



PDF Download  
3706599.3719885.pdf  
27 February 2026  
Total Citations: 0  
Total Downloads: 669

 Latest updates: <https://dl.acm.org/doi/10.1145/3706599.3719885>

WORK IN PROGRESS

## Exploring AI-Driven Affective Avatars for Autistic Adults and Adults with Social Anxiety in Virtual Meetings

**TIFFANY D. DO**, Drexel University, Philadelphia, PA, United States

**MARTEZ E MOTT**, Microsoft Research, Redmond, WA, United States

**JOHN TANG**, Microsoft Research, Redmond, WA, United States

**SASA JUNUZOVIC**, Microsoft Research, Redmond, WA, United States

**ANN PARADISO**, Microsoft Research, Redmond, WA, United States

**EDWARD CUTRELL**, Microsoft Research, Redmond, WA, United States

Open Access Support provided by:

Microsoft Research

Drexel University

Published: 25 April 2025

[Citation in BibTeX format](#)

CHI EA '25: Extended Abstracts of the  
CHI Conference on Human Factors in  
Computing Systems  
April 26 - May 1, 2025  
Yokohama, Japan

Conference Sponsors:  
SIGCHI

# Exploring AI-Driven Affective Avatars for Autistic Adults and Adults with Social Anxiety in Virtual Meetings

Tiffany D. Do  
Drexel University  
Philadelphia, Pennsylvania, USA  
Microsoft Research  
Redmond, Washington, USA  
tiffany.do@drexel.edu

Martez E Mott  
Microsoft Research  
Redmond, Washington, USA  
mamott@microsoft.com

John Tang  
Microsoft Research  
Microsoft  
Mountain View, California, USA  
johntang@microsoft.com

Sasa Junuzovic  
Microsoft Research  
Redmond, Washington, USA  
sasa.junuzovic@microsoft.com

Ann Paradiso  
Microsoft Research  
Redmond, Washington, USA  
annpar@microsoft.com

Edward Cutrell  
Microsoft Research  
Redmond, Washington, USA  
cutrell@microsoft.com

## Abstract

Virtual meetings often exclude individuals who rely on text-based communication, such as autistic adults and those with social anxiety. This paper introduces a prototype that converts typed text into emotive avatars using LLM technology, which convey emotional tone through modulated vocal and facial expressions. We reflect on design choices for using LLMs in accessible meetings and discuss insights from our semi-structured interviews with 18 autistic adults and adults with social anxiety. Our qualitative analysis revealed the following key insights: 1) Participants found the avatars helpful in alleviating challenges like masking and exhaustion, with some noting that the avatars enhanced their communication, increasing participation and confidence; 2) While they valued the avatars' affective capabilities, including both vocal and facial animations, they were sensitive to inaccuracies in vocal expression; and 3) Participants desired more personalized control over the avatars' affect to balance societal expectations with authentic expression.

## CCS Concepts

• **Human-centered computing** → **Accessibility**.

## Keywords

virtual avatars, virtual meetings, autistic adults

## ACM Reference Format:

Tiffany D. Do, Martez E Mott, John Tang, Sasa Junuzovic, Ann Paradiso, and Edward Cutrell. 2025. Exploring AI-Driven Affective Avatars for Autistic Adults and Adults with Social Anxiety in Virtual Meetings. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3706599.3719885>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*CHI EA '25, Yokohama, Japan*

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1395-8/25/04

<https://doi.org/10.1145/3706599.3719885>

## 1 Introduction

Virtual meetings are vital in today's digital landscape, facilitated by platforms like Zoom and Microsoft Teams for education, healthcare, and training. However, these platforms can unintentionally exclude individuals who struggle with verbal communication [11, 33, 39]. Many autistic adults<sup>1</sup> prefer text-based channels due to cognitive fatigue and social anxiety [7, 18, 33], which can hinder their participation in meetings where text chat is often undervalued [1]. For example, Zolyomi et al. [43] found that autistic users feel less empowered in video calls and prefer text communication. A recent survey indicates that sensory stress in video calls poses unique challenges for autistic individuals, suggesting that options to turn off video and use text chat could enhance accessibility [42]. However, relying solely on text can diminish social presence [24], negatively affecting satisfaction [8] and performance [41].

Our paper seeks to address these challenges by presenting a prototype that enhances accessibility through the transformation of text-based communication into a video conferencing format. The first author, who is neurodivergent, directly shaped the avatar design and research methodology through lived experience. Our prototype features AI-Driven Affective (AIDA) avatars, powered by large language models (LLMs), specifically designed for autistic individuals and adults with social anxiety who prefer text communication [7, 18, 43]. These avatars translate users' text into dynamic facial expressions and emotive text-to-speech (TTS) voices, allowing for a richer, more inclusive virtual meeting experience. When users input text, the LLM analyzes the content in context to generate appropriate emotional responses from the avatar.

To better understand how AIDA avatars might support autistic adults and those with social anxiety, we conducted an exploratory study with 18 participants. The focus was on one-on-one virtual meetings, a setting especially relevant to the autistic community [3, 32, 38]. Each participant customized an avatar and engaged in a simulated first-time meeting with an experimenter [31]. Throughout the study, we assessed participants' perceptions of AI-driven affect across both auditory and visual dimensions, with each participant experiencing four distinct avatars, each with varying facial

<sup>1</sup>We use people-first language when describing autistic individuals according to recommendations by the NIH. <https://www.nih.gov/about-nih/what-we-do/science-health-public-trust/perspectives/writing-respectfully-person-first-identity-first-language>

animations and vocal tones. Qualitative interview results revealed diverse participant perspectives, with 72% (N=13) open to using AI-driven avatars. Many found the avatars helpful in reducing masking effort, with some reporting enhanced communication, increased participation, and confidence. This paper contributes by 1) introducing a prototype that transforms text into emotive avatars for one-on-one meetings, and 2) highlighting design preferences for avatar emotional rendering.

## 2 Related Work

### 2.1 Virtual Meeting Experiences of Autistic Adults and Adults with Social Anxiety

Recent AI advancements have sought to enhance video conferencing for autistic users. For example, Begel et al. [4] used AI to provide feedback on facial expressions. In contrast, our work focuses on converting text into expressive speech to support communication. Research has highlighted the challenges autistic individuals face in virtual meetings (e.g., [1, 11, 33, 39]). Tang [33] found that neurodivergent adults often turn off their cameras during online meetings, either to conserve cognitive energy or to avoid stigma. However, this practice can reduce engagement, as their presence is limited to a static image. Buchholz et al. [7] stress the importance of offering multiple communication options, such as text chat, for those with communicative or cognitive disabilities.

Virtual meetings intensify accessibility issues for those who prefer non-verbal communication. Alharbi et al. [1] and Howard and Sedgewick [18] note that autistic individuals often favor written communication due to challenges with verbal expression and sensory overload, which can lead to communication avoidance. Das et al. [11] argue that these barriers in video calls represent essential access needs, underscoring the importance of inclusive meeting environments that address diverse communication preferences. We focused on using AI to enhance avatar voice expression, particularly through prosody and affective tone, inspired by advancements in emotional Text-to-Speech (TTS) technology. Research shows that flat TTS voices, even with neural models, can cause discomfort [13]. Epstein [14] suggested that expressive voices could reduce "Zoom Fatigue" and improve virtual meeting experiences. However, individuals with cognitive and communicative disabilities may struggle with tone regulation and may resort to "masking" a neurotypical tone [34], which is exhausting [9]. By leveraging AI for both voice and facial expressions, we aimed to create a more comfortable and engaging virtual meeting experience for users with these challenges.

### 2.2 Virtual Avatars as Proxies in Virtual Meetings

Our exploration of avatars in virtual meetings builds on prior research showing their effectiveness as proxies [19, 23, 30]. Avatars are becoming increasingly common in platforms like Microsoft Teams, Zoom, and Meta's Horizon Workrooms, reflecting their growing role in virtual meetings. Yasuoka et al. [40] and Watanabe [37] noted the acceptance of avatars in professional settings, underscoring their value in virtual workspaces. For people with

disabilities, avatars offer a more relational and flexible form of representation. Tang [33] highlighted that individuals with cognitive disabilities prefer avatars over static images for video alternatives, aligning with the "lifechanging" impact Davis [12] described. Our study aims to further explore how avatars can enhance inclusivity in virtual meetings.

Avatar use in meetings is influenced by context and individual preferences. Research suggests that users often modify their avatars based on societal expectations and meeting types, such as workplace versus social settings [5, 17, 27, 35, 36]. Junuzovic et al. [20] and Inkpen and Sedlins [19] stressed the importance of realistic avatars that align with users' intentions in professional environments. Our study focuses on AI-driven facial expressions, as they are key to fostering interpersonal connection, more so than other body movements [26]. Research suggests avatars can enhance social presence beyond physical expressions, using augmented behaviors like stronger eye contact or more intense smiles [16, 25, 26, 28, 29].

## 3 AI-Driven Affective (AIDA) Avatar Design

We built AIDA avatars using Microsoft's Mesh Avatars<sup>2</sup>. Our system allowed users to customize various aspects of their avatar's appearance, including body shape, skin tone, eye color, and hairstyle. Additionally, our avatars are equipped with standard idle animations, such as blinking, nodding, and subtle swaying movements to enhance their visual realism, similar to the system presented by Ma et al. [21]. An example image of a personalized avatar is depicted in Figure 1. AIDA avatars' speech was generated by Azure Cognitive Services Text-to-Speech<sup>3</sup> emotive voices, which were gender-matched to each participant's self-identified gender. Participants who identified as non-binary were given the choice to select either a male or female voice.

Users interacted with our system by inputting text, which was processed by a large language model (LLM), specifically GPT-4 (0314 version). Using a prompt, GPT-4 was instructed to segment the text by emotion based on contextual information and conversation history, with each segment assigned a primary emotion from a predefined list: Neutral, Happy, Sad, Angry, Fear, Surprise, and Disagreement. The entire conversation history was maintained as context in the model without pruning, given the short experiment. Each text segment's emotion was mapped to the corresponding TTS emotion, using the Aria-Neural voice for female avatars and the Jason-Neural voice for male avatars, with the emotional style degree set at 0.6<sup>4</sup>. This value was chosen based on initial pilot studies, balancing noticeable emotional expression without exaggeration. Simultaneously, the avatar's facial expression and lipsync was dynamically adjusted to match the emotion using viseme-driven lip animation (blendshape animations) [21]. The avatars were presented through a Unity application.

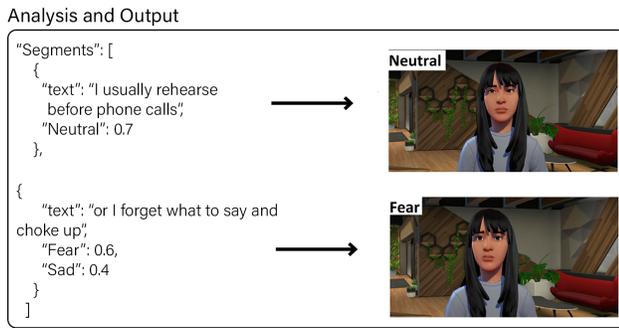
## 4 Methods

We recruited participants to engage in a virtual meeting using our avatars. During the sessions, participants communicated with a

<sup>2</sup><https://www.microsoft.com/en-us/mesh> [Accessed 2025-02]

<sup>3</sup><https://learn.microsoft.com/en-us/azure/ai-services/speech-service/text-to-speech> [Accessed 2025-02]

<sup>4</sup><https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-synthesis-markup-voice> [Accessed 2025-02]



**Figure 1: An example of how user input is transformed into an emotive avatar. In addition to facial expressions, the avatar would also speak in a tone of voice fitting the emotion. This example was based on participant responses; however, we omitted actual participant responses to ensure privacy and confidentiality.**

single member of the research team through Zoom’s chat window. Their typed responses were processed and transformed into speech by our AIDA avatars. We explored the design space of facial and vocal expressions using four combinations of affect: *No Affect* (no facial or vocal affect), *Face Affect* (facial affect only), *Voice Affect* (vocal affect only), and *Full Affect* (both facial and vocal affect). Participants were intentionally kept unaware of the differences between these conditions. However, we did not prevent participants from commenting on differences throughout the session. The sessions were recorded and transcribed to facilitate analysis.

#### 4.1 Procedure

The study procedure was approved by an Institutional Review Board (IRB) before implementation. Each session lasted approximately 1 hour and 15 minutes and was conducted remotely via Zoom. Prior to the session, participants completed a survey about their age, gender, and self-reported disabilities, as well as their virtual meeting experiences. Participants joined the Zoom call from their location of choice to interact with the experimenter, using the AIDA application as their video feed to appear as their avatar. During the session, participants customized their avatar’s appearance. After customization, participants engaged in a virtual conversation using their avatars under four affective rendering conditions, assigned through a Latin square design to vary the presentation order. Participants answered conversation questions adapted from the “Fast Friends procedure” [2], effective in building camaraderie in remote settings [10]. Four questions, one from each set (see Appendix A), were randomly selected for each condition. Participants were encouraged to respond naturally and to observe their avatars as they verbalized the responses. The experimenter communicated verbally, with questions also presented in chat for accessibility. Each discussion lasted about 3 minutes. After the conversation, participants participated in a semi-structured interview (see Appendix B for questions). This procedure was repeated for all four affect combinations. Finally, participants took part in a short exit interview, which

asked them to compare the conditions and overall thoughts. Finally, participants received a \$75 USD gift card as a gratuity.

#### 4.2 Participants

We recruited participants from disability advocacy centers and coalitions serving individuals with disabilities, as well as through online listings for accessibility-related organizations (e.g., mailing lists supporting students or employees with disabilities). Eighteen participants (seven women, seven men, four non-binary) took part in our study. Table 1 summarizes their age, gender, and self-described disabilities. Eleven participants identified as autistic (ASD), while eleven reported social anxiety or generalized anxiety disorders affecting their virtual meeting participation. Additionally, two participants identified as selectively (verbally) mute. During the study, most participants answered semi-structured interview questions using a mix of verbal and written communication; however, three participants (P12, P17, P10) communicated almost solely through written text due to difficulties verbalizing their responses. Twenty participants initially took part in our study. However, two participants were removed by request and technical difficulties.

#### 5 Results

We conducted qualitative analysis using inductive techniques [15] to explore participants’ perceptions of avatars with and without AI-driven affect. We employed Braun and Clarke’s six-phase framework for thematic analysis [6], conducting a grounded thematic analysis of the semi-structured interview transcripts. This systematic approach enabled us to identify patterns, themes, and insights that complement our quantitative results. First, we thoroughly reviewed each transcript to familiarize ourselves with the data. Next, we applied template coding to identify key themes, categorizing responses based on their relevance (e.g., perception of affect, customization, trust in AI) and refining them by specific attributes (e.g., discomfort with affect levels, desire for selective affect). The first author collaborated closely with three coauthors to iteratively code the data, achieving consensus on the identified themes, as recommended by McDonald for CSCW and HCI practices [22]. Throughout the analysis, the remaining authors provided guidance on interpreting the results. Appendix C presents the codes and illustrative quotations for each theme.

##### 5.1 Interview Observations

*5.1.1 Perceived Mental Offloading of Masking and Augmented Communication Abilities.* Through our interviews, we gauged participants’ comfort levels with AI-driven control over their avatar’s affect. Most participants (72%, N=13) conveyed their willingness to embrace AI-driven affect. Several expressed enthusiasm for the technology, noting the comfort it provided in reducing the mental burden of masking, as P15 stated: “I also feel like [the AIDA avatars] could be really helpful to prevent being exhausted in meetings from having to mask with the voice and facial expression. I hope that this would become more normalized.” This observation is notable, since previous research highlights that masking is mentally, physically, and emotionally draining. Participants appreciated the AIDA avatars for automatically masking their facial and vocal affect, with one participant (P04) stating that the avatars “nailed a neurotypical

**Table 1: Demographics of all participants who participated in the study.**

P#	Age	Gender	Described Disability	P#	Age	Gender	Described Disability
P02	18-24	M	ASD	P03	18-24	M	ASD/ADHD
P06	55-64	NB	ASD	P04	25-34	M	ASD, Stuttering
P07	18-24	M	ASD	P12	35-44	M	Social Anxiety
P08	35-44	W	ASD	P14	25-34	M	Social Anxiety
P10	18-24	W	ASD	P16	25-34	W	Social Anxiety
P01	18-24	W	ASD, Generalized Anxiety	P17	18-24	W	Social Anxiety
P05	25-34	W	ASD, Generalized Anxiety	P15	18-24	NB	Situational Mutism, Social Anxiety
P09	18-24	NB	ASD, Social Anxiety	P18	25-34	W	Situational Mutism, Social Anxiety
P11	18-24	NB	ASD, Social Anxiety	P13	25-34	W	Speech impediment, Social Anxiety

perspective." Interestingly, while participants noted that the AI-processed affect did not closely align with their natural personality, they favored its inclusion to conform to social norms, which can be mentally straining. P17 noted: *"I like the expressivity, but I also think that it doesn't reflect how I would sound. I think I would speak the same way throughout, regardless of the subject matter... an avatar that's more expressive than me seems better."*

This observation highlights the tension between societal norms and individual expression. Although participants felt the AI-processed affect didn't match their natural expressions, they preferred it to blend in with expectations. However, even within this exploratory study, where detailed customization was not implemented, some participants felt that using an avatar with AI-driven affect enhanced their communication abilities. P16 states, *"I did truly feel like the avatar was helping me feel more comfortable with sharing more information... So it did really help. And that was really exciting."* The majority of participants reported feeling more at ease and confident in expressing their thoughts, aligning with earlier findings related to masking. They perceived a sense of liberation, feeling less constrained in conveying opinions without the cognitive load of masking their expressions. However, participants also emphasized their preference for this technology in contexts where they felt the need to mask, such as work or virtual learning settings. Conversely, a smaller subset of participants (28%, N=5) expressed discomfort with AI-driven affect for any context, citing a lack of trust in the AI's ability to reflect their intentions and perceiving the affect as impersonal: *"I wouldn't feel good about technology representing [me], because it's just so wildly impersonal. And then it's not just not your face, but it's also not your voice."* –P05

**5.1.2 Accuracy of Vocal Affect Influences User Experience.** In our interviews, we discovered that the accuracy of vocal affect played a crucial role for our participants. Over half of the participants specifically mentioned their discomfort with the AI-processed vocal affect, and placed emphasis on the importance of accurate vocal expressiveness: A number of participants emphasized the significance of precise vocal expression, with one individual noting that "tone speaks louder than words." Overall, participants conveyed a sense of dissatisfaction, stating that a vocal affect that was accurate "most of the time" was still unacceptable, particularly highlighting the discomfort they experienced during the few instances when the tone did not align with their intended communication. Some participants noted that they would rather have vocal affect be flat rather than be wrong, as P13 noted: *"I think most of the time voice*

*did a very good job. But when the voice doesn't match up, I think that is a bigger problem than if the face doesn't match up right?"*

In general, participants observed that a flat vocal affect, while not ideal, was more comfortable than a wrong vocal affect. Additionally, a majority of participants (72%, N=13) also explicitly stated during the interviews that the AI-driven vocal affect felt too intense in general, causing discomfort. Interestingly, our chosen vocal affect aimed to capture a mid-range of emotion for the TTS voices, rated at 0.6 on a 0-1 scale. However, several autistic participants remarked that their natural vocal and facial expressions naturally tend to be less dynamic, and the disparity between their genuine selves and avatar might have contributed to their discomfort. Additionally, several autistic participants explicitly mentioned that the volume modulation (e.g., voice becoming louder when the tone is angry) was occasionally uncomfortable for them. One autistic participant likened the escalating vocal volume to a "jumpscare," highlighting the jarring effect it could have. Autistic participants expressed that they would prefer vocal volume to remain constant, even if it was not similar to how vocal emotions are naturally expressed.

**5.1.3 Facial Affect is Valued, Even With Inaccuracies.** Conversely, only two participants explicitly mentioned that the facial expressions felt exaggerated, which may indicate that the AI-driven facial expressions felt more appropriate. We also found that participants were also uncomfortable with avatars with no facial affect. Several participants emphasized that a lack of facial affect would not resonate with their self-identities and how they wished to be presented. Participants highlighted that the absence of affect in both facial expressions and voice contributed to their negative experience during no-affect conditions. However, participants particularly noted that a lack of facial affect would make them apprehensive on how others perceive them and preferred to have emotive facial expressions, even if the accuracy is not perfect. For example, during the AI-processed facial affect conditions, P16 said: *"Although I dislike that I can't control exactly what the facial expression looks like, I would feel more comfortable with [AI-processed face affect]. I feel like in general, people expect that from me."*

In line with this observation, P14 mentioned that, *"the wrong expression is better than no expression"*. This insight highlights how facial expressions were valued, regardless of complete accuracy. Interestingly, some participants in the "Face Affect" conditions reported a multimodal illusion, perceiving affect in the voice, when

only facial expressions had affect. It is important to note that participants were not told what the avatars were capable of, and were not told of the difference between conditions.

**5.1.4 Desire for Tailored Affect and Authentic Identity.** Notably, the avatars had a set intensity level for both vocal tone and facial expressions, and all participants had the same AI-driven system to determine the appropriate affect based on context. Most participants explicitly noted that they would want more control over the affect expressed by the AI in order to reflect their unique identity and behaviors, such as selectively choosing to express certain emotions. For example, some participants preferred not to express "happy" emotions, or wanted these emotions to be less intense: *"I thought the sad emotions were appropriate. I thought the happy emotions weren't really like my attitude. I just don't seem that happy. –P15"*

On the other hand, several other participants wanted to disable negative emotions, such as anger. For example, P13 mentioned, *"I think expressions were good up until the anger. I just don't like the anger emotion in general in my life. So... seeing the text that I had written, with the anger emotion underlying it felt so weird"*. These observations emphasize that while participants valued the opportunity to alleviate the burden of masking, as previously noted, they also desired a balance by incorporating a degree of authentic expression. Our qualitative analysis also unveiled participants' preferences for customizing their affect. Participants suggested various ideas, including the utilization of sliders to modify the intensity of the affect (P15), text override tags (P03, P18, P06), and on-screen override buttons (P08). P16 expressed the desire to directly adjust facial expressions through a graphical user interface, using menus and sliders. This need for personal representation also caused some participants to dislike the notion of AI-driven affect entirely. A smaller yet notable subset of participants (28%, N=5) expressed discomfort with the notion of AI-driven affect, indicating their reluctance to use similar technologies. Our qualitative analysis unveiled two prominent factors contributing to this sentiment: a lack of trust in the AI's ability to accurately reflect their intentions and a perception of AI-driven affect as impersonal. As stated by P05: *"I wouldn't feel good about technology representing [me], because it's just so wildly impersonal. And then it's not just not your face, but it's also not your voice."*

**5.1.5 Discomfort with Affect Transitions.** More than half of the participants (61%, N=11) highlighted the absence of smooth transitions between emotions as a drawback impacting their experience, providing valuable insights for enhancing AI-driven affect. In the current system, emotions can rapidly transition between extremes (e.g., happy to sad), leading to discomfort. Participants characterized the lack of transitions as "a noticeable contrast" (P08), "intimidating" (P12), "jarring" (P03), and "disruptive" (P05). Additionally, participants noted that the absence of transitions affected their experience adversely, both in terms of facial expressions and vocal tone, indicating that both would be important to their experience. Several participants mentioned that the avatar switched emotions too quickly, which may be a methodological artifact of the "fast friends" prompts, since natural conversation may not change topic as quickly. Nevertheless, it is important for future work to understand how to better transition between contrasting emotions.

**5.1.6 Preferred avatar prototype.** During the exit interview, participants had the opportunity to express their preferences for avatars. The selections were as follows: *Face Affect* (nine participants), *No Affect* (seven participants), *Voice Affect* (two participants), and *Full Affect* (one participant). Most participants chose a single avatar, while one participant opted for two avatars, indicating a preference for distinct avatars for different contexts. The rest of the participants felt that while the context wouldn't change their avatar choice, it would influence whether they'd use an avatar at all (such as for professional engagements but not with friends). Reflecting their interview responses, the majority favored the *Face Affect* avatar, valuing the inclusion of facial expressions even if the emotion was not entirely faithful to their intentions.

## 6 Implications for Future Work

### 6.1 Potential Implications of AI-driven Affect and Masking

Avatars with AI-driven affect may unintentionally promote "masking" or social camouflaging [9], raising ethical concerns. Our research, which involved autistic adults in our design and development team, aligns with the growing emphasis on user-centered design [4]. This study aimed to empower autistic adults and those with social anxiety, who often face challenges in virtual meeting environments. Chapman et al. [9] examined masking's link to mental health factors such as stigma and rejection, highlighting its impact on mental well-being due to the physical, mental, and emotional strain on autistic individuals. Masking is a complex and context-dependent behavior, especially among autistic adults, and cannot be simplified. Buchholz et al. [7] emphasized the importance of allowing individuals with cognitive or communicative disabilities the freedom to choose their preferred mode of remote communication. This is especially relevant for autistic adults, who may choose to mask only in certain situations, such as with strangers [18]. It is crucial that accessibility options, including text modality transformations, are participant-selected to maximize freedom and comfort in online interactions. We acknowledge that avatars with AI-driven affect may reinforce the perception that masking is necessary. However, our user study revealed participants' eagerness to adopt technology similar to AIDA avatars, as it helps alleviate the burden of masking. Several participants even noted that affective avatars reduced the cognitive demands of masking while improving social interaction. These findings suggest that further development in this area could promote inclusivity in virtual meetings.

### 6.2 Implications for Inclusive Avatar Design

**6.2.1 Importance of Accurate Vocal Affect.** Our study highlights the significance of accurate vocal affect in affective avatar design for text modality transformation. Participants were more sensitive to misaligned vocal affect than facial expressions, with some emphasizing that vocal tone was more important than facial accuracy. This suggests that neural vocal affect in its current form may be inadequate for virtual meetings, or that voice customization and personalization play a crucial role. While participants could customize their avatar's visual appearance, they were limited to stock male or female voices. Another factor contributing to discomfort with vocal affect was volume modulation. Our emotive

text-to-speech (TTS) voices varied in volume based on emotion, with certain tones (e.g., happiness, anger, surprise) being louder and others (e.g., sadness, fear) softer. Several autistic participants highlighted difficulties with auditory processing, noting that the volume fluctuations were uncomfortable. Maintaining consistent volume levels for avatar speech, particularly for autistic adults, is an important design recommendation to create a more predictable and comfortable environment.

**6.2.2 Using Customization to Balance Authentic Expression and Societal Pressures.** Our study revealed that while participants appreciated AI-driven affect, they also wanted flexibility in customizing their avatar's emotional expressions to balance authenticity with societal pressures to conform to neurotypical norms. Our avatars featured predefined emotions and intensity levels to enhance expressiveness, but one participant pointed out that this approach primarily reflected a neurotypical perspective. Participants were enthusiastic about using these avatars to reduce the challenges of masking during virtual meetings but also wanted the option to fine-tune emotional expressions to better reflect their authentic selves. Participants suggested various methods for personalizing avatar emotions, including linear sliders for adjusting visual representation and override buttons for immediate control, and text tags for specific emotional modifications. Some adjustments, such as defining the spectrum of emotions the avatar could express, were suggested as part of the initial avatar creation phase, while others, like text tag overrides, could be used in real-time.

## 7 Conclusion

Our research investigates the use of LLM technology to drive emotive avatars, aiming to empower autistic adults and adults with social anxiety in virtual meetings, where text communication is often preferred due to cognitive stressors. We developed AIDA avatars, a system that translates user text into emotional expressions via personalized, AI-driven avatars. Our qualitative analysis revealed participants' enthusiasm for the technology, as it helped address challenges like masking or fatigue, and some even reported improved communication abilities. However, participants also emphasized the need for more personalized control over emotions to match their personalities. Interestingly, they placed greater importance on accurate vocal affect than facial affect, pointing to areas for future research.

## References

- [1] Rahaf Alharbi, John Tang, and Karl Henderson. 2023. Accessibility Barriers, Conflicts, and Repairs: Understanding the Experience of Professionals with Disabilities in Hybrid Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–15. <https://doi.org/10.1145/3544548.3581541>
- [2] Arthur Aron, Edward Melinat, Elaine N Aron, Robert Darrin Vallone, and Renee J Bator. 1997. The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and social psychology bulletin* 23, 4 (1997), 363–377.
- [3] Gazi F Azad, Jill Locke, Margaret Mary Downey, Ming Xie, and David S Mandell. 2015. One-to-one assistant engagement in autism support classrooms. *Teacher education and special education* 38, 4 (2015), 337–346.
- [4] Andrew Begel, John Tang, Sean Andrist, Michael Barnett, Tony Carbary, Piali Choudhury, Edward Cutrell, Alberto Fung, Sasa Junuzovic, Daniel McDuff, Kael Rowan, Shibashankar Sahoo, Jennifer Frances Waldern, Jessica Wolk, Hui Zheng, and Annuska Zolyomi. 2020. Lessons Learned in Designing AI for Autistic Adults. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Virtual Event Greece, 1–6. <https://doi.org/10.1145/3373625.3418305>
- [5] Marion Boberg, Petri Piippo, and Elina Ollila. 2008. Designing avatars. In *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts*. ACM, Athens Greece, 232–239. <https://doi.org/10.1145/1413634.1413679>
- [6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- [7] Margret Buchholz, Ulrika Ferm, and Kristina Holmgren. 2018. “That is how I speak nowadays” – experiences of remote communication among persons with communicative and cognitive disabilities. *Disability and Rehabilitation* 40, 12 (June 2018), 1468–1479. <https://doi.org/10.1080/09638288.2017.1300340>
- [8] Saniye Tugba Bulu. 2012. Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education* 58, 1 (2012), 154–161.
- [9] Louise Chapman, Kieran Rose, Laura Hull, and William Mandy. 2022. “I want to fit in... but I don't want to change myself fundamentally”: A qualitative exploration of the relationship between masking and mental health for autistic teenagers | Elsevier Enhanced Reader. <https://doi.org/10.1016/j.rasd.2022.102069>
- [10] William J. Chopik and Jeewon Oh. 2022. Implementing the Fast Friends Procedure to Build Camaraderie in a Remote Synchronous Teaching Setting. *Teaching of Psychology* (Feb. 2022), 00986283211065746. <https://doi.org/10.1177/00986283211065746>
- [11] Maitraye Das, John Tang, Kathryn E. Ringland, and Anne Marie Piper. 2021. Towards Accessible Remote Work: Understanding Work-from-Home Practices of Neurodivergent Professionals. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–30. <https://doi.org/10.1145/3449282>
- [12] Donna Z. Davis and Karikarn Chansiri. 2019. Digital identities – overcoming visual bias through virtual embodiment. *Information, Communication & Society* 22, 4 (March 2019), 491–505. <https://doi.org/10.1080/1369118X.2018.1548631>
- [13] Tiffany D. Do, Ryan P. McMahan, and Pamela J. Wisniewski. 2022. A New Uncanny Valley? The Effects of Speech Fidelity and Human Listener Gender on Social Perceptions of a Virtual-Human Speaker. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA.) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 424, 11 pages. <https://doi.org/10.1145/3491102.3517564>
- [14] Helen-Ann Brown Epstein. 2020. Virtual meeting fatigue. *Journal of Hospital Librarianship* 20, 4 (2020), 356–360.
- [15] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods* 5, 1 (2006), 80–92. <https://doi.org/10.1177/160940690600500107> arXiv:<https://doi.org/10.1177/160940690600500107>
- [16] Mar Gonzalez-Franco, Anthony Steed, Steve Hoogendyk, and Eyal Ofek. 2020. Using Facial Animation to Increase the Enfacement Illusion and Avatar Self-Identification. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (May 2020), 2023–2029. <https://doi.org/10.1109/TVCG.2020.2973075>
- [17] Eegy Han, Mark R Miller, Cyan DeVeaux, Hanseul Jun, Kristine L Nowak, Jeffrey T Hancock, Nilam Ram, and Jeremy N Bailenson. 2023. People, places, and time: a large-scale, longitudinal study of transformed avatars and environmental context in group interaction in the metaverse. *Journal of Computer-Mediated Communication* 28, 2 (Jan. 2023), zmac031. [https://doi.org/10.1093/jcmc/zmac031\\_eprint](https://doi.org/10.1093/jcmc/zmac031_eprint); <https://academic.oup.com/jcmc/article-pdf/28/2/zmac031/48520441/zmac031.pdf>
- [18] Philippa L Howard and Felicity Sedgewick. 2021. ‘Anything but the Phone!': Communication Mode Preferences in the Autism Community. *Autism* 25, 8 (Nov. 2021), 2265–2278. <https://doi.org/10.1177/13623613211014995>
- [19] Kori M. Inkpen and Mara Sedlins. 2011. Me and my avatar: exploring users' comfort with avatars for workplace communication. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, Hangzhou China, 383–386. <https://doi.org/10.1145/1958824.1958883>
- [20] Sasa Junuzovic, Kori Inkpen, John Tang, Mara Sedlins, and Kristie Fisher. 2012. To see or not to see: a study comparing four-way avatar, video, and audio conferencing for work. In *Proceedings of the 17th ACM international conference on Supporting group work*. ACM, Sanibel Island Florida USA, 31–34. <https://doi.org/10.1145/2389176.2389181>
- [21] Fang Ma, Ju Zhang, Lev Tankelevitch, Payod Panda, Torang Asadi, Charlie Hewitt, Lohit Petikam, James Clemons, Marco Gillies, Xueni Pan, et al. 2024. Nods of Agreement: Webcam-Driven Avatars Improve Meeting Outcomes and Avatar Satisfaction Over Audio-Driven or Static Avatars in All-Avatar Work Videoconferencing. *arXiv preprint arXiv:2412.13265* (2024).
- [22] Nora McDonald, Sarita Schoenbeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–23. <https://doi.org/10.1145/3359174>
- [23] Britta Nordin Forsberg and Kathrin Kirchner. 2021. The Perception of Avatars in Virtual Reality During Professional Meetings. In *HCI International 2021 - Posters*, Constantine Stephanidis, Margherita Antona, and Stavroula Ntoa (Eds.), Vol. 1420.

- Springer International Publishing, Cham, 290–294. [https://doi.org/10.1007/978-3-030-78642-7\\_39](https://doi.org/10.1007/978-3-030-78642-7_39) Series Title: Communications in Computer and Information Science.
- [24] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI* 5 (Oct. 2018), 114. <https://doi.org/10.3389/frobt.2018.00114>
- [25] Soo Youn Oh, Jeremy Bailenson, Nicole Krämer, and Benjamin Li. 2016. Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments. *PLOS ONE* 11, 9 (Sept. 2016), e0161794. <https://doi.org/10.1371/journal.pone.0161794>
- [26] Catherine Oh Kruzic, David Kruzic, Fernanda Herrera, and Jeremy Bailenson. 2020. Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments. *Scientific Reports* 10, 1 (Nov. 2020), 20626. <https://doi.org/10.1038/s41598-020-76672-4>
- [27] Arushi Puri, Steven Baker, Thuong N Hoang, and Romina Carrasco Zuffi. 2017. To be (me) or not to be?: photorealistic avatars and older adults. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction*. ACM, Brisbane Queensland Australia, 503–507. <https://doi.org/10.1145/3152771.3156166>
- [28] Daniel Roth, Gary Bente, Peter Kullmann, David Mal, Chris Felix Purps, Kai Vogeley, and Marc Erich Latoschik. 2019. Technologies for Social Augmentations in User-Embodied Virtual Reality. In *25th ACM Symposium on Virtual Reality Software and Technology*. ACM, Parramatta NSW Australia, 1–12. <https://doi.org/10.1145/3359996.3364269>
- [29] Daniel Roth, Constantin Kleinbeck, Tobias Feigl, Christopher Mutschler, and Marc Latoschik. 2018. Beyond Replication: Augmenting Social Behaviors in Multi-User Virtual Realities. <https://doi.org/10.1109/VR.2018.8447550>
- [30] N. Sadat Shami, Li-Te Cheng, Steven Rohall, Andrew Sempere, and John Patterson. 2010. Avatars meet meetings: design issues in integrating avatars in distributed corporate meetings. In *Proceedings of the 16th ACM international conference on Supporting group work - GROUPE '10*. ACM Press, Sanibel Island, Florida, USA, 35. <https://doi.org/10.1145/1880071.1880078>
- [31] Susan Sprecher. 2021. Closeness and other affiliative outcomes generated from the Fast Friends procedure: A comparison with a small-talk task and unstructured self-disclosure and the moderating role of mode of communication. *Journal of Social and Personal Relationships* 38, 5 (2021), 1452–1471.
- [32] Antoine Tanet, Annik Hubert-Barthelemy, Graciela C Crespin, Nicolas Bodeau, David Cohen, Catherine Saint-Georges, and GPIS Study Group. 2016. A developmental and sequenced one-to-one educational intervention for autism Spectrum disorder: a randomized single-blind controlled trial. *Frontiers in Pediatrics* 4 (2016), 99.
- [33] John Tang. 2021. Understanding the Telework Experience of People with Disabilities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–27. <https://doi.org/10.1145/3449104>
- [34] Siobhan Tierney, Jan Burns, and Elizabeth Kilbey. 2016. Looking behind the mask: Social coping strategies of girls on the autistic spectrum. *Research in Autism Spectrum Disorders* 23 (2016), 73–83.
- [35] Stefano Triberti, Ilaria Durosini, Filippo Aschieri, Daniela Villani, and Giuseppe Riva. 2017. Changing Avatars, Changing Selves? The Influence of Social and Contextual Expectations on Digital Rendition of Identity. *Cyberpsychology, Behavior, and Social Networking* 20, 8 (Aug. 2017), 501–507. <https://doi.org/10.1089/cyber.2016.0424>
- [36] Asimina Vasalou, Adam N. Joinson, and Jeremy Pitt. 2007. Constructing my online self: avatars that increase self-focused attention. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 445–448. <https://doi.org/10.1145/1240624.1240696>
- [37] Kentaro Watanabe. 2023. Augmented telework with avatar technology: impact on workplace and required actions. In *A Research Agenda for Workplace Innovation*, Peter R.A. Oeij, Steven Dhondt, and Adela J. McMurray (Eds.). Edward Elgar Publishing, Chapter 4, 51–66. [https://ideas.repec.org/h/elg/eechap/20504\\_4.html](https://ideas.repec.org/h/elg/eechap/20504_4.html)
- [38] Paul Wehman, Stephanie Lau, Alissa Molinelli, Valerie Brooke, Katie Thompson, Chandler Moore, and Michael West. 2012. Supported Employment for Young Adults with Autism Spectrum Disorder: Preliminary Data. *Research and Practice for Persons with Severe Disabilities* 37, 3 (2012), 160–169. <https://doi.org/10.2511/027494812804153606> arXiv:<https://doi.org/10.2511/027494812804153606>
- [39] Shaomei Wu. 2023. “The World is Designed for Fluent People”: Benefits and Challenges of Videoconferencing Technologies for People Who Stutter. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–17. <https://doi.org/10.1145/3544548.3580788>
- [40] Mika Yasuoka, Tomomi Miyata, Momoko Nakatani, Yuki Taoka, and Nana Hamaguchi. 2023. How Remote-Controlled Avatars Are Accepted in Hybrid Workplace. In *Distributed, Ambient and Pervasive Interactions*, Norbert A. Streitz and Shin’ichi Konomi (Eds.). Springer Nature Switzerland, Cham, 295–307.
- [41] Mika Yasuoka, Marko Zivko, Hiroshi Ishiguro, Yuichiro Yoshikawa, and Kazuki Sakai. 2022. Effects of Digital Avatar on Perceived Social Presence and Co-presence in Business Meetings Between the Managers and Their Co-workers. In *International Conference on Collaboration Technologies and Social Computing*. Springer, 83–97.
- [42] Kengo Yuruki and Masahiko Inoue. 2023. Stress and benefits of video calling for people with autism spectrum disorders. *PLOS ONE* 18, 4 (04 2023), 1–13. <https://doi.org/10.1371/journal.pone.0283526>
- [43] Annuska Zolyomi, Andrew Begel, Jennifer Frances Waldern, John Tang, Michael Barnett, Edward Cutrell, Daniel McDuff, Sean Andrist, and Meredith Ringel Morris. 2019. Managing Stress: The Needs of Autistic Adults in Video Calling. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–29. <https://doi.org/10.1145/3359236>

## A Conversation Questions

Selected conversation questions adapted from [2]. One question was randomly selected from each set, per condition.

### Small Talk

- What was the best gift you’ve ever received and why?
- How often do you get a hair cut? Have you ever had a really bad haircut experience?
- What foreign country would you like to visit? What attracts you to this place?
- If you had to move from your current state, where would you go? What would you miss about your state?
- What is your favorite holiday? Why?
- What is the best TV show you’ve seen? Tell me about it.
- Where did you go to high school? What was your high school like?
- Do you like to get up early or stay up late? Is there anything funny that has resulted in this?
- What did you do this summer?
- What is the best book you’ve read? Tell me about it.

### Set 1

- For what in your life do you feel most grateful?
- If you could change anything about the way you were raised, what would it be?
- If you could wake up tomorrow having gained any one quality or ability, what would it be?
- Before making a phone call, do you ever rehearse what you are going to say? Why?
- What would constitute the “perfect” day for you?
- Given the choice of anyone in the world, whom would you want as a dinner guest?

### Set 2

- If a crystal ball could tell you the truth about yourself, your life, the future, or anything else, what would you want to know?
- Is there something that you’ve dreamed of doing for a long time? Why haven’t you done it?
- If you knew that in one year you would die suddenly, would you change anything about the way you are now living? Why?
- Can you describe a memory where you were sad?
- Can you describe a memory where you were angry?
- What is your most treasured memory?
- What is the greatest accomplishment of your life?
- How close and warm is your family? Do you feel your childhood was happier than most other people’s?
- What do you value most in a friendship?

### Set 3

- Share with your interviewer an embarrassing moment in your life.
- Can you share an experience when emotions played a significant role in your life? What factors or events contributed to those emotions?
- Are there any topics that you believe should be approached with sensitivity and caution when it comes to humor? Why?
- Your house, containing everything you own, catches fire. After saving your loved ones and pets, you have time to safely make a final dash to save any one item. What would it be? Why?

- If you were to become close friends with someone, please share what would be important for them to know.

## **B Interview Questions**

### **Post-Avatar Interview (conducted after each avatar prototype)**

- (1) Did the expression of the avatar's voice match your expectations based on the text you provided? Why or why not?
- (2) Did the expression of the avatar's face match your expectations based on the text you provided? Why or why not?
- (3) How did you feel about viewing the avatar's expressions and emotions of your words?
- (4) How comfortable would you feel with using this avatar in a real meeting? How would context impact your decision? For example, a meeting with your friends compared to a meeting with your coworkers.
- (5) How would you feel about interacting with someone else who is using this type of avatar?

### **Exit Interview**

- (1) If you had another conversation, what "avatar mode" would you choose for that conversation? What are the reasons you would choose that "avatar mode"?
- (2) How much control would you want over your avatar's voice and facial expressions?
- (3) In which ways would you prefer to control these aspects?
- (4) Is there another aspect or attribute that you think is important to convey through your avatar? (e.g., hand gestures, prosody, vocal pace or quality, body language)
- (5) Would you like to say anything more about the avatars?
- (6) Is there anything else we should have asked you about?

## C Codebook

Codes	Illustrative Quotes
<b>Perceived Mental Offloading of Masking and Augmented Social Ability</b>	
AI-driven affect decreases mental burden of masking (72%, N=13)	"I also feel like [the AIDA avatars] could be really helpful to prevent being exhausted in meetings from having to mask with the voice and facial expression. I hope that this would become more normalized." –P15
Perceptions of augmented communication abilities (39%, N=7)	"I like that the avatar was more expressive than I would be when talking during the last 3 questions. I felt it conveyed the feeling one would get when talking about an embarrassing moment better than I would have." –P17
<b>Importance of Vocal Affect Accuracy</b>	
Affective conditions have too much vocal affect (67%, N=12)	"I definitely feel like the voice emotion was too much. I guess I'm glad that they matched just.... I think that it needs to dial it back a bit." –P02
Flat conditions have too little vocal affect (44%, N=8)	"If it was Siri talking to me, I'd be like semi fine with it. But I don't know, as a representation with me? I don't think it matches." –P14
Discomfort with vocal affect inaccuracies	"I didn't feel like the voice affect matched what I was trying to say. But I would rather it be flat than be wrong." –P05
Discomfort with volume modulation (28%, N=5)	"When he answered the question about having one quality or ability, for example, he spoke a bit too loudly for my sake." –P12
<b>Importance of Facial Affect</b>	
Flat conditions have too little facial affect (44%, N=8)	"I would want it to smile more because that would be reflective of my habits, even if there's no good reason for smiling." –P17
Desire for facial expressions, even with inaccuracies (50%, N=9)	"Although I dislike that I can't control exactly what the facial expression looks like, I would feel more comfortable with [AI-processed face affect]. I feel like in general, people expect that from me." –P16
Facial affect induces multimodal perception (33%, N=6)	"I thought the [Face Affect] avatar was pretty cool that it emoted both verbally and with the face. But it wasn't like overly expressive." –P15
<b>Desire for Tailored Affect and Identity</b>	
Desire for selective affect (67%, N=12)	"I think expressions were good up until the anger. I think that was really put me off... in a meeting you usually don't get angry right? Like you shouldn't get angry. So I think, like everything, because everything else is pretty good. Actually, I would feel comfortable using this [if anger was disabled]." –P13
Ways to control customization (44%, N=8)	"I would want some sort of override flag for every message that it'll do it automatically, whatever AI says. But if I want it to be a specific emotion, I can flag it." –P18
Discomfort with voice mismatch (33%, N=6)	"I think it's not that it's not necessarily a bad thing, but it's pretty feminine. And I personally don't identify very well with it." –P11
<b>Affect Transitions</b>	
Discomfort with affect transitions (61%, N=11)	"I think it would be good to have a tone. It's more of just like it felt like one sentence was high, and one it was low, so it felt extreme." –P08