Janis Grotz. 1977. The measurement of trust and its relationship to self-disclosure. *Human Communication Research* 3, 3 (1977), 250–257.

[37] Jason Wu, Sayan Ghosh, Mathieu Chollet, Steven Ly, Sharon Mozgai, and Stefan Scherer. 2018. Nadia: Neural network driven virtual human conversation agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 173–178.

[38] Mohan Zalake, Fatemeh Tavassoli, Kyle Duke, Thomas George, Francois Modave, Jordan Neil, Janice Krieger, and Benjamin Lok. 2021. Internet-based tailored virtual human health intervention to promote colorectal cancer screening: design guidelines from two user studies. *Journal on Multimodal User Interfaces* 15, 2 (2021), 147–162.