More Kawaii than a Real-Person Live Streamer: Understanding How the Otaku Community Engages with and Perceives Virtual YouTubers

Zhicong Lu
City University of Hong Kong
Kowloon, Hong Kong
zhicong.lu@cityu.edu.hk

Chenxinran Shen
University of Toronto
Toronto, ON, Canada
elise.shen@mail.utoronto.ca

Jiannan Li
University of Toronto
Toronto, ON, Canada
jiannanli@dgp.toronto.edu

Hong Shen
Carnegie Mellon University
Pittsburgh, PA, USA
hongs@cs.cmu.edu

Daniel Wigdor
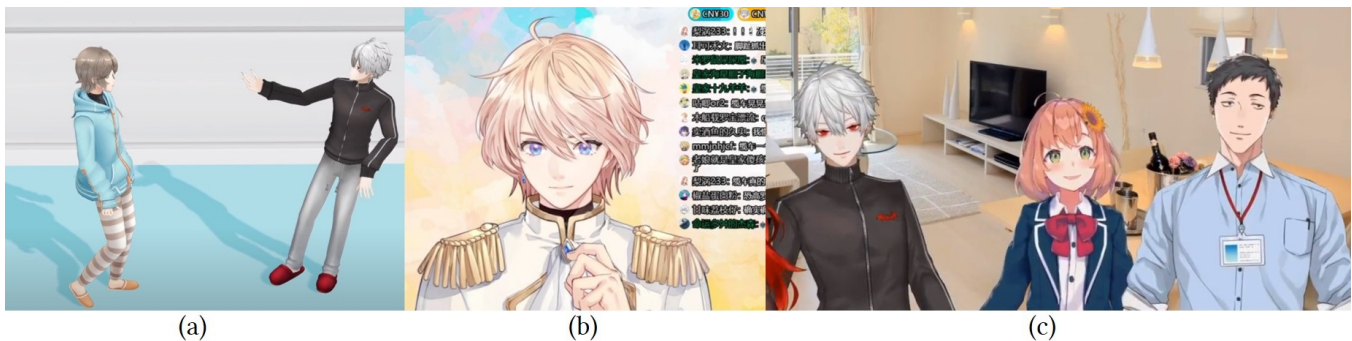University of Toronto
Toronto, ON, Canada
daniel@dgp.toronto.edu

Figure 1: Screenshots of VTuber live streaming: (a) Kanae co-streaming with VTuber Kuzuha using their 3D avatars; (b) VTuber 罗伊Roi streaming by themselves; (c) Multiple VTubers co-streaming together using their 2D avatars.

ABSTRACT

Live streaming has become increasingly popular, with most streamers presenting their real-life appearance. However, Virtual YouTubers (VTubers), virtual 2D or 3D avatars that are voiced by humans, are emerging as live streamers and attracting a growing viewership in East Asia. Although prior research has found that many viewers seek real-life interpersonal interactions with real-person streamers, it is currently unknown what makes VTuber live streams engaging or how they are perceived differently than real-person streamers. We conducted an interview study to understand how viewers engage with VTubers and perceive the identities of the voice actors behind the avatars (i.e., Nakanohito). The data revealed that Virtual avatars bring unique performative opportunities which result in different viewer expectations and interpretations of VTuber behavior. Viewers intentionally upheld the disembodiment of VTuber

avatars from their voice actors. We uncover the nuances in viewer perceptions and attitudes and further discuss the implications of VTuber practices to the understanding of live streaming in general.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *Empirical studies in HCI*.

KEYWORDS

virtual idols, live streaming, social media, user engagement, virtual YouTuber

1 INTRODUCTION

Sharing live video of real life experiences [39, 41, 58], live events [22, 57], creative activities [14, 40, 43], educational content [13, 42, 43], or playing video games [23] is becoming increasingly widespread

around the world, largely due to the emergence of live streaming on Twitch, YouTube, TikTok, etc. The live chat and virtual gifting features of live streaming enable viewers to interact with streamers and other viewers via text, emoji, or other modalities in real-time. This high degree of interactivity facilitates the building of the streamer's community and results in live streaming being an engaging medium for viewers to have fun and acquire information on demand.

Prior research on live streaming has mostly focused on *real-person* live streams, finding that the "realness" afforded by live streaming is the key factor that makes them engaging and popular [22, 23, 43]. However, the emerging trend of *virtual YouTubers* (VTubers) in East Asia seems to contradict this. Virtual YouTubers or virtual streamers are a unique category of content creators because they are virtual 2D or 3D avatars that are voiced by a human (Figure 1). Such streamers are becoming celebrities on YouTube, NicoNico [12], and Bilibili [4], with dedicated fan bases and corporate sponsorship deals. VTubers are becoming so popular that many companies in Japan and China are investing a large amount of resources in "virtual talent" and have led to the establishment of talent agencies to manage these avatars. Such digital avatar public figures are also not a phenomenon only constrained to East Asia, since Instagram has seen emerging *virtual influencers* such as Lil Miquela, who has over 2.7 million followers [6].

Although VTubers have been gaining traction in East Asia since 2016, there is a limited understanding of why and how viewers become engaged while watching VTubers through live streaming, a medium which emphasizes the *authentic* nature of live video and *real-time, interpersonal* interactions. It is also unknown how viewers perceive the differences between virtual-avatar streamers and real-person streamers, and how they perceive the identity and presence of the voice actor or *Nakanohito* who is behind the avatar. Uncovering specific motivations and viewing practices of VTuber live streaming could expand mainstream live streaming research in HCI, which has mostly focused on real-person streamers, to studying virtual-avatar streamers and the emerging fusion of live streaming and the subcultures of anime and comics, i.e., the Otaku community. Such knowledge could also inform the design of future live streaming interfaces that support both real-person streamers and virtual-avatar streamers.

To better understand the socio-technological phenomenon of VTubers and its genre- and community-specific practices, opportunities, and challenges, we conducted an interview-based study with 21 VTuber viewers from the Otaku community who had extensive experience watching VTuber live streams (i.e., they had watched VTuber for over a year and watched at least once per week). The interviews revealed that through the mediation of virtual avatars, streamers or Nakanohitos can engage viewers, build communities, and gain viewers' support without disclosing their real life appearance or identity. Virtual avatars bring unique performative opportunities which result in different viewer expectations and interpretations of VTuber behavior. For example, viewers felt more distance between themselves and the virtual avatars and showed tolerance to VTubers' offensive language and 'stupid' behavior compared to real-person streamers. The analysis also revealed that although viewers intentionally maintained beliefs about Nakanohito's disembodiments from their avatars so that viewers could maintain

perfect images of VTubers in their minds, they still cared about Nakonohitos when labor disputes arose. Drawing on our findings, we discuss the blurry and ever-changing line between the real and virtual in live streaming, labor relationships between VTuber agencies and Nakanohitos, and the need for future streaming platforms to support identity management, transparency, and fairness.

Thus, our contributions to HCI are: i) a nuanced description of the practices and motivations of VTubers' communities and how they engage in these practices within the social and cultural contexts of China, ii) an understanding of viewers' perceptions of VTubers versus real-person streamers and the beliefs viewers hold about the voice actors behind virtual avatars, iii) the implications of VTubers practices to live streaming and avatar-based social interactions.

2 BACKGROUND AND RELATED WORK

We first describe the background of VTubers, and then review research into avatar-based virtual idols and social interactions in virtual worlds, followed by a review of live streaming motivations and practices.

2.1 Background

Over the past 30 years, Japanese manga, anime, and video games have gained widespread popularity in East Asia and more generally, worldwide. This has given rise to a billion dollar industry and a unique Otaku subculture [26] that is characterized by a strong enthusiasm for manga, anime, and related subcultures. Concurrently, Japanese idols have also cultivated large and passionate fan bases, due in large part to their manufactured images and personalities [19]. VTubers have emerged as an innovative cultural phenomenon as these trends have begun to intertwine.

VTubers originated in Japan and have rapidly gained international popularity since their first appearance in 2016. By mid-January 2019, there were more than 10,000 active VTubers around the globe [1], and more than 600 VTubers had over 10,000 followers and had attracted millions of views [2]. The influence of VTubers has quickly grown beyond the Otaku subculture. The first VTuber, Kizuna Ai, for example, was chosen as the ambassador of an international culture campaign by the Japanese National Tourism Organization in 2018 [67].

A VTuber is an animated virtual avatar that delivers performances in live video streams or recorded videos. The avatar is often voiced by an actor, who is referred to as a *Nakanohito* (中の人) in Japanese. VTubers typically stream with half-body 2D avatars, which are created with tools such as Live2D [44] that capture the actor's facial movements and drive the avatar's facial expressions (Figure 1). The motions of other body parts of the avatar can be triggered within such programs using commands sent from desktop computers. VTubers who have access to full-body motion capture systems can perform with 3D avatars. This allows them to have a higher range of motion with the avatar's body. Vtubers also often read comments that viewers post in-stream and respond to them, similar to the practices of real-person streamers. VTubers often have a unique persona, e.g., Kizuna Ai has the appearance of a 16-year-old brunette animated girl, who was a recently-developed advanced artificial intelligence and her words and actions were

naive, imitating the behavior of an AI agent, although her performance was actually driven by several Nakanohitos.

VTubers can be operated by an individual or an agency. Some for-profit VTuber agencies, such as Nijisanji [48] and Hololive [25], have more than 20 VTubers under contract in parallel. So far, corporate-operated VTubers have attracted larger viewership than individual-operated ones. In January 2020, for example, 94 out of the 100 most followed VTubers on YouTube were run by agencies [2]. This is likely because corporate-run VTubers are overall better supported because they have professionally designed and engineered avatars, use more expensive motion capture equipment to generate more realistic avatar movement, and have stronger promotion and marketing.

The Nakanohitos (i.e., voice actors) of individual-operated VTubers typically have full ownership over their VTuber projects. In contrast, VTuber agencies own their VTubers and claim that they are their intellectual property. Employees of these agencies have limited control over changes to the overall operation of their avatar, such as future program planning or if another Nakanohito will take over their avatar, which can potentially lead to labor disputes.

China is one of the most important VTuber markets. A large number of VTubers operate official channels on BiliBili, a video platform that is highly popular among young generations [4]. On BiliBili, there are VTubers designed and managed by Chinese and Japanese individuals and agencies [68]. To overcome the language barriers between streamers and viewers, Japanese-speaking VTubers stream with real-time Chinese subtitles, or use simpler words to communicate with Chinese fans. Dedicated VTuber fans also curate stream highlights and form online social groups on social networking sites and forums to discuss VTuber-related topics, e.g., the National Geographic of Azeroth (NGA) forum [50]. By contributing a study of viewers of VTubers in China, we aim to capture both how viewers engage with VTubers and how VTubers are appropriated in China, where live streaming is a predominate activity for Internet users [43].

2.2 Avatar-Based Virtual Idols and Social Interactions in Virtual Worlds

The history of avatar-based virtual idols can be traced back to the mid-1990s. The first two avatar-based idols, Kyoko Date and Yuki Terai, were created in Japan in 1996-1997. In 2007, Hatsune Miku was released by Crypton Future Media. Building on Yamaha's software engine Vocaloid, and illustrated as a 16-year-old female character, Miku soon became an international phenomenon. Scholars have examined avatar-based idols such as Miku from a feminist perspective, arguing that the female representation and the essentializing of gender perpetuate female stereotypes, even in virtual worlds [33, 34]. Others have examined the fan communities around avatar-based idols, noting that although being virtual has limited idols' capacities to emulate real people, independence from a living body has cultivated a new type of relationship between fans and the idol [5] — amateur and professional musicians, animators, artists, and fans formed a distributed global network and recreated and remixed Miku for cultural expression through large-scale collaborations and peer-production, which differed from the corporate

control of Miku's franchise [35, 66]. Despite the international success of recent avatar-based idols like Miku, these virtual idols have arguably never achieved real-time interactivity with their fans. Our work contributes to this line of research by focusing specifically on VTuber live streaming, where VTubers interact with viewers in real time, harnessing both the spontaneity and unpredictability of live performances.

Also relevant is avatar-based social interactions in virtual worlds (VWs), which has been a key theme in HCI [15]. Past research has offered insights into user perception in game-oriented VWs such as World of Warcraft (e.g., [11, 45]) and social-oriented VWs such as Second Life (e.g., [46, 47]). Users usually enter these VWs through virtual bodies (i.e., avatars) that are created and controlled by them, and interact with each other using their avatar in a variety of social activities. Through online interactions in VWs, users construct their identities that might be their ideal selves (i.e., *Ego Ideals*), something different than they truly are (i.e., *Role Players*), or even multiple identities [8, 21, 47]. The constructed identity and the avatar appearance both influence social interactions in VWs. For example, prior work has found that human-like and more realistic avatars with faces tend to enable more positive social interactions [51, 65], and people treat other virtual avatars more intimately if they have commonality [3]. Agents with real/physical body (i.e. *embodiment*) in virtual reality also increase rapport and nonverbal realism [30]. In some VWs, users can deliver live performances [53], such as musical concerts, and establish fan bases [10, 28]. Though relevant, avatar-based interactions in VWs often do not reach the same level of expressiveness and interactivity of VTuber live streaming, and perception of VTuber differs from virtual character perception in VW because viewers are not in the same VW as the streamer. We are inspired by the discussion of identity and emobodiment in this line of research to explore VTuber live streaming.

2.3 Live Streaming Motivations and Practices

Prior work about live streaming in HCI has found that the interactivity and the sociality of live streams are what make live streaming engaging [23, 43]. Due to the ubiquitous nature of live streaming platforms, live streaming has been leveraged by many users as a way to have fun, to learn, to get inspirations, to improve their skills, and to build online communities and form social relationships with community members [59].

The domain of video games has been a popular topic within live streaming. Streamers on Twitch often broadcast their own live gameplay experience or spectate on others' gameplay. They also share gameplay skills and inform viewers of interesting video games thus using live streaming to cultivate a community of gamers who share similar interests [23, 29, 52, 55, 63]. Novel user interfaces that extend communication channels between viewers and streamers during video game live streams have also been designed and developed [20, 36]. Although these projects found increased viewer engagement in live streams, their findings were largely confined to video game live streams.

Live streamers in other domains have also leveraged live streaming to connect and engage community members, such as for performance (e.g., singing), education [7, 24], knowledge sharing [42], programming [13], cultural practices [40], and creative activities

[43]. These aforementioned studies have found different motivations, demographics, practices, and engagement levels within different live streaming communities, highlighting the need to further understand live streaming user groups to improve the design of live streaming interfaces to support a variety of user needs.

The affordances of live streaming (e.g., live chats and emojis) is relevant to the Hyperpersonal Model in computer-mediated communication (CMC) [54, 60, 61]. VTubers can be viewed as message senders who selectively present themselves with virtual avatars, and viewers can be viewed as receivers who interpret the content and form impressions of VTubers. The Hyperpersonal Model suggests that senders often have a number of advantages in their interactions with receivers, compared to traditional face-to-face communications (e.g., the ability to modify self-presentation to develop an optimized image and relative anonymity). As a result, receivers tend to over interpret the cues inherent in senders' self-presentations because the cues in text-based CMC are limited compared to face-to-face interactions. VTuber viewers cannot see the appearance and non-verbal behaviors of Nakanohitos, resulting in reduced cues.

The mainstream live streaming research mostly focused on live streams of *real-person* streamers who live stream with their real-life appearance. This work contributes to the broader live streaming research in HCI by understanding how viewers of the Otaku community engage with and perceive *virtual characters* of VTubers.

3 METHOD

To understand the motivational and situational contexts of viewers from the Otaku community who watch VTuber live streams, and how they engage with and perceive VTubers, we conducted a qualitative interview-based study with 21 viewers who had been watching live streams of VTubers at least once a week for at least a year (Table 1). The study protocol was informed by online archival data and approved by the institutional review board (IRB).

3.1 Interviewee Recruitment

Because VTubers mostly attract viewers from the Otaku community [68], potential "dedicated" viewers were purposefully approached through either personal connections or VTuber discussion board on the National Geographic of Azeroth (NGA) forum [50], which is the largest VTuber discussion board in China, in terms of both the number of active users and threads [56]. Four interviewees were from a WeChat group of anime, comics, and gaming (ACG) fans that one of the authors joined, and were invited to participate through direct messaging on WeChat. The other 17 interviewees were recruited from the NGA forum. A recruitment advertisement was posted in Chinese on the forum soliciting those who wanted to participate to contact the research team through WeChat or e-mail. All those who self-identified as viewers and who watched at least one VTuber's live stream once a week with over one year of watching experience were invited to participate. Since this study valued the individual experiences of viewers and there is no previous work exploring this specific type of streaming, purposive sampling was helpful to explore factors that could influence viewers' perceptions. Due to this recruitment method, however, the recruited interviewees skewed male and were mostly students. This aligns with anecdotal evidence that the research team had about VTuber viewer demographics.

3.2 Background Material Collection

To broaden the perspectives on the VTuber phenomenon and scaffold the interview protocol design, we collected extensive background material from online resources before and during the interview process. The primary source of information was the VTuber discussion board of the NGA forum [50]. The users of the board vary from casual viewers to highly dedicated fans who offer significant time or monetary support to their beloved VTubers. Such user diversity on this board allowed us to sample a broad range of opinions. Furthermore, the board archives some in-depth discussions and heated debates about VTubers. These threads helped the research team stay up-to-date with the evolving VTuber industry and identify relevant themes for the interviews.

One author followed the posts in this discussion board daily between October 2019 and December 2019, following the top active threads reflecting viewers' perceptions of VTubers, their discussions about VTuber's live streams, Nakanotitos and VTuber agencies, and important changes in the VTuber industry. We used keywords including popular VTubers' names and agencies, and notions like VTuber and Nakanohito to search for threads to follow. Altogether 1059 posts from across 52 forum threads were sampled. Strauss' open coding method was used [9] to analyze the posts in the threads and identified preliminary themes, which were incorporated into the interview protocol.

3.3 Interview Protocol

Semi-structured interviews were conducted with 21 dedicated viewers of VTubers. The interviews were conducted remotely using video or audio calls in November and December of 2019 and March and April of 2020. Each interview lasted approximately 60 minutes and interviewees were provided with a 50 CNY honorarium for their time. The interviews included questions about viewer's motivations to watch VTubers' content, what types of content they watched or enjoyed watching, what they liked and disliked about VTubers' live streams, how they interacted with VTubers and other viewers, and so on. Interviewees were also primed to think about differences between VTubers and real-person streamers and their perceptions towards the roles that Nakanohitos play within virtual live streaming. Because there were many cases where corporate-operated VTubers were mistreated, e.g., Nakanohitos being replaced without a fair reason, we asked interviewees to reflect on these issues during the interview.

To build rapport with interviewees, we contacted them 2 weeks prior to the interviews and asked them to share interesting videos of their favorite VTubers. All interviewees sent at least one video prior to the interviews. Two authors reviewed the videos and noted main themes. During the interviews, we then asked questions tailored to the content they had shared, e.g., how VTubers performed and engaged viewers, what topics they chatted about, unexpected dramatic moments, etc. Interviews were conducted in Mandarin, audio-taped, and transcribed by the transcription service iflyrec.cn after removing all personally identifiable information.

Table 1: Summary of participants interviewed. Preferred platforms: YouTube (YT), Bilibili (BB), Niconico (NN)

ID	Age	Sex	Occupation	Education level	Years watching VTuber	Times watching VTuber /week	Preferred platforms
1	27	M	Software developer	Bachelor's	1	>7	YT, Twitch
2	20	M	Student	College student	1	3-7	YT, BB
3	21	M	Contract clerk	College student	2	3-7	YT, BB
4	18	F	Student	High school student	1.5	>7	YT, BB
5	16	F	Student	High school student	>2	>7	YT, BB
6	19	M	Student	College student	>2	3-7	YT, BB
7	20	M	Student	College student	1	3-7	YT, BB
8	22	M	Student	College student	1.5	3-7	YT, BB
9	18	M	Student	College student	2	1-3	BB
10	20	M	Student	College student	2	1-3	YT, BB
11	17	M	Student	High school student	1.5	3-7	YT, BB, NN
12	26	M	Office clerk	Bachelor's	2	1-3	YT, BB
13	20	M	Student	College student	1	3-7	YT, BB
14	20	M	Student	College student	2	3-7	BB
15	19	M	Student	College student	2	3-7	YT, BB
16	22	M	Student	College student	1.5	1-3	BB
17	21	M	Student	College student	2	3-7	BB
18	21	M	Student	College student	2	3-7	YT, BB
19	23	M	Office clerk	Bachelor's	1.5	1-3	YT, BB
20	22	M	Freelancer	Bachelor's	>2	3-7	BB
21	23	F	Self-employed	Bachelor's	>2	>7	BB

3.4 Data Analysis

The interview transcripts were analyzed using an open coding method [9]. Two native Mandarin-speaking authors coded the transcripts individually using an inductive approach and met to discuss disagreements to gain consensus about codes. All the codes were then translated into English and were discussed by the research team using affinity diagramming to find emerging themes. All codes were transcribed on sticky notes and then arranged in a random order. We then iteratively rearranged the notes into a hierarchy of themes and reached a consensus about viewer behaviors, engagement, and perceptions of VTuber live streaming. The collected background material was also used to supplement our knowledge about the subculture of VTubers' communities and the overall context for the codes.

Misogyny and sexism sentiment was present in many interviewee's responses. The misogyny might be rooted from both the Otaku subculture and the virtual idol industry. Sexism seems prevalent in the Otaku subculture, where cuteness is desired or sexually attractive for the community [26]. The virtual idol industry also actively exploits and commercializes sexism [5]. We fully acknowledge the existing gender biases in the existing VTuber communities, which might contribute to the behaviors and interactions observed in this study. The multi-gender research team has contextualized interviewee responses to ensure the misogynistic undertones are eliminated and reflected upon.

4 FINDINGS

4.1 Viewer Motivation

All interviewees described themselves as dedicated members of the Otaku subculture and were enthusiastic consumers of Japanese manga, anime, and video games. They found that VTubers, most of whom have anime-character-style appearances and tend to deliver content related to the Otaku subculture [26], aligned with their tastes and interests. They were attracted to VTuber live streaming because VTuber appearances and specially-designed persona appealed to them, and VTuber live streaming gave them opportunities to interact with their favorite anime characters, e.g., "*Mea is so kawaii that I will regret if I miss her show*" (P19) and "*VTuber streaming turns an unreal character into reality, so that you can chat and interact with characters you like, which is very cool*" (P15). They also attributed their dedicated viewing behaviors (several times per week) to their strong interests in following a long series of anime or manga, e.g., "*I watch her live streams almost every time she streams. It's just like following every episode of an anime*" (P3). Several interviewees also noted curiosity and creativity as a main motivation to watch VTuber, because "*vtubers' content could be more creative and beyond imagination than real-person streamers*" (P17). Other motivations included to seek relaxation and entertainment, to pass time, and to experience a sense of company and community, similar to prior findings [23, 43, 59].

4.2 Streaming Content and Interactions

Interviewees watched VTubers' live streams about similar content as in real-person live streams. However, compared to prior research, there were unique content provided by VTubers' and ways that viewers engaged with live streams.

4.2.1 Main Activities: Singing, Chatting, and Gaming. Interviewees reported that singing, chatting, and gaming were the three most common VTuber activities they watched. In comparison to similar performances of real-person streamers [43], VTubers' content

tended to align more with manga and anime subcultures, e.g., *"VTubers all have some specialties in anime, manga, and video game. There is more depth to it so is there more resonance."* (P7). There was also a strong emphasis on creative program planning to ensure that the streams are entertaining, e.g.,*"VTubers always impress us with unimaginable program design*" (P12). and *"it was a (board game) campaign for newbies as if they played together"* (P4).

Interviewees also noted several novel streaming opportunities that virtual avatars afford. For example, during a gaming co-streaming session, two Vtubers switched their avatars. Interviewees were excited by the mismatches between avatars and the Nakanohitos, e.g., "*Familiar souls, new bodies, what fun it is!*"(P17), and "*Their avatars do things that they have never done before*"(P20). P19 mentioned another VTuber who ran a show where most characters on the stage were represented by variants of the VTuber avatar that were all voiced by the same Nakanohito.

In general, interviewees did not have a strong preference on the 3D avatar streams over 2D ones, although appreciating 3D avatar's ability to leverage full-body motions to achieve a higher degree of expressiveness. However, interviewees also recognized that current 3D motion capture technologies were costly for individual VTubers with unstable financial source, so 2D avatars were *"good enough for chatting"* (P18).

4.2.2 Content about the Daily Lives of Nakanohitos. Interviewees enjoyed that Nakanohitos shared their personal daily lives and talked about real-world events during live streams. Most interviewees are dedicated anime watchers and are aware that their favorite anime characters only exist in fantasized anime worlds. The daily life content that VTubers share, however, brings them closer to viewers than typical anime characters, which VTubers ironically appear to be, e.g., "*Kagura Mea talks about how she (the Nakanohito) has been doing and comments on different events, like the recent typhoon in Japan … Her experiences feel real, (it is) something that has a sense of commonalities with the real world … She feels much closer than a typical anime character*" (P3). This feeling was echoed by P8, *"before, because of the pressure of monetization, (you can feel) the authenticity of Mea's Nakanohito, a 'corporate slave' who could not escape from pressure even after getting off her work"*. In this example of Kagura Mea, although she is a corporate-run VTuber, her sharing of daily lives made viewers perceive more 'authenticity' from her and feel closer to her. Our interviewees also valued having a sense of community and company, i.e., "*(it is) like someone sharing her life just like chat between friends, (which is) more relieving than interesting*"(P5).

4.2.3 Co-streaming. Co-streaming is also mentioned by several interviewees as an engaging and popular form of VTuber live streaming, similar to real-person co-streaming [37]. In some instances, the avatars of multiple VTubers perform on one virtual stage together (Figure 1c), whereas in other cases, VTubers stream with other real-person streamers. In the latter situation, VTubers and real-person streamers are typically shown side-by-side in a composite video stream, with occasional exceptions where VTubers appear on a display in the physical world. As mentioned by four interviewees, one of the most popular co-streaming pairs is VTuber Takatsuki Ritsu and real-person streamer Suzuki Yuyuuta. They sometimes co-stream live performances together, acting as if they were in romantic or platonic relationship. Interviewees enjoyed watching the pair support and interact with each other intimately.

Many viewers discovered new VTubers to watch because they had previously co-streamed with their favorite VTubers. Although the performances that VTubers co-stream are mostly composed of singing, chatting, and gaming, interviewees highlighted the interactions between streamers as what made VTuber streams engaging. P6 explained that *"there is a lot of new content that you do not see in single-streamer sessions, like the exchange of banter"*. Some interviewees enjoyed displays of friendship or inter-personal relationships between VTubers, as P7 described, "*I like seeing a group of streamers in close relationships with each other, like watching an ensemble cast TV show.* ".

4.2.4 Technological Limitations Affecting Experiences. Interviewees found that the technologies currently available for virtual streaming limit the diversity of content that is created and constrain streamers' abilities to engage with viewers.

They first desired for avatars to have richer and more natural facial expressions. The most popular tool for animating 2D VTubers, Live2D, plays pre-recorded animations of different parts of an avatar. Since there are only a few animatable sections of a 2D avatar's face (i.e., the eyebrows, eyes, and mouth), this limited set of available facial expressions made VTuber's verbal communications look rigid. P6 remarked, *"I hope that VTubers can have richer facial expressions, such that in conversations, they look more like real people talking naturally… facial expression changes as she talks"*.

Viewers also found that most performances of VTubers were confined to the space that a Nakanohito had in front of their desktop computer and involved little full-body movement or interaction with the physical world. Motion capture technology can drive the full-body animation of 3D avatars and expand the repertoire of VTubers' content to include dancing and motion sensing games, but it is costly and not free of limitations. As P8 commented, *"VTubers rely on face capture and therefore have to look at the camera all the time. Some large companies have 3D motion capture, but there are still many things that it cannot capture and only visible on real-person streams"*. Within virtual streaming, interaction with the physical world, such as picking up and interacting with real objects, remains a substantial technical challenge.

4.2.5 Viewer Interactions During and Beyond Streaming. VTuber viewer interactions share many common elements with those within real-person streams [43, 63] but also bear subtle differences. While viewers were willing to provide monetary support to VTubers, they paid more attention to the usage of cultural tokens, such as memes, to reinforce their subculture identities. All interviewees reported that they had supported VTubers by purchasing YouTube Super Chat and Bilibili channel badges through a hierarchical membership system. Their main motivation was to show support for streamers who provided enjoyable content, similar to real-person stream viewers [43], e.g., *"When I am really happy and excited while watching the (VTuber's) program, I will pay the minimum amount to encourage her"* (P7).

Our interviews revealed the important role of memes in VTuber communities. Members of these communities frequently created memes out of dramatic moments of VTubers' performances with screenshots of live streams. These memes often took graphical or

Figure 2: Memes of VTubers. Names of VTubers with meme text translated, if available, from left to right: Nekomiya Hinata (What are the bosses talking about?), Tsukino Mito, Sasaki Saku (Arrogant), Natsuiro Matsur, Kizuna Ai (Eh-yo-yo), Minato Aqua and Kagura Nana (How could crying help?).

textual forms, or a mixture of both (Figure 2). On the NGA VTuber discussion board, there were community-maintained threads dedicated to introducing popular VTuber memes to newcomers. Further, memes were widely referenced and recreated in the comments of live streams and fan group discussions. Viewers often engaged in a particular type of group meme, playfully termed "persecution", which involved the burst posting of mischievous memes about a streamer during their live stream. "Persecution", as a group action, adds to the overall entertaining atmosphere of a stream while reinforcing the collective identity of the viewers [16]. In VTuber communities, members sought to strengthen the group's identity by learning meme knowledge and rejecting meme usage that they deemed inappropriate. P13 reported that *"I started to watch more streams so that I can understand the most recent memes used in the fan group cha"*. For these dedicated viewers, knowing the correct meaning and usage of memes can be a form of *cultural capital* to differentiate them from outsiders [49].

4.3 Perceived Differences between VTuber and Real-Person Streamers

All interviewees had watched both real-person streamers and VTubers, especially those promoted by the platform, and they tended to watch VTubers more than real-person streamers. When asked explicitly about the differences they perceived between VTubers and real-person streamers, most interviewees first reported no significant differences between the two. However, upon reflection, most of them implicitly mentioned several key differences they perceived.

4.3.1 Different Expectations for Streamers' Behaviors. Interviewees noted that because a VTuber's avatar is in a 2-D or 3-D "virtual" form, they unconsciously adopted a mindset that the VTuber was "in the virtual world". They felt that when watching VTubers, they had different expected norms for the virtual world, similar to their expectations for the virtual worlds in games, anime, or comics. "*If a human streamer says that she is from a different space, I would feel it is fake and strange. However, if a VTuber says this, I won't think it strange. I think she is in the virtual world, so all she says is about the virtual world, where almost nothing is impossible.* "(P5) Because viewers have different expectations for what could happen in the virtual world, many behaviors that might be embarrassing, unacceptable, or inappropriate in real life could be normal, acceptable, and appropriate in VTubers' contexts. For example, viewers seem more tolerant of VTubers' dirty or offensive language or "stupid" behaviors, e.g., *"I think sometimes she behaves in a stupid way, and*

her facial expression is dull. However, I know she is a virtual character, so such behaviors are acceptable. They make her cute. If a human streamer behaves like that, I will feel it embarrassing and too strange." (P2)

Several interviewees commented on streamers' inclinations to solicit virtual gifts from viewers, a unique reward mechanism prevalent within live streaming in China. They disliked some real-person streamers' solicitation behaviors for virtual gifts and were jealous of, and had judgements about, streamers who earned much money, e.g., *"I don't like to watch real-person streamers, such as beautiful girls singing. I think they are snobbish. They emphasize the leaderboard of top gift-senders too much. They only ask for money from viewers without putting effort into their streams."*(P3).

However, VTubers' solicitation behaviors seemed more acceptable to interviewees. Several interviewees mentioned Kagura Mea, a VTuber who has a persona of being enthusiastic to make money and often asks for gifting during her streams. Her acts of solicitation were seen as acceptable, or even favored, by her fans, e.g., *"I think her solicitation is mostly that she acts like her persona. She is not really snobbish or money-mad. When she asks for gifts, it is cute, and I like it"* (P8). Interviewees also seemed more willing to purchase virtual gifts for their favorite VTubers, e.g., *"Purchasing virtual gifts for VTubers seems like purchasing virtual goods in games. It feels less like tipping to a real person"* (P3).

4.3.2 Sense of Distance. Interviewees reported that they felt more distant from VTubers than real-person streamers when watching live streams. This sense of distance partly originated from their virtual mediation with the avatar, wherein viewers could not directly see the real facial expressions or other non-verbal cues of the Nakanohito controlling the avatar, as noted by P3, "*Watching VTubers' live streams is like watching the performance of an actor wearing a mask. Although I know there is a human behind the avatar, it feels different and she becomes more distant*".

The sense of distance is further enhanced by the persona of the VTuber. Many VTubers are designed with a specific identity, e.g., exotic, vampire, devil, princess, fairy, alien, or futuristic, and the visual appearances of these avatars aligned with these identities. Due to these unique identities and appearances, VTubers were perceived less like a human, and thus, more distant from viewers, as described by P13, "*When I am watching VTubers and commenting in real time, and my comments get read and responded by VTubers, I think it is different from getting responses from a real-person streamer. It is like interacting with something virtual, from a different world, even if there's a human behind the character.*"

This sense of distance causes viewers to pay more attention to the actual content of live streams than the appearance or other characteristics of the Nakanohito, e.g., "*The avatar creates an atmosphere where I pay more attention to the content and performance, and isolate the performance, content, and stories from the real person behind the VTuber*" (P5).

4.3.3 Public Personas and Identities. Interviewees perceived that unlike real celebrities or streamers whose public personas are more implicit, the public personas of VTubers are more explicit and obvious. They reported that every time a new VTuber begins to stream, a self introduction video is produced and widely shared online to explicitly provide basic information about the persona of the VTuber.

The video often includes background stories of the virtual character, his/her persona, and characteristics. Such videos are an important on-boarding experience for viewers and drum up initial interests in a VTuber. However, interviewees believed that there were diverse practices with respect to the degree to which streamers' behaviors aligned with a designed persona and also that personas were often temporally dynamic.

The personas that VTubers have created tend to follow common-place anime or manga persona templates that seem simple and "flat", combining personality elements that are known to be attractive to subculture community members. As the Japanese philosopher Azuma Hiroki commented, anime personas are often constructed out of the 'moe (*affection*) database' [17]. VTubers usually have one core characteristic that stood out to ensure that their persona was less complicated than a real human, which often makes them unique and popular, e.g., "*Nekomiya Hinata has a persona of a cute girl with cat ears, who is skillful in playing FPS (first person shooter) games. She really impressed me in her famous stream where she killed 16 players in PUBG*" (P6).

Some VTubers strictly aligned with such original ideal personas, thus transforming a Nakanohito into a role player. Kizuna Ai is an example of such a role-playing VTuber and is presented as a cute, cheerful and occasionally careless girl that is commonly seen in anime and manga. Viewers can easily grasp the simple persona of this VTuber and the impression of cuteness that she exudes accumulates and is reinforced with every exposure a viewer has to her stream.

Other VTubers' personas can be more flexible, e.g., diverging from the originally designed persona, with some even disclosing the Nakanohito's characteristics or real life. The discrepancy between the designed persona and the presented image of the VTuber has become more and more prevalent for emerging VTubers. The contrast between the initial and evolving personas can make a VTuber's content more fun to watch, e.g., "*Many new VTubers start from an interesting introduction video about the persona, which attracts many viewers' attention. However, after they officially started streaming, everyone is astonished, because they do not act like the persona at all*" (P5).

The personality elements that compose the designed personas, while well-accepted, may feel familiar or even repetitive to some long-time VTuber viewers. Interviewees noted that some VTubers intentionally have a conflicting personality to generate "*drama and conflict*" (P11) to attract viewers. Others noted that this discrepancy is due to the differences between Nakanohito's personalities and designed personas. Through streaming regularly, Nakanohitos' personalities and styles are often revealed unconsciously, gradually transforming into the VTuber's persona, e.g., "*I can see some of the Nakanohito's real reactions in some cases, and it tells me about the style or personality of her*" (P15).

4.3.4 Situational Changes in Persona. Public personas also undergo situational changes with viewers' preferences. While some VTubers gain initial popularity due to the discrepancies between their designed and presented personas, this becomes a challenge when they begin to gain traction. As they become more popular, a presented persona may not continue to attract followers because the discrepancies might only appeal to a niche group of viewers. In this case, many VTubers have to adjust their personas so that they cater to a broader audience and avoid losing followers. This happened with Kagura Mea, e.g., "*Mea was famous for her dirty language and absurd behaviors. However, now she's more popular than before, she intentionally reduces the use of dirty language to avoid attracting trolls*"(P7). Explicit persona changes also provide opportunities to generate interactions between viewers and VTubers, e.g., "*We often make fun of the VTuber by mentioning her persona when she behaves inconsistently with her persona. She will seriously explain her behaviors with excuses. It looks very absurd and beyond the reality for us, and hard to explain* "(P21).

Due to the performative nature of VTubers, it is hard to determine whether deviations from a VTuber's designed persona result from intentional planning to attract viewership or from the Nakanohito's personality blending in. The highly commercial and viewership-driven atmosphere among VTubers also results in their identities being more fluid between the ends of the spectrum and open to changes to maintain viewer appeal.

4.3.5 Presence through Multiple Channels. Interviewees also mentioned that VTubers extend their virtual lives and presence into the real world using many different channels. Interviewees provided several examples of VTubers' growth into anime, comics, garage kits (i.e., model figures portraying anime characters), video games, novels, and traditional radio broadcasting channels. For example, several popular VTubers have anime and comics that tell stories about their characters and their imaginary lives. Video games that enable fans to role play as a friend of their favorite VTuber and interact with them are also popular in Japan. VTubers also travel to offline events to meet their fans, i.e., they are shown on a public large display and interact with fans in real time.

Most interviewees were interested in such artefacts and perceived them as a way to stay connected with their favorite VTubers in real life, in the real world. They also regarded these artefacts (e.g., garage kit) as a way to extend VTubers "physically" in the real world, and they desired to purchase them to support their favorite VTubers, e.g., "*I think through buying a garage kit of her, I get a copy of her in the real world which can be with me. It is also my way to support her*" (P8).

4.4 Perceptions of Nakanohito Disembodiment and Nakanohito's Labor

A Nakanohito (中の人) is the person performing behind an avatar who provides the VTuber's voice and controls their movements. A proficient Nakanohito must have a pleasant voice and talent for generating interesting conversations with viewers during live streams. Although they play a pivotal role in VTubers' performances and interactions with viewers, a VTuber's Nakanohito remains invisible throughout most live streams. Most Nakanohitos do not disclose their personal identity to the public, due to privacy concerns or corporations' policies. Even for those who do disclose some of their personal information to the public, VTubers' identities are mostly determined by the persona of the virtual avatar that was initially set by the company or the streamer rather than the Nakanohito's identity. For example, Tsukudani Norio (佃煮のりお) is a female manga artist, who is the Nakanohito of a popular male VTuber

Inuyama Tamaki (犬山たまき). Although viewers know that a female is portraying a male character, many viewers are still attracted to the persona of the virtual avatar. They enjoy watching a unique show wherein Tsukudani Norio and Inuyama Tamaki stream simultaneously talking to each other, as if they were co-streaming (Figure 3).

The degree of disembodiment between Nakanohitos and their virtual avatars is unique in virtual live streaming, as it brings forth challenges related to identity management and user engagement.

4.4.1 Intentionally Maintaining Nakanohito and Avatar Disembodiment. Interviewees noted that the VTuber community agreed that personal information about Nakanohitos is sensitive and viewers should avoid talking about it online. For example, on the VTuber sub-forum of NGA, the first rule was that discussions about the identity of Nakanohitos and disclosing their personal information was prohibited (Figure 4). These community members actively maintain the disembodiment of VTubers by limiting the spread of Nakanohito's personal information. Most interviewees reported that although they were curious about Nakanohitos, they had little interest in knowing their identities and they denounced behaviors relating to prying into Nakanohitos' privacy, e.g., *"I tend not to know or ask about Nakanohitos' identities. Their privacy should be respected"* (P8). Some interviewees noted that rules about not discussing Nakanohito's personal information are required or suggested by corporations or other stakeholders, because *"their strategies are to make Nakanohito more invisible to avoid risks or trouble. Prohibiting discussions about Nakanohito could be good for their*

Figure 3: Tsukudani Norio talking with her own VTuber avatar in a live stream. Left: Tsukudani Norio talking in her real identity; Right: Tsukudani Norio talking as VTuber Inuyama Tamaki with her own face covered.
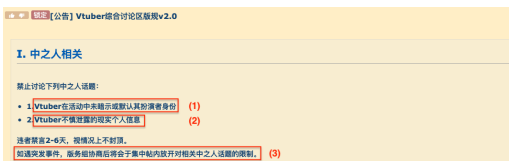

Figure 4: Rules of NGA's VTuber subforum. Translation of text: (1) Discussing a Nakanohitoe's identity is prohibited, unless it has been disclosed implicitly in activities; (2) Discussing accidentally leaked Nakanohito personal information is prohibited; (3) Under certain circumstances, discussions about certain Nakanohito-related topics may be allowed by the forum committee.

operation and management" (P13). Interestingly, most community members supported and followed these rules even if they knew that they were for corporate interests.

4.4.2 Maintaining the Disembodiment of Nakanohitos to Maintain Perfect Impressions. Some interviewees intentionally supported disembodiment for the sake of maintaining the perfect image of the virtual avatar in their mind, e.g., *"In some cases, if I know that the Nakanohito is totally different from her persona, for example, if her style is not as cute as the avatar. I may get very disappointed and feel uncomfortable when watching her videos or streams, even abandon her"* (P8). This misogyny sentiment may root from preferences for cuteness in the Otaku community [26], the psychological distance they perceived with the VTuber, and that they perceived the VTuber more as a character than as a real person.

Because Nakanohitos are behind the scenes and only disclose or alter their voice when streaming, viewers have limited information to infer about Nakanohitos' real personalities and identities. On the other hand, virtual avatars often have well-designed persona settings and background stories which often leave "perfect impressions" on viewers. As time goes on, viewers may begin to believe that the Nakanohito is also as perfect as the virtual avatar. For some viewers, seeing the real side of the Nakanohito would ruin the fantasy in their mind, e.g.,

> *"VTuber is perfect in the virtual world. We should keep a distance from the Nakanohito to appreciate her performance. Watching up close will ruin the fantasy. Someone tried to dig out information about Kagura Mea's Nakanohito. He found a photo of her that does not align with his imagination. He was so disappointed that he never watched her streams again. But actually the Nakanohito did not post her photos online."* (P10)

4.4.3 Indifference towards Corporations Replacing Nakanohitos. The disembodiment of Nakanohito and virtual avatars is also reflected by indifferent attitudes towards replacing Nakanohitos. All interviewees mentioned cases where Nakanohitos were replaced by agencies, e.g., Kizuna Ai's case. When asked to reflect on how they would feel if their favorite VTuber's Nakanohito were replaced, most interviewees did not think it would matter.

Some interviewees cared most about the quality of the content they experienced with a VTuber, which was often the outcome of teamwork, rather than a Nakanohito's own efforts. In these cases, the influence of Nakanohitos on their experiences was limited, e.g., *"Many shallow viewers will not notice that the Nakanohito is replaced if the quality of the streams and personas remain unchanged... If the team is professional, it is achievable even if Nakanohito is replaced"*(P4).

Some interviewees thought that replacing the Nakanohito with a more experienced one would provide better quality content. They cared more about the Nakanohito's skill set than the person, e.g., *"The Nakanohito is essential for a VTuber, but she is not the entire project. If the new one is interesting enough, I will keep watching her live streams"* (P6).

Some interviewees considered the possibility of replacing Nakanohito as a unique feature. For example, P11 noted that having

different Nakanohitos for one VTuber was similar to a person journeying through different life stages, which could bring new levels of freshness and engagement, *"I prefer to think in this way that VTuber can grow over time just like human ... Different Nakanohitos represent the VTuber in different life stages, but it is the same VTuber. I will accept this if the company transfers Nakanohitos smoothly".*

4.4.4 Viewers Care about Fairness for Nakanohitos' Labor. Although most interviewees could accept the replacement of a Nakanohito, they felt that they were owed a proper reason for this change in personnel. In particular, they cared about Nakanohitos' working conditions and fair treatment.

In 2019 April, four Nakanohitoes in the Game Club project accused the Unlimited operating company of unfair treatment, including overtime work and sexual harassment. The company soon replaced all four Nakanohitos. In response to this, the Game Club project was eventually terminated due to a boycott from the viewer community. On the other hand, the viewers can accept the replacement if there are appropriate reasons. For example, the Nakanohito of one of ReVdol's VTubers resigned for personal reasons, and viewers mostly supported her choice.

All interviewees mentioned that they would not accept a new Nakanohito without confirming that the original Nakanohito was treated fairly. For example, P4 mentioned that *"It's an awful thing to change Nakanohito only because the company wants to make more money.".* P3 stated that, *"we won't forgive the company if they abuse a Nakanohito".*

5 DISCUSSION

Our study uncovered many novel facets of VTuber live streaming. We now reflect and highlight our key findings, situate our results with prior research, and discuss insights and implications for HCI.

5.1 Multi-Layer Identity and the Line between Virtual and Real in Live Streaming

Our findings suggest that VTuber live streaming is a form of performance where the virtual and the real worlds converge. However, the boundaries between real and virtual selves are ever-changing, thus influencing social interactions.

These results first provide evidence that in VTuber live streaming, where viewers watch performances of Nakanohitos through the mediation of a virtual avatar, viewer's interpretations of the avatar depend on a number of factors, and do not always align with the Hyperpersonal Model of CMC [60, 61]. Viewers' motivations to watch VTuber may be an important factor. On one hand, some viewers who are more attracted by the virtual avatar and its persona may be more strict about the consistency between VTuber's behaviors and personas. They may often over-interpret cues from an avatar who has a simple and straight-forward persona and unique appearance. Viewers can thus form a close-to-perfect impression of a VTuber and follow her like fans, which aligns with the Hyperpersonal Model. These fans seem to maintain a unique tie to the avatar, similar to fandom in the anime domain. On the other hand, some viewers regard VTubers as friends and enjoy when VTubers share details about their daily life with them, either fictional or from the Nakanohito's real life. For these viewers, consistency between VTuber's behaviors and persona is less important because

they care more about the personality or even the 'character' of the Nakanohito, which is the presented self of the Nakanohito behind the anime mask of the virtual avatar. This is often not the intention of VTubers and does not align with the Hyperpersonal Model. Such *multiple layers of identity* may be the reason why some viewers may have conflicting interpretations of the avatar and Nakanohito. Viewers have agency to selectively interpret and creatively engage with VTubers' multiple identities.

The content that VTubers live stream is a mixture of virtual and real. VTubers build their virtual world persona by providing content based on their skills and experiences in the real world, resulting in a sense of distance between the VTuber and their viewers. This persona constructs the virtual identity of the VTuber, which echoes prior research [38]. We further show that viewers' interactions in live-streams and social networks (e.g. fan groups) also affect VTubers' identity construction. The unpredictability of live streams can contribute to more stories that viewers co-experience with a VTuber and these stories gradually merge into a VTuber's persona. Viewers' interactions within the community, encouraged by the creation and spreading of memes, can further add to VTubers' online identities.

Because the line between virtual and real in VTuber streaming is dynamic and blurry, viewers often do not consider a virtual avatar and the Nakanohito separately, but rather, regard the VTuber as a character in the virtual world and thus less subject to real world social norms. This mindset makes viewers more tolerant to VTubers' inappropriate behaviors and may contribute to viewers' indifferent attitudes towards agencies replacing Nakanohitos. However, where the virtuality-reality boundary of VTubers lies varies on individual viewers and may be shifted by external influences. Occasionally, the complexity of reality cuts in and breaks the illusion. A notable example occurred when the virtual "side" collapsed when viewers found out, through collective actions within the community (e.g., boycotts), that agencies were mistreating Nakanohitos and replacing them without an acceptable reason. This echoes findings of Kou et al. [31] about the importance of collective actions in seeking transparency and fairness in spaces mediated by virtual avatars.

5.2 Corporations Roles and the Labor of (Virtual) Live Streamers

Previous literature has discussed labor issues in the live streaming industry. For example, Woodcock and Johnson [64] used the theoretical lenses of affective and immaterial labor to explore a range of streamer activities on Twitch (e.g., building para-social intimacy with spectators, soliciting donations, etc.), and called for more efforts to study live streaming as part of the landscape of contemporary digital work.

Contributing to this line of research, these results showed how dedicated viewers perceive and care about the labor in live streaming, i.e., the often hidden and invisible labor of Nakanohitos. Unlike real-person streamers who can build their personal brand and be rewarded [43, 63], it is hard for Nakanohitos to conduct personal branding, as they are mostly behind the scenes of a stream. For VTubers, the brand and emotional attachments from viewers are mostly for the virtual avatar. This could be why some viewers had indifferent attitudes towards corporations replacing Nakanohitos.

Additionally, most of the time, the intellectual property of the avatar and the technologies required for producing VTuber content belongs to a corporation, which places a Nakanohito at a disadvantaged position. A corporation could change Nakanohitos at any time, without much cost. This is an extreme example of *consumer capitalism* in pop culture in East Asia, in that a virtual idol's lack of an autonomous existence "makes her an interchangeable and disposable image commodity" [18]. This also reminds us that as the live streaming industry grows and becomes more lucrative, new forms of labor exploitation may also emerge. Currently, we have a limited understanding of the labor contributed by different types of streamers, including Nakanohito, and of the policy options that might be leveraged to regulate the industry and to protect labor rights.

5.3 VTuber Live Streaming Challenges

By exploring the emerging VTuber live streaming, we unveiled several unique challenges that virtual live streamers have and compare our findings to prior research to provide insights into the professionalism of live streamers.

5.3.1 Constructing Virtual Identities. Dedicated viewers perceived that VTubers construct their identities dynamically through stories and performances to attract viewers with different tastes or to intentionally reveal parts of a Nakanohito's personality or life to generate contrast and amuse viewers. These results echo findings about the identity reassembling of fashion bloggers on Second Life, in that "a virtual body is only a slice of the continuum of ongoing identity reassemblin" [38]. VTubers' online identities are mediated through ever-changing avatar bodies and Nakanohito's real identities.

Live streaming viewers also have a profound impact on VTubers' identities. Their comments during live streams create moments that complement and transform streamers' existing personas and their interactions go beyond the live streams, e.g., spreading memes created during live streams in group chats and other fan communities. VTubers' personalities usually begin with simple designs, but evolve to be more complicated through the interactions they have with viewers. This creates challenges for VTubers. As their virtual identities become more complex and viewers' expectations change dynamically, Nakanohitos cannot effectively maintain a consistent persona or image, which will potentially lead them to lose fans.

5.3.2 Changeable or Multiple Nakanohitos. VTubers are unique in that the avatar and the Nakanohito are intentionally disembodied by both the community and a corporation. Because of this, VTubers could have Nakanohitos replaced or even have multiple Nakanohitos performing different content behind the same avatar together or separately. For example, two Nakanohitos could stream together, with one focusing on playing a game and another on interacting with viewers. Different Nakanohitos could also take turns delivering content that they have expertise in to viewers. It is surprising that viewers are accepting these practices, as long as the quality of the content remains untainted and the process is transparent. These practices, however, bring challenges to VTubers' teams. Changing Nakanohito sometimes also changes the talking style or other traits of the VTuber, which may again leads to a loss of fans.

5.3.3 Content Diversity. As discussed in 4.2.4, VTubers can not perform in an expressive manner as real-person streamers can due to technology constraints. Firstly, VTubers cannot provide content involving physical interactions with the real world which is the stage for many popular streaming genres, such as handcrafting tutorials [40] or travel experience sharing [41]. While there have been attempts by VTubers to co-stream with real-person streamers, the presence of VTubers can only be fixed to computer displays in physical environments. Moreover, in contrast to the characters in other computer graphic experiences such as games or animated movies, VTuber avatars rarely interact with the virtual environments they reside in, missing out on opportunities for enhanced narratives and storytelling.

5.4 Design Implications

Our findings about VTuber relates to the ongoing discussion in HCI and CSCW about inclusive live streaming, AI-mediated communication [27], and deception on social media. With virtual avatars, people who are unwilling to disclose their disabilities or imperfections may feel more comfortable broadcasting their talents and sharing stories through live streaming by constructing a different identity. With more and more advanced AI technology, automatically generating the appearance of different virtual idols, creating character designs and background stories, or even altering voices, will become much easier. Some companies have already begun creating fully AI-powered TV program hosts that use machine learning to simulate the voice, facial expressions, and gestures of real-life hosts [32]. We now detail several design considerations which should inform the design, implementation, and exploration of future live streaming services.

5.4.1 Design for Identity Management. Our results highlight that VTubers construct their identities dynamically, often evolving them over time based on characteristics of Nakanohitos or viewers community participation. Further, VTubers could have different Nakanohitos or even multiple Nakanohitos who collaborate to perform. Due to this potential complexity, maintaining a consistent identity will remain challenging for VTubers but is crucial to maintaining a stable fan base. There is thus a design opportunity to provide collaborative identity managing tools for VTubers, which could provide Nakanohitos with guidelines and actionable strategies to maintain a consistent identity during live streams. For example, when viewers send comments that risk revealing the identity of a Nakanohito during a live stream, the tool could alert the Nakanohito about the risk and guide them to respond to the comments in manner consistent with their constructed identity. Such tools may also need to tailor their output to different Nakanohitos if there are multiple Nakanohitos performing for one VTuber. While future natural language processing may make these features automatic, current streamers could rely on an "assistant" to help them flag risky comments and provide suggestions for maintaining identity, similar to the role of moderators [62].

5.4.2 Design for Transparency and Fairness. Replacing Nakanohito or having multiple Nakanohitos is a unique dynamic within VTuber live streaming. While interviewees were fine with this when the

quality of content remained consistent and the process was transparent and fair to Nakanohito, it was hard for viewers to determine if a Nakanohito was replaced without notice, and if they found out after the fact, they have been known to boycott the corporation. It is thus necessary for corporations to make the invisible labor of Nakanohitos more visible by disclosing information such as whether the Nakanohitos are treated fairly by the corporation, for example, not working overtime or being replaced without a convincing reason. The health or emotional states of Nakanohitos can also be made known to the viewers to some extent although care must be taken to protect Nakanohito's privacy.

5.4.3 Design for both the Virtual and the Real. Our findings also revealed the opportunities that can occur when blurring the line between the real and the virtual through VTubers' live streaming. VTubers' live streams suggest creative performance forms that could inspire real-person live streamers. Interviewees showed great enthusiasm towards watching multiple VTubers co-located in one virtual world. Most current real-person co-streaming, however, only displays the juxtaposition of multiple streamers' videos on the screen. This suggests a design opportunity to break geographic barriers and bring the co-streaming experience of VTubers to real-person streamers, using technologies such as markerless motion capture. Future Mixed Reality technologies could enhance viewers' experiences of VTuber live streams. Augmented reality could bring human-sized 3D avatars of VTubers to the real world. Virtual reality could immerse the viewers in the the virtual world where the VTuber performs. Live streaming that blurs the real and the virtual has the potential to be further combined with specific application areas, such as cultural heritage, tourism, and knowledge sharing, to better engage broader viewership.

5.5 Limitations and Future Work

We note several limitations of our study that should be considered when interpreting this work. Our study drew from the interview data of a sample of dedicated viewers that skewed young and male. Future research is needed to expand and deepen our understanding of the VTuber community with a larger and more diverse sample, even in different countries. Due to the nature of this study, causal conclusions or generalizability of the findings could not be made. Interviewing different stakeholders, such as Nakanohitos of VTubers, their moderators, other supporting roles, and managers of corporations could probably reveal more insights about this phenomenon. Alternative methodology where participants are invited to engage with a particular set of VTubers with their reactions to VTubers analyzed in a controlled experiment could also be explored in future research.

6 CONCLUSION

This work examined how the Otaku community engages with and perceive VTubers. The results revealed that the content and viewer motivations of VTuber live streaming were similar to real-person streamers. Viewers, however, perceived VTubers differently from real-person streamers by holding different expectations for VTuber's behaviors and feeling a stronger sense of distance towards them. Viewers also intentionally maintained the disembodiment of Nakanohitos from avatars to preserve VTubers' perfect images,

though they still cared about the Nakanohito with respect to labor fairness and transparency. By identifying unique challenges that viewers and VTubers encounter within the realm of live streaming, e.g., collaborative identity management and limited interactions with the real world, these results could inspire the design of future live streaming platforms and enable the community to better understand the influence and effects of VTubers or even AI-powered virtual streamers, on audiences.

REFERENCES

[1] 株式会社ユザロカル. 2020. バチャルYouTuber、本日1万人を突破（ユザロカルベ）｜株式会社ユザロカル. https://www.userlocal.jp/press/20200115vi/. [Online; accessed 16-January-2020].

[2] 株式会社ユザロカル. 2020. ファン数ランキング(1ページ)｜バチャルYouTuberランキング. https://virtual-youtuber.userlocal.jp/document/ranking. [Online; accessed 16-January-2020].

[3] Jeremy N Bailenson, Andrew C Beall, Jack Loomis, Jim Blascovich, and Matthew Turk. 2004. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators & Virtual Environments* 13, 4 (2004), 428–441.

[4] Bilibili. 2020. Bilibili. https://www.bilibili.com/.

[5] Daniel Black. 2012. The virtual idol: Producing and consuming digital femininity. In *Idols and celebrity in Japanese media culture.* Palgrave Macmillan, London, UK, 209–228.

[6] Raymond Blanton and Darlene Carbajal. 2019. Not a Girl, Not Yet a Woman: A Critical Case Study on Social Media, Deception, and Lil Miquela. In *Handbook of Research on Deception, Fake News, and Misinformation Online.* IGI Global, Hershey, PA, USA, 87–103.

[7] Di Laura Chen, Dustin Freeman, and Ravin Balakrishnan. 2019. Integrating Multimedia Tools to Enrich Interactions in Live Streaming for Language Learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* ACM, ACM, New York, NY, USA, 438.

[8] Mark Childs. 2011. *Identity: A Primer.* Springer London, London, 13–31. https://doi.org/10.1007/978-0-85729-361-9_2

[9] Juliet Corbin and Anselm Strauss. 1998. *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Sage Publications, Inc, Thousand Oaks, CA, USA.

[10] Donna Z Davis. 2010. *Exploring the media effects of three-dimensional online immersive worlds and how they impact both virtual and real social capital.* University of Florida.

[11] Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J Moore. 2006. Alone together?: exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI conference on Human Factors in computing systems.* ACM, ACM, New York, NY, USA, 407–416.

[12] Dwango. 2020. Niconico. https://www.nicovideo.jp/.

[13] Travis Faas, Lynn Dombrowski, Alyson Young, and Andrew D. Miller. 2018. Watch Me Code: Programming Mentorship Communities on Twitch.Tv. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article Article 50 (Nov. 2018), 18 pages. https://doi.org/10.1145/3274319

[14] C Ailie Fraser, Joy O Kim, Alison Thornsberry, Scott Klemmer, and Mira Dontcheva. 2019. Sharing the Studio: How Creative Livestreaming can Inspire, Educate, and Engage. In *Proceedings of the 2019 on Creativity and Cognition.* ACM, ACM, New York, NY, USA, 144–155.

[15] 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 (CHI EA '20).* Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3382923

[16] Noam Gal, Limor Shifman, and Zohar Kampf. 2016. "It Gets Better": Internet memes and the construction of collective identity. *New Media & Society* 18, 8 (2016), 1698–1714. https://doi.org/10.1177/1461444814568784

[17] Patrick W Galbraith. 2009. Moe: Exploring virtual potential in post-millennial Japan. *electronic journal of contemporary japanese studies* 9, 3 (2009).

[18] Patrick W Galbraith. 2012. Idols: The image of desire in Japanese consumer capitalism. In *Idols and celebrity in Japanese media culture.* Springer, New York, NY, USA, 185–208.

[19] Patrick W Galbraith and Jason G Karlin. 2012. Introduction: The mirror of idols and celebrity. In *Idols and celebrity in Japanese media culture.* Springer, New York, NY, USA, 1–32.

[20] Seth Glickman, Nathan McKenzie, Joseph Seering, Rachel Moeller, and Jessica Hammer. 2018. Design Challenges for Livestreamed Audience Participation Games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18).* ACM, New York, NY, USA, 187–199. https://doi.org/10.1145/3242671.3242708

[21] 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 (2020), 2023–2029.

[22] Oliver L. Haimson and John C. Tang. 2017. What Makes Live Events Engaging on Facebook Live, Periscope, and Snapchat. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 48–60. https://doi.org/10.1145/3025453.3025642

[23] William A. Hamilton, Oliver Garretson, and Andruid Kerne. 2014. Streaming on Twitch: Fostering Participatory Communities of Play Within Live Mixed Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1315–1324. https://doi.org/10.1145/2556288.2557048

[24] William A. Hamilton, Nic Lupfer, Nicolas Botello, Tyler Tesch, Alex Stacy, Jeremy Merrill, Blake Williford, Frank R. Bentley, and Andruid Kerne. 2018. Collaborative Live Media Curation: Shared Context for Participation in Online Learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 555, 14 pages. https://doi.org/10.1145/3173574.3174129

[25] Hololive. 2020. Hololive. https://www.hololive.tv/.

[26] Mizuko Ito, Daisuke Okabe, and Izumi Tsuji. 2012. *Fandom unbound: Otaku culture in a connected world.* Yale University Press, NEw Haven, CT, USA.

[27] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T Hancock, and Mor Naaman. 2019. AI-Mediated Communication: How the Perception that Profile Text was Written by AI Affects Trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, ACM, New York, NY, USA, 239.

[28] Phylis Johnson. 2010. *Second Life, Media, and the Other Society.* Peter Lang.

[29] Lucas Kempe-Cook, Stephen Tsung-Han Sher, and Norman Makoto Su. 2019. Behind the Voices: The Practice and Challenges of Esports Casters. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article Paper 565, 12 pages. https://doi.org/10.1145/3290605.3300795

[30] Kangsoo Kim, Arjun Nagendran, Jeremy N. Bailenson, Andrew Raij, Gerd Bruder, Myungho Lee, Ryan Schubert, Xin Yan, and Gregory F. Welch. 2017. A Large-Scale Study of Surrogate Physicality and Gesturing on Human–Surrogate Interactions in a Public Space. *Frontiers in Robotics and AI* 4 (2017), 32. https://doi.org/10.3389/frobt.2017.00032

[31] Yubo Kou, Xinning Gui, Shaozeng Zhang, and Bonnie Nardi. 2017. Managing disruptive behavior through non-hierarchical governance: Crowdsourcing in League of Legends and Weibo. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 62.

[32] Lily Kuo. 2018. World's first AI news anchor unveiled in China. Retrieved June 7, 2019 from https://www.theguardian.com/world/2018/nov/09/worlds-first-ai-news-anchor-unveiled-in-china

[33] Ka Yan Lam. 2016. The Hatsune Miku Phenomenon: More Than a Virtual J-Pop Diva. *The Journal of Popular Culture* 49, 5 (2016), 1107–1124.

[34] Linh K Le. 2014. Examining the rise of Hatsune Miku: the first international virtual idol. *The UCI Undergraduate Research Journal* 13, 1 (2014), 1–12.

[35] Alex Leavitt, Tara Knight, and Alex Yoshiba. 2016. Producing Hatsune Miku: Concerts, Commercialization, and the Politics of Peer Production. In *Media Convergence in Japan*. Kinema Club, New Haven, CT, USA, 200–229.

[36] Pascal Lessel, Alexander Vielhauer, and Antonio Krüger. 2017. Expanding Video Game Live-Streams with Enhanced Communication Channels: A Case Study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1571–1576. https://doi.org/10.1145/3025453.3025708

[37] Jie Li, Xinning Gui, Yubo Kou, and Yukun Li. 2019. Live Streaming as Co-Performance: Dynamics between Center and Periphery in Theatrical Engagement. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW, Article 64 (Nov. 2019), 22 pages. https://doi.org/10.1145/3359166

[38] Christine Liao. 2011. Virtual fashion play as embodied identity re/assembling: Second Life fashion bloggers and their avatar bodies. In *Reinventing ourselves: contemporary concepts of identity in virtual Worlds*. Springer, London, UK, 101–127.

[39] Danielle Lottridge, Frank Bentley, Matt Wheeler, Jason Lee, Janet Cheung, Katherine Ong, and Cristy Rowley. 2017. Third-Wave Livestreaming: Teens' Long Form Selfie. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 20, 12 pages. https://doi.org/10.1145/3098279.3098540

[40] Zhicong Lu, Michelle Annett, Mingming Fan, and Daniel Wigdor. 2019. "I Feel It is My Responsibility to Stream": Streaming and Engaging with Intangible Cultural Heritage through Livestreaming. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article Paper 229, 14 pages. https://doi.org/10.1145/3290605.3300459

[41] Zhicong Lu, Michelle Annett, and Daniel Wigdor. 2019. Vicariously Experiencing it all without Going Outside: A Study of Outdoor Livestreaming in China. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–28.

[42] Zhicong Lu, Seongkook Heo, and Daniel Wigdor. 2018. StreamWiki: Enabling Viewers of Knowledge Sharing Live Streams to Collaboratively Generate Archival Documentation for Effective In-Stream and Post-Hoc Learning. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), Article 112. https://doi.org/10.1145/3274381

[43] Zhicong Lu, Haijun Xia, Seongkook Heo, and Daniel Wigdor. 2018. You Watch, You Give, and You Engage: A Study of Live Streaming Practices in China. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article Paper 466, 13 pages. https://doi.org/10.1145/3173574.3174040

[44] Tetsuya Nakajo. 2020. Live2D. https://www.live2d.com/en/.

[45] Bonnie Nardi. 2010. *My life as a night elf priest: An anthropological account of World of Warcraft.* University of Michigan Press, Ann Arbor, MI, USA.

[46] Carman Neustaedter and Elena Fedorovskaya. 2009. Capturing and sharing memories in a virtual world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, ACM, New York, NY, USA, 1161–1170.

[47] Carman Neustaedter and Elena Fedorovskaya. 2009. Presenting Identity in a Virtual World through Avatar Appearances. In *Proceedings of Graphics Interface 2009 (GI '09)*. Canadian Information Processing Society, CAN, 183–190.

[48] Nijisanji. 2020. Nijisanji. https://nijisanji.ichikara.co.jp/.

[49] Asaf Nissenbaum and Limor Shifman. 2017. Internet memes as contested cultural capital: The case of 4chan's /b/ board. *New Media & Society* 19, 4 (2017), 483–501. https://doi.org/10.1177/1461444815609313

[50] National Geographic of Azeroth Forum. 2020. National Geographic of Azeroth Forum: VTuber Discussion Board. https://bbs.nga.cn/thread.php?fid=-60204499.

[51] 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 (09 2016), 1–18. https://doi.org/10.1371/journal.pone.0161794

[52] Anthony J. Pellicone and June Ahn. 2017. The Game of Performing Play: Understanding Streaming As Cultural Production. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4863–4874. https://doi.org/10.1145/3025453.3025854

[53] Linden Research. 2014. *Live Performances - Second Live Wiki.* Retrieved May 23, 2020 from http://wiki.secondlife.com/wiki/Live_Performances

[54] Jeff T. Sheng and Sanjay R. Kairam. 2020. From Virtual Strangers to IRL Friends: Relationship Development in Livestreaming Communities on Twitch. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 94 (Oct. 2020), 34 pages. https://doi.org/10.1145/3415165

[55] Stephen Tsung-Han Sher and Norman Makoto Su. 2019. Speedrunning for Charity: How Donations Gather Around a Live Streamed Couch. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article Article 48 (Nov. 2019), 26 pages. https://doi.org/10.1145/3359150

[56] SimilarWeb. 2020. Bbs.nga.cn Analytics - Market Share Stats & Traffic Ranking. https://www.similarweb.com/website/bbs.nga.cn. [Online; accessed 16-January-2020].

[57] John Tang, Gina Venolia, Kori Inkpen, Charles Parker, Robert Gruen, and Alicia Pelton. 2017. Crowdcasting: Remotely Participating in Live Events Through Multiple Live Streams. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 98 (Dec. 2017), 18 pages. https://doi.org/10.1145/3134733

[58] John C. Tang, Gina Venolia, and Kori M. Inkpen. 2016. Meerkat and Periscope: I Stream, You Stream, Apps Stream for Live Streams. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4770–4780. https://doi.org/10.1145/2858036.2858374

[59] TL Taylor. 2018. *Watch me play: Twitch and the rise of game live streaming*. Vol. 24. Princeton University Press, Princeton, NJ, USA.

[60] Joseph B Walther. 1996. Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication research* 23, 1 (1996), 3–43.

[61] Joseph B Walther. 2011. Theories of computer-mediated communication and interpersonal relations. *The handbook of interpersonal communication* 4 (2011), 443–479.

[62] Donghee Yvette Wohn. 2019. Volunteer Moderators in Twitch Micro Communities: How They Get Involved, the Roles They Play, and the Emotional Labor They Experience. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, ACM, New York, NY, USA, 160.

[63] Donghee Yvette Wohn, Guo Freeman, and Caitlin McLaughlin. 2018. Explaining viewers' emotional, instrumental, and financial support provision for live streamers. In *Proceedings of the 2018 CHI conference on human factors in computing systems (CHI '18)*. ACM, ACM, New York, NY, USA, 474.

[64] Jamie Woodcock and Mark R Johnson. 2019. The Affective Labor and Performance of Live Streaming on Twitch.tv. *Television & New Media* 20 (2019), 1527476419851077. Issue 8.

[65] Nick Yee, Jeremy N Bailenson, and Kathryn Rickertsen. 2007. A Meta-Analysis of the Impact of the Inclusion and Realism of Human-like Faces on User Experiences in Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/1240624.1240626

[66] Yiyi Yin. 2018. Vocaloid in China: Cosmopolitan music, cultural expression, and multilayer identity. *Global Media and China* 3, 1 (2018), 51–66. https://doi.org/10.1177/2059436418778600

[67] Japan National Tourism Organization New York. 2018. *JNTO to Launch 'Come to Japan' Campaign With Kizuna AI, the World's First Virtual YouTuber*. Retrieved May 31, 2020 from https://www.prnewswire.com/news-releases/jnto-to-launch-come-to-japan-campaign-with-kizuna-ai-the-worlds-first-virtual-youtuber-300608037.html

[68] 依光流. 2018. 冯提莫不是B站直播的杀招，*Vtuber* 才是. Retrieved May 31, 2020 from https://www.ifanr.com/1325444