

MIT Open Access Articles

AI Comes Out of the Closet: Using AI-Generated Virtual Characters to Help Individuals Practice LGBTQIA+ Advocacy

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Pillis, Daniel, Pataranutaporn, Pat, Maes, Pattie and Sra, Misha. 2024. "AI Comes Out of the Closet: Using AI-Generated Virtual Characters to Help Individuals Practice LGBTQIA+ Advocacy."

As Published: 10.1145/3640543.3645213

Publisher: ACM

Persistent URL: <https://hdl.handle.net/1721.1/154389>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution-Noncommercial



AI Comes Out Of The Closet: Using AI-Generated Virtual Characters to Help Individuals Practice LGBTQIA+ Advocacy

Daniel Pillis*
 Pat Pataranutaporn*
 pillis@media.mit.edu
 patpat@media.mit.edu
 Massachusetts Institute of Technology
 Boston, Massachusetts, USA

Pattie Maes
 Massachusetts Institute of Technology
 Boston, Massachusetts, USA
 pattie@media.mit.edu

Misha Sra
 University of California, Santa
 Barbara
 Santa Barbara, California

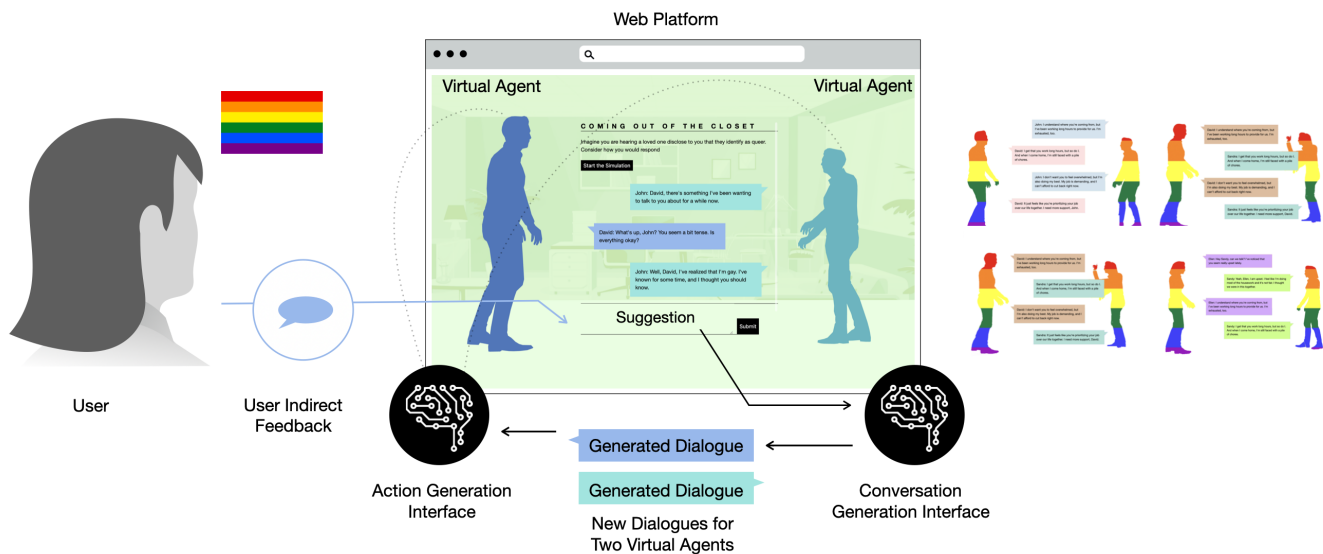


Figure 1: This paper introduces the concept of the "LGBTQIA+ Advocacy Simulator," a pedagogical tool that utilizes virtual characters and large language models to create a space for individuals to actively engage with and explore social scenarios and practice LGBTQIA+ advocacy.

ABSTRACT

Despite significant historical progress, discrimination and social stigma continue to impact the lives of LGBTQIA+ individuals. The use of AI-generated virtual characters offers a unique opportunity to facilitate advocacy by engaging individuals in simulated conversations that can foster understanding, education, and empathy. This paper explores the potential of AI simulations to help individuals practice LGBTQIA+ advocacy, while also acknowledging the need for ethical considerations and addressing concerns about oversimplification or perpetuation of stereotypes. By combining technological innovation with a commitment to inclusivity, we aim

to contribute to the ongoing struggle for equality in both the legal framework and the hearts and minds of the community. We present a study evaluating virtual characters driven by generative conversational AI simulating the social interactions surrounding “coming out of the closet”, a rite of passage associated with LGBTQIA+ communities. In our study, virtual characters embodied as queer individuals engage with users in a text-based conversation simulation paired with visual representations. We investigate how the interactions between the virtual characters and a user influence the user’s comfort, confidence, empathy and sympathy. The AI simulation includes distinct visual personas deployed in a series of conditions. We present findings from our deployments involving 307 users. Finally, we discuss the design implications of our work on the potential future of embodied, self-actuated and openly LGBTQIA+ intelligent agents.

*Both authors contributed equally to this research.



This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

IUI '24, March 18–21, 2024, Greenville, SC, USA
 © 2024 Copyright held by the owner/author(s).
 ACM ISBN 979-8-4007-0508-3/24/03
<https://doi.org/10.1145/3640543.3645213>

CCS CONCEPTS

• **Human-centered computing** → **Interaction design theory, concepts and paradigms; Empirical studies in interaction**

design; Empirical studies in HCI; HCI theory, concepts and models; HCI theory, concepts and models.

KEYWORDS

LGBTQIA+ · Drama Management · AI Actor · Virtual Characters · Player Modelling · Believable Characters · Choice-Based Narrative · Interactive Theatre

ACM Reference Format:

Daniel Pillis, Pat Pataranutaporn, Pattie Maes, and Misha Sra. 2024. AI Comes Out Of The Closet: Using AI-Generated Virtual Characters to Help Individuals Practice LGBTQIA+ Advocacy. In *29th International Conference on Intelligent User Interfaces (IUI '24), March 18–21, 2024, Greenville, SC, USA*. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3640543.3645213>

1 INTRODUCTION

“Either you have tolerance to open community or you don’t, and you don’t get to pick and choose.”

- Sam Altman, 2002, discussing coming out of the closet

When OpenAI CEO Sam Altman was 17 years old, despite the objections of others, a speaker was invited to his high school for National Coming Out Day [3]. In spite of an outspoken attempt by his peers to prevent the talk, Altman instead decided to give a speech in support of the speaker. Altman, who at the time had only recently come out of the closet to his family, gave a speech to the entire student body of his conservative high school—a speech of advocacy, bravery, and openness—from which the above quote is drawn. Altman’s advocacy for tolerance towards ‘openness’, now embroidered into the contemporary culture of artificial intelligence, is a quintessentially queer aspect of the LGBTQIA+ social movement—a sentiment which nonetheless requires persistent resilience in the face of all deterrents.

Despite significant cultural and legal progress, discrimination and social stigma continue to impact the lives of LGBTQIA+ individuals. In recent years, there has been increased recognition of the discrimination and challenges faced by the LGBTQIA+ community. Just in 2023 alone, at least 417 bills were introduced in the US which threatened LGBTQIA+ rights according to the ACLU [17]. Statistics reveal alarming rates of discrimination, with more than one in three LGBTQIA+ Americans experiencing some form of discrimination in the past year [49][40], including a majority of transgender individuals. Discrimination not only has negative psychological impacts, but also affects the economic and mental well-being of LGBTQIA+ Americans. In order to avoid discrimination, many LGBTQIA+ individuals feel compelled to hide personal relationships or alter other aspects of their lives. This highlights the urgent need for LGBTQIA+ advocacy to promote acceptance, equality, and inclusivity for all.

The practice of advocacy plays a crucial role in effecting change and supporting marginalized communities. Scholars define advocacy as taking action and speaking out to support an idea, need, person, or group [35, 37]. In the context of LGBTQIA+ advocacy, collective action has proven vital for addressing the systemic injustices faced by LGBTQIA+ individuals. However, many factors often prevent the widespread adoption of advocacy, particularly if situations are not available for individuals to explore the practice of advocacy for others. Previous research has identified factors such

as perceived injustice, efficacy, and social identity as key determinants of an individual’s decision to engage in collective action. Understanding and addressing these factors can enable effective strategies for advocacy, empowering individuals to advocate for the rights and well-being of the LGBTQIA+ community.

In this work, we propose the use of interactive AI simulations as a unique opportunity to create safe spaces for individuals to practice LGBTQIA+ advocacy. By utilizing AI-generated virtual characters designed as conversational user interfaces, individuals can engage in realistic conversations that simulate real-life experiences that may otherwise be intimidating, harmful, or place LGBTQIA+ individuals at risk. Crucially, interactive virtual simulations allow individuals to develop their advocacy skills, gain confidence in expressing their ideas, and better understand the experiences and perspectives of the LGBTQIA+ community. Moreover, AI simulations can provide a judgment-free and accessible environment where users can make mistakes, learn from them, and in the process, refine their approach to advocacy.

We present a study that evaluates virtual characters driven by generative conversational AI, simulating social interactions surrounding the experience of “coming out of the closet.” Through text-based conversation simulations paired with visual representations, we investigate how these interactions influence the users’ perception and engagement in LGBTQIA+ advocacy. Our study, involving 307 participants, provides insights into the potential of these interfaces for simulating queer social interactions and enhancing LGBTQIA+ potential. By discussing the design implications of our work, we hope to pave the way for future advancements in embodied, self-actuated, and openly LGBTQIA+ virtual characters.

We present: 1) the development of the LGBTQIA+ Advocacy Simulator, a system that utilizes virtual characters and large language models to create a space for individuals to actively engage, practice, and explore conflict scenarios on an accessible platform designed as a web-based system that can be shared and customized; 2) two unique approaches for LGBTQIA+ advocacy: ‘First Person’, which allows the user to play the character, and ‘Third Person’, which allows the user to step back and direct the interaction, and 3) our study, which presents empirical evidence that the ‘Third Person’ simulator significantly improves participants’ confidence, empathy and sympathy when confronted with a “coming out of the closet” interaction. These findings highlight the effectiveness of LGBTQIA+ Advocacy Simulator as a pedagogical tool for enhancing participants’ abilities to navigate LGBTQIA+ interactions in the workplace[5].

2 BACKGROUND: DISCRIMINATION AT WORK

Overwhelmingly, research highlights the need for greater support systems to create more inclusive and accepting workplace environments for LGBTQIA+ individuals. In 2021, the Human Rights Campaign (HRC) released a report [6] highlighting the challenges faced by LGBTQIA+ individuals in the workplace. It noted that up to 30% of LGBTQIA+ employees experienced reduced engagement due to feeling unwelcome at work. Shockingly, 46% of LGBTQIA+ workers reported remaining closeted at work, citing reasons such

as concerns about being stereotyped (38%), making colleagues uncomfortable (36%), and fear of losing connections with coworkers (31%)[6]. Being openly out at work increased the risk of discrimination, with a 10.9% incidence rate compared to 2.2% for those who were not out. Also in 2021, the UCLA School of Law Williams Institute conducted a survey of 935 LGBTQIA+ adults[49] to assess workplace discrimination and harassment experiences. The study revealed that nearly half (45.5%) of LGBT workers at some point in their lives faced unfair treatment at work, including being fired, not hired, or harassed due to their sexual orientation or gender identity. Within the past five years, over 31% of respondents reported experiencing discrimination or harassment, and 8.9% mentioned being fired or not hired because of their sexual orientation or gender identity in the past year. Many respondents had resorted to concealing their LGBTQIA+ identity and altering their appearance to avoid discrimination, and a significant number considered leaving their jobs due to unfair treatment. While both LGBTQIA+ and non-LGBTQIA+ workers report comfort in discussing their personal lives with colleagues, there's a double standard when it comes to LGBTQIA+ personal lives. LGBTQIA+ employees often feel discomfort from coworkers when discussing their sexual orientation or gender identity. A significant portion of non-LGBTQIA+ workers still believe it's unprofessional to discuss these topics in the workplace. [6].

2.1 Coming Out at Work

Noted LGBTQIA+ historian George Chauncey[3] discusses how "coming out" within the gay community preceded its modern connotation. Originally, the term "coming out" didn't refer to gay individuals coming out to the wider world, but instead specifically referred to private relationships, in which gay individuals would "come out" to other gay people. The phrase, historically rooted in the context of debutante balls, originally referred to the rites of passage young women underwent to be introduced to society. Before World War II, a select group of gay men entered this tradition, "coming out" at drag balls, a practice patterned after the debutante balls seen in major cities.[3]. Notably, significant figures in modern technology and STEM have participated in this practice. In 2014, Apple CEO Tim Cook became the first CEO of a major company to openly come out as gay. Cook's international announcement sparked significant discussions, particularly in regions where LGBTQIA+ rights still face challenges[1]. More recently, OpenAI CEO Sam Altman discussed coming out in high school, a place he discussed as "not the kind of place where you would really stand up and talk about being gay, and that was okay." [7]. Thankfully, individuals such as Cook and Altman lead a positive trend towards pride and openness in tech (with their privilege intact) to spearhead conversations towards creating safe workplaces for all.

In the article "Working Closets: Mapping Queer Professional Discourses and Why Professional Communication Studies Need Queer Rhetorics"[18], author Matthew B. Cox proposes the idea of "working closets," a concept which refers to how LGBTQIA+ professionals navigate and succeed in professional settings while managing their sexual orientation or gender identity. This process is described as inherently "complicated", "layered" and comprised

of "networked relations and interactions" that are complex and non-linear. Nowhere is this more relevant than in careers in STEM and tech-related fields, which have been reported as having significant cultural concerns for being safe spaces for LGBTQIA+ employees.

A recent study by the Institution of Engineering and Technology (IET)[2] revealed that a significant portion of LGBTQIA+ individuals avoided pursuing careers in science, technology, and engineering due to concerns about discrimination and bullying. The research highlighted the prevalence of homophobic bullying and outdated attitudes, leading many LGBTQIA+ professionals to remain closeted. Approximately 33% of gay engineers surveyed felt that their sexuality hindered career progression, and a substantial 53% of LGBTQIA+ workers reported hearing derogatory jokes about gay and lesbian people in the workplace. Additionally, 1 in 5 LGBT employees mentioned being pressured to conform to gender stereotypes in their attire[2].

"Queer in STEM: Workplace Experiences Reported in a National Survey of LGBTQIA+ Individuals in Science, Technology, Engineering, and Mathematics Careers"[58] explores the workplace experiences of LGBTQIA+ individuals in STEM professions. While discussions about STEM often emphasize global competitiveness and educational preparation, this research specifically highlights the importance of individual identity factors in these fields. It suggests that marginalized or minoritized groups may face challenges related to job satisfaction, career success, and workplace productivity. To create supportive environments in STEM, it is suggested that it is essential to understand the interplay between STEM professional norms and the broader social group expectations with which individuals may reckon. This highlights a potential tension that influences the need for openness and safety in the workplace, as one truly must bring one's whole self to work in order to be authentic, satisfied and productive.

3 RELATED WORK

Our work on understanding the potential of an LGBTQIA+ Advocacy Simulator is situated within three domains of prior research: (1) a theory of advocacy, which we explore with artificially intelligent virtual characters, (2) simulations that evoke and analyze systems of bias, and (3) the social circumstances of being openly gay in the workplace.

3.1 Theory of Advocacy

Scholars defined advocacy as speaking out and taking action to support an idea, need, person, or group and to effect change in a social system [35, 37]. "Advocates use cognitive, emotional, and behavioral strategies to influence others' attitudes, behaviors, and/or decisions for the benefit of specific individuals (oneself or others) or groups" [37]. Advocacy carried out on behalf of a particular group, such as the LGBT community, is termed "collective action"[52].

Faculty activism plays a vital role in shared governance within higher education institutions, but the underlying dynamics are not well-understood. The author Lori Messenger[37] conducted oral history interviews of 30 university faculty members who were actively involved in advocating for LGBT-supportive policies on their respective campuses. Messenger identified various forms of advocacy, ranging from low-risk actions like signing petitions, to

high-risk actions with potential legal, financial, social, or physical costs. Factors influencing advocacy include the strength of conviction, self-confidence, transformational skills, and the presence of a 'politicized' social identity within the affected group. Additionally, situational factors, such as the availability of resources, support from others (particularly those in positions of privilege), opportunities for voice, popularity of causes, and positive reinforcement were also seen to foster advocacy. Messinger's study concluded that engagement in advocacy is more successful when individuals have institutional support and a culture that encourages questioning and autonomy. The combination of these theories provides a framework for our approach to simulating LGBTQIA+ advocacy.

3.2 Bias Simulations

Reducing social stigmas through the use of simulations has been an oft explored and successful HCI practice. Lee et al. (2023) [32] explored the impact of chatbot-based social contact on reducing mental illness stigma. They designed two chatbots that provided either first-person or third-person narratives about mental illness and conducted a mixed methods study to evaluate their effects. Their findings indicated that participants in both chatbot groups exhibited a decrease in the belief that individuals are personally responsible for their mental illnesses and an increase in their intentions to help. Similarly, Obremski et al. (2023) [4] address the issue of racial bias, both implicit and explicit, in modern society and its negative impact on interactions with foreigners. They proposed a novel approach to reduce racial bias using Intelligent Virtual Agents (IVAs) in Virtual Reality (VR). Two IVAs, representing mixed-cultural and mono-cultural backgrounds, engaged participants in a virtual pub quiz. This study demonstrated a significant decrease in both implicit and explicit bias scores, before and after the intervention.

Nag and Yalçın (2020) [39] discuss the use of gender stereotypes in the design of virtual agents, which often rely on visual, behavioral, and verbal cues to influence user interactions. They conducted an iterative design process to create male, female, and androgynous agents with minimal visual differences. Their results found that traditional gender stereotypes were not as effective or influential as previously believed.

These works showcase the potential of virtual agents as a tool to mitigate real-world biases, and laid the foundation for our subsequent investigation.

The use of interactive simulations for storytelling to investigate bias is a research area with significant impact. Kasibante, Rubegni, and Jaccheri (2022)[28] addressed the issue of gender stereotypes in children by exploring the design of a digital storytelling platform called "Generalize" to detect stereotypes at an early age. In their study, they developed a tool through a co-design process involving experts in children-computer interaction and tested it with adults, such as including guardians, parents, and teachers. While the study had a limited sample size, the results suggested that 'Generalize' effectively facilitated the detection of gender stereotypes.

Similarly, Martin and Gonzalez (2010) [36] investigated the influence of individuals' religious and political affiliations on conflict

resolution strategies in an interactive computer game called 'Peace-Maker'. Their research highlighted the potential of interactive computer games for studying cross-cultural interactions in complex and uncertain environments.

3.3 LGBTQIA+ Representation in Tech

LGBTQIA+ representation in tech is a vital area in significant need of development and support. In a 2011 paper published in 'Interacting with Computers', Ann Light discusses the integration of digital technologies into identity formation and the need to maintain flexibility in self-presentation within these tools. Light [33] drew from Queer Theory to explore ways to resist the rigid formalization of identity in computer systems through a process she calls "queering". DeVito et al. (2021)[20] discussed the evolving role of queer researchers in the field of Queer Human-Computer Interaction (HCI). They highlight the need for better support for queer researchers engaged in HCI research that extends beyond queer populations. Bilimoria and Stewart [12] examined the academic climate for lesbian, gay, bisexual, and transgender (LGBT) faculty in science and engineering at two research universities. Their study, involving 14 faculty members who identify as lesbian or gay, explored the challenges LGBTQIA+ individuals face in their workplaces. While some faculty experience overt hostility, the more common issues include invisibility, interpersonal discomfort, and pressure to conceal their sexuality. The study proposes a model detailing the career consequences of the academic environment for sexual minority faculty in science and engineering, and provides a compelling rationale for future institutional changes to create more inclusive campus climates for LGBTQIA+ faculty across all disciplines.

3.4 Coming Out

The LGBTQIA+ ritual of "coming out" has been studied previously in an HCI context. Dym et al.[22] investigated how online communities, particularly transformative fandom communities with a significant LGBTQIA+ membership, play a vital role in supporting LGBTQIA+ individuals during the process of coming out. The study, involving interviews with 31 LGBTQIA+ fan-fiction authors, revealed that these online spaces create "community narratives" helping individuals establish safety while exploring their identities, challenging stereotypes, and supporting others. The research offers insights that can guide the design of online support spaces for vulnerable populations. Furthermore, a 19-month ethnographic study by Mary L. Gray [25] examined how rural youth use digital media to shape their identities, particularly in the context of "coming out." The study focused on the unique challenges faced by rural LGBTQIA+ youth who must negotiate the politics of visibility and family ties in their daily lives. Gray's approach emphasizes the need for a nuanced understanding of media engagement, one that goes beyond merely consuming specific media texts.

3.5 Virtual Characters

The ability to generate and interact with dynamic, responsive narratives is of importance for interfaces that wish to use stories effectively for education, well-being, or entertainment [9, 42, 43]. The utilization of virtual characters [10, 11, 15, 16, 24, 38, 41, 44, 54] for storytelling has shown promising results. These characters are

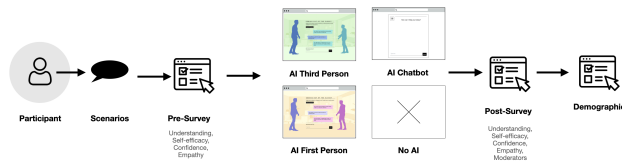


Figure 2: Our experimental study was conducted online and involved a total of 307 participants aged between 18-60 years old, all of whom were recruited via the Prolific platform. The study was pre-registered prior to its launch

realistic digital representations of a person, created through machine learning algorithms such as generative adversarial networks (GANs) [27, 57, 59], neural radiance fields [26, 34, 56], and self-supervised motion transfer networks [45]. AI-generated characters have been used for various purposes, including generating facial animations [26, 27, 34, 56, 57, 59], preserving privacy in documentaries and interviews [14], dubbing films [47], and reanimating historical images [46, 48]. With the recent advancements in AI character generation, it is now possible to create highly realistic representations of users. Additionally, progress in AI-generated conversations based on generative models and natural language processing (NLP) allows for simulating conversations. Many recent large language models (LLMs) based on widely available transformer architectures have increased in model size, training data size, and their capabilities [13, 21, 50]. This adoption of LLMs has significantly broadened the scope of integration between humans and computers, expanding it far beyond the area of human-computer interaction to further the burgeoning realm of human-AI interaction [8, 30, 31, 51, 53, 55].

By incorporating elements from advocacy, virtual characters, the coming out process, and studies on workplace behavioral science, this study introduces the LGBTQIA+ Advocacy Simulator, a unique system that utilizes virtual characters and large language models to create a simulation ‘safe space’ where individuals can actively engage with and practice the experience advocating for colleagues who come out.

4 CONCEPT

LGBTQIA+ Advocacy Simulator, as conceived in this study, is a novel form of interactive performance that unfolds on a virtual stage with virtual characters mediated by active human participation. The overarching conceptualization of LGBTQIA+ Advocacy Simulator revolves around the potential use of interactive virtual characters as an empowering tool to encourage reflection, learning, and LGBTQIA+ awareness through active interventions in AI-driven simulations. LGBTQIA+ Advocacy Simulator intends not to replace or undermine the human roles in advocacy, but to expand its pedagogical reach into the digital realm. We believe that the human element remains critical, with an equally significant role in this context. This approach illuminates our philosophy of using AI for “human augmentation,” [23, 29] emphasizing the role of technology in extending individuals’ capabilities in driving societal and personal transformations.

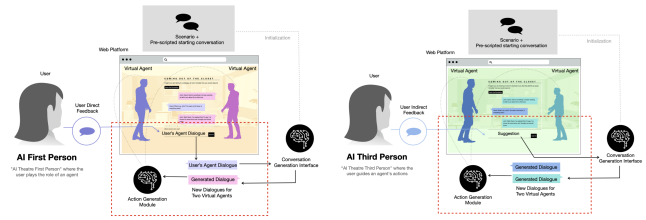


Figure 3: The front-end system of our LGBTQIA+ Advocacy Simulator is web-based, featuring a conversational interface and two virtual characters embodying 3D virtual humans engaged in an interaction. The system was developed using Three.js to render a 3D environment online. The front end is connected through a back-end system containing two AI modules for conversation generation and action generation, both of which drive the animations of the virtual characters. Both modules use GPT-4, an LLM widely adopted for its capability to generate human-like conversations from input prompts via an API call.

5 METHODOLOGY

Our research methodology involved the development of LGBTQIA+ Advocacy Simulator and conducting an online experimental study with the aim of examining how LGBTQIA+ Advocacy Simulator empowers individuals in their interpersonal conflict resolution. The research questions guiding our inquiry are as follows:

- RQ1: How may an LGBTQIA+ Advocacy Simulator impact individuals’ empathy and sympathy, comfort, and confidence when interacting with LGBTQIA+ colleagues coming out of the closet?
- RQ2: How do different modes of interaction within LGBTQIA+ Advocacy Simulator (First Person vs Third Person) affect the main outcomes and influence people’s perceptions?
- RQ3: What are the qualitative experiences of participants engaging with the LGBTQIA+ Advocacy Simulator, and how do these experiences contribute to their overall understanding and attitudes towards the experiences of LGBTQIA+ individuals in the workplace?

Our experimental study was conducted online and involved a total of 307 participants aged between 18-60 years old, all of whom were recruited via the Prolific platform. The study was pre-registered prior to its launch.

5.1 Technical Implementation

In demonstrating our vision, we developed three separate conditions of the experimental simulation, each of which explores the experience of LGBTQIA+ advocacy using unique variations of AI interfaces. We outline our development here.

5.1.1 The Front-end & Back-end systems. The front-end system of LGBTQIA+ Advocacy Simulator is web-based, featuring a conversational interface and two virtual characters embodying 3D virtual humans engaged in an interaction. The system was developed using the web programming language Three.js to render an interactive 3D environment online. The front end is connected through a back-end

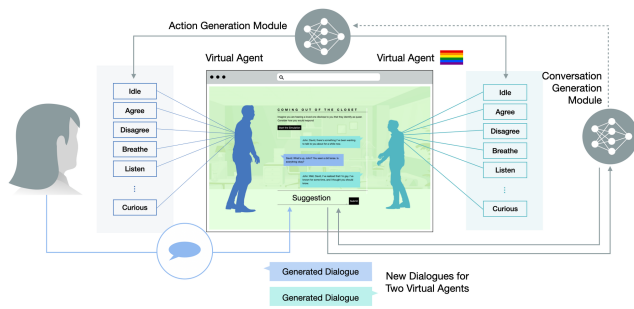


Figure 4: A description of how the action generation module works, showing each of the movements that can be triggered by the action generation module

system containing two AI modules for conversation generation and action generation, both of which drive the animations of the virtual characters. We implemented the public OpenAI API to manage the dialogue. Both interfaces use GPT-4, a large language model widely adopted for its capability to generate human-like conversations from input prompts via an API call.

5.1.2 Conversation Generation Module. The conversation generation module is prompted to continue the dialogue given the previous dialogue in a manner appropriate to the scenario and the interaction condition. The interaction starts with a pre-scripted starting dialogue created by the authors. In the ‘First Person’ condition, the user controls the dialogue for one virtual character, thus the conversation generation module only generates dialogue for the remaining virtual character. In contrast, in the ‘Third Person’ mode, the user takes charge of the entire interaction, and the conversation generation module generates dialogues for both virtual characters, following the user’s instructions.

5.1.3 Action Generation Module. The action generation module (AGM), which drives the animation from the user, receives the dialogue from the conversation generation module (CGM). Based on the nature of the text provided, the CGM decides upon the most appropriate animation from the fourteen pre-loaded animation tracks in the virtual characters: agree, disagree, breathe, chat, curious, focus, idle, nagging, relax, stand, listen, argue, serious. These animation tracks were curated from open-source online databases. We employed animation blending technology to combine multiple animations, creating performances composed of a series of seamlessly integrated movements.

5.1.4 Virtual Agents. The virtual characters, also curated from open-source sources, come in various genders and body types. We deliberately used lush colors with an emphasis on bright pastels and soft compositions to create a flattened abstraction, symbolizing humans without explicitly picturing their gender identity or race. The characters each have bones and animation tracks pre-loaded in them and have been exported as .glb files for web rendering in ThreeJS.

5.1.5 User’s Interaction. When the system is loaded, it begins by displaying the premise of the scenario and a pre-scripted starting

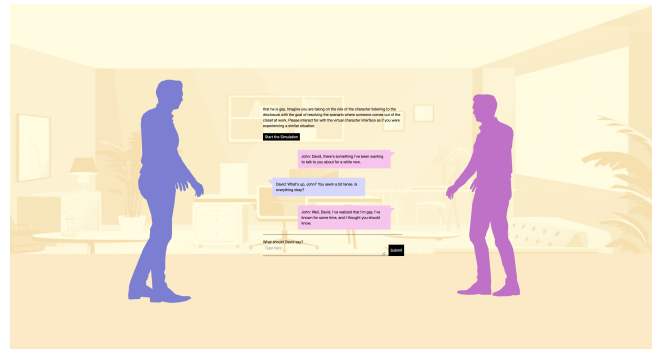


Figure 5: The first condition, ‘First Person’ interface.

dialogue. The user is presented with the initial premise through the web interface. As the participant begins interacting with the system, starting dialogue loads on the screen in text bubbles, and the action generation module, powered by LLMs, triggers the animation in the corresponding virtual characters. When the pre-scripted dialogue is complete, depending on the condition, a message appears, either inviting the participant to pretend to be a character and respond, or to provide direction on what the virtual character should do next. The participant then types the response into the response area. The input received from the participant is sent to the conversation generation module to generate new dialogue, which in turn triggers the action generation module. GPT-4 then interprets the whole context of the participant’s text, the original text, and a looping prompt, generating new exchanges between the virtual characters based on the participant’s text entry.

5.2 Experimental Protocol

In our study, each participant in the experimental conditions was randomly assigned to interact with one of two experimental conditions or one of two baselines, used as control conditions. We describe each condition below in detail.

- **First Person**

In the ‘First Person’ approach, the participant takes on the role of one of the virtual characters witnessing one individual coming out to a colleague. The participant speaks as the character by directly typing in their response. By taking on the role of a virtual character, participants are able to role play how they may respond to a similar situation in their real life, using their natural speech. Virtually projecting oneself into another character has been shown to decrease discriminatory biases and increase reasoning skills [19].

- **Third Person** The second scenario, the ‘Third Person’, the participant guides the behavior of a virtual character in the same interaction. A participant interacts with the same situation and is enabled to suggest the actions of one of the characters, guiding the behavior of the character without directly supplying dialogue. In this scenario, the user’s input functions as a prompt for the LLM to produce new dialogue as opposed to the first condition where the user sends input of direct speech. This indirect control allows the participants

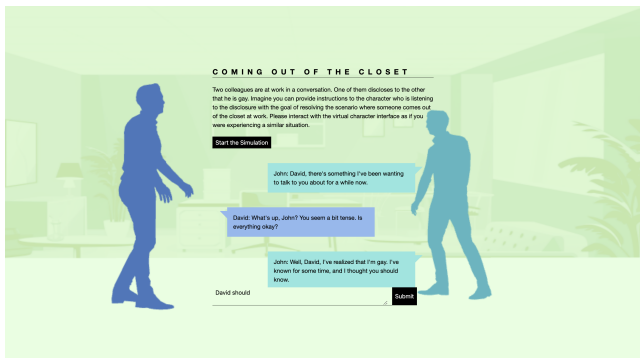


Figure 6: The third condition, 'Third Person' interface.

to approach the problem from a distance and observe the interactions between the two virtual characters with a certain level of detachment.

- **AI Chatbot** In the AI chatbot condition, a participant interacts with a single AI chatbot without any representation of a virtual character. This serves as the baseline and is effectively the same interface as when people consult with a regular chatbot, such as ChatGPT, for advice or suggestions. In this condition, the participant responds to the same coming out scenario (without any visualization of virtual characters) and provides text-based interactions within a standard conversational user interface.
- **No AI: (Control Condition)** In the fourth condition, our second baseline and control, participants respond to the same scenario (without an AI agent) and provide a general response to the coming-out premise in a text entry box. Participants did not interact with any experimental condition or an AI agent, and went directly to the text entry summary. This text entry summary was also present as the last interaction in each of the other three conditions.

The study conducted was a three-part, online, between-subjects study, executed using Qualtrics. The first part entailed pre-intervention instructions given to the participants. First, participants were asked to read a description of the 'Coming out of the Closet' scenario. Subsequently, participants were given a questionnaire probing their initial reaction to the main outcome variable measured by seven statements organized into three clusters: composite-empathy-sympathy, composite-comfort, and composite-confidence.

The second part of the study involved randomly placing participants into one of the four conditions. Participants were requested via a text prompt to interact with the simulation as if they were experiencing a similar situation in their day-to-day lives. We suggested each participant interact with the interface for at least five minutes, or until they felt the interaction was resolved (although there was no strict time limit imposed on the interactions).

In the third and final part, participants were asked the same set of questions for the main outcome variables, provided prior to their interaction with the intervention. The experiment concluded with a series of questions used as moderators, alongside standardized ones used to understand participant demographics and their opinions about the interface.

5.3 Scenarios

In this study, we designed a scenario around an LGBTQIA+ interaction situated in a workplace environment. The scenario provided the context and initiated a conversation for the participants, but was left deliberately open-ended. The base scenario dialogue is replicated below:

John: "David, there's something I've been wanting to talk to you about for a while now."

David: "What's up, John? You seem a bit tense. Is everything okay?"

John: "Well, David, I've realized that I'm gay. I've known for some time, and I thought you should know."

5.4 Measurement

5.4.1 Main Outcome Variables. Prior to interacting with the study, participants were presented with a description of the scenario and asked to consider their first impression, i.e. how they would approach a colleague coming out of the closet at work if confronted with a similar situation. After reviewing the statement, each participant was prompted to respond to a questionnaire. They were asked to report how much they agreed/disagreed with each on a 7-point Likert-scale (1 = "Strongly disagree" and 7 = "strongly agree"). Questions were grouped together based on three different composite scores: "Composite-empathy-sympathy (4,6)", "Composite-confidence(2, 3, 5)", and "Composite-comfort (1, 7)".

5.5 Statistical Analysis

We calculated the delta from pre and post-intervention ratings on clustered groupings of the main outcome variables, organized into three themes: comfortability in the scenario, empathy and sympathy for the scenario, and confidence in their ability to approach the scenario. We also calculated the mean for each of the moderators. In terms of statistical analysis, we compared the delta between the main outcome variables between experimental conditions. We conducted a homogeneity test using a Levene test to assess whether the samples were from populations with equal variances. If the samples were not homogeneous, we ran a Welch analysis of variance (ANOVA) and a Tukey post-hoc test. If the samples were homogeneous, we ran a basic ANOVA test and a Tukey post-hoc test. For the moderators, we ran linear regression to see if the moderators affect the main outcome variables.

5.6 Participant

For our evaluation, we recruited a total of 307 online participants. The participant pool had a balanced sex distribution and a mean age of 35.3 years. 147 of the participants identified themselves as female, 151 as male, 6 as non-binary, 2 opted to not disclose and 1 opted to self-describe. 223 of the participants identified as Straight/Heterosexual, 42 identified as Bisexual, 11 identified as Gay/Lesbian/Homosexual, 8 as Asexual, and 14 preferred not to describe or specify. All participants were from the United States and fluent in English. For participating in the study, each respondent received a payment. Participation lasted around 15 minutes.

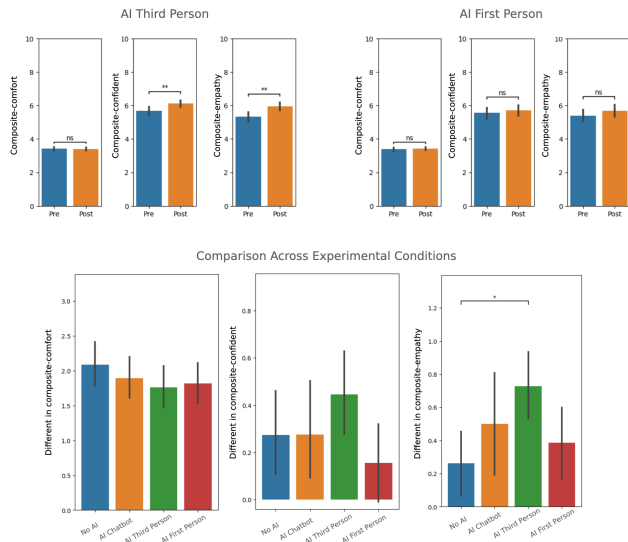


Figure 7: We observed a significant difference in participants' empathy and sympathy, with the Third Person condition showing the highest increase. p-value annotation legend: ns: $p \leq 1.00e+00$ *: $1.00e-02 < p \leq 5.00e-02$ **: $1.00e-03 < p \leq 1.00e-02$ *: $1.00e-04 < p \leq 1.00e-03$ ****: $p \leq 1.00e-04$**

5.7 IRB

This research was reviewed and approved by the local ethics review board, the Committee on the Use of Humans as Experimental Subjects.

6 RESULTS

We outline our findings of the main outcome variables, moderators, as well as participant's feedback on the system. First, participants were randomly assigned to each condition with the equal distribution: First Person ($n = 71$), Third Person ($n = 77$), AI Chatbot ($n = 75$), and No AI Control ($n = 84$). The conversation turns for each experimental condition was as follows: First Person ($M = 13.85$, $SD = 10.72$), Third Person ($M = 12.29$, $SD = 7.30$), AI Chatbot ($M = 9.06$, $SD = 4.51$).

6.1 Main Findings

We investigated the impact of the degree to which experiencing a simulated interaction with a virtual character coming out of the closet impacts the participant's comfort level within the given situation, the participant's confidence in providing support to the person, the participant's confidence in making the other party feel supported, the participant's capacity for empathy toward the other person in the scenario, the participant's sense of capability and effectiveness in managing the situation, the participant's feelings of sympathy regarding the scenario, and the participant's perception of the situation's difficulty, as well as how different modes of interaction (Third Person vs First Person) affect the main outcomes and perceptions. We measured these variables before and after interaction, and our results are derived from the delta of participants'

answers. We observed a significant difference in participants' empathy and sympathy, with the 'Third Person' condition showing the highest increase.

6.1.1 Empathy and Sympathy. Participants' assessment of how empathetic and sympathetic they felt about the scenario was evaluated using two self-report questions, rating the following on a seven point Likert scale, 'I would feel empathetic about the other person in the scenario' and 'I felt sympathetic about the scenario'. An analysis of variance (ANOVA) showed a significant effect of the different experimental conditions on participants' report of being empathetic or sympathetic ($p = 0.043$). The mean values for the different conditions were as follows: No AI ($M = 0.26$, $SD = 0.96$), First Person ($M = 0.39$, $SD = 0.97$), AI Chatbot ($M = 0.50$, $SD = 1.35$), Third Person ($M = 0.73$, $SD = 0.92$).

Post-hoc analysis using the Tukey Test reveals that participants in different conditions exhibited varying levels of change in their responses. AI Chatbot ($M = 0.50$, $SD = 1.35$) showed a statistically significant increase in their responses ($p = 0.011$) when compared to No AI ($M = 0.26$, $SD = 0.96$). AI Chatbot ($M = 0.50$, $SD = 1.35$) also demonstrated a substantial increase compared to 'Third Person' ($M = 0.727$, $SD = 0.916$).

'First Person' ($M = 0.387$, $SD = 0.968$) exhibited a moderate increase compared to both 'Third Person' ($M = 0.727$, $SD = 0.916$) and 'No AI' ($M = 0.262$, $SD = 0.965$), with effect sizes of -0.359 and 0.129 , respectively. In summary, the participants in the different conditions showed varying levels of response increase, with 'AI Chatbot' displaying the highest increase, followed by 'First Person' and 'Third person', while 'No AI' had the lowest increase.

6.1.2 Confidence. Participants' assessment of their confidence in approaching the scenario was assessed using three self-report Likert statements, "I felt confident that the person would feel supported", "I felt confident that I would be able to support the person" and "I felt capable and effective handling the scenario".

An analysis of variance (ANOVA) revealed no statistically significant impact of the different experimental conditions on participants' comprehension of the conflict scenario ($p = 0.20$). The mean values for the different conditions were as follows: No AI ($M = 0.27$, $SD = 0.82$), First Person ($M = 0.15$, $SD = 0.71$), Third person ($M = 0.44$, $SD = 0.77$), and Chatbot ($M = 0.27$, $SD = 0.95$). It is important to note that while the sample sizes were not equal, homogeneity was met as indicated by Levene's test ($s_{levene} = 0.45$, $p_{levene} = 0.71$). The basic ANOVA results showed that there was no significant difference in participants' understanding across the experimental conditions ($s_{bANOVA} = 1.57$, $p_{bANOVA} = 0.20$).

6.1.3 Comfort. Participants' assessment of how 'comfortable' they felt about speaking to a work colleague coming out of the closet was evaluated using two self-report measures on a 7-point Likert scale, 'I felt comfortable in this situation' and 'I felt confident that I would be able to support the person'. An analysis of variance (ANOVA) showed no significant effect of the 'First Person', 'Third Person' or 'Chatbot' conditions on participants' report of being comfortable ($p = 0.354$). The mean and standard deviation for each sample were as follows: 'First Person' had a mean of 3.54 with a standard deviation of approximately 0.48 , 'Third Person' had a mean of 3.48 with a standard deviation of approximately 0.53 , and 'Chatbot' had a mean

of 3.57 with a standard deviation of approximately 0.58, while the control condition, 'No AI', had a mean of 3.64 with a standard deviation of approximately 0.68.

6.2 Comparison Across Experimental Conditions

6.2.1 Empathy & Sympathy. We performed tests on four different samples: 'No AI', 'First Person', 'Third Person', and 'Chatbot'. The sample sizes are considered sufficient for analysis. Although the sample sizes are not equal, a test for homogeneity suggests that the variances among the samples are similar.

A basic analysis of variance (ANOVA) test was conducted, and it yielded a statistically significant result (p -value = 0.043). This indicates that there is a significant difference in means among the samples.

A Tukey test was conducted to compare the means of different sample pairs. The results of the Tukey test revealed the following. When comparing the 'Third Person' condition against the 'First Person': (p = 0.211) there was no significant difference. When comparing No AI vs. 'AI Chatbot': (p = 0.493) there was no significant difference. However, when comparing the 'No AI' condition against the 'Third Person': (p = 0.030), there was a significant difference.

Upon post-hoc analysis using the Tukey Test, it was observed that participants in various conditions exhibited differing levels of change in their responses. Specifically, 'AI Chatbot' (M = 0.50, SD = 1.35) displayed a statistically significant increase in their responses (p = 0.011) compared to No AI (M = 0.26, SD = 0.96). Furthermore, 'AI Chatbot' (M = 0.50, SD = 1.35) demonstrated a notable increase in comparison to 'Third Person' (M = 0.73, SD = 0.916).

In summary, the basic ANOVA test found a significant difference in means among the samples, and the Tukey test identified a significant difference between the 'No AI' and 'Third Person' samples.

6.2.2 Comfort. The sample sizes are deemed sufficient for analysis. Although the sample sizes are not equal, a test for homogeneity indicates that the variances among the samples are similar. A basic analysis of variance (ANOVA) test was performed, but it did not yield a statistically significant result (p -value = 0.443). The mean values for the different conditions were as follows: 'No AI' (M = 2.089, SD = 1.477), 'First Person' (M = 1.817, SD = 1.334), 'Third Person' (M = 1.760, SD = 1.314), and 'AI Chatbot' (M = 1.893, SD = 1.313). This suggested that there is no significant difference in means among the samples.

Post-hoc analysis using the Tukey Test indicates that there were no statistically significant differences in the responses between the different conditions. The p -values for all pairwise comparisons were above the significance threshold of 0.05. Thus, it can be concluded that there were no significant differences in responses between the conditions.

6.2.3 Difference for Confidence. The data shows the results of running statistical tests on four different samples: 'No AI', 'First Person', 'Third Person', and 'Chatbot'. The sample sizes are sufficient for the analysis, and homogeneity is met, indicating that the variances of the samples are similar. A basic analysis of variance (ANOVA) test

was conducted, and the result was not statistically significant (p -value = 0.196). This suggests that there is no significant difference in the means of the samples.

The mean values for the different conditions were as follows: 'No AI' (M = 0.274, SD = 0.828), 'First Person' (M = 0.155, SD = 0.710), 'Third Person' (M = 0.446, SD = 0.775), and 'AI Chatbot' (M = 0.276, SD = 0.958). Post-hoc analysis using the Tukey Test indicates that there were no statistically significant differences in the responses between the different conditions. The p -values for all pairwise comparisons were above the significance threshold of 0.05. Thus, it can be concluded that there were no significant differences in responses between the conditions.

6.3 Moderators

After interacting with the intervention simulation, participants were then asked to fill out additional survey questions which were used as moderators, comprised of custom and standardized questions. The standardized questions were drawn from the Lesbian, Gay, and Bisexual Knowledge and Attitudes Scale (LGB-KAS), the Connectedness to the LGBT Community Scale Items, Origin, and Factor Loadings Scale, and the General Attitudes towards Artificial Intelligence Scale. Below, we report the result of our inquiry into how moderating factors such as knowledge of the LGBTQIA+ community, perceived realism of the simulation, connectedness, and attitudes toward AI affected the outcomes of the interaction.

6.3.1 Perceived Realism of The AI Simulation. Linear regression reveals that the perceived realism of the AI simulation moderated the outcomes in term of confidence (p = 0.0001, slope = 0.1622, Intercept = -0.5076, r = 0.2538, SE = 0.0416), but not empathy/sympathy (p = 0.0893, slope = 0.0970, intercept = 0.0623, r = 0.1141, SE = 0.0568) or comfortability (p = 0.8120, slope = -0.0066, intercept = 3.5617, r = -0.0160, SE = 0.0276). This indicates that the more realistic the simulation is, the greater there is a likelihood of users feeling confidence in their ability to interact with LGBTQIA+ simulations. An open challenge for future design is how to use simulations such as this to encourage more empathy in participants.

6.3.2 Attitude towards AI. Regarding participants' attitudes toward artificial intelligence, participants were asked to rate the following statements: "There are many beneficial applications of Artificial Intelligence," "I am impressed by what Artificial Intelligence can do," "I think Artificial Intelligence is dangerous," "Artificial Intelligence can have positive impacts on people's wellbeing," "Artificial Intelligence is exciting," and "I think artificially intelligent systems make many errors".

Linear regression revealed that Moderator-AI did not moderate any of the outcome variables: empathy-sympathy (p = 0.2289, slope = 0.0836, Intercept = 0.1271, r = 0.0809, SE = 0.0693), confidence (p = 0.1223, slope = 0.0807, Intercept = -0.1049, r = 0.1038, SE = 0.0520), and comfort (p = 0.3938, slope = 0.0286, Intercept = 3.3870, r = 0.0574, SE = 0.0335).

6.3.3 LGB-KAS-Knowledge. To assess participants' familiarity with LGBTQ+ issues, participants were questioned about the following statements: (1) "I am knowledgeable about the history and mission of the PFLAG organization," (2) "I am familiar with the significance of the Stonewall Riot to the Gay Liberation Movement," (3) "I am

aware of the work of the National Gay and Lesbian Task Force," (4) "I could educate others about the history and symbolism behind the 'pink triangle'," and (5) "I feel qualified to educate others about how to be affirmative regarding LGBTQIA+ issues."

Linear regression revealed that LGB-KAS-Knowledge did not moderate any of the outcome variables: empathy-sympathy ($p = 0.2326$, slope = -0.0570 , Intercept = 0.7069 , $r = -0.0803$, SE = 0.0476), confidence ($p = 0.6058$, slope = -0.0185 , Intercept = 0.3494 , $r = -0.0347$, SE = 0.0359), and comfort ($p = 0.4667$, slope = 0.0168 , Intercept = 3.4808 , $r = 0.0490$, SE = 0.0230).

6.3.4 LGBTQIAConnectedness. To assess participants' connection to the LGBTQIA+ community, participants were asked to respond to the following statements: (1) "You feel you're a part of the LGBTQIA+ community," (2) "Participating in the LGBTQIA+ community is a positive thing for you," (3) "You feel a bond with the LGBTQIA+ community," (4) "You are proud of the LGBTQIA+ community," (5) "It is important for you to be politically active in the LGBTQIA+ community," and (6) "You really feel that any problems faced by the LGBTQIA+ community are also your own problems."

Linear regression reveals that feelings of connectedness to the LGBTQIA community did not moderate any of the outcome variables: empathy-sympathy ($p = 0.4362$, slope = -0.0332 , Intercept = 0.6723 , $r = -0.0524$, SE = 0.0425), confidence ($p = 0.9533$, slope = 0.0019 , Intercept = 0.2886 , $r = 0.0039$, SE = 0.0320), or comfort ($p = 0.0568$, slope = 0.0390 , Intercept = 3.3766 , $r = 0.1278$, SE = 0.0204).

6.3.5 Internalized Affirmativeness. To assess participants' attitudes and behaviors related to LGBTQIA+ issues, participants were asked to respond to the following statements: (1) "Feeling attracted to another person of the same sex would not make me uncomfortable," (2) "I would display a symbol of gay pride (pink triangle, rainbow, etc.) to show my support of the LGBTQIA+ community," (3) "I have close friends who are LGBTQIA+," (4) "I conceal my positive attitudes toward LGBTQIA+ people when I am with someone who is homophobic," (5) "I feel qualified to educate others about how to be affirmative regarding LGBTQIA+ issues," and (6) "I would attend a demonstration to promote LGBTQIA+ civil rights."

6.4 Qualitative Analysis

Here we report the participants' subjective experience during the scenario in each of the four conditions. To document the self-report of the participants, we asked each participant to provide a text description of how they would handle a scenario related to a colleague coming out of the closet at work. Participants in the condition generally reported positive reactions to the realistic nature of the AI interactions, the interface experience, and the interactions with the AI.

In the 'No AI' condition, participants gave direct responses to the situation. Consistent with prior research, participants repeatedly commented that they did not feel that the colleague should have shared the 'personal information' because it was not relevant to the workplace.

Notable examples from each condition are reproduced below:

In the 'No AI' condition: "I would emphasize remaining on work-related tasks if we are not making good progress in our work.

I would possibly inquire during a break why such a thing would come up in conversation."

"I'd advise the colleague who is not gay to steer the conversation elsewhere after an acknowledgment of the statement. I don't think sexual orientation should be a conversation within the workplace."

"I would question why the coworker feels like I should know this, first off. Also, I'd be sure to let them know right off the bat that I don't endorse that sort of lifestyle. I think it's sinful."

"Firstly, nobody should be disclosing sexual orientation or sexual preferences at work in a professional setting. To be a person that engages in conversation regarding sexual acts in a workplace, it says that you do not possess a high social intelligence. Secondly, the better way to address a situation like this is to not at all. Nobody cares or should care about people's sex acts and preferences."

In the 'Third Person' condition,

"It is probably not okay to discuss matters like this at work."

"State to the person who has come out, 'This is an inappropriate work conversation. I will inform my supervisor of this incidence. Please do not mention this to me again.'"

In the 'First Person' condition,

"I told him I don't care who he sleeps with. It's not related to work."

"I suggested that there is no reason to reveal your sexual preferences in the workplace. He responded about people judging him and he was struggling with identity. I told him announcing his sexuality is different from identity questions and reminded him that it is against company policy to date or have romantic relationships with coworkers,

"I wanted to avoid talking about it"

"I would be kind to him and not disrespect him or treat him differently at work, but I would not be able to socialize with him outside of work because I have values that differ from his."

In the 'Chat bot' condition:

"We will be informally discouraging employees from sharing overly personal information while they are on the clock. We cannot control what they share on their own time, such as during breaks, but during actual work periods we will ask that conversation be kept to work-related subjects or general subjects."

"I would remain respectful and not tell my colleagues about my religious beliefs. Work requires not to discriminate based on sexual orientation."

"My approach was to see what the chatbot would offer as a solution and see if the chatbot would be able to adapt to conflicting solutions. The chatbot had only one solution which was to force me to accept the person's ideology which I was not going to do. I had to bring the conversation to an end as the chatbot was biased."

"I disclosed how uncomfortable I felt discussing sexual matters in the workplace and will now seek employment elsewhere."

These responses are consistent with prior sociological studies which have shown the 'double standard' when it comes to LGBTQIA+ personal lives, as LGBTQIA+ employees often feel discomfort from coworkers when discussing their sexual orientation or gender identity. As cited above, a significant portion of non-LGBTQIA+ workers still believe it's unprofessional to discuss these topics in the workplace[4], yet bringing one's whole self to work demands it.

7 DISCUSSION

Reducing stigma using simulations has been an oft explored and successful HCI practice. Lee et al. (2023) [32] explored the impact of chatbot-based social contact on reducing mental illness stigma. They designed two chatbots that provided either first-person or third-person narratives about mental illness and conducted a mixed methods review. Initially, our study revealed that participants exhibited a pre-existing understanding of the proposed scenarios, and the Advocacy Simulator did not substantially alter this pre-existing comprehension. This phenomenon can be attributed, in part, to the fact that individuals already possessed a high level of familiarity with the scenario in question, as evidenced by their consistently high scores in the post-intervention questionnaire.

Our analysis further shows the impact of the LGBTQIA+ Advocacy Simulator. The study demonstrates that while the Advocacy Simulator may not necessarily impact the understanding of the conflict scenario, they can significantly improve participants' confidence and self-efficacy in devising solutions for the scenarios.

We observed a significant difference in participants' sympathy towards the situation, with LGBTQIA+ Advocacy Simulator Third Person condition showing the highest increase. This could be due to the nature of the Third Person approach, which allows participants to observe the interactions between the two virtual agents from a certain level of detachment. This distance may provide them with a broader perspective and a clearer understanding of the dynamics at play in the conflict scenario. By not directly embodying one of the characters, participants are freed from the constraints of personal biases or emotions that may hinder their decision-making.

Our results improve our understanding of using approaches to advocacy to study increasingly capable and dynamic virtual agents. Furthermore, the study found a significant increase in participants' confidence in their solution, with the Advocacy Simulator First Person condition showing the highest improvement. This heightened engagement can be attributed to the personal transformation experienced by the user, as they become deeply involved in the character's perspective. By taking on the role of the central character in the conflict scenario and successfully resolving the case, the simulation becomes personal to the user. As a result, their confidence levels are bolstered, leading to a greater belief in the effectiveness of their chosen solution.

We found that the believability of the AI simulation had a moderating effect on the outcome. This suggests that the more realistic the simulation, the more likely users are to feel sympathy and empathy. Therefore, future design considerations should focus on how to use simulation to increase sympathy and empathy. Understanding these moderating variables can guide the design and implementation of future LGBTQIA+ Advocacy Simulator systems to optimize their effectiveness in promoting individuals to deal with interpersonal conflicts confidently.

7.1 Limitations

While our study shows promising results and a significant trend, there are several limitations that we need to acknowledge. First, the study was limited to self-reported measures of confidence and self-efficacy. Although these measures are widely used and accepted in psychological research, we acknowledge that future studies could

benefit from including more objective measures, such as independent evaluations of the quality of conflict resolution strategies proposed by participants. Second, the script of the virtual agents was pre-defined, which may have limited the potential for unique, improvised interactions. More dynamically generated conversations could allow for a broader spectrum of responses, further personalizing the experience for the user. Third, future work may delve into more complex, nuanced conflicts that reflect the real-world challenges individuals face in their personal and professional lives. This would provide a more immersive and potentially impactful experience. The system might also allow the user to provide their own conflict for the simulation. Finally, as our results demonstrate, the experience of empathy from an AI agent is not yet resolved enough to enable participants to feel that the interactions conveyed compelling affective interactions to the users.

7.2 Ethical Considerations

When considering the ethical implications of deploying the LGBTQIA+ Advocacy Simulator, several points must be raised. Firstly, the LGBTQIA+ Advocacy Simulator is not created with the intention of eclipsing or undervaluing human participation actual, lived advocacy. Instead, its objective is to extend these practices, largely enriched by human involvement, into virtual landscapes. Its potential role as a supplement rather than a substitute for interpersonal office experiences of coming out of the closet is particularly beneficial for individuals with limited diversity training and access to resources related to LGBTQIA+ identity. This aligns with our philosophy of "human augmentation," which underscores the significantly influential role technology can play in elevating individuals' potential for instigating both societal and personal advancements. Importantly, the AI system developed in this paper offers a high degree of accessibility, as it can conveniently be accessed online through a web browser. This means that individuals can engage in LGBTQIA+ advocacy simulations from the comfort of their own homes, without the need for specialized software or equipment. This accessibility opens up opportunities for a wider range of individuals to participate in advocacy practices and engage in meaningful conversations about LGBTQIA+ issues. By leveraging the power of technology and making it readily accessible, we aspire to empower individuals to become effective advocates for LGBTQIA+ rights and create lasting change within their communities.

Secondly, accuracy and fairness in the AI models are significant considerations. Inaccurate responses or unfair biases in the virtual agents might lead to harmful effects or misinformation to the users. Therefore, the training data must be carefully curated, and the AI models must be thoroughly tested to ensure that they provide reliable responses. To this end, further research and development are required to minimize AI bias. Finally, there is a significant ethical responsibility in the frame of designing the 'coming out of the closet' scenarios, ensuring that it is respectful, does not perpetuate harmful stereotypes, and are sensitive to the different cultural contexts in which users may come from. Despite these potential ethical challenges, the LGBTQIA+ Advocacy Simulator offers a promising use of AI technology to empower individuals in their growth as advocates for others. However, a thoughtful

and ethical approach will be crucial in ensuring its appropriate implementation and use.

Finally, given the widespread debate about the emerging ubiquity of generative AI, it is important to acknowledge the criticisms and limitations of AI simulations in this context. Some argue that AI simulations may oversimplify the complexities and nuances of real-life experiences, potentially diminishing the true struggles and challenges faced by the LGBTQIA+ community. Additionally, there are concerns about the potential for AI simulations to perpetuate stereotypes or reinforce biases, if not developed and implemented with careful consideration. It is crucial to approach the application of AI in LGBTQIA+ advocacy with ethical guidelines, critical evaluation, and ongoing feedback from diverse voices within the community to ensure the effectiveness, integrity, and inclusivity.

8 CONCLUSION

In this paper, we explored the use of AI-generated virtual characters to help individuals practice LGBTQIA+ advocacy. As a member of the LGBTQIA+ community, the author of this paper has a personal understanding of the importance and complexity of advocating for LGBTQIA+ rights. This lived experience provides a unique perspective and informs the research conducted in this study. By drawing on personal insights and experiences, the author strives to ensure that the AI simulations developed in this paper are sensitive, inclusive, and representative of the diverse range of experiences within the LGBTQIA+ community.

In conclusion, this study has pioneered a novel perspective on human-AI interaction by introducing the concept of LGBTQIA+ Advocacy Simulator to create interfaces that enable greater appreciation, understanding, empathy and sympathy for LGBTQIA+ community members. Recognizing the potential of virtual characters as a rehearsal for advocacy in real life, we have blended technology and the self to address social issues that are often challenging to simulate. Our pre-registered, randomized experiment with 307 participants has shown that our design of the LGBTQIA+ Advocacy Simulator can significantly improve participants' feelings of sympathy and empathy when encountering members of the LGBTQIA+ community in the workplace. In particular, we demonstrate promising results that illuminate the potential differences of a first person/third person approach to advocacy with virtual agents. Moreover, our web-based system's accessibility adds value by enabling widespread usage, opening possibilities for future research and application in varied societal contexts. However, an important consideration for future work will be to further scrutinize the potential limitations and ethical implications of LGBTQIA+ Virtual Characters. As we continue to advance on the path of AI-enhanced interactions, it is important to remember that while AI simulations of real interactions are indeed a potent tool, it's the "human" that holds the key to positive change and individual empowerment. By discussing the ethical complexity and potential societal implications of our work, we hope to pave the way for future advancements in embodied, self-actuated, and openly LGBTQIA+ virtual characters.

REFERENCES

- [1] [n.d.]. Apple CEO Tim Cook comes out as gay - The New York Times. <https://www.nytimes.com/video/multimedia/10000003207243/apple-ceo-tim-cook-comes-out-as-gay.html>. (Accessed on 10/10/2023).
- [2] [n.d.]. Diversity in Tech | Attracting LGBT into Tech - Diversity in Tech. <https://www.diversityintech.co.uk/attracting-lgbt-into-tech>. (Accessed on 10/10/2023).
- [3] [n.d.]. National Coming Out Day and 'Come Out of the Closet' Origin | Time. <https://time.com/4975404/national-coming-out-day-closet-metaphor-history/>. (Accessed on 10/10/2023).
- [4] [n.d.]. Reducing Racial Bias by Interacting with Virtual Agents: An Intervention in Virtual Reality. <https://www.southampton.ac.uk/~eg/AAMAS2023/pdfs/p747.pdf>. (Accessed on 10/10/2023).
- [5] [n.d.]. The State of the LGBTQ Community in 2020 - Center for American Progress. <https://www.americanprogress.org/article/state-lgbtq-community-2020/>. (Accessed on 10/10/2023).
- [6] [n.d.]. The Cost of the Closet and the Rewards of Inclusion – hrc.org. <https://www.hrc.org/resources/the-cost-of-the-closet-and-the-rewards-of-inclusion>. [Accessed 10-10-2023].
- [7] [n.d.]. Who Is OpenAI's Sam Altman? Meet the Oppenheimer of Our Age. <https://nymag.com/intelligencer/article/sam-altman-artificial-intelligence-openai-profile.html>. (Accessed on 10/10/2023).
- [8] Salema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.
- [9] Prithviraj Ammanabrolu and Mark O. Riedl. 2021. Situated Language Learning via Interactive Narratives. <http://arxiv.org/abs/2103.09977>
- [10] Zahra Ashktorab, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. 2021. Effects of Communication Directionality and AI Agent Differences in Human-AI Interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 238, 15 pages. <https://doi.org/10.1145/3411764.3445256>
- [11] Yulong Bian, Chenglei Yang, Dongdong Guan, Sa Xiao, Fengqiang Gao, Chia Shen, and Xiangxu Meng. 2016. Effects of Pedagogical Agent's Personality and Emotional Feedback Strategy on Chinese Students' Learning Experiences and Performance: A Study Based on Virtual Tai Chi Training Studio. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 433–444. <https://doi.org/10.1145/2858036.2858351>
- [12] Diana Bilimoria and Abigail J. Stewart. 2009. "Don't Ask, Don't Tell": The Academic Climate for Lesbian, Gay, Bisexual, and Transgender Faculty in Science and Engineering. *NWSA Journal* 21, 2 (2009), 85–103. <http://www.jstor.org/stable/20628175>
- [13] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [14] Sean Burns. 2020. Documentary 'Welcome To Chechnya' Takes Us Behind The Lines Of A Genocide. <https://www.wbur.org/artery/2020/06/30/welcome-to-chechnya-review/>, Last accessed on 2020-06-24.
- [15] Jessy Ceha and Edith Law. 2022. Expressive Auditory Gestures in a Voice-Based Pedagogical Agent. 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 163, 13 pages. <https://doi.org/10.1145/3491102.3517599>
- [16] Alan Cheng, Lei Yang, and Erik Andersen. 2017. Teaching language and culture with a virtual reality game. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 541–549.
- [17] Annette Choi. [n.d.]. Record number of anti-LGBTQ bills have been introduced this year | CNN Politics – cnn.com. <https://www.cnn.com/2023/04/06/politics/anti-lgbtq-plus-state-bill-rights-dg/index.html>. [Accessed 10-10-2023].
- [18] Matthew B. Cox. 2018. Working Closets: Mapping Queer Professional Discourses and Why Professional Communication Studies Need Queer Rhetorics. *Journal of Business and Technical Communication* 33, 1 (Sept. 2018), 1–25. <https://doi.org/10.1177/1050651918798691>
- [19] Laura Day. 2002. "Putting Yourself in Other People's Shoes": The use of Forum theatre to explore refugee and homeless issues in schools. *Journal of Moral Education* 31, 1 (2002), 21–34. <https://doi.org/10.1080/03057240120111418>
- [20] Michael Ann DeVito, Caitlin Lustig, Ellen Simpson, Kimberley Allison, Tee Chuan-romanee, Katta Spiel, Amy Ko, Jennifer Rode, Brianna Dym, Michael Muller, Morgan Klaus Scheurman, Ashley Marie Walker, Jed Brubaker, and Alex Ahmed. 2021. Queer in HCI: Strengthening the Community of LGBTQIA Researchers and Research. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM. <https://doi.org/10.1145/3411763.3450403>
- [21] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [22] Brianna Dym, Jed R. Brubaker, Casey Fiesler, and Bryan Semaan. 2019. "Coming Out Okay". *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–28. <https://doi.org/10.1145/3359256>

- [23] Umer Farooq, Jonathan Grudin, Ben Shneiderman, Pattie Maes, and Xiangshi Ren. 2017. Human computer integration versus powerful tools. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1277–1282.
- [24] Guo Freeman, Samaneh Zamanifard, Divine Maloney, and Alexandra Adkins. 2020. My body, my avatar: How people perceive their avatars in social virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.
- [25] Mary L. Gray. 2009. Negotiating Identities/Queering Desires: Coming Out Online and the Remediation of the Coming-Out Story. *Journal of Computer-Mediated Communication* 14, 4 (July 2009), 1162–1189. <https://doi.org/10.1111/j.1083-6101.2009.01485.x>
- [26] Yudong Guo, Keyu Chen, Sen Liang, Yong-Jin Liu, Hujun Bao, and Juyong Zhang. 2021. Ad-nerf: Audio driven neural radiance fields for talking head synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5784–5794.
- [27] Fa-Ting Hong, Longhao Zhang, Li Shen, and Dan Xu. 2022. Depth-Aware Generative Adversarial Network for Talking Head Video Generation. *arXiv preprint arXiv:2203.06605* (2022).
- [28] Rahimu Kasibante, Elisa Rubegni, and Letizia Jaccheri. 2022. Designing a Digital StoryTelling platform to detect gender stereotypes. In *Interaction Design and Children*. ACM. <https://doi.org/10.1145/3501712.3535292>
- [29] Hideki Koike, Jun Rekimoto, Junichi Ushiba, Shinichi Furuya, and Asa Ito. 2021. Human augmentation for skill acquisition and skill transfer. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–3.
- [30] Vivian Lai, Samuel Carton, Rajat Bhatnagar, Q Vera Liao, Yunfeng Zhang, and Chenhao Tan. 2022. Human-ai collaboration via conditional delegation: A case study of content moderation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [31] Mina Lee, Percy Liang, and Qian Yang. 2022. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–19.
- [32] Yi-Chieh Lee, Yichao Cui, Jack Jamieson, Wayne Fu, and Naomi Yamashita. 2023. Exploring Effects of Chatbot-based Social Contact on Reducing Mental Illness Stigma. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM. <https://doi.org/10.1145/3544548.3581384>
- [33] Ann Light. 2011. HCI as heterodoxy: Technologies of identity and the queering of interaction with computers. *Interacting with Computers* 23, 5 (Sept. 2011), 430–438. <https://doi.org/10.1016/j.intcom.2011.02.002>
- [34] Xian Liu, Yinghao Xu, Qianyi Wu, Hang Zhou, Wayne Wu, and Bolei Zhou. 2022. Semantic-Aware Implicit Neural Audio-Driven Video Portrait Generation. *arXiv preprint arXiv:2201.07786* (2022).
- [35] Manuel London. 2010. Understanding social advocacy: An integrative model of motivation, strategy, and persistence in support of corporate social responsibility and social entrepreneurship. *Journal of management development* 29, 3 (2010), 224–245.
- [36] Jolie M. Martin and Cleotilde Gonzalez. 2010. The cultural determinants of strategic bias. In *Proceedings of the 3rd international conference on Intercultural collaboration*. ACM. <https://doi.org/10.1145/1841853.1841876>
- [37] Lori Messenger. 2011. A Qualitative Analysis of Faculty Advocacy on LGBT Issues on Campus. *Journal of Homosexuality* 58, 9 (Oct. 2011), 1281–1305. <https://doi.org/10.1080/00918369.2011.605740>
- [38] Maryam Mustafa, Stefan Guthe, Jan-Philipp Tauscher, Michael Goesele, and Marcus Magnor. 2017. How human am I? EEG-based evaluation of virtual characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5098–5108.
- [39] Procheta Nag and Özge Nilay Yalçın. 2020. Gender Stereotypes in Virtual Agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. ACM. <https://doi.org/10.1145/3383652.3423876>
- [40] Abby Ohlheiser. 2020. The lonely reality of Zoom funerals. <https://www.technologyreview.com/2020/04/13/999348/covid-19-grief-zoom-funerals/>.
- [41] Pat Pataranutaporn, Valdemar Danry, Lancelot Blanchard, Lavanay Thakral, Naoki Ohsugi, Pattie Maes, and Misha Sra. 2023. Living Memories: AI-Generated Characters as Digital Mementos. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 889–901.
- [42] Pat Pataranutaporn, Valdemar Danry, Joanne Leong, Parinya Punpongsonan, Dan Novy, Pattie Maes, and Misha Sra. 2021. AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence* 3, 12 (2021), 1013–1022.
- [43] Mark O. Riedl and Vadim Bulitko. 2012. Interactive Narrative: An Intelligent Systems Approach. *AI Magazine* 33, 3 (2012), 67–77.
- [44] Susanne Schmidt, Sven Zimmermann, Celeste Mason, and Frank Steinicke. 2022. Simulating Human Imprecision in Temporal Statements of Intelligent Virtual Agents. 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 422, 15 pages. <https://doi.org/10.1145/3491102.3517625>
- [45] Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe. 2020. First order motion model for image animation. *arXiv preprint arXiv:2003.00196* (2020).
- [46] Jason Steinhauer. 2022. Nostalgia on Demand. In *History, Disrupted*. Springer, 33–41.
- [47] Dean Takahashi. 2020. Deepdub uses AI to localize voice dubbing for foreign language films. <https://venturebeat.com/2020/12/16/deepdub-uses-ai-to-localize-dubbing-for-foreign-language-films/>. Accessed: 2021-04-15.
- [48] The Dali Museum. 2020. Dali lives (via artificial intelligence). <https://thedali.org/exhibit/dali-lives/>, Last accessed on 2020-06-24.
- [49] Loyal | thisisloyal.com. [n. d.]. LGBT People's Experiences of Workplace Discrimination and Harassment – williamsinstitute.law.ucla.edu. <https://williamsinstitute.law.ucla.edu/publications/lgbt-workplace-discrimination/>. [Accessed 10-10-2023].
- [50] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, et al. 2022. Llama: Language models for dialog applications. *arXiv preprint arXiv:2201.08239* (2022).
- [51] Takane Ueno, Yuto Sawa, Yeongdae Kim, Jacqueline Urakami, Hiroki Oura, and Katie Seaborn. 2022. Trust in human-AI interaction: Scoping out models, measures, and methods. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.
- [52] Martijn van Zomeren, Tom Postmes, and Russell Spears. 2008. Toward an integrative social identity model of collective action: A quantitative research synthesis of three socio-psychological perspectives. *Psychological Bulletin* 134, 4 (July 2008), 504–535. <https://doi.org/10.1037/0033-2909.134.4.504>
- [53] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From human-human collaboration to Human-AI collaboration: Designing AI systems that can work together with people. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–6.
- [54] Isaac Wang, Jesse Smith, and Jaime Ruiz. 2019. Exploring Virtual Agents for Augmented Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300511>
- [55] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.
- [56] Shunyu Yao, RuiZhe Zhong, Yichao Yan, Guangtao Zhai, and Xiaokang Yang. 2022. DFA-NeRF: Personalized Talking Head Generation via Disentangled Face Attributes Neural Rendering. *arXiv preprint arXiv:2201.00791* (2022).
- [57] Fei Yin, Yong Zhang, Xiaodong Cun, Mingdeng Cao, Yanbo Fan, Xuan Wang, Qingyan Bai, Baoyuan Wu, Jue Wang, and Yujiu Yang. 2022. StyleHEAT: One-Shot High-Resolution Editable Talking Face Generation via Pretrained StyleGAN. *arXiv preprint arXiv:2203.04036* (2022).
- [58] Jeremy B. Yoder and Allison Mattheis. 2015. Queer in STEM: Workplace Experiences Reported in a National Survey of LGBTQA Individuals in Science, Technology, Engineering, and Mathematics Careers. *Journal of Homosexuality* 63, 1 (Aug. 2015), 1–27. <https://doi.org/10.1080/00918369.2015.1078632>
- [59] Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, and Victor Lempitsky. 2019. Few-shot adversarial learning of realistic neural talking head models. In *Proceedings of the IEEE/CVF international conference on computer vision*. 9459–9468.

Received 9 October 2023; accepted 14 December 2023